Prediktivni model tehničke ispravnosti vozila temeljem tehničkih pregleda

Šiško, Igor

Doctoral thesis / Disertacija

2015

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Zagreb, Faculty of Mechanical Engineering and Naval Architecture / Sveučilište u Zagrebu, Fakultet strojarstva i brodogradnje**

Permanent link / Trajna poveznica: https://urn.nsk.hr/urn:nbn:hr:235:361568

Rights / Prava: In copyright / Zaštićeno autorskim pravom.

Download date / Datum preuzimanja: **2024-03-29**

Repository / Repozitorij:

Repository of Faculty of Mechanical Engineering and Naval Architecture University of Zagreb







Sveučilište u Zagrebu

Fakultet strojarstva i brodogradnje

Igor Šiško

PREDIKTIVNI MODEL TEHNIČKE ISPRAVNOSTI VOZILA TEMELJEM TEHNIČKIH PREGLEDA

DOKTORSKI RAD

Zagreb, 2015.



Sveučilište u Zagrebu

Fakultet strojarstva i brodogradnje

Igor Šiško

PREDIKTIVNI MODEL TEHNIČKE ISPRAVNOSTI VOZILA TEMELJEM TEHNIČKIH PREGLEDA

DOKTORSKI RAD

Mentor:

Prof. dr.sc. Dragutin Lisjak

Zagreb, 2015.



University of Zagreb

Faculty of mechanical engineering and naval architecture

Igor Šiško

PREDICTIVE ROADWORTHINESS OF VEHICLES MODEL BASED ON PERIODICAL TECHNICAL INSPECTIONS

DOCTORAL THESIS

Supervisor:

Prof. dr.sc. Dragutin Lisjak

Zagreb, 2015.

PODACI ZA BIBLIOGRAFSKU KARTICU

UDK: **629.3.017**

Ključne riječi: **Osobni automobil/vozilo, tehnički pregled, tehnička ispravnost, prediktivni model, sigurnost cestovnog prometa**

Znanstveno područje: **TEHNIČKE ZNANOSTI**

Znanstveno polje: **strojarstvo**

Institucija u kojoj je rad izrađen: **Fakultet strojarstva i brodogradnje, Sveučilišta u Zagrebu**

Mentor rada: **dr.sc. Dragutin Lisjak, izvanredni profesor**

Broj stranica: 191

Broj slika: 42

Broj tablica: 54

Broj korištenih bibliografskih jedinica: 71

Datum obrane: 2. srpanj 2015.

Povjerenstvo: dr. sc. Biserka Runje, redoviti profesor - predsjednik, dr. sc. Dragutin Lisjak, izvanredni profesor - mentor, dr. sc. Ivica Veža, redoviti profesor - član, dr. sc. Tomislav Šarić, redoviti profesor - član, dr. sc. Zoran Jurković, izvanredni profesor - član.

Institucija u kojoj je rad pohranjen: **Fakultet strojarstva i brodogradnje, Sveučilišta u Zagrebu**
Nacionalna i sveučilišna knjižnica

ZAHVALA

Osobito se zahvaljujem mentoru dr. sc. Dragutinu Lisjaku za savjetovanje i usmjeravanje koje mi je pružio tijekom izrade ovog rada, a posebno za pruženu podršku i razumijevanje usprkos svim otežavajućim okolnostima koje su bile prisutne.

Koristim priliku da se zahvalim i predsjednici povjerenstva dr. sc. Biserki Runje te članovima povjerenstva - dr. sc. Ivici Veži, dr. sc. Tomislavu Šariću i dr. sc. Zoranu Jurkoviću na vrijednim, konstruktivnim i dobronamjernim savjetima tijekom izrade ovog doktorskog rada.

Također se zahvaljujem dr. sc. Nedeljku Štefaniću, dr. sc. Goranu Đukiću i dr. sc. Goranu Zovaku na cjelokupnom angažmanu i podršci pruženoj u prethodnoj fazi izrade ovog rada.

Posebno se zahvaljujem Hrvatskom autoklubu, čiji sam zaposlenik od 1998. godine, na povjerenju iz kojeg je proizašla odluka o financiranju mojeg poslijediplomskog studija.

Igor Šiško, dipl. ing.

"Previše volim zvijezde da bih se plašio noći.".....*Galileo Galilei*

...mojem Mark-u

SADRŽAJ

- **Popis slika** iii
- **Popis tablica** v
- **Popis kratica** viii
- **Predgovor** ix
- **Sažetak rada i ključne riječi** x

1. UVOD 1

2. TEHNIČKI PREGLEDI VOZILA 10

2.1. SUSTAV TEHNIČKIH PREGLEDA VOZILA U RH 10

2.2. TEHNIČKI PREGLEDI VOZILA U KONTEKSTU ZAKONODAVSTVA EUROPSKE UNIJE 16

3. POSLOVNA INTELIGENCIJA 22

3.1. UVOD U POSLOVNU INTELIGENCIJU 22

3.2. METODE OTKRIVANJA ZNANJA 25

3.2.1. *Rudarenje podataka* 25

3.2.2. *Stabla odlučivanja* 28

3.2.3. *CHAID i Exhaustive CHAID algoritmi* 34

3.2.4. Perceptronska neuronska mreža 48

3.2.5. Pareto analiza 63

3.2.5.1. *Pareto analiza u funkciji izrade prediktivnog modela tehničke ispravnosti vozila temeljem tehničkih pregleda* 64

4. MODELIRANJE PREDIKTIVNIH MODELA 73

4.1. RAZRADA METODOLOGIJE OTKRIVANJA ZNANJA 73

4.2. DEFINIRANJE VARIJABLI I PARAMETARA 76

4.3. METRIKA ZA USPOREDBU PREDIKTIVNIH MODELA 84

4.4. IZRADA I USPOREDBA PREDIKTIVNIH MODELA 87

4.4.1. Prediktivni modeli za osobne automobile VW Golf 87

4.4.1.1. *Analiza prediktivnog potencijala modela temeljenog na Exhaustive CHAID algoritmu* 100

4.4.1.2. *Poboljšanje prediktivnog potencijala modela temeljenog na Exhaustive CHAID algoritmu* 102

4.4.1.3. *Analiza nakon provedenog poboljšanja modela temeljenog na Exhaustive CHAID algoritmu* 103

4.4.1.4. *Alternativni prediktivni model temeljen na algoritmu perceptronskih neuronskih mreža* 110

4.4.2. Prediktivni modeli za osobni automobili Fiat Punto 112

4.4.2.1. *Analiza prediktivnog potencijala modela temeljenog na Exhaustive CHAID algoritmu* 125

4.4.2.2. *Poboljšanje prediktivnog potencijala modela temeljenog na Exhaustive CHAID algoritmu* 126

4.4.2.3. *Analiza nakon provedenog poboljšanja modela temeljenog na Exhaustive CHAID algoritmu* 126

4.4.2.4. *Alternativni prediktivni model temeljen na algoritmu perceptronskih neuronskih mreža* 134

4.4.3. Prediktivni modeli za osobne automobile Opel Astra 136

4.4.3.1. *Analiza prediktivnog potencijala modela temeljenog na Exhaustive CHAID algoritmu* 147

4.4.3.2. *Poboljšanje prediktivnog potencijala modela temeljenog na Exhaustive CHAID algoritmu* 148

4.4.3.3. *Analiza nakon provedenog poboljšanja modela temeljenog na Exhaustive CHAID algoritmu* 148

4.4.3.4. *Alternativni prediktivni model temeljen na algoritmu perceptronskih neuronskih mreža* 155

5. VALIDACIJA MODELA 158

6. INTERPRETACIJA REZULTATA 161

6.1. OGRANIČENJA MODELA 168

6.2. PRIJEDLOG DALJNJEG ISTRAŽIVANJA 170

7. ZAKLJUČAK 173

8. LITERATURA 179

PRILOG 1 184

ŽIVOTOPIS 190

POPIS SLIKA

Slika 2.1. - *Shematski prikaz sustava tehničkog pregleda vozila u RH*..... 10

Slika 2.2. - *Shematski prikaz kretanja osobnog automobila po tehnološkoj liniji*..... 13

Slika 2.3. - *Slijed faza procesa tehničkog pregleda vozila*..... 14

Slika 2.4. - *Shematski prikaz slijeda direktiva EU iz domene tehničkih pregleda vozila*..... 18

Slika 3.1. - *Okvir razvoja procesa rudarenja podataka*..... 26

Slika 3.2. - *Princip stabla odlučivanja*..... 31

Slika 3.3. - *Nomenklatura stabla odlučivanja*..... 32

Slika 3.4. - *Model umjetnog neurona*..... 49

Slika 3.5. – *Prikaz backpropagation modela neurona*..... 49

Slika 3.6. – *Aktivacijske funkcije log-sigmoid tipa*..... 49

Slika 3.7. – *Aktivacijske funkcije tan-sigmoid tipa*..... 50

Slika 3.8. – *Aktivacijske funkcije lin tipa*..... 50

Slika 3.9. – *Shema jednoslojne unaprijedne BP mreže*..... 51

Slika 3.10. – *Shema dvoslojne unaprijedne backpropagation mreže*..... 51

Slika 3.11. – *Gradijent ili nagib tangente funkcije y=f(x)*..... 52

Slika 3.12. – *Funkcije greške E=f(w)*..... 53

Slika 3.13. – *Prikaz SSE površine kao funkcije od W i B*..... 54

Slika 3.14. – *Značenje koeficijenta brzine učenja*..... 58

Slika 3.15. – *Primjer površine greške sa putanjom konvergencije*..... 58

Slika 3.16. – *Problem lokalnog i globalnog minimuma*..... 59

Slika 3.17. – *Problem overfittinga mreže*..... 61

Slika 3.18. – *Shematski prikaz early stopping metode*..... 63

Slika 3.19. – *Načelo 80/20*..... 63

Slika 3.20. – *Dijagramski prikaz rezultata Pareto analize* 70

Slika 4.1. - *Esencijalni mehanizam prediktivnog modela* 73

Slika 4.2. – *Novo-razvijeni algoritam prediktivnog modela tehničke ispravnosti osobnih automobila*..... 74

Slika 4.3. – *Prikaz sadržaja čvorova u okviru stabla odlučivanja* 88

Slika 4.4. - *Stablo odlučivanja trening uzorka prediktivnog modela za VW Golf*..... 91

Slika 4.5. - *Dijagram dobiti za ciljanu kategoriju "pad" za VW Golf* 97

Slika 4.6. - *Dijagram indeksa za ciljanu kategoriju "pad" za VW Golf* 97

Slika 4.7. - *Stablo odlučivanja trening uzorka prediktivnog modela za VW Golf nakon provođenja radnje za poboljšanje modela*..... 104

Slika 4.8. - *Prikaz nezavisnih varijabli prema procjeni razine značaja za VW Golf (neuronska mreža)*..... 111

Slika 4.9. - *Stablo odlučivanja trening uzorka prediktivnog modela za Fiat Punto*.. 118

Slika 4.10. - *Dijagram dobiti za ciljanu kategoriju "pad" za Fiat Punto* 122

Slika 4.11. - *Dijagram indeksa za ciljanu kategoriju "pad" za Fiat Punto* 122

Slika 4.12. - *Stablo odlučivanja trening uzorka prediktivnog modela za Fiat Punto nakon provođenja radnje za poboljšanje modela* 127

Slika 4.13. - *Prikaz nezavisnih varijabli prema procjeni razine značaja za Fiat Punto (neuronska mreža)* 135

Slika 4.14. - *Stablo odlučivanja trening uzorka prediktivnog modela za Opel Astra* 141

Slika 4.15. - *Dijagram dobiti za ciljanu kategoriju "pad" za Opel Astra* 144

Slika 4.16. - *Dijagram indeksa za ciljanu kategoriju "pad" za Opel Astra*..... 144

Slika 4.17. - *Stablo odlučivanja trening uzorka prediktivnog modela za Opel Astra nakon provođenja radnje za poboljšanje modela* 149

Slika 4.18. - *Prikaz nezavisnih varijabli prema procjeni razine značaja u okviru algoritma neuronske mreže za osobne automobile Opel Astra* 156

POPIS TABLICA

Tablica 3.1. - *Trgovačke marke osobnih automobila sukladno kriteriju 1* 65

Tablica 3.2. - *Tipovi osobnih automobila sukladno kriteriju 2*..... 66

Tablica 3.3. – *Tipovi osobnih automobile - utjecajni čimbenici* 68

Tablica 4.1. – *Primjer dijela ekstrahiranih podataka za VW Golf* 76

Tablica 4.2. – *Pregled varijabli korištenih u algoritmima*..... 80

Tablica 4.3. – *Pregled definiranih parametara Exhaustive CHAID algoritma u poboljšanom prediktivnom modelu* 83

Tablica 4.4. - *Primjer matrice konfuzije (kontingencijske tablice)*..... 86

Tablica 4.5. - *Derivacije matrice konfuzije* 86

Tablica 4.6. - *Postavke i zbirni rezultati modela za VW Golf (Exhaustive CHAID)*.. 87

Tablica 4.7. - *Dobit po završnim čvorovima za VW Golf*..... 95

Tablica 4.8. - *Rizik za VW Golf* 98

Tablica 4.9. - *Klasifikacija za VW Golf (Exhaustive CHAID)* 98

Tablica 4.10. – *Prikaz vrijednosti MCC za VW Golf (Exhaustive CHAID)* 99

Tablica 4.11. - *Izvod dijela prediktivnih vrijednosti za VW Golf*..... 100

Tablica 4.12. - *Zbirni rezultati modela za VW Golf nakon provođenja radnje za poboljšanje (Exhaustive CHAID)* 105

Tablica 4.13. - *Rizik za VW Golf nakon provođenja radnje za poboljšanje*..... 105

Tablica 4.14. - *Klasifikacija za VW Golf nakon provođenja radnje za poboljšanje (Exhaustive CHAID)* 106

Tablica 4.15. - *Prikaz vrijednosti MCC za VW Golf nakon provođenja radnje za poboljšanje (Exhaustive CHAID)* 107

Tablica 4.16. - *Prikaz poslovnih pravila za poboljšani Exhaustive CHAID prediktivni model VW Golf* 107

Tablica 4.17. - *Parametri strukture neuronske mreže za VW Golf*..... 110

Tablica 4.18. - *Klasifikacija za osobne automobile VW Golf (neuronska mreža)*.... 111

Tablica 4.19. – *Prikaz vrijednosti MCC za VW Golf (neuronska mreža)*..... 112

Tablica 4.20. - *Postavke i zbirni rezultati modela za Fiat Punto (Exhaustive CHAID)* 113

Tablica 4.21. - *Dobit po završnim čvorovima za osobne automobile Fiat Punto* 120

Tablica 4.22. - *Rizik za Fiat Punto* 123

Tablica 4.23. - *Klasifikacija za Fiat Punto (Exhaustive CHAID)* 123

Tablica 4.24. - *Prikaz vrijednosti MCC za Fiat Punto (Exhaustive CHAID)* 124

Tablica 4.25. - *Izvod dijela prediktivnih vrijednosti za Fiat Punto* 124

Tablica 4.26. - *Zbirni rezultati modela za Fiat Punto nakon provođenja radnje za poboljšanje (Exhaustive CHAID)* 128

Tablica 4.27. - *Rizik za Fiat Punto nakon provođenja radnje za poboljšanje* 129

Tablica 4.28. – *Klasifikacija za Fiat Punto nakon provođenja radnje za poboljšanje (Exhaustive CHAID)* 129

Tablica 4.29. - *Prikaz vrijednosti MCC za Fiat Punto nakon provođenja radnje za poboljšanje (Exhaustive CHAID)* 130

Tablica 4.30. - *Prikaz poslovnih pravila za poboljšani Exhaustive CHAID prediktivni model Fiat Punto* 131

Tablica 4.31. - *Parametri strukture neuronske mreže za Fiat Punto* 134

Tablica 4.32. – *Klasifikacija za Fiat Punto (neuronska mreža)* 135

Tablica 4.33. – *Prikaz vrijednosti MCC za Fiat Punto (neuronska mreža)* 136

Tablica 4.34. – *Postavke i zbirni rezultati modela za Opel Astra (Exhaustive CHAID)* 137

Tablica 4.35. - *Dobit po završnim čvorovima za osobne automobile Opel Astra* 142

Tablica 4.36. - *Rizik za Opel Astra* 145

Tablica 4.37. - *Klasifikacija za Opel Astra (Exhaustive CHAID)* 145

Tablica 4.38. - *Prikaz vrijednosti MCC za Opel Astra (Exhaustive CHAID)* 148

Tablica 4.39. - *Izvod dijela prediktivnih vrijednosti za Opel Astra* 146

Tablica 4.40. - *Zbirni rezultati modela za Opel Astra nakon provođenja radnje za poboljšanje (Exhaustive CHAID)* 150

Tablica 4.41. - *Rizik za Opel Astra nakon izvršene radnje za poboljšanje* 151

Tablica 4.42. - *Klasifikacija za Fiat Punto nakon izvršene radnje za poboljšanje (Exhaustive CHAID)* 151

Tablica 4.43. - *Prikaz vrijednosti MCC za Opel Astra nakon provođenja radnje za poboljšanje (Exhaustive CHAID)* 152

Tablica 4.44. - *Prikaz poslovnih pravila za poboljšani Exhaustive CHAID prediktivni model Opel Astra*..... 153

Tablica 4.45. - *Parametri strukture neuronske mreže za Opel Astra* 155

Tablica 4.46. - *Klasifikacija za Opel Astra (neuronska mreža)*..... 156

Tablica 4.47. - *Prikaz vrijednosti MCC za Opel Astra (neuronska mreža)* 157

Tablica 5.1. - *Usporedba uspješnosti prediktivnih modela*..... 158

Tablica 5.2. - *Usporedba skupova podataka* 160

Tablica 6.1. - *Usporedba rezultata istraživanja između obrađenih osobnih automobila*..... 162

Tablica 6.2. - *Usporedba kritičnih nezavisnih varijabli u kontekstu "pada" osobnih automobila na tehničkom pregledu*..... 165

POPIS KRATICA

RH – Republika Hrvatska

CHAID – engl. *Chi Square Automatic Interaction Detection*

STP – stanica za tehnički pregled vozila

MUP - Ministarstvo unutarnjih poslova

CVH – Centar za vozila Hrvatske d.d.

HAK – Hrvatski autoklub

M1 – M1 kategorija vozila (osobni automobili)

L – L kategorija vozila (vozila s dva ili tri kotača i kvadricikli)

ADAC - njem. *Allgemainer Deutscher Automobilclub*

CITA - franc. *Comité international de l'inspection technique automobile*

OBD - engl. *On-Board Diagnostics*

MiD – engl. *Minor Defects*

MaD - engl. *Major Defects*

DD - engl. *Dangerous Defects*

BI – engl. *Business Intelligence*

CART - engl. *Classification and Regression Trees*

QUEST - eng. *Quick, Unbiased, Efficient Statistical Tree*

ID3 - engl. *Iterative Dichotomiser 3*

AID – engl. *Automatic Interaction Detection*

THAID – engl. *THeta Automatic Interaction Detection*

BAV - engl. *Bicuspid Aortic Valve*

TP – tehnički pregled vozila

BEZ KAT – vozilo pogonjeno benzinom bez katalizatora ili s nereguliranim katalizatorom

REG KAT - vozilo pogonjeno benzinom s reguliranim katalizatorom

DIZEL – vozilo pogonjeno dizelskim gorivom

SQL – engl. *Structured Query Language*

MCC – Matthews-ov koeficijent korelacije

cca – latin. *Circa*

PREDGOVOR

Podaci o vozilu dobiveni tijekom provođenja tehničkog pregleda, odnosno tijekom obveznog procesa u sklopu kojeg se vozilo podvrgava nizu tehničkih ispitivanja radi provjere tehničke ispravnosti pojedinih uređaja/sklopova te vozila u cjelini, predstavljaju vrijedan resurs informacija koje, korištene na odgovarajući način, mogu pridonijeti poboljšanju sigurnosti prometa na cestama.

Namjera autora ovog doktorskog rada jest da ponudi primjenjiv prediktivni model tehničke ispravnosti osobnih automobila temeljem rezultata tehničkih pregleda čiji je potencijal znanstveno dokazan, te koji s obzirom na svoju univerzalnost ima veliku mogućnost primjene na sva vozila, a rezultati modela u smislu predikcije tehničke ispravnosti vozila imaju značaj iz aspekta sigurnosti prometa na cestama i iz aspekta gospodarstva (prodaja vozila).

PREDIKTIVNI MODEL TEHNIČKE ISPRAVNOSTI VOZILA TEMELJEM TEHNIČKIH PREGLEDA

Sažetak rada

Tehnički pregled vozila djelatnost je od općeg (nacionalnog) interesa u kontekstu provjere tehničke ispravnosti vozila, koja ima izravni utjecaj na razinu sigurnosti prometa na cestama. Povratna informacija glede rezultata tehničkog pregleda vozila, posebno u situaciji kada se na vozilu utvrde tehničke neispravnosti koje za posljedicu imaju negativan rezultat vozila na tehničkom pregledu ("pad" vozila na tehničkom pregledu), od iznimnog je značaja za proizvođača i vlasnika/korisnika vozila, kako u pogledu podizanja razine tehničke ispravnosti vozila u eksploataciji, tako i u pogledu gospodarenja istim. Cilj ovog znanstvenog istraživanja je izraditi prediktivni model tehničke ispravnosti osobnih vozila temeljem rezultata tehničkih pregleda, primjenom metoda poslovne inteligencije (rudarenja podataka, stabla odlučivanja, CHAID algoritmi, neuronske mreže).

Ključne riječi: Osobni automobil/vozilo, tehnički pregled, tehnička ispravnost, prediktivni model, sigurnost cestovnog prometa

PREDICTIVE ROADWORTHINESS OF VEHICLES MODEL BASED ON PERIODICAL TECHNICAL INSPECTIONS

Summary

Technical vehicles inspection procedure is of general (national) interest in the context of the roadworthiness of vehicles, which has a direct impact on the security of the Road Traffic safety. Feedback regarding the results of periodical technical inspections (PTI) of vehicles, especially in a situation where the vehicle malfunctions are determined, result in a negative outcome in the vehicle technical inspection, is of great importance for the manufacturers and owners/users of vehicles, both in the context of raising the level of roadworthiness of vehicles in exploitation and in the same context, management. Main goal of this scientific research is to build a predictive roadworthiness of passenger vehicles model based on the results of technical inspections, by using methods of Business Intelligence (Data Mining, Decision trees, CHAID algorithms, neural networks).

Keywords: personal automobile/vehicle, periodical technical inspection, roadworthiness, predictive model, road traffic safety

1. UVOD

Sigurnost prometa na cestama je nacionalni interes RH. „Vozilo“, odnosno njegova tehnička ispravnost, jedan je od tri ključna čimbenika koji direktno utječu na sigurnost prometa na cestama, uz „Vozača“ i „Cestu“. Proteklih nekoliko desetljeća razvoj i broj vozila na europskim prometnicama, pa tako i na hrvatskim prometnicama, poprimio je velike razmjere. Logično je da s povećanjem broja vozila u prometu dolazi i do većeg broja prometnih nesreća. S obzirom da je održavanje tehničke ispravnosti vozila u uvjetima eksploatacije od primarne važnosti za sigurnost svih sudionika u prometu na cestama država je propisala obvezu kontrole tehničke ispravnosti vozila, odnosno zakonski definirala sustav tehničkih pregleda i registracije vozila.

Osnovni smisao postojanja tehničkog pregleda vozila je uspostava periodične kontrole tehničke ispravnosti vozila kako bi se na taj način što je moguće bolje održavao postojeći vozni park u ispravnom stanju, tj. kako bi se što bolje održavali vitalni sklopovi i dijelovi koji utječu na sigurnost u prometu te kako bi se smanjio utjecaj vozila kao jednog od tri osnovna čimbenika koji uzrokuju prometne nesreće. Uvođenjem ispitivanja ispušnih plinova (EKO testa) smisao tehničkog pregleda proširen je i na ekološku dimenziju, jer se uvođenjem periodičke obveze provjere sadržaja ispušnih plinova utječe na vlasnike vozila da redovno servisiraju svoja vozila te na taj način za vrijeme cjelokupnog životnog vijeka vozila održavaju emisiju ispušnih plinova u propisanim granicama.

Rezultat tehničkog pregleda vozila u smislu utvrđivanja njegovog tehničkog stanja, odnosno činjenice je li njegova tehnička ispravnost zadovoljavajuća ili je utvrđena jedna ili vise tehničkih neispravnosti, predstavlja značajan podatak u kontekstu sigurnosti cestovnog prometa. Od posebnog (direktnog) značaja u kontekstu sigurnosti cestovnog prometa jest negativni rezultat tehničkog pregleda vozila, odnosno situacija u kojoj vozilo “pada” na tehničkom pregledu zbog utvrđene tehničke neispravnosti ili više njih (što u slučaju da se ne radi o prvoj registraciji vozila za posljedicu ima neproduljenje valjanosti prometne dozvole vozila), zbog čega predmetno vozilo više ne smije sudjelovati u prometu na cestama (sve dok se predmetne tehničke

neispravnosti ne otklone, odnosno dok se tehnička ispravnost predmetnog vozila ponovno ne utvrdi u sklopu tehničkog pregleda).

Međutim, postoji potreba tranzicije rezultata tehničkih pregleda vozila iz konteksta podataka u kontekst visokovrijednih informacija s dimenzijom predikcije, čime će se dodatno potencirati njihov značaj ne samo iz aspekta sigurnosti prometa na cestama nego i iz drugih aspekata, što je i tema ovog doktorskog rada.

Tehnički podaci o vozilima koja su pristupila tehničkim pregledima te rezultati tehničkih pregleda vozila, strukturirani u vidu velikih baza podataka, predstavljaju veliki potencijal za analizu tehničke ispravnosti vozila kako u kontekstu sigurnosti prometa na cestama, tako i u kontekstu prodaje vozila (marketinški kontekst).

U kontekstu navedenih činjenica definiran je problem istraživanja:

Analiza rezultata tehničkih pregleda vozila [1], u pravilu se svode na izvještajnu statističku obradu podataka iz koje ne proizlaze relevantne informacije u pogledu predikcije tehničke ispravnosti vozila.

Sukladno problemu istraživanja definiran je predmet istraživanja:

Predmet istraživanja predstavljaju rezultati tehničkih pregleda osobnih (kategorija M1, prema homologacijskoj klasifikaciji Europske unije [2], [3]) automobila (cca 1.400.000 vozila godišnje), koja su pristupili tehničkom pregledu u razdoblju od 01.01. do 31.12. 2009. godine, odnosno u razdoblju od 01.01. do 31.12. 2010. godine, strukturirani od strane *HAK*-a u vidu informatičke baze podataka.

Hipoteza i cilj istraživanja:

Kroz istraživanje se dokazuje hipoteza istraživanja da je, na temelju načela poslovne inteligencije, putem uporabe metoda rudarenja podataka i stabla odlučivanja, a nakon

kompetentnog tumačenja dobivenih rezultata, moguće otkriti zakonitosti među promatranim varijablama, odnosno na temelju postojećeg stanja predvidjeti buduće stanje.

Cilj ovog istraživanja jest izrada teorijski utemeljenog modela za predviđanje tehničke ispravnosti osobnih automobila temeljem rezultata tehničkih pregleda i temeljem tehničkih karakteristika osobnih automobila. Konačni cilj predstavlja razvoj mehanizma za pružanje povratnih informacija iz procesa tehničkog pregleda osobnih automobila - posebno u situaciji kada se na osobnom automobilu utvrde tehničke neispravnosti koje za posljedicu imaju negativan rezultat osobnog automobila na tehničkom pregledu – državnim tijelima nadležnim za sigurnost cestovnog prometa, proizvođačima automobila (čime se stvaraju preduvjeti za provođenje projektne modifikacije na automobilima u cilju postizanja boljih sigurnosnih i ekoloških karakteristika, odnosno radi podizanja kvalitete automobila) te vlasnicima/korisnicima osobnih automobila (čime se stvaraju preduvjeti za donošenje optimalne odluke glede nabave osobnog automobila, te objektivne i utemeljene informacije u kontekstu održavanja osobnog automobila), posljedice čega je u konačnici podizanje razine tehničke ispravnosti osobnog automobila te sigurnosti prometa na cestama.

Dosadašnja istraživanja:

Izrada prediktivnih modela koji se temelje na metodama poslovne inteligencije (rudarenja podataka), rasprostranjena je u nizu znanstvenih područja kao što su medicina, obrazovanje, ekonomija, kriminalistika, itd. U sklopu ovog rada dan je osvrt na neka od provedenih istraživanja s posebnim naglaskom na univerzalni aspekt metodologije izrade prediktivnih modela.

Autor ovog rada u procesu istraživanja literature iz predmetnog područja, uključujući pretragu baze znanstvenih radova *Science Direct*, nije naišao na znanstvene radove iz područja tehničkih pregleda vozila, odnosno znanstvene radove u sklopu kojeg su istraživani prediktivni modeli tehničke ispravnosti vozila temeljem rezultata tehničkih pregleda.

Međutim, provedena su određena stručna istraživanja u kontekstu tehničke ispravnosti vozila, u sklopu kojih su korištene klasične statističke metode.

Bilprovningen, švedska organizacija ovlaštena za provođenje tehničkih pregleda vozila, provela je istraživanje [4] u sklopu kojeg je ocijenjeno 138 tipova osobnih automobila u ovisnosti o zabilježenim tehničkim neispravnostima tijekom tehničkog pregleda, te u ovisnosti o povratnim informacijama korisnika vozila prikupljenih od strane švedske agencije za zaštitu potrošača te direktno od samih korisnika vozila putem internetskog upitnika. Analiziraju se vrijednosti postotnih iznosa tehničkih neispravnosti po pojedinom tipu osobnog automobila unutar sljedećih kategorija tehničkih neispravnosti:
- tehničke neispravnosti kao posljedica tipa osobnog automobila (neispravnosti za koje se s velikom sigurnošću može pretpostaviti da nisu uzrokovane neprimjerenim održavanjem vozila od strane korisnika vozila, odnosno neispravnosti za koje se može pretpostaviti da su posljedica projektiranja vozila ili materijala upotrebljenog u konstrukciji predmetnog sklopa/dijela vozila),
- tehničke neispravnosti kao posljedica neadekvatnog održavanja osobnog automobila od strane korisnika vozila (neispravnosti na čiji nastanak korisnik vozila ima veliki utjecaj),
- tehničke neispravnosti u pogledu ekologije (sastav ispušnih plinova).

Na temelju rezultata tehničkih pregleda svih osobnih automobila određuje se vrijednost tzv. općeg prosječnog osobnog automobila u smislu njegovih karakteristika. Do ocjene pojedinog tipa osobnog automobila dolazi se na osnovu usporedbe postotnih iznosa tehničkih neispravnosti po kategorijama tehničkih neispravnosti tog tipa osobnog automobila s postotnim iznosima tehničkih neispravnosti tzv. općeg prosječnog osobnog automobila.

*ADAC (*njem. *Allgemainer Deutscher Automobilclub)*, njemački nacionalni autoklub, proveo je istraživanja [5], [6], [7] i [8] u svezi s tehničkim neispravnostima na osobnim automobilima (te mini busevima i lakim teretnim vozilima) detektiranim tijekom pružanja intervencija u sklopu pomoći na cesti. Osobni automobili podijeljeni su prema veličini/vrijednosti, te prema godinama proizvodnje, u pet kategorija: mikro automobili, mali automobili, automobili srednje niže klase, automobili srednje klase i automobili

srednje više klase. Tehničke neispravnosti podijeljene su u četrnaest kategorija, u ovisnosti u skupini sklopova/uređaja na kojima su iste detektirane. Na temelju podataka o detektiranim tehničkim neispravnostima na osobnim automobilima tijekom pružanja intervencija u sklopu pomoći na cesti, određuje se prosječna vrijednost osobnog automobila za pojedinu kategoriju osobnog automobila. Svaki pojedini tip osobnog automobila ocjenjuje se na osnovu usporedbe s prosječnom vrijednošću za pojedinu kategoriju osobnog automobila.

Europska komisija (glavno izvršno tijelo koje *de facto* predstavlja Vladu Europske unije) je u srpnju 2012. godine pokrenula inicijativu za promjene u legislativi Europske unije koja se odnosi na područje tehničkih pregleda, pod nazivom "Cestovna sigurnost: Paket valjanosti za sudjelovanje u cestovnom prometu - Stroži tehnički pregled vozila radi spašavanja života" (engl. *Road Safety: The Roadworthiness Package - Tougher vehicle testing rules to save lives*). Predmetna inicijativa za posljedicu bi imala postrožavanje standarda tehničkog pregleda vozila u nizu segmenata (ciklus obveznog tehničkog pregleda vozila, kriteriji za utvrđivanje tehničkih neispravnosti na vozilima, kriteriji za osoblje koje provodi tehničke preglede vozila, kriteriji za uređaje pomoću kojih se provode tehnička ispitivanja na vozilima tijekom tehničkog pregleda, itd.). Za potrebe predmetne inicijative Europska komisija izradila je analizu stanja i postojeće legislative Europske unije iz područja tehničkih pregleda vozila [9], u sklopu koje su izneseni argumenti za postrožavanje standarda. Navedeni argumenti temelje se na sljedećim podacima (područje Europske unije):

- pet mrtvih osoba na europskim prometnicama svakog dana u prometnim nesrećama koje su povezane s tehničkim neispravnostima na vozilima,
- tehničke neispravnosti na vozilima (osim motocikala) uzrokuju 6% prometnih nesreća s procijenjenim posljedicama od 2.000 poginulih osoba godišnje,
- tehničke neispravnosti na motociklima uzrokuju 8% prometnih nesreća u kojima sudjeluju motocikli,
- 10% vozila u prometu u Ujedinjenom Kraljevstvu i SR Njemačkoj u određenom trenutku eksploatacije (vožnje) ima određenu tehničku neispravnost zbog koje predmetno vozilo ne bi "prošlo" tehnički pregled,
- više od 36.000 prometnih nesreća (što je u financijskom kontekstu procijenjeno na 5,6 milijardi Eura) uvjetovano je tehničkim neispravnostima na vozilima.

Europska komisija se u kontekstu prethodno iznesenih podataka poziva na istraživanja provedena od strane *CITA-e* (franc. *Comité international de l'inspection technique automobile)*, međunarodne asocijacije organizacija koje provode tehnički pregled vozila [10], *Ministarstva za transport Ujedinjenog Kraljevstva* [11], *Državnog zavoda za statistiku SR Njemačke* [12], te *DEKRA*-e [13] i [14], itd.

Također, autor ovog rada je u procesu istraživanja literature, uključujući pretragu baze znanstvenih radova *Science Direct,* naišao na nekoliko znanstvenih radova koji se bave tehničkom ispravnošću vozila i ostalim aspektima sigurnosti u širem kontekstu prometa na cestama.

Prema [15], na području Južne Afrike provedeno je istraživanje o utjecaju tehničkih neispravnosti na vozilima na nastanak prometnih nesreća. Izvršena je analiza podataka o prometnim nesrećama prikupljenih od strane nadležnih tijela, te podataka prikupljenih direktnim terenskim istraživanjem putem cestovne kontrole tehničke ispravnosti vozila.

U pogledu istraživanja koje je predmet ovog rada indikativno je istraživanje [16] provedeno od strane katedre za industrijsko inženjerstvo fakulteta (ISIK University, Faculty of Engineering) u Istanbulu, Turska, koje se bavi analizom tehničkih neispravnosti osobnih automobila u funkciji modeliranja jamstvenog roka. U predmetnom istraživanju posebno se ističe pouzdanost kao jedan od najvažnijih atributa proizvoda (automobila) za potencijalne potrošače, prilikom njihovog donošenja odluke o kupnji određenog proizvoda (automobila) između niza različitih alternativnih ponuda. Naglašena je današnja praksa većine proizvođača, koja je uvjetovana intenzivnim konkurentskih uvjetima na tržištu, da za svoje automobile nude značajno dulje jamstvene rokove od onih zakonski minimalno propisanih (od 2 do 3 godine, ovisno o državi u kojoj se prodaja vrši). Većina proizvođača automobila upravo produljenje jamstvenog roka koristi kao marketinško oruđe za prodaju, pa se tako navodi da je proizvođač Hyundai produžio jamstveni rok za svoje automobile namijenjene američkom tržištu s 5 godina/100.000 milja na 10 godina/100.000 milja, te da je proizvođač Honda produžio jamstveni rok za svoje automobile namijenjene turskom tržištu s 3 godina/100.000 km na 6 godina/150.000 km.

Nadalje, u predmetnom istraživanju [16] se konstatira da prodaja rabljenih automobila u većini razvijenih zemalja predstavlja izrazito veliki udio u sklopu ukupnog automobilskog tržišta pa se tako navodi podatak o 11 milijuna prodanih novih automobila (osobni i laki teretni automobil) za razliku od 36,5 milijuna prodanih rabljenih automobila, u Sjedinjenim američkim državama u 2008 godini. Posebno se ističe činjenica da pitanje jamstvenog roka (te pitanje opće pouzdanosti kupnje rabljenog automobila, koja u pravilu nije upitna kod kupnje novog automobila) u kontekstu prodaje rabljenih automobila ima još veći značaj nego u kontekstu prodaje novih automobila, s obzirom da kod prodaje novih automobila postoji zakonska obveza u svezi minimalnog jamstvenog roka, dok kod prodaje rabljenih automobila takva obveza ne postoji. Kod prodaje rabljenih automobila pitanje jamstvenog roka u potpunosti ovisi isključivo o prodajnoj politici prodavača automobila (pod prodavačem automobila podrazumijeva se isključivo pravna osoba koja je registrirana za djelatnost prodaje rabljenih automobila).

Nadalje, u predmetnom istraživanju [16] se posebni značaj pridaje proračunavanju troškova jamstva u funkciji modeliranja optimalnih rokova jamstva, s obzirom na činjenicu da prodavači rabljenih vozila s jedne strane moraju voditi računa o svojoj konkurentnosti na tržištu (što podrazumijeva što je moguće dulji jamstveni rok za prodani automobil), a s druge strane moraju voditi računa o isplativosti poslovanja. U tom kontekstu navodi se niz istraživanja [17], [18], [19], [20] i [21] koja se bave analizom financijskog i tehničkog aspekta (tehnički aspekt podrazumijeva nastanak neispravnosti na proizvodu, u funkciji starosti i razine eksploatacije proizvoda) predmetne problematike. Nadalje se navodi nekoliko modela (politika) jamstava koji se primjenjuju u praksi te koji podrazumijevaju različite opcije na relaciji prodavač vozila - potrošač. Ključni aspekt za optimalno proračunavanje troškova jamstva rabljenih automobila, a time posljedično i za optimalno modeliranje rokova jamstva, predstavlja razina vjerodostojnosti, odnosno kvaliteta podataka o tehničkim neispravnostima na automobilima, u funkciji starosti i razine eksploatacije automobila.

U konačnici, autor ovog rada je u procesu istraživanja literature, uključujući pretragu baze znanstvenih radova *Science Direct,* naišao na nekoliko znanstvenih radova iz drugih znanstvenih područja (medicina, ekonomija, kriminalistika, obrazovanje, itd.) u

sklopu kojih je obrađena metoda stabla odlučivanja i CHAID algoritam. Predmetni znanstveni radovi su opisani u sklopu ovog rada, a posebno u točci 3.3.1.2.

Metodologija i plan istraživanja primijenjeni u ovom radu:

Tijekom istraživanja postavljenog problema i izrade ovog rada upotrijebljene su metode:

- komparacije (uspoređivanje prikupljenih podataka i informacija),
- analize i sinteze (analiziranje prikupljenih podataka i informacija, te interpretacija dobivenih rezultata),
- kompilacije (korištenje poznatih matematičkih algoritama te programskih aplikacija),
- modeliranje (izrađivanje prediktivnog modela tehničke ispravnosti vozila/osobnih automobila).

Plan istraživanja u provedbenom aspektu podrazumijeva tri temeljne faze istraživanja:

1.faza - metodom Pareto analize primijenjenom na rezultate tehničkih pregleda onih osobnih automobila kod kojih su u navedenom razdoblju utvrđene tehničke neispravnosti radi čega su ista zabilježila negativan rezultat na tehničkom pregledu (cca 310.000 osobnih automobila), došlo se do spoznaje o trgovačkim markama i tipovima osobnih automobila koji dominiraju u kontekstu utvrđene tehničke neispravnosti.

2.faza – uporabom metoda poslovne inteligencije (rudarenje podataka, stabla odlučivanja, CHAID algoritmi, neuronske mreže) na rezultate tehničkih pregleda onih trgovačkih marki i tipova osobnih automobila koji su u sklopu 1. faze metodom Pareto analize utvrđeni kao ključni, došlo se do predikcije tehničke ispravnosti ciljane skupine osobnih automobila u budućem razdoblju.

3.faza – interpretacija rezultata istraživanja – spoznaja o kritičnim trgovačkim markama i tipovima osobnih vozila, odnosno spoznaja o parametrima kritičnosti glede tehničke neispravnosti.

Očekivani znanstveni doprinos istraživanja:

Očekivani znanstveni doprinos istraživanja podrazumijeva dokazivanje potencijala primjenjivosti izrađenog modela u predikciji tehničke ispravnosti osobnih automobila, temeljem rezultata tehničkih pregleda i tehničkih karakteristika osobnih automobila, a radi podizanja razine sigurnosti prometa na cestama. Očekuje se da će rezultati istraživanja pokazati da postoji prepoznatljiv odnos između promatranih podataka, a primjena odabranih metoda daje uvid u potencijal njihove primjene u kontekstu sigurnosti prometa na cestama te omogućava jasniju spoznaju o korelacijama područja koja su predmet ovog istraživanja.

2. TEHNIČKI PREGLEDI VOZILA

2.1. SUSTAV TEHNIČKIH PREGLEDA VOZILA U RH

Tehnički pregled vozila, kao djelatnost od općeg interesa, propisana je člankom 255. *Zakona o sigurnosti prometa na cestama* (Narodne novine 67/2008, 74/2011, 80/2013, 92/2014) (u daljnjem tekstu – Zakon) [22], dok je kao nadležno tijelo državne uprave za predmetnu djelatnost određeno Ministarstvo unutarnjih poslova (u daljnjem tekstu – *MUP*).

Ministar unutarnjih poslova, u suglasnosti s ministrom nadležnim za poslove prometa donosi propise o tehničkim pregledima vozila. Temeljni podzakonski propis iz područja tehničkih pregleda vozila predstavlja *Pravilnik o tehničkim pregledima vozila* (Narodne novine 148/2008, 36/2010, 52/2013, 111/2014, 122/2014).

Također, ministar unutarnjih poslova, na osnovu rezultata javnog natječaja a temeljem članka 273. *Zakona*, donosi rješenje o ovlašćivanju stručne organizacije za poslove organiziranja i jedinstvenog provođenja tehničkih pregleda vozila, organiziranja poslova registracije vozila u stanicama za tehnički pregled vozila (u daljnjem tekstu - STP) i izdavanja pokusnih pločica, kao javne ovlasti.



Slika 2.1. Shematski prikaz sustava tehničkog pregleda vozila u RH

Centar za vozila Hrvatske d.d. (u daljnjem tekstu - *CVH*) i *Hrvatski autoklub* (u daljnjem tekstu - *HAK*) predstavljaju stručnu organizaciju iz članka 273. *Zakona* temeljem *Rješenja*[1] ministra unutarnjih poslova.

Na slici 2.1. prikazana je temeljna legislativa i nositelji sustava tehničkih pregleda u RH.

CVH i *HAK* su, temeljem članka 273. *Zakona,* donijeli niz općih akata kojim su uredili odnose od interesa za jedinstveni sustav tehničkih pregleda i registracije vozila:

- *Pravilnik o stručnom nadzoru nad radom stanica za tehnički pregled vozila* (Narodne novine 33/2009, 1/2012, 6/2012),
- *Pravilnik o normativima rada i jedinstvenim elementima vrednovanja rezultata rada u stanicama za tehnički pregled vozila* (Narodne novine 33/2009),
- *Pravilnik o uvjetima koje mora ispunjavati stanica za tehnički pregled vozila* (Narodne novine 33/2009),
- *Pravilnik o informatičkom sustavu i jedinstvenomu programskom rješenju za obradu podataka o tehničkim pregledima i registraciji vozila u stanicama za tehnički pregled vozila* (Narodne novine 33/2009).

Sukladno odredbama općih akata *HAK* obavlja poslove stručnog nadzora nad radom STP, dok *CVH* obavlja ostale stručno-tehničke poslove.

MUP obavlja upravni i inspekcijski nadzor nad radom STP sukladno članku 268. Zakona, te nadzor nad zakonitošću općih akata kao i upravni i inspekcijski nadzor nad radom stručne organizacije sukladno članku 273. Zakona.

U sustavu tehničkih pregleda vozila u 2015. godini posluje 169 STP (146 STP u 2009. i 2010. godini).

[1] *Rješenje o davanju javnih ovlasti stručnim organizacijama za obavljanje poslova organiziranja i jedinstvenog provođenja tehničkih pregleda vozila, organiziranja poslova registracije vozila u stanicama za tehničke preglede vozila i izdavanja pokusnih pločica*, od dana 05. kolovoza 1993. godine, broj: 511-01-72-7693/7-93.

Sukladno *Pravilniku o tehničkim pregledima vozila* [23], tehnički pregled vozila mogu obavljati samo ovlašteni stručni i kvalificirani ispitivači, koristeći za to propisanu opremu i uređaje u za tu namjenu posebno izgrađenom i opremljenom radnom prostoru – STP-u. STP je posebno izgrađena i opremljena tehničko-tehnološka cjelina, u kojoj stručni i ovlašteni djelatnici obavljaju poslove tehničkog pregleda vozila.

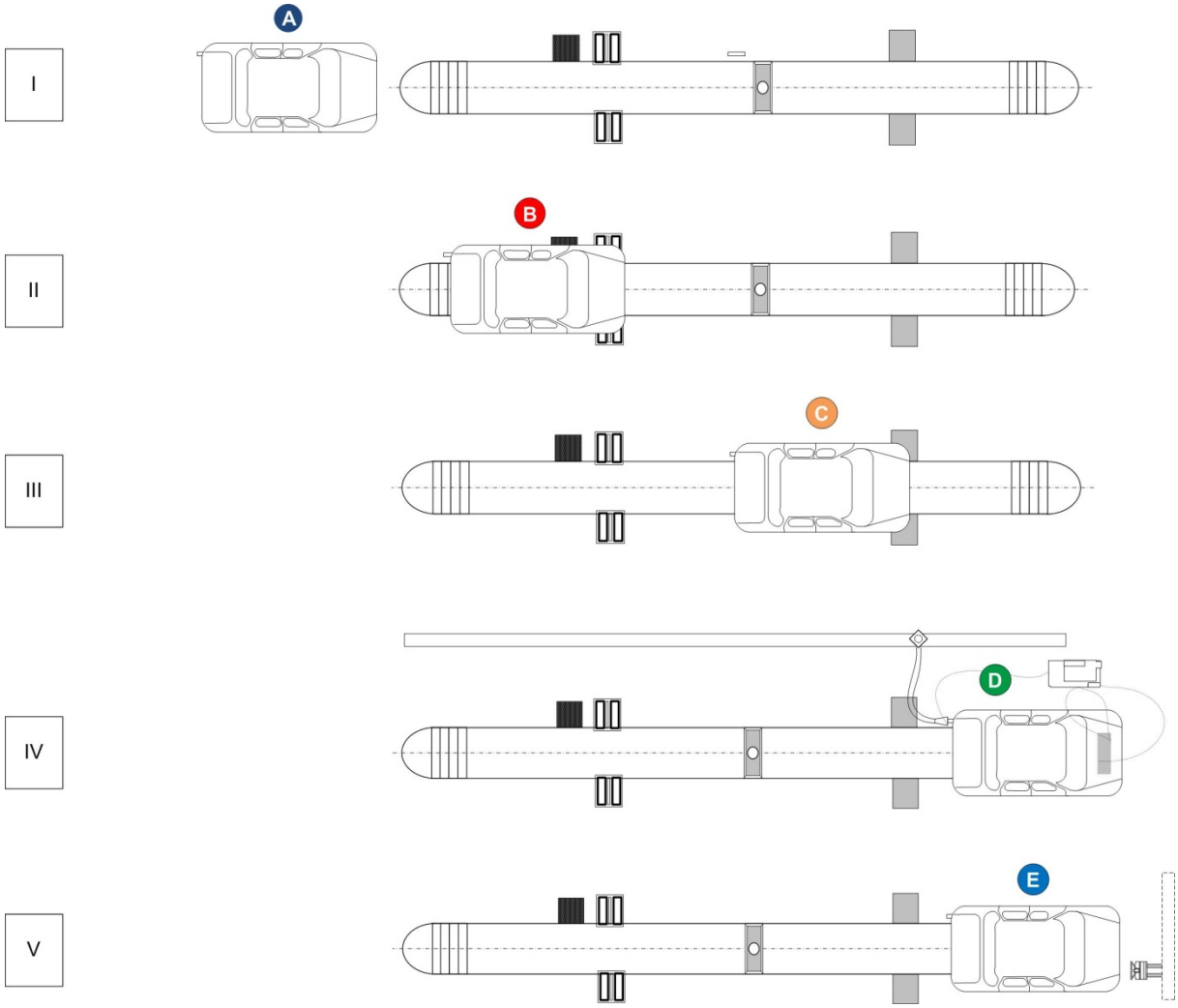
Tehnički pregled može se obavljati izvan STP samo za traktore i vozila posebnih tehničkih karakteristika (dimenzije, masa, itd.), te na otocima (ako na otoku ne postoji stanica za tehnički pregled vozila), uz prethodno odobrenje nadležne policijske uprave.

Tehnički pregledi vozila mogu biti redovni, izvanredni i preventivni. Redovni tehnički pregled vozila predstavlja unaprijed propisani slijed postupaka ispitivanja ispravnosti sklopova i dijelova vozila. Ostali tehnički pregledi vozila, izvanredni i preventivni, obavljaju se na istovjetan način kao i redovni tehnički pregled vozila, ali samo na onim sklopovima, odnosno dijelovima vozila koji podliježu izvanrednom ili preventivnom tehničkom pregledu vozila.

Za potrebe ovog rada, predmet istraživanja su redovni tehnički pregledi osobnih automobila (kategorija M1[2]), izvršeni u razdoblju od 01.01. do 31.12.2009. godine, odnosno u razdoblju od 01.01. do 31.12.2010. godine.

Na slikama 2.2. i 2.3. prikazano je kretanje osobnog automobila po tehnološkoj liniji stanice za tehnički pregled vozila tijekom tehničkog pregleda, te pojedine faze predmetnog procesa. Valja napomenuti da se u izloženom primjeru radi o reprezentativnom tehničkom pregledu osobnog automobila, odnosno najčešćem tijeku tehničkog pregleda osobnog automobila u smislu redoslijeda provođenja pojedinih faza, odnosno radnji. Redoslijed provođenja pojedinih faza, odnosno radnji tehničkog pregleda osobnog automobila (kao i tehničkih pregleda drugih vrsta vozila) može biti različit u različitim stanicama za tehnički pregled vozila, kao i raspored ispitnih uređaja i opreme na tehnološkoj liniji.

[2] Sukladno članku 2. *Pravilnika o tehničkim uvjetima vozila u prometu na cestama* (Narodne novine 51/2010, 84/2010, 145/2011, 140/2013, 85/2014) osobni automobil kategorije M1 jest motorno vozilo za prijevoz osoba koje osim sjedala za vozača ima još najviše 8 sjedala.

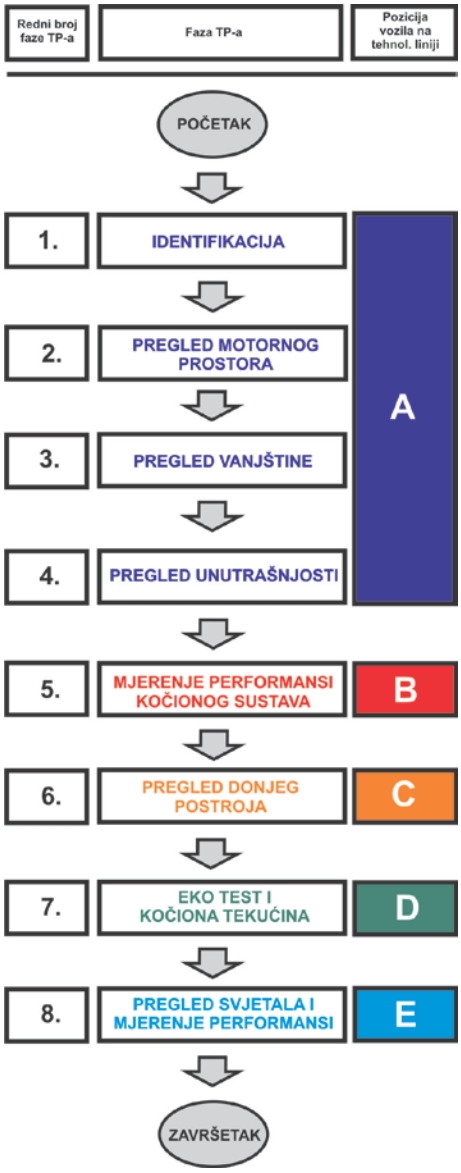


Slika 2.2. Shematski prikaz kretanja osobnog automobila po tehnološkoj liniji

Tehnički pregled osobnog automobila može se podijeliti u osam faza – identifikaciju (A), pregled motornog prostora (A), pregled vanjštine (A), pregled unutrašnjosti (A), mjerenje performansi kočnog sustava (B), pregled donjeg postroja (C), provođenje EKO testa i kontrola kočne tekućine (D), te pregled i mjerenje performansi svjetala (E).

Tijekom redovnog tehničkog pregleda osobnog automobila obvezno se kontrolira ispravnost uređaja/sklopova (17 cjelina) i njihovih dijelova (134 pozicije).
Činjenica koja je specifična za sustav tehničkih pregleda RH i koja se direktno reflektira na postavke istraživanja koje je predmet ovog rada, kako u idejnom tako i u provedbenom aspektu, proizlazi iz odredbi članka 21. *Pravilnika o tehničkim pregledima vozila* [23] u kojima je propisano da neispravnost uređaja/sklopova,

odnosno njihovih dijelova, predstavlja razlog za ne-prolazak (“pad”) predmetnog vozila (osobnog automobila) na tehničkom pregledu.



Slika 2.3. Slijed faza procesa tehničkog pregleda vozila

Specifičnost navedenih odredbi očituje se u činjenici da tehničke neispravnosti na uređajima/sklopovima, odnosno njihovim dijelovima koje su razlog za ne-prolazak (“pad”) predmetnog vozila (osobnog automobila) na tehničkom pregledu nisu stupnjevane, odnosno bilo kakva tehnička neispravnost (tehnička neispravnost bilo kojeg stupnja) uvjetuje “pad” predmetnog vozila (osobnog automobila) na tehničkom pregledu.

Nadalje, istim odredbama navedenog Pravilnika propisano je da vozilo (osobni automobil) kod kojeg je tehnička neispravnost utvrđena na bilo kojem od 134 dijelova uređaja/sklopova, osim na parkirnim svijetlima, svijetlima registracijske pločice i žutim rotacijskim ili treptavim svijetlima (drugim riječima, praktički na svim dijelovima uređaja/sklopova koji se na vozilu/osobnom automobilu kontroliraju), za posljedicu ima "pad" na tehničkom pregledu.

Iako se općenito može konstatirati da je tehnički pregled vozila u RH u smislu operativnog procesa usklađen s legislativom Europske unije koja propisuje predmetno područje to nije slučaj u kontekstu definiranja tehničkih neispravnosti, odnosno definiranja situacija koje za posljedicu imaju ne-prolazak ("pad") vozila na tehničkom pregledu.

Naime, u sklopu legislative Europske unije koja se odnosi na područje tehničkih pregleda vozila tehničke neispravnosti uređaja/sklopova odnosno njihovih dijelova su stupnjevane u ovisnosti o karakteristikama istih u kontekstu procijenjene razine utjecaja na sigurnost prometa na cestama, o čemu će posebno biti riječi u nastavku ovog rada.

Odredbama članka 256. *Zakona* [22] propisan je obvezni ciklus redovnog tehničkog pregleda vozila (tzv. 0/2/1/1/1... ciklus). Nova vozila moraju biti podvrgnuta tehničkom pregledu tijekom mjeseca u kojem istječe rok od 24 mjeseca od prvog tehničkog pregleda ("nulta" godina). Vozila stara dvije ili više godina moraju biti podvrgnuta tehničkom pregledu tijekom svakog 12. mjeseca od posljednjeg redovnog tehničkog pregleda.

2.2. TEHNIČKI PREGLEDI VOZILA U KONTEKSTU ZAKONODAVSTVA EUROPSKE UNIJE

Propisi Europske unije koji se odnose na područje tehničkih pregleda podrazumijevaju dvije cjeline:

- „klasični“ tehnički pregled vozila koji podrazumijeva kontrolu tehničke ispravnosti vozila u stanicama za tehnički pregled vozila,
- cestovnu kontrolu tehničke ispravnosti gospodarskih vozila.

Načelno, postoje dvije vrste propisa – direktive (engl. *Directives*) i preporuke (engl. *Recommendations*).

Direktive se dijele u dvije kategorije:

1. Osnovne direktive Europskog parlamenta i Vijeća Europske unije,
2. Dopunske (amandmanske) direktive Europske komisije.

Na slici 2.4. prikazana je dinamika donošenja propisa iz područja „klasičnog“ tehničkog pregleda vozila.

Osnovna direktiva koja je trenutno na snazi (navedenoj direktivi su iz novijeg doba prethodile osnovne direktive iz 1996. i 2009. godine) iz područja „klasičnog“ tehničkog pregleda vozila jest direktiva iz 2014. godine *(engl. Directive 2014/45/EU of the European Parliament and of the Council of 3 April 2014 on periodic roadworthiness tests for motor vehicles and their trailers and repealing Directive 2009/40/EC)*.

Osnovna direktiva iz 1996. godine nadopunjavana je nizom dopunskih direktiva i to uglavnom u kontekstu ispitivanja štetnih sastojaka ispušnih plinova:

- *Commission Directive 1999/52/EC of 26 May 1999,*
- *Commission Directive 2001/9/EC of 12 February 2001,*
- *Commission Directive 2001/11/EC of 14 February 2001,*
- *Commission Directive 2003/27/EC of 3 April 2001.*

Usvajanjem osnovne direktive iz 2009. godine *(Directive 2009/40/EC of the European Parlament and of then Council of 6 May 2009 on roadworthiness tests for motor vehicles and their trailers)* sve prethodne direktive stavljene su van snage. Osnovna direktiva iz 2009. godine nadopunjena je dopunskom direktivom:

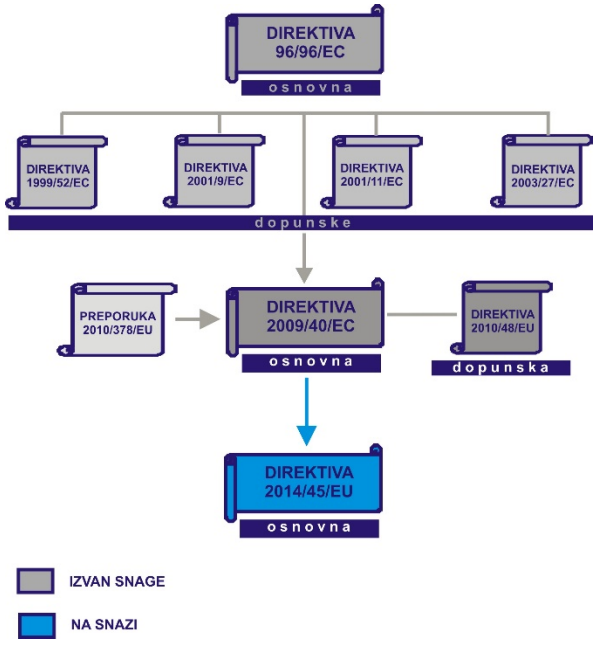
- *Commission Directive 2010/48/EU of 5 July 2010 adapting to technical progress Directive 2009/40/EC of the European Parlament and of the Council roadworthiness tests for motor vehicles and their trailers.*

Također, osim navedene dopunske direktive, Europska komisija je po prvi puta bila donijela i preporuku, kojom se na operativnoj razini razrađivala predmetna direktiva:

- *Commission Recommendation 2010/378/EU of 5 July 2010 on the assessment of defects during roadworthiness testing in accordance with Directive 2009/40/EC of the European Parlament and of the Council of 6 May 2009 on roadworthiness tests for motor vehicles and their trailers.*

Osnovnom direktivom *2009/40/EC,* te njezinom nadopunom *2010/48/EU,* Europska komisija bila je definirala proces tehničkog pregleda vozila u provedbenom smislu, te su bile utvrđene stavke (u smislu uređaja/sklopova i njihovih dijelova) na vozilu koje podliježu obveznoj kontroli tijekom tehničkog pregleda.

Putem preporuke *2010/378/EU* Europska komisija je bila definirala smjernice za ocjenjivanje (stupnjevanje) nedostataka (u tri stupnja značaja) odnosno tehničkih neispravnosti utvrđenih na vozilima u sklopu kontrole uređaja/sklopova i njihovih dijelova tijekom tehničkog pregleda.



Slika 2.4. Shematski prikaz slijeda direktiva EU iz domene tehničkih pregleda vozila

Usvajanjem nove osnovne direktive iz 2014. godine *(Directive 2014/45/EU of the European Parliament and of the Council of 3 April 2014 on periodic roadworthiness tests for motor vehicles and their trailers and repealing Directive 2009/40/EC)* [24] sve prethodne direktive stavljene su van snage.

Direktiva iz 2014. godine donijela je niz značajnih novosti u sustav tehničkih pregleda vozila Europske unije, te su u sklopu iste definirani i krajnji rokovi do kojih su države članice Europske unije predmetne novosti obvezne implementirati u nacionalna zakonodavstva u području tehničkih pregleda vozila. Neke od značajnijih novosti su sljedeće:

- uvodi se obvezan tehnički pregled za traktore (T5 kategorija) čija je najveća brzina veća od 40 km/h i koji se uglavnom koriste na javnim cestama u komercijalne svrhe cestovnog prijevoza tereta,
- uvodi se mogućnost slobodnog izbora metoda ispitivanja ispušnih plinova (metoda sondom ili OBD metoda) za vozila s Euro 6 motorima,
- omogućava se državama članicama da u sklopu nacionalnih zakonodavstava propišu obvezan tehnički pregled vozila u situacijama kada su nakon prometnih nesreća oštećeni glavni sigurnosni dijelovi vozila (kotači, ovjes, zračni jastuci,

upravljač, kočnice, itd.), kada su sigurnosni i okolišni sustavi i dijelovi vozila promijenjeni ili preinačeni, kada dođe do promjene imatelja potvrde o registraciji vozila, kada vozila dosegnu prijeđenu kilometražu veću od 160.000 km, odnosno u situacijama kada je sigurnost prometa na cestama ugrožena,

- propisuje se obvezna kategorizacija utvrđenih tehničkih neispravnosti na vozilu po skupinama, u skladu s njihovim intenzitetom (stupnjem značaja) odnosno u skladu s razinom njihovog negativnog utjecaja na sigurnost prometa na cestama,
- propisuje se obvezno uvođenje sustava obučavanja osoblja stanica za tehnički pregled vozila (nadzornici tehničke ispravnosti vozila).

Krajnji rok za provedbu prethodno navedenih novosti je 20. svibanj 2018.

- obvezuje se Europska komisija glede izrade analize o uvođenju obveznog tehničkog pregleda lakih prikolica i vozila s dva i tri kotača.

Krajnji rok za provedbu prethodno navedene novosti je 30. travanj 2019.

- obvezuje se Europska komisija glede izrade europske elektroničke platforme s tehničkim informacijama o vozilima (podaci o tehničkim pregledima, prijeđenoj kilometraži, homologaciji, itd.) s ciljem razmjene podataka u sklopu koje će se i proizvođači vozila obvezati za davanje takvih podataka besplatno ili po razumnoj cijeni.

Krajnji rok za provedbu prethodno navedene novosti je 30. travanj 2020.

- uvodi se sustav elektroničkog prijenosa rezultata tehničkih ispitivanja.

Krajnji rok za provedbu prethodno navedene novosti je 20. svibanj 2021.

- propisuju se minimalni zahtjevi za objekte stanica za tehnički pregled vozila, uređaje i opremu,
- propisuje se obvezno uvođenje sustava stručnog nadzora nad radom stanica za tehnički pregled vozila i minimalni zahtjevi za nadzorna tijela.

Krajnji rok za provedbu prethodno navedenih novosti je 20. svibanj 2023.

Za potpuno sagledavanje zakonodavnog sustava Europske unije potrebno je uzeti u obzir činjenicu da legislativa Europske unije, konkretno, direktive Europskog parlamenta i Vijeća propisuju minimalne zahtjeve koje pojedina država članica mora ispuniti u kontekstu predmetnog područja, u ovom slučaju tehničkih pregleda i registracije vozila, dok je istovremeno, svakoj državi članici ostavljen prostor da u sklopu svog nacionalnog zakonodavstva nadogradi, odnosno postroži univerzalne minimalne zahtjeve, uvažavajući svoje specifične okolnosti, tradiciju i interese (može se utvrditi da se takva politika čak i potiče od strane zakonodavstva Europske unije).

Što se tiče utjecaja svih prethodno navedenih novosti europskog zakonodavstva u području tehničkih pregleda vozila na hrvatski sustav tehničkih pregleda vozila valja konstatirati da je većina tih novosti odavno implementirana kako u zakonodavstvu tako i u praksi Republike Hrvatske.

Od svih navedenih značajnih novosti koje je donijela direktiva *2014/45/EU* u hrvatski sustav tehničkih pregleda vozila još uvijek nije implementirana kategorizacija utvrđenih tehničkih neispravnosti na vozilu s obzirom na procijenjeni potencijalni utjecaj tehničkih neispravnosti na sigurnost vozila u pogledu prometa na cestama (riječ je o identičnoj kategorizaciji koja je prvotno bila definira u sklopu preporuke *2010/378/EU* Europske komisije),

Predmetna kategorizacija tehničkih neispravnosti podrazumijeva tri kategorije, odnosno tri stupnja značaja tehničkih neispravnosti na vozilu utvrđenih tijekom tehničkog pregleda:

- **Manje tehničke neispravnosti (engl. *Minor Defects* – MiD)**

  Tehničke neispravnosti koje nemaju presudni utjecaj na sigurnost vozila i druge manje neispravnosti. Vozilo ne treba nužno ponovno pregledati (u sklopu tehničkog pregleda), s obzirom da je razumno očekivati da će utvrđene neispravnosti biti otklonjene bez odgađanja.

- **Veće tehničke neispravnosti (engl. *Major Defects* – MaD)**

  Tehničke neispravnosti koje mogu umanjiti sigurnost vozila ili ugroziti ostale sudionike u prometu, kao i ostale značajnije neispravnosti. Za daljnju uporabu vozila u prometu bez otklanjanja utvrđenih neispravnosti važe posebni uvjeti. Nadležna tijela država članica moraju donijeti postupak za utvrđivanje uvjeta pod kojima vozilo smije sudjelovati u prometu do ponovljenog uspješnog tehničkog pregleda.

- **Opasne tehničke neispravnosti (engl. *Dangerous Defects* – DD)**

  Tehničke neispravnosti koje predstavljaju neposrednu opasnost po sigurnost cestovnog prometa na način da vozilo ni pod kojim uvjetima ne smije sudjelovati u cestovnom prometu. Drugim riječima, utvrđivanje ove kategorije neispravnosti podrazumijeva trenutno isključivanje iz prometa predmetnog vozila na način da se istom "skinu" registarske pločice na licu mjesta, u stanici za tehnički pregled vozila.

U tijeku je implementiranje predmetne kategorizacije u sustav tehničkih pregleda vozila Republike Hrvatske i njena puna implementacija se očekuje i prije krajnjeg zakonskog roka, odnosno prije 20. svibnja 2018.

Nadalje, osnovnom direktivom *2014/45/EU* Europska komisija propisala je obvezni minimalni ciklus redovnog tehničkog pregleda vozila, u ovisnosti o kategoriji vozila. Za vozila kategorije M1 (osobni automobil) propisan je minimalan ciklus - tzv. 4/2/2/2... ciklus, odnosno nova vozila moraju biti podvrgnuta tehničkom pregledu četiri godine nakon prve registracije, te nakon toga svake dvije godine.

Svakoj državi članici ostavljen je prostor da uvažavajući svoje specifičnosti (tradicija, stanje sigurnosti cestovnog prometa, struktura voznog parka, itd.) propiše učestaliji ciklus obveznog redovnog tehničkog pregleda.

3. POSLOVNA INTELIGENCIJA

3.1. UVOD U POSLOVNU INTELIGENCIJU

Moderno poslovanje, koje je obilježeno strelovitim razvojem informatičke tehnologije tijekom protekla dva desetljeća, uvjetovano je dinamikom tržišne utakmice, koja je svakim danom sve intenzivnija.

Oduvijek je uspješno poslovanje bilo posljedica kontinuiranog donošenja dobrih poslovnih odluka a temeljem pouzdanih informacija koje su dobivane iz niza (ulaznih) podataka.

Međutim, okolnosti poslovanja u praktički svakom kutku današnjeg svijeta su se dramatično promijenile. Iako prethodno navedeno načelo uspješnog poslovanja u suštini vrijedi i danas, intenzivnost aktualne tržišne utakmice očituje se u potrebi donošenja ispravnih poslovnih odluka u što je moguće kraćem vremenu (često u ekstremno kratkom roku), ali temeljenih na informacijama često dvojbene pouzdanosti, dobivenih iz ogromnih količina često nestrukturiranih i heterogenih podataka. Globalizacijom svijeta, odnosno otvaranjem novih tržišta - Kine, Indije, Rusije, Brazila, itd. – stvorena je višestruka alternativa u smislu konkurencije donedavno dominirajućim tržištima tzv. Zapadne Europe i SAD-a, praktički u svakom aspektu poslovanja, što za direktnu posljedicu ima radikalizaciju globalne tržišne utakmice, koja je u kontekstu svjetske gospodarske recesije još izraženija.

U takvim nepovoljnim okolnostima djelovanja pogotovo postoji potreba za poslovnim pristupom koji jamči veliku brzinu rješavanja novonastalih problema, odnosno veliku brzinu adaptacije na novonastale uvjete, veliku brzinu učenja novih znanja i sposobnost stvaranja novih ideja.

Pod pojmom “poslovne inteligencije” (engl. *Business Intelligence* – BI) podrazumijeva se upravo takav poslovni pristup. U literaturi koja se bavi ovim područjem postoji nekoliko različitih definicija pojma poslovne inteligencije.

Prema [25] poslovna inteligencija predstavlja proces u sklopu kojeg se podaci iz internih i eksternih izvora integriraju u svrhu generiranja relevantnih informacija krucijalnih za donošenje poslovnih odluka, koje za konačnu posljedicu imaju unaprjeđenje strateške poslovne pozicije na tržištu, odnosno stvaranje prednosti na istom. Drugim riječima radi se o procesu u sklopu kojeg informacije proizašle iz prikupljenih podataka, koji su obrađeni raznim alatima poslovne inteligencije, dobivaju "stratešku" poslovnu dimenziju. Iz te perspektive proizlazi pojam tzv. inteligentne tvrtke, odnosno one tvrtke koja koristi koncept poslovne inteligencije u cilju donošenja bržih i mudrijih poslovnih odluka nego što to čini njezina konkurencija. U tom kontekstu ključna je validacija informacija, a vrijednost određene informacije proteže se u kontinuumu, dok se njezina uporaba širi interno u tvrtki te eksterno prema okruženju tvrtke (klijenti, partneri, itd.). Vrijednost određene informacije raste s kvadratom broja korisnika koji s njom raspolažu, umnoženo s brojem poslovnih područja u sklopu kojih korisnici djeluju. Kvadrat broja korisnika proizlazi iz tzv. Metcalfe-ovog zakona mreže prema kojem je vrijednost (računalne) mreže direktno proporcionalna kvadratu njenih korisnika.

Prema [26], [27], [28] i [29] posebno se ističe tehnološki aspekt poslovne inteligencije u smislu da koncept poslovne inteligencije podrazumijeva uporabu niza alata i tehnika za prikupljanje, prijenos, pohranjivanje, analitičku obradu i prezentiranje podataka, odnosno proizašlih informacija, na način da predmetne informacije omogućavaju donošenje optimalnih poslovnih odluka.

Prema [9] koncept poslovne inteligencije podrazumijeva četiri temeljne dimenzije:

- poslovna dimenzija – podrazumijeva određivanje upravljačkih metoda i tehnika u kontekstu stjecanja znanja,

- funkcionalna dimenzija – podrazumijeva određivanje funkcionalnog opsega subjekta u kojem se provodi koncept poslovne inteligencije,

- tehnološka dimenzija – podrazumijeva određivanje informacijskih alata i metoda za izgradnju koncepta poslovne inteligencije,

- organizacijska dimenzija – podrazumijeva metodologiju implementacije koncepta poslovne inteligencije.

Prema [29], [30], [31] i [32] poslovna inteligencija predstavlja ranije prikriveno znanje koje se otkriva iz operativnih, rutinski prikupljanih podataka primjenom odgovarajućih računsko-logičkih metoda, podržanih informatičkom tehnologijom. Posljedica primjene koncepta poslovne inteligencije nije stvaranje veće količine podataka, odnosno informacija dobivenih njihovom obradom, već isključivo generiranje boljih, kvalitetnijih podataka iz kojih proizlaze pouzdane informacije potrebne za donošenje optimalnih poslovnih odluka.

Nadalje, primijenjeni koncept poslovne inteligencije pravovremeno pruža korisnicima samo one informacije koje su im potrebne te ih iskazuje na način koji najviše odgovora korisnicima. Drugim riječima, koncept poslovne inteligencije osigurava okvir za reduciranje kvantitete u kvalitetu, te na taj način potiče pozitivne suštinske promjene u okruženju u kojem se primjenjuje.

Pojednostavljeno, koncept poslovne inteligencije temelji se na načelu: pravi podaci + pravo vrijeme + prava osoba (pravovremeno obrađeni kvalitetni podaci interpretirani od strane kompetentne osobe).

Prema [30], [31] i [32] uočavaju se četiri osnovne kategorije kvalitete podataka koje imaju direktni utjecaj na donošenje poslovnih odluka:

- standardiziranost – podrazumijeva ujednačenost u izrazu te standardizaciju dokumenata nositelja poslovnih informacija,

- podudarnost – podrazumijeva identifikaciju podudarnih podataka iz brojnih izvora i njihovu eliminaciju,

- verificiranost – podrazumijeva postupak usporedbe nekog podatka s poznatim izvorom kako bi se utvrdila relevantnost tog podatka,

- potencijal proširenja – podrazumijeva mogućnost dodavanja novih podataka postojećem skupu podataka.

U literaturi koja se bavi područjem poslovne inteligencije, a koja je proučavana u sklopu ovog rada, navode se razni procesi poslovne inteligencije. Međutim, svim tim izvorima zajednička je činjenica da proces rudarenja podataka ističu kao krucijalni.

3.2. METODE OTKRIVANJA ZNANJA

3.2.1. Rudarenje podataka

Pod pojmom rudarenja podataka (engl. *Data Mining*) se, prema [30], podrazumijeva proces pronalaženja zakonitosti između velikog broja podataka (koji su najčešće nestrukturirani), odnosno sustava, interaktivan i iterativan proces izvođenja i prikazivanja korisnog, implicitnog i inovativnog znanja iz podataka.

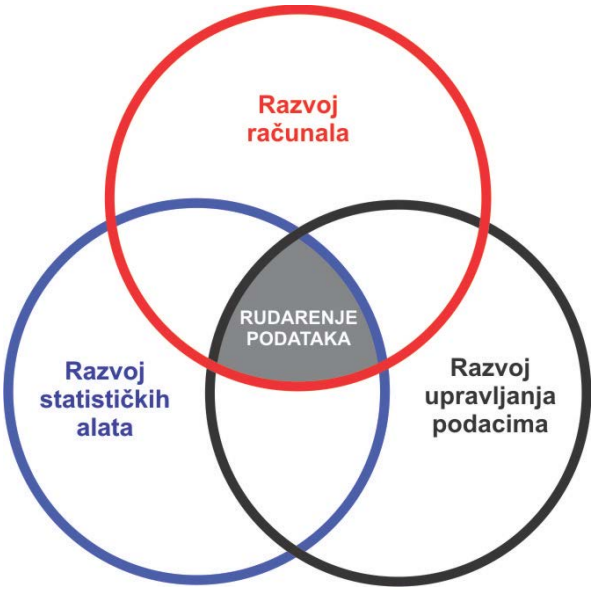
Odnosi između podataka u smislu zakonitosti, trendova i matrica ponašanja u pravilu su skriveni, te je njihovo otkrivanje u velikim količinama podataka praktički nemoguće uporabom tradicionalnih statističkih metoda obrade podataka kao što su, na primjer, normalna, binomna i Poissonova razdioba, razni oblici regresijske analize, itd.

Značaj rudarenja podataka proizlazi i iz činjenice da se radi o univerzalnom procesu primjenjivom u širokom rasponu područja interesa – od pravnih područja preko industrijskih procesa do astronomije i medicine. Dapače, sve metode rudarenja podataka proizašle su iz različitih područja kao što su statistika, teorija informacija, umjetna inteligencija, itd. Njihov inicijalni razvoj zabilježen je u zadnjim desetljećima dvadesetog stoljeća kada je i definirana sintagma – rudarenje podataka.

U skladu s prethodno navedenim, prema [33], rudarenje podataka predstavlja interdisciplinarni istraživački proces u sklopu kojeg se podrazumijeva integriranje različitih metodologija.

Prema [34] metode rudarenja su rezultat dugog procesa istraživanja i razvoja statističkih aplikacija. Taj proces započeo je s prvim slučajevima pohranjivanja poslovnih podataka u računala, a nastavio se usporedno s poboljšanjima u pristupu podacima, te, u zadnje vrijeme, s generiranjem tehnologija koje korisnicima omogućuju navigaciju kroz podatke u realnom vremenu. Ključni čimbenik glede operativne primjene procesa rudarenja podataka u realnom (praktičnom) okruženju proizlazi iz činjenice da su tri temeljne tehnologije dosegle odgovarajuću razinu razvoja, što je prikazano na slici 3.1.:

- tehnologija za masovno prikupljanje podataka (razvoj upravljanja podacima),
- tehnologija izrade "snažnih" multiprocesorskih računala (razvoj računala),
- tehnologija izrade algoritamskih programskih aplikacija za rudarenje podataka (razvoj statističkih alata).



Slika 3.1. – Okvir razvoja procesa rudarenja podataka [34]

Upravo algoritamske programske aplikacije omogućuju potpunu i kompleksnu matematičku, financijsku i statističku obradu podataka radi analize korelacija, trendova, itd., a u kontekstu predikcije budućih događaja i procesa, odnosno otkrivanja Skrivenih znanja iz raspoloživih podataka.

Međutim, za otkrivanje prikrivenog znanja nije presudna moćna tehnologija dostupna u današnje doba nego ispravna interpretacija dobivenih rezultata, od strane stručnjaka iz predmetnog područja, što je jedan od ključnih čimbenika načela poslovne inteligencije.

U literaturi koja se bavi područjem rudarenja podataka definirana su dva tipa rudarenja podataka. Prema [30] i [35] razlikuje se rudarenje podataka u kontekstu verifikacije određene hipoteze (usmjereno rudarenje podataka), te rudarenje podataka u kontekstu otkrivanja novih znanja. Kod rudarenja podataka u kontekstu verifikacije određene hipoteze cilj je izvršiti provjeru glede utemeljenosti određene ideje ili dojma u pogledu značaja korelacija između određenih podataka, dok je kod rudarenja podataka u kontekstu otkrivanja novih znanja cilj pronaći korelacije između određenih podataka za koje se prvotno ne pretpostavlja da postoje.

Prema [35] područje primjene rudarenja podataka proteže se kroz široki raspon djelatnosti i to u vidu šest skupina operacija:

- klasifikacija (tip: usmjereno rudarenje podataka) – podrazumijeva ispitivanje karakteristika podataka te njihovo pozicioniranje u jednu od predifiniranih klasa,

- procjenjivanje (tip: usmjereno rudarenje podataka) – podrazumijeva predviđanje vrijednosti podataka u sklopu klasifikacije istih,

- predikcija (tip: usmjereno rudarenje podataka) – identična operacija kao klasificiranje ili procjenjivanje osim toga što se klasificiranje atributa (prognoziranje) provodi s obzirom na zadanu ciljanu varijablu,

- asocijativno grupiranje (tip: rudarenje podataka u kontekstu otkrivanja novog znanja) – podrazumijeva određivanje asocijativnih pravila povezivanja više podataka,

- klasteriranje (tip: rudarenje podataka u kontekstu otkrivanja novog znanja) – podrazumijeva klasificiranje podataka bez unaprijed određenih klasa,

- profiliranje (oba tipa rudarenja podataka su moguća) – podrazumijeva opisivanje interakcija među podacima.

U literaturi koja se bavi područjem rudarenja podataka, a koja je proučavana u sklopu ovog rada, navode se sljedeće metode rudarenja podataka u smislu metoda koje su najzastupljenije - klasične statističke metode (klasifikacijske metode, metode klasteriranja, regresijske metode), neuronske mreže, stabla odlučivanja, metode (asocijativnih) analiza veza i genetski algoritmi.

U ovom radu korištena je metoda stabla odlučivanja.

3.2.2. Stabla odlučivanja

Pod pojmom "stabla odlučivanja" (engl. *Decision Trees*) podrazumijeva se metoda rudarenja podataka koja za cilj ima klasifikaciju atributa s obzirom na zadanu ciljanu varijablu, odnosno kreiranje prediktivnih (prognostičkih) modela, te ova metoda u skladu s klasifikacijom s prethodne stranice primarno potpada u skupinu operacija predikcije. Stabla odlučivanja temelje imaju u klasičnim statističkim metodama kao što je linearna regresija, te imaju ishodište u istom znanstvenom području kao neuronske mreže.

Prema [28], [35], [36] i [37] metoda stabla odlučivanja je od svih navedenih metoda rudarenja podataka najlakša za interpretaciju (metoda neuronskih mreža je okarakterizirana kao najteža za interpretaciju) u simboličkom i vizualnom smislu. Stabla odlučivanja relativno su jednostavna za provođenje, jednostavna su za shvaćanje i jednostavna su za uporabu u smislu interpretacije, s obzirom na sposobnost integriranja velikog broja prediktora u jednostavan prikaz "korak-po-korak". Istovremeno, stabla odlučivanja integriraju različite oblike mjerenja – kvalitativno i kvantitativno mjerenje. Također, stabla odlučivanja za razliku od većine drugih metoda rudarenja podataka imaju sposobnost prilagođavanja različitim promjenama u

ponašanju podataka, u smislu amortiziranja raznih nepredviđenih interakcija koje negativno djeluju na izbalansiranost i linearnost podataka.

Stabla odlučivanja pružaju izvrsne mogućnosti kompetentnim osobama (npr. analitičarima, menadžerima, itd.) da predstave kompleksne situacije donošenja odluka, a posebno onda kada postoji mogućnost da se donošenje odluke podijeli u niz manjih situacija odlučivanja, koje se u vremenskom slijedu naslanjaju jedna na drugu, odnosno kada se radi o lančano vezanim odlukama (jedna vrsta odluke proizlazi iz donošenja druge vrste odluke).

U skladu s prethodno navedenim proizlazi zaključak da je stablom odlučivanja moguće prikazati anatomiju izrazito kompleksnih odluka, odnosno projicirati razradu cjelokupne situacije odlučivanja – sve raspoložive inačice odluke (na više razina donošenja odluke), povezane neizvjesnosti s mogućom provedbom svake od inačica (na svakoj promatranoj razini) i mogući ishodi i posljedice svake od akcija koje čine inačice odluke (koji se mogu kvantificirati).

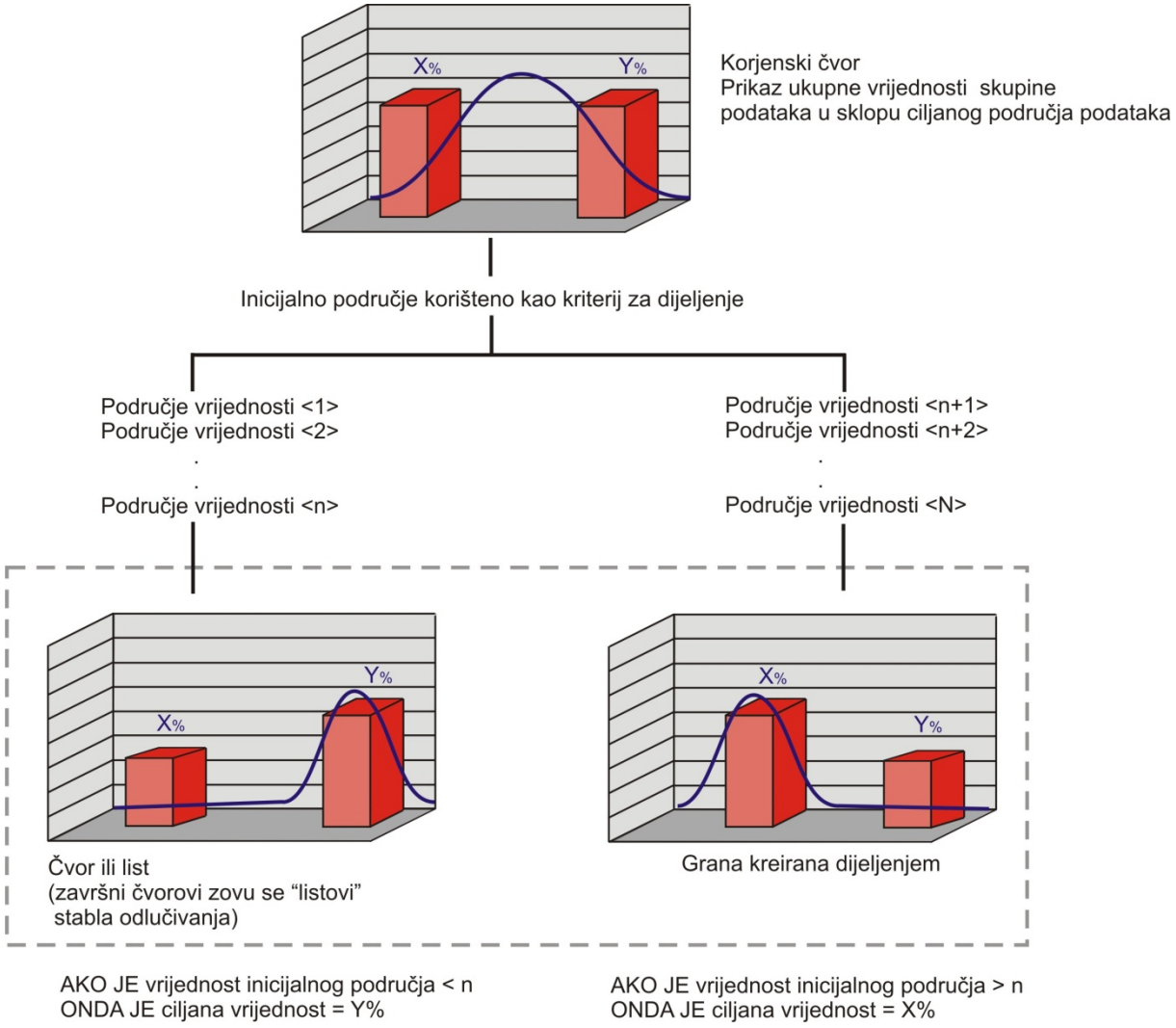
Nadalje, stabla odlučivanja pomažu pri utvrđivanju neophodnih informacija u odlučivanju te usmjeravaju primjenu kriterija maksimizacije očekivane korisnosti (informacijske dobiti) pri izboru jedne od raspoloživih inačica odluke, u skladu s načelom tzv. povratne indukcije. Stabla odlučivanja služe kompetentnim osobama koje donose odluku kao potpora razumijevanju osnovnog problema izbora, procjeni raspoloživih inačica odluke i obuhvaćanju neizvjesnih događaja koji utječu na ishode i posljedice inačica odluke.

Stabla odlučivanja produkt su matematičkih algoritama koji koriste različite načine dijeljenja skupina podataka u segmente u vidu grana. Stabla odlučivanja ishodište imaju u početnom, tzv. korijenskom čvoru (engl. *Root node*). Naziv skupine (područja) podataka koji su predmet obrade vidljiv je u korijenskom čvoru kao i sadržana distribucija vrijednosti. Način grananja ispod korijenskog čvora, odnosno tzv. pravilo odlučivanja (engl. *decision rule*) temelji se na metodi koja izdvaja međurelacije između objekta obrade (koji služi kao ciljano područje skupine podataka) i jednog ili više područja podataka koja služe kao inicijalna područja za kreiranje grana. Iz vrijednosti inicijalnih područja proizlazi procjena očekivane vrijednosti u ciljanom području. Ciljano

područje još se naziva - ishodom (engl. *outcome*), odzivom/odgovorom (engl. *response*) ili zavisnim područjem (engl. *dependent field*), odnosno zavisnom varijablom (engl. *dependent variable*).

Prema [28], na slici 3.2. grafički je prikazan princip stabla odlučivanja. Na temelju izdvojenih međurelacija proizlazi jedno ili više pravila odlučivanja pomoću kojih se opisuju međurelacije između inicijalnih (nezavisnih) varijabli i ciljanih (zavisnih) varijabli, i to u vidu stabla odlučivanja. Stablo odlučivanja je grafički način interpretacije mreže međurelacija koje egzistiraju u kontekstu inicijalnih (nezavisnih) varijabli i ciljanih (zavisnih) varijabli. Na temelju pravila odlučivanja vrši se predikcija vrijednosti novih ili do tada skrivenih opservacija u kontekstu inicijalnih (nezavisnih) varijabli, dok istovremeno vrijednosti u kontekstu ciljanih (zavisnih) varijabli mogu izostati. Također, u dijelu literature koja obrađuje predmetno područje inicijalne (nezavisne) varijable nazivaju se prediktivnim varijablama.

Svakom pravilu odlučivanja pridružen je zapis ili opservacija iz određene skupine podataka koja je sadržana u pojedinom čvoru, a temelji se na vrijednosti jednog od područja u sklopu određene skupine podataka. Područje vrijednosti na temelju kojeg se kreira pravilo odlučivanja naziva se inicijalnim područjem. Pravila dijeljenja primjenjuju se jedno za drugim a za posljedicu imaju kreiranje mreže grana koje imaju formu stabla odlučivanja.

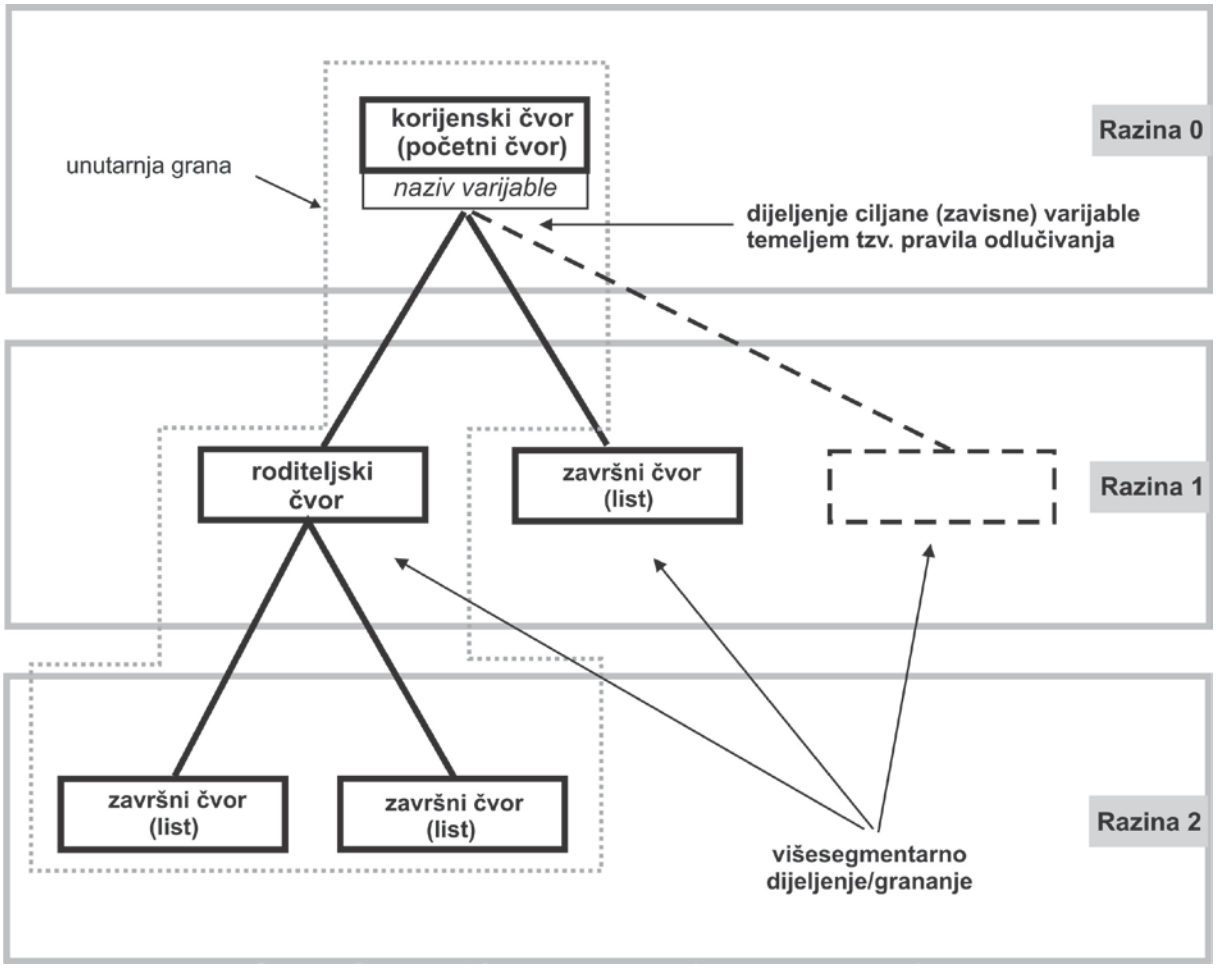


Slika 3.2. – Princip stabla odlučivanja [28]

Prema [28], na slici 3.3. prikazana je nomenklatura stabla odlučivanja. Grane (segmenti) stabla odlučivanja sastavljene su od čvorova (engl. *Nodes*).

Čvor sa svojim nastavkom (novim čvorom koji proizlazi iz postojećeg čvora) čini granu. U pojedinoj literaturi koja se bavi područjem stabla odlučivanja čvorovi iz kojih proizlaze novi čvorovi nazivaju se i "roditeljskim" čvorovima (engl. *Parent Node*), dok se novi čvorovi nazivaju "dječjim" čvorovima (engl. *Child Node*). Po logici stvari, korijenski čvor, koji je početni čvor, jest čvor bez "roditeljskog" (prethodnog) čvora.

Navedeni čvorovi, koji nisu na dnu stabla odlučivanja, smatraju se čvorovima odluke/stanja.



Slika 3.3. – Nomenklatura stabla odlučivanja [28]

Čvorovi koji se pojavljuju na dnu stabla odlučivanja, te iz kojeg nema daljnjeg grananja novih čvorova nazivaju se završnim čvorovima (engl. *Terminal Node*), “listovima” (engl. *Leaf*), odnosno čvorovima posljedice. Po logici stvari, završni čvorovi su istovremeno i "dječji" čvorovi s obzirom da se granaju iz "roditeljskih" (prethodnih) čvorova.

Za svaki “list”, odnosno završni čvor, pravilo odlučivanja osigurava jedinstven način na koji se podaci klasificiraju u klasu koja je definirana upravo u sklopu predmetnog završnog čvora. Svi čvorovi, uključujući završne čvorove, funkcioniraju po međusobno isključivim pravilima pridruživanja što za posljedicu ima da zapisi ili opservacije zabilježene u “roditeljskoj” skupini podataka mogu postojati samo u pojedinom čvoru.

Jednom kada su određena (determinirana) pravila odlučivanja, moguće ih je koristiti za predikciju novih vrijednosti (čvorova) utemeljenih na novim, odnosno do tada

neotkrivenim podacima. U sklopu prediktivnog modeliranja prediktivna vrijednost proizlazi iz pravila odlučivanja.

U sklopu istraživanja kojim se bavi ovaj rad, u kontekstu metode stabla odlučivanja a prema prethodno opisanom principu, osobni automobili predmetnih marki i tipova pripadaju dvjema skupinama, te su u skladu s time označeni kao tehnički ispravni (odnosno koji su "prošli" tehnički pregled) i tehnički neispravni (odnosno koji su "pali" na tehničkom pregledu). Cilj prediktivnog modela koji je predmet ovog rada jest pronalazak klasifikacijskih varijabli (atributa) koje najbolje razdvajaju skup tehnički ispravnih od tehnički neispravnih osobnih automobila. U tom kontekstu algoritam stabla odlučivanja započinje korijenskim čvorom koji sadrži skup tehnički ispravnih i tehnički neispravnih osobnih automobila, nakon čega se donose odluke (odnosno pronalaze svi mogući ishodi) s ciljem dobivanja kritične (najkorisnije) varijable i odgovarajuće granične vrijednosti koja najbolje vrši razdvajanje podskupova tehnički ispravnih i tehnički neispravnih osobnih automobila. Podaci se dijele prema svim zadanim kriterijima u grane, pri čemu se odabire kriterij koji podatke dijeli u skupine koje su homogenije od početne skupine podataka. Procedura se ponavlja sve dok podatke nije moguće dalje dijeliti u skupine koje su homogenije od početnih podataka.

Načelno, prema [38] proces konstruiranja stabla odlučivanja može se opisati u tri osnovna koraka:

1. izgradnja logičkog modela – stabla odlučivanja s ulaznim podacima koji podrazumijevaju sve čvorove i grane u kronološkom redoslijedu,
2. računanje očekivanih vrijednosti odluka postupkom računanja unatrag – računanje unatrag započinje na završnim čvorovima i kreće prema unatrag, do početnog čvora odluke, te se svakom čvoru pridodaje ekvivalentna očekivana vrijednost, i to:
   - na završnom čvoru izračunata je konačna vrijednost predmetne alternative,
   - završnom čvoru pridružuju se očekivane vrijednosti izračunate kao:

$$EV_{i-1} = \sum_j p_j EV_{i,}\ i \in \{1,2,\dots,n\}, j \in \{1,2,\dots,m\} \quad (3.1)$$

gdje je:

EV_{i-1}- očekivana vrijednost u čvoru i-1,

p_j – vjerojatnost grane j koja izlazi iz čvora i-1,

- na čvoru odluka očekivana pridružena vrijednost jednaka je najvećoj od algoritmom prethodno izračunatih očekivanih vrijednosti sljedećih čvorova u stablu odlučivanja.

3. pronalaženje optimalnog puta postupkom računanja unaprijed – nakon što su se u prethodnom koraku izračunale očekivane vrijednosti za svaki čvor, može se prepoznati optimalan put u stablu odlučivanja, računanjem prema unaprijed, od početnog čvora odluke.

U literaturi koja se bavi područjem stabla odlučivanja, a koja je proučavana u sklopu ovog rada, navode se sljedeći algoritmi stabla odlučivanja – CHAID algoritam (engl. *Chi-square Automatic Interaction Detection*) odnosno njegova unaprijeđena inačica Exhaustive CHAID algoritam, algoritmi perceptronskih neuronskih mreža, CART algoritama (engl. *Classification and regression Trees*), QUEST (engl. *Quick, Unbiased, Efficient Statistical Tree*) ID3 algoritmi (engl. *Iterative Dichotomiser 3*), C4.5 algoritmi (ekstenzija ID3 algoritama), itd.

U ovom radu korišten je Exhaustive CHAID algoritam te algoritam perceptronske neuronske mreže.

3.2.3. CHAID i Exhaustive CHAID algoritmi

Prema [39] CHAID algoritam razvijen je krajem sedamdesetih godina prošlog stoljeća od strane Gordona V. Kassa kao unaprijeđena verzija algoritama razvijenih polovicom prošlog stoljeća – AID algoritma (engl. *Automatic Interaction Detection*) i THAID algoritma (engl. *THeta Automatic Interaction Detection*), na način da CHAID algoritam predstavlja *hi*-kvadrat (engl. *chi-square*) verziju ove dva algoritma.

Prema [28] unaprjeđenje CHAID algoritma u odnosu na AID i THAID algoritme proizlazi iz sljedećeg:

- identificiranje samo onih međurelacija (veza) koje su značajno različite u odnosu na nasumične događaje,
- statističko podešavanje provodi se na polariziranim varijablama koje su kandidati za dijeljenje, odnosno daljnje grananje u vidu čvorova,
- daljnji rast, odnosno grananje stabla prestaje u trenutku kada pojedini čvor više ne prolazi test značajnosti (odnosno isti nije značajno različit u odnosu na nasumične događaje).

CHAID algoritam jedan je od algoritama zavisnosti iz područja multivarijacijske analize podataka kod koje je zavisna varijabla kategorijska varijabla, dok nezavisne (prediktivne) varijable mogu biti kategorijske ili metričke varijable. U slučaju da se radi o metričkim varijablama iste se automatski preoblikuju u kategorijske varijable.

Prema [28], [34] i [37] neke od temeljnih prednosti CHAID algoritma u odnosu na druge algoritme metode stabla odlučivanja jest izrazita preglednost vizualne interpretacije, te sposobnost obrade velike skupine podataka u kratkom vremenu. Također, ovaj algoritam u potpunosti je kompatibilan s platformama SQL baza podataka.

Prema [40] i [41], uporabom CHAID algoritma postiže se veći prediktivni potencijal nego uporabom drugih algoritama predikcije u sklopu metode stabla odlučivanja (CART, QUEST, itd.)

Osnovni zadatak CHAID algoritma jest automatsko otkrivanje međurelacija (veza) između varijabli, odnosno izdvajanje nezavisnih varijabli na temelju χ^2 (*hi*-kvadrat) testa [42]. Prema [43] i [44] pomoću *hi*-kvadrat testa ispituju se razlike između opaženih i očekivanih frekvencija, što znači da osnovni podaci koji su predmet istraživanja mogu biti i mjerne vrijednosti, ali *hi*-kvadrat test ispituje samo njihove frekvencije. Pod opaženim frekvencijama podrazumijevaju se nizovi/serije događaja koji su prikupljeni nekim mjerenjem (nulta hipoteza), dok se pod očekivanim frekvencijama podrazumijevaju nizovi/serije događaja koji prema subjektivnim predviđanjima imaju sasvim drugi raspored (frekvenciju) u odnosu na opažene frekvencije. Očekivane frekvencije proizlaze iz postavljene hipoteze.

Nulta hipoteza se odbacuje (odnosno prihvaća se postavljena hipoteza) ukoliko se *hi*-kvadrat testom utvrdi postojanje statistički značajne razlike između opaženih i očekivanih frekvencija.

U sklopu CHAID algoritma, na temelju *hi*-kvadrat testa, provodi se rekurzivno dijeljenje zadane skupine podataka na specifične podskupine primjenom algoritma koji minimizira varijacije zavisne varijable unutar specifične podskupine, odnosno koji ih istovremeno maksimizira između specifičnih podskupina. Rekurzivno dijeljenje provodi se u tri koraka – integriranje (engl. *merging*), razdvajanje (engl. *splitting*) i zaustavljanje (engl. *stopping*) – dok je krajnji rezultat stablo odlučivanja, odnosno CHAID stablo. CHAID stablo se razvija uzastopnim ponavljanjem navedenih koraka na svakoj specifičnoj podskupini – čvoru stabla, s početkom na inicijalnom (korijenskom) čvoru. Stablo predstavlja cjelinu skupa podataka, prva razina grana (čvorovi – specifične podskupine) prikazuje kategorije nezavisne varijable koja je najjači prediktor zavisne varijable, odnosno pripada određenoj skupini. CHAID analizom se automatski, putem iterativne procedure, određuje kako će se grupirati vrijednosti prediktivnih varijabli, tako da se na stablu pojavljuju oni čvorovi koji se statistički značajno razlikuju.

Unaprijeđena inačica CHAID algoritma - Exhaustive CHAID algoritam – razvijena je 1991. godine od strane D. Biggsa [45]. Exhaustive CHAID algoritam od svoje originalne verzije razlikuje se u smislu proširenog rekurzivnog dijeljenja u koraku integriranja (*merging*) sličnih parova podskupina podataka sve dok ne preostane samo jedan par, te poboljšanog proračuna Bonferroni množitelja kod postupka Bonferroni prilagođavanja (korekcije). S obzirom da se radi o istom algoritmu, odnosno s obzirom da su svi ostali koraci CHAID algoritma i Exhaustive CHAID algoritma identični osim koraka, pseudokod je razrađen na primjeru CHAID algoritma, dok su posebno izdvojene različitosti u koraku integriranja (*merging*) i u kontekstu Bonferroni prilagođavanja (korekcije).

Prema [39], [45] i [46] pseudokod CHAID algoritama je sljedeći:

Oznake

Y	Zavisna (ciljana) varijabla
$X_m, m = 1, \dots, \mathrm{M}$	Skupina svih nezavisnih varijabli.
$\hbar = \{\mathbf{x}_n, y_n\}_{n=1}^{N}$	Cijeli uzorak za učenje
w_n	Značajnost slučaja povezan sa slučajem *n*.
f_n	Značajnost frekvencija povezan sa slučajem *n*. Ne-cjelobrojna pozitivna vrijednost zaokružuje se na najbliži cijeli broj.

CHAID ALGORITAM

Integriranje (*merging*)

Za svaku nezavisnu varijablu *X* integriraju se kategorije koje nemaju značaj (koje su ne-signifikantne). Svaka konačna kategorija od *X* rezultirati će jednim dječjim čvorom ako je *X* upotrijebljen za razdvajanje (*splitting*) čvora. U koraku u kojem se vrši integriranje također se proračunava prilagođena (korigirana) *p* – vrijednost koja se koristi u koraku razdvajanja. Koraci:

1. Ako *X* ima samo jednu kategoriju, dolazi do zaustavljanja i podešavanja prilagođene (korigirane) *p* – vrijednosti na vrijednost 1.
2. Ako *X* ima dvije kategorije, prelazi se na korak 8.
3. Inače, pronalazi se par kategorija od *X* (dopušten par kategorija ordinalne nezavisne varijable predstavljaju dvije susjedne kategorije, a za nominalnu nezavisnu varijablu bilo koje dvije kategorije) koje su najmanje različite u smislu značajnosti (signifikantnosti), odnosno koje su najsličnije. Najsličniji par je onaj par čiji statistički test daje najveću *p* – vrijednost uzimajući u obzir zavisnu varijablu *Y*.
4. Za par koji ima najveću *p* – vrijednost vrši se provjera da li je njihova *p* – vrijednost veća od alfa-razine $\alpha_{integriranje}$ *(alpha_merge)* koju je specificirao korisnik. Ako jeste, taj se par integrira u jedinstvenu kategoriju. U tom slučaju je formirana nova skupina kategorija. Ako nije, prelazi se na korak 7.
5. (opcija) Ako se novoformirana kategorija sastoji od tri ili više originalnih kategorija, tada se pronalazi najbolje binarno razdvajanje unutar formirane kategorije čija je *p* – vrijednost najmanja. Ovakvo se binarno razdvajanje

provodi ako njezina *p* – vrijednost nije veća od alfa-razine $\alpha_{\text{razdvajanje-}integriranje}$ *(alpha_spli-merge).*

6. Prelazi se na korak 2.
7. (opcija) Bilo koja kategorija koja ima previše opservacija (što proizlazi iz usporedbe s minimalnom veličinom segmenta koji je specificiran od strane korisnika) se integrira s drugom kategorijom koja joj je najsličnija prema izmjerenoj najvećoj *p* – vrijednosti.
8. Prilagođena (korigirana) *p* – vrijednost je izračunata za integrirane kategorije primjenom metode Bonferroni prilagođavanja (korekcije).

Razdvajanje (*splitting*)

„Najbolje“ razdvajanje za svaku nezavisnu varijablu proizlazi iz koraka koji podrazumijeva integriranje. U okviru koraka razdvajanja provodi se selektiranje u smislu koja će nezavisna varijabla biti upotrebljena za najbolje razdvajanje čvora. Predmetno selektiranje se postiže na temelju međusobnog uspoređivanja prilagođenih (korigiranih) *p* – vrijednosti povezanih sa svakom pojedinom nezavisnom varijablom. Prilagođena (korigirana) *p* – vrijednost proizlazi iz koraka integriranja.

1. Provodi se selektiranje nezavisne varijable koja ima najmanju prilagođenu (korigiranu) *p* – vrijednost (npr. najznačajniju odnosno najviše signifikantnu).
2. Ako je prilagođena (korigirana) *p* – vrijednost manja ili jednaka alfa-razini $\alpha_{razdvajanje}$ *(alpha_split)* koju je specificirao korisnik, provodi se razdvajanje čvora temeljem predmetne nezavisne varijable. Inače, ne provodi se razdvajanje i čvor se smatra završnim čvorom.

Zaustavljanje (*stopping*)

U okviru koraka zaustavljanja provjerava se da li bi proces rasta stabla trebao biti zaustavljen u skladu sa sljedećim pravilima glede zaustavljanja:

1. U slučaju da čvor postane čist; to je završetak, svi slučajevi u čvoru imaju identične vrijednosti zavisne varijable, te u čvoru više nema razdvajanja.
2. U slučaju da svi slučajevi u čvoru imaju identične vrijednosti za svaku nezavisnu varijablu, u čvoru više nema razdvajanja.

3. U slučaju da trenutna dubina stabla dosegne najveću vrijednost dubine koja je specificirana od strane korisnika, proces daljnjeg rasta stabla biti će obustavljen.
4. U slučaju da je veličina čvora manja od najmanje vrijednosti veličine čvora koja je specificirana od strane korisnika u čvoru više nema razdvajanja.
5. U slučaju da razdvajanje čvora rezultira „dječjim“ čvorom čija je veličina čvora manja od najmanje vrijednosti veličine „dječjeg“ čvora koja je specificirana od strane korisnika, „dječji“ čvorovi koji imaju premalo slučajeva (što proizlazi iz usporedbe s ovim minimumom) će se integrirati s najsličnijim čvorovima prema izmjerenoj najvećoj *p* – vrijednosti. Međutim, ako je rezultirajući broj „dječjih“ čvorova 1, u čvoru više nema razdvajanja.

EXHAUSTIVE CHAID ALGORITAM

Integriranje (*merging*)

1. Ako *X* ima samo jednu kategoriju, podešava se prilagođena (korigirana) *p* – vrijednost na vrijednost 1.
2. Postavlja se *index* = 0. *P* – vrijednost se proračunava na osnovu kategorija *X* u tom trenutku. Proizlazi da je *p* – vrijednost *p(index) = p(0)*
3. Inače, pronalazi se par kategorija od *X* koje su najmanje različite u smislu značajnosti (signifikantnosti), odnosno koje su najsličnije. Ovo može biti određeno pomoću para čiji statistički test daje najveću *p* – vrijednost uzimajući u obzir zavisnu varijablu *Y*.
4. Par koji ima najveću *p* – vrijednost integrira se u jedinstvenu kategoriju.
5. (opcija) Ako se novoformirana kategorija sastoji od tri ili više originalnih kategorija, tada se traži najbolje binarno razdvajanje unutar formirane kategorije čija je *p* – vrijednost najmanja. Ako je predmetna *p* – vrijednost veća od one u novoformiranoj kategoriji provodi se binarno razdvajanje unutar formirane kategorije.
6. Ažurira se *index* = *index* + 1, te se *p* – vrijednost proračunava na osnovu kategorija *X* u tom trenutku. Označava se *p(index)* kao *p* – vrijednost.

7. Koraci 3 i 6 se ponavljaju dok ne preostanu samo dvije kategorije. Zatim se između svih indeksa pronalazi skupina kategorija kod koje je *p(index)* najmanji.
8. (opcija) Bilo koja kategorija koja ima previše opservacija (što proizlazi iz usporedbe s minimalnom veličinom segmenta koji je specificiran od strane korisnika) se integrira s drugom kategorijom koja joj je najsličnija prema izmjerenoj najvećoj *p* – vrijednosti.
9. Prilagođena (korigirana) *p* – vrijednost je izračunata primjenom metode Bonferroni prilagođavanja (korekcije) na način kako je objašnjeno u nastavku.

Za razliku od originalnog CHAID algoritma nije potrebna alfa-razina specificirana od strane korisnika *(alpha_split-merge* ili *alpha_merge).* Jedino je tijekom koraka razdvajanja potrebna alfa-razina $\alpha_{razdvajanje}$ *(alpha_split).*

Proračunavanje *p* -vrijednosti

Proračunavanje (neprilagođene) *p* – vrijednosti ovisi o tipu zavisne varijable.
Korak integriranja ponekad podrazumijeva *p* – vrijednost za par *X* kategorija, a ponekad podrazumijeva *p* – vrijednost za sve *X* kategorije. Kada je riječ o *p* – vrijednosti za par *X* kategorija relevantan je samo dio podataka u predmetnom čvoru. Neka *D* označava relevantne podatke. Pretpostavlja se da *D* sadržava I kategorija od *X*, te J kategorija od *Y*.

Proračunavanje p -vrijednosti za kontinuiranu (metričku) zavisnu varijablu:
Ako je zavisna varijabla Y kontinuirana (metrička), provodi se ANOVA F test (analiza varijance) koji testira da li je značenje Y za različite kategorije X identično. ANOVA F test proračunava *F*-statističku vrijednost iz čega proizlazi p –vrijednost kao

$$\mathrm{F} = \frac{\sum_{i=1}^{I}\sum_{n\in D} w_n f_n I(x_n=i)(\bar{y}_i-\bar{y})^2/(I-1)}{\sum_{i=1}^{I}\sum_{n\in D} w_n f_n I(x_n=i)(y_n-\bar{y}_i)^2/(N_f-I)}, \tag{3.2}$$

$$p = \Pr\left(F(I-1, N_f - I) > F\right), \tag{3.3}$$

gdje

$$\overline{y_i} = \frac{\sum_{n\in D} w_n f_n y_n I(x_n=i)}{\sum_{n\in D} w_n f_n I(x_n=i)}, \bar{y} = \frac{\sum_{n\in D} w_n f_n y_n}{\sum_{n\in D} w_n f_n}, N_f = \sum_{n\in D} f_n\ , \qquad (3.4)$$

i $F\big(I-1, N_f - I\big)$ je slučajna varijabla koja prati F – distribuciju sa stupnjevima slobode I i $N_f - I$.

Proračunavanje p -vrijednosti za nominalnu zavisnu varijablu:

Ako je zavisna varijabla *Y* nominalna (kategorijska) testira se nul-hipoteza nezavisnosti *X* i *Y*. Da bi se provelo ovo testiranje formira se kontingencijska tablica koju čine razredi *Y* u stupcima i kategorije nezavisne varijable *X* u redovima. Procjenjuje se očekivana frekvencija polja u kontekstu nul-hipoteze. Opažene frekvencije polja i očekivane frekvencije polja se koriste za proračunavanje Pearson-ovog *hi*-kvadrata ili omjera vjerojatnosti. Na temelju jednog od prethodno navedenih čimbenika izračunava se *p* – vrijednost.

Slijede Pearson-ov *hi*-kvadrat i omjer vjerojatnosti

$$X^2 = \sum_{j=1}^{J}\sum_{i=1}^{I} \frac{\left(n_{ij} - \hat{m}_{ij}\right)^2}{\hat{m}_{ij}}, \qquad (3.5)$$

$$G^2 = 2\sum_{j=1}^{J}\sum_{i=1}^{I} n_{ij}\ \ln\big(n_{ij}/\hat{m}_{ij}\big), \qquad (3.6)$$

gdje je $n_{ij} = \sum_{n\in D} f_n\, I(x_n = i \wedge y_n = j)$ opažena frekvencija polja i $\hat{m}_{ij}$ je procijenjena očekivana frekvencija polja za polje $(x_n = i, y_n = j)$ nezavisnog modela kako slijedi. Odgovarajuća *p* – vrijednost je dana pomoću $p = Pr(\chi_d^2 > X^2)$ za Pearson-ov *hi*-kvadrat test ili $p = Pr(\chi_d^2 > G^2)$ za test omjera vjerojatnosti, gdje χ_d^2 prati *hi*-kvadrat distribuciju sa stupnjevima slobode $d = (J-1)(I-1)$.

Procjena očekivanih frekvencija polja bez i značajnosti slučaja:

$$\widehat{m}_{ij} = \frac{n_{i.}n_{.j}}{n_{..}}, \tag{3.7.}$$

gdje

$$n_{i.} = \sum_{j=1}^{J_1} n_{ij} \ , n_{.j} = \sum_{i=1}^{I_1} n_{ij}, n = \sum_{j=1}^{J_1}\sum_{I=1}^{I_1} n_{ij}. \tag{3.8}$$

Procjena očekivanih frekvencija polja sa značajnosti slučaja:

Ako je značajnost slučaja specificirana onda je očekivana frekvencija polja pod nul-hipotezom nezavisnosti u obliku

$$m_{ij} = \ \overline{w}_{ij}^{-1}\alpha_i\beta_j, \tag{3.9}$$

gdje su α_i i β_j parametri koje treba procijeniti i

$$\overline{w}_{ij} = \frac{w_{ij}}{n_{ij}}, w_{ij} = \sum_{n\in D} w_n f_n I(x = i \wedge y_n = j). \tag{3.10}$$

Na temelju sljedeće iterativne procedure parametri procjenjuju $\hat{\alpha}_i, \hat{\beta}_j$ te $\widehat{m}_{ij}$

1. $k = 0,\ \alpha_i^{(0)} = \beta_j^{(0)} = 1, m_{ij}^{(0)} = \overline{w}_{ij}^{-1}.$ (3.11)
2. $\alpha_i^{(k+1)} = \frac{n_i}{\sum_j \overline{w}_{ij}^{-1}\beta_j^{(k)}} = \alpha_i^{(k)}\frac{n_i}{\sum_j m_{ij}^{(k)}}.$ (3.12)
3. $\beta_j^{(k+1)} = \frac{n_{.j}}{\sum_i \overline{w}_{ij}^{-1}\alpha_i^{(k+1)}}.$ (3.13)
4. $m_{ij}^{(k+1)} = \ \overline{w}_{ij}^{-1}\alpha_i^{(k+1)}\beta_j^{(k+1)}.$ (3.14)
5. Ako je $max_{i,j}\left|m_{ij}^{(k+1)} - m_{ij}^{(k)}\right| < \varepsilon,$ dolazi do zaustavljanja i izlaza $\alpha_i^{(k+1)}, \beta_j^{(k+1)}\ i\ m_{ij}^{(k+1)}$ kao konačnih procjena $\hat{\alpha}_i, \hat{\beta}_j, \widehat{m}_{ij}$. Inače, $k = k + 1$, te se prelazi na korak 2.

Proračunavanje p -vrijednosti za ordinalnu zavisnu varijablu:

Ako je zavisna varijabla *Y* ordinalna kategorijska varijabla nul-hipoteza nezavisnosti od *X* i *Y* se testira putem tzv. modela utjecaja redova (koji podrazumijeva da redovi predstavljaju kategorije od *X* a stupci razrede od *Y*), prema Goodmanu (1979.). Dvije skupine očekivanih frekvencija polja $\hat{m}_{ij}$ (pod hipotezom nezavisnosti) i $\hat{\hat{m}}_{ij}$ (pod hipotezom da podaci slijede model utjecaja redova) se procjenjuju. Omjer vjerojatnosti i *p* – vrijednost su

$$H^2 = 2\sum_{i=1}^{I}\sum_{j=1}^{J}\hat{\hat{m}}_{ij}\ \ln\left(\hat{\hat{m}}_{ij}/\hat{m}_{ij}\right), \tag{3.15}$$

$$p = Pr(\chi^2_{I-1} > H^2) \tag{3.16}$$

Procjena očekivanih frekvencija polja putem modela utjecaja redova:

Za model utjecaja redova potrebni su rezultati za razrede od *Y*. Sukladno početnim postavkama (*default*) poredak razreda od *Y* se koristi u smislu rezultata za razrede. Korisnici mogu specificirati svoje vlastite skupine rezultata. Rezultati se postavljaju na početku stabla i naknadno se ne mijenjaju. Neka je s_j rezultat razreda *j* od *Y*, *j* = 1,...,*J*. Očekivana frekvencija polja putem modela utjecaja razreda je dana s

$$m_{ij} = \bar{w}_{ij}^{-1}\alpha_i\beta_j\gamma_i^{(s_j-\bar{s})}, \tag{3.17}$$

gdje

$$\bar{s} = \left.\sum_{j=1}^{J} w_{.j}s_j\right/\sum_{j=1}^{J} w_{.j} \tag{3.18}$$

u čemu su $w_{.i} = \sum_i w_{ij}, \alpha_i, \beta_j\ i\ \gamma_i$ nepoznati parametri koje treba procijeniti.

Procijenjeni parametri $\hat{\hat{\alpha}}_i, \hat{\hat{\beta}}_j, \hat{\hat{\gamma}}_i\ do\ \hat{\hat{m}}_{ij}$ su rezultat sljedeće iterativne procedure.

1. $k = 0,\ \alpha_i^{(0)} = \beta_j^{(0)} = \gamma_i^{(0)} = 1, m_{ij}^{(0)} = \bar{w}_{ij}^{-1}.$ (3.19)

2. $\alpha_i^{(k+1)} = \dfrac{n_{.j}}{\sum_j \bar{w}_{ij}^{-1}\beta_j^{(k)}\left(\gamma_i^{(k)}\right)^{(s_j-\bar{s})}} = \alpha_i^{(k)}\dfrac{n_{i.}}{\sum_j m_{ij}^{(k)}}.$ (3.20)

3. $\beta_j^{(k+1)} = \dfrac{n_{.j}}{\sum_i \bar{w}_{ij}^{-1}\alpha_i^{(k+1)}\left(\gamma_i^{(k)}\right)^{(s_j-\bar{s})}}.$ (3.21)

4. $m_{ij}^{*} = \overline{w}_{ij}^{-1}\alpha_i^{(k+1)}\beta_j^{(k+1)}\left(\gamma_i^{(k)}\right)^{(s_j-\bar{s})}, G_i = 1 + \frac{\sum_j (s_j-\bar{s})(n_{ij}-m_{ij}^{*})}{\sum_j (s_j-\bar{s})^2 m_{ij}^{*}}.$ (3.22)

5. $\gamma_i^{(k+1)} = \begin{cases} \gamma_i^{(k)} G_i & G_i > 0 \\ \gamma_i^{(k)} & \text{inače} \end{cases}$ (3.23)

6. $m_{ij}^{(k+1)} = \overline{w}_{ij}^{-1}\alpha_i^{(k+1)}\beta_j^{(k+1)}\left(\gamma_i^{(k+1)}\right)^{(s_j-\bar{s})}.$ (3.24)

7. Ako je $max_{i,j}\left|m_{ij}^{(k+1)} - m_{ij}^{(k)}\right| < \varepsilon$, dolazi do zaustavljanja i izlaza $\alpha_i^{(k+1)}, \beta_j^{(k+1)}, \gamma_i^{(k+1)} i\ m_{ij}^{(k+1)}$ kao konačnih procjena $\hat{\hat{\alpha}}_i, \hat{\hat{\beta}}_j, \hat{\hat{\gamma}}_i, \hat{\hat{m}}_{ij}$. Inače, $k = k + 1$, te se prelazi na korak 2.

Bonferroni prilagođavanja (korekcije)

Prilagođena (korigirana) *p* – vrijednost se proračunava na način da se p – vrijednost pomnoži s Bonferroni množiteljem. Bonferroni množitelj vrši prilagođavanje (korigiranje) za višestruka testiranja.

CHAID

Neka se pretpostavi da nezavisna varijabla ima I kategorija te da je reducirana na r kategorija nakon provođenja koraka integriranja. Bonferroni množitelj B je broj mogućih načina da I kategorije mogu biti integrirane u r kategorija. Za $r = I$, $B = 1$. Za 2 <= r < I koriste se sljedeća jednadžba

$$B = \begin{cases} \binom{I-1}{r-1} \\ \sum_{\vartheta=0}^{r-1}(-1)^{\vartheta}\frac{(r-\vartheta)^I}{\vartheta!(r-\vartheta)!} \\ \binom{I-2}{r-2} + r\binom{I-2}{r-1} \end{cases} \quad (3.25)$$

Exhaustive CHAID

Exhaustive CHAID iterativno integrira dvije kategorije sve dok ne preostanu samo dvije kategorije. Bonferroni množitelj B je zbroj mogućih načina integriranja dvije kategorije u svakoj iteraciji.

$$B = \begin{cases} \frac{I(I-1)}{2} \\ \frac{I(I^2-1)}{2} \\ \frac{I(I-1)}{2} \end{cases} \quad (3.26)$$

Vrijednosti koje nedostaju

Ukoliko zavisna varijabla slučaja nedostaje ona neće biti korištena u analizi. Ukoliko sve nezavisne varijable slučaja nedostaju, predmetni slučaj se ignorira. Ukoliko značajnost slučaja nedostaje ili je nula ili je negativna predmetni slučaj se također ignorira. Ukoliko značajnost frekvencija nedostaje ili je nula ili je negativna predmetni slučaj se također ignorira

Inače, vrijednosti koje nedostaju tretirati će se kao prediktivna kategorija. Za ordinalne nezavisne varijable algoritam prvotno generira „najbolju“ skupinu kategorija koristeći sve ne-nedostajaće informacije iz podataka. Sljedeće, algoritam prepoznaje kategoriju koja je najsličnija kategoriji koja nedostaje. U konačnici, algoritam odlučuje o integriranju kategorije koja nedostaje s njoj najsličnijom kategorijom ili o zadržavanju kategorije koja nedostaje kao zasebne kategorije. Proračunavaju se dvije *p* – vrijednosti, jedna za skupinu kategorija koje su formirane nakon integriranja kategorije koje nedostaju s njima najsličnijim kategorijama i druga za skupinu kategorija koje su formirane nakon pridruživanja kategorija koje nedostaju kao zasebnih kategorija. Provodi se opcija koja daje najmanju *p* – vrijednost.

Za nominalne nezavisne varijable kategorija koja nedostaje tretira se kao ostale kategorije u analizi.

Primjena CHAID algoritma u kontekstu klasificiranja i izrada prediktivnih modela je izrazito širokopojasna te se susreće u istraživanjima u nizu znanstvenih područja, kao što je medicina (klasifikacija i prognoza razvoja bolesti), obrazovanje (predikcija uspješnosti učenika), ekonomije/marketinga (segmentacije tržišta), kriminalistike, itd.

U sklopu šireg područja sigurnosti prometa na cestama istraživanje [47] bavi se utjecajem vozila (u smislu predmeta udara), odnosno ceste, na ozljede glave pješaka i biciklista. Predmetnim istraživanjem obrađeno je 205 prometnih nesreća te je utvrđeno da je u većini prometnih nesreća (110) uzrok ozljede glave udarac o cestu, no da je većini najtežih ozljeda glave uzrok udar vozila. Uporabom CHAID algoritma, u ovisnosti o 74 prediktivne varijable, utvrđeno je da je nekoliko čimbenika povezano s izvorom ozljede - vrsta interakcije između vozila u naletu i pješaka/biciklista, životna dob pješaka/biciklista i vrsta ozljede.

Nadalje, u sklopu šireg područja sigurnosti prometa na cestama istraživanje [48] bavi se problematikom sudara vlakova i vozila, na tzv. pasivnim željezničko-cestovnim prijelazima. Uporabom CHAID i CART algoritama, u ovisnosti o 12 prediktivnih varijabli, izrađen je prediktivni model sudara vlakova i vozila.

U području medicine, istraživanje [49] bavi se opisom karakteristika i identifikacijom prediktora u kontekstu progresivne dilatacije aorte, kod pacijenata s urođenom bikuspidnom aortnom valvulom (engl. *bicuspid aortic valve* – BAV). U slučaju 280 odraslih pacijenata s izoliranim BAV sindromom podvrgnutih ehokardiografiji, uporabom CHAID i CART algoritama, razvijeni su prediktivni modeli dilatacije u inicijalnom aspektu oboljenja, te u kontekstu njegove progresije. Također su istraživani prediktori aneurizma s kirurškom indikacijom.

Nadalje, u području medicine, istraživanje [50] bavi se primjenom modernih informacijskih tehnologija za analizu podataka i predviđanje u medicini. Uporabom nekoliko algoritama metode stabla odlučivanja, uključujući CHAID algoritam, izvršena je predikcija ishoda liječenja i ishoda bolesti 100 pacijenata koji su doživjeli moždani udar u ovisnosti o 8 čimbenika rizika (spol, starost, krvni tlak, razina šećera u krvi, inzult u obitelji, veličina hematoma, neurološki skorovi), te 251 pacijenta kod kojih je praćena demencija u ovisnosti o 29 prediktivnih varijabli.

U području obrazovanja, istraživanje [51] bavi se analizom uspješnosti (učinaka) učenika srednjih škola. Istraživanjem je obrađeno 772 učenika te je uporabom CHAID algoritma izgrađen prediktivni model njihove uspješnosti u procesu srednjoškolskog obrazovanja, u ovisnosti o 7 prediktivnih varijabli.

U području marketinga, istraživanje [52] bavi se segmentacijom tržišta. Uporabom logističke regresije i CHAID algoritma obrađeno je 400 posjetitelja restorana u vidu ankete o razini zadovoljstva dobivenom uslugom. Izgrađen je prediktivni model ponovnih posjetitelja, u ovisnosti o 7 prediktivnih varijabli.

U području turizma istraživanje [53] bavi se izradom prediktivnog modela glede naplate turističkih pristojbi namijenjenih zaštiti okoliša, uporabom CHAID algoritma, u ovisnosti o 7 prediktivnih varijabli.

U području gospodarstva istraživanje [54] bavi se izradom prediktivnog modela solventnosti malih poslovnih subjekata (obrta) u Egiptu, uporabom nekoliko algoritama metode stabla odlučivanja uključujući CHAID algoritam, u ovisnosti o 6 prediktivnih varijabli.

U području financija istraživanje [37] bavi se izradom prediktivnog modela rizika od bankrota tvrtki (model sustava za rano upozoravanje) uporabom CHAID algoritma, u ovisnosti o 47 prediktivnih varijabli.

U području kriminalistike istraživanje [55] bavi se izradom prediktivnog modela kriminaliteta (provale, prevare, krađe, pljačke, prijetnje) u kontekstu maloprodaje i proizvodnje, uporabom CHAID algoritma, u ovisnosti o 14 prediktivnih varijabli.
U području zaštite na radu istraživanje [56] bavi se izradom prediktivnog modela za situacije u kojima je ugrožena sigurnost radnika u građevinarstvu zbog mogućnosti strujnog udara, uporabom CHAID algoritma, u ovisnosti o 6 prediktivnih varijabli.

U području zaštite okoliša istraživanje [57] bavi se izradom prediktivnog modela glede atmosferske bioakumulacije metala u mahovini, uporabom CHAID algoritma, u ovisnosti o 12 prediktivnih varijabli.

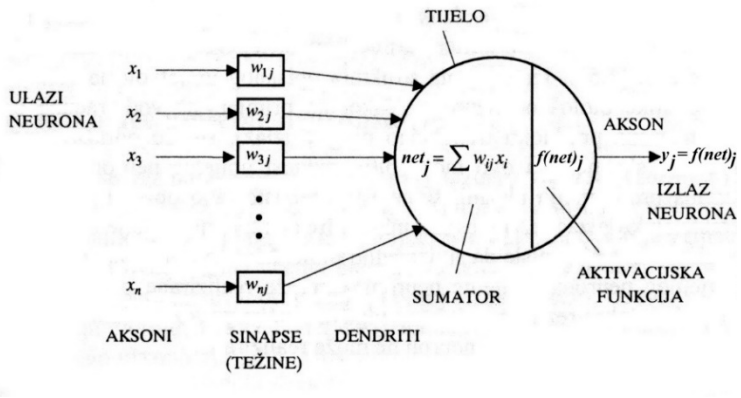
3.2.4. Perceptronska neuronska mreža

Prema [58], originalna ideja za umjetnu neuronsku mrežu generirala se iz niza pokušaja modeliranja biofiziologije mozga čovjeka, s ciljem razumijevanja i objašnjenja kako isti funkcionira. Ta ideja podrazumijeva kreiranje modela sposobnog procesuirati (prihvaća, obrađuje, generira, pohranjuje i prenosi) informacije, analogno aktivnostima mozga čovjeka. William James postavlja 1890. godine, sljedeću tvrdnju: Aktivnost bilo koje točke mozga čovjeka predstavlja zbroj tendencija svih ostalih točaka da se *"prazne"* (ispaljuju) u nju. Ova temeljna tvrdnja poslužila je za izgradnju osnovne strukture umjetnog neurona. Naime, ako se određena točka u mozgu čovjeka zamijeni neuronom, onda se aktivnost neurona može modelirati kao zbroj *"otežanih" ulaza* neurona. Otežani ulazi jesu ulazi pomnoženi određenim faktorima koji se nazivaju *težine* neurona. Prema tome aktivnost umjetnog neurona ovisi o : *broju ulaza* (veza) iz okoline (okruženja) neurona, *intenzitetu tih veza* (iznosu težinskih faktora), te o *pragu osjetljivosti* koji stanje neurona mora dosegnuti prije nego što "ispali impuls" preko svog izlaza u okolinu neurona koju čine ostali neuroni umjetne neuronske mreže. Danas se neuronske mreže primjenjuju se u područjima kao što su: klasificiranju signala, kontroli vođenju sustava, robotici, području identifikacije složenih dinamičkih sustava, u medicinskoj dijagnostici*, itd.*

Povratna (engl. Backpropagation -BP) neuronska mreža

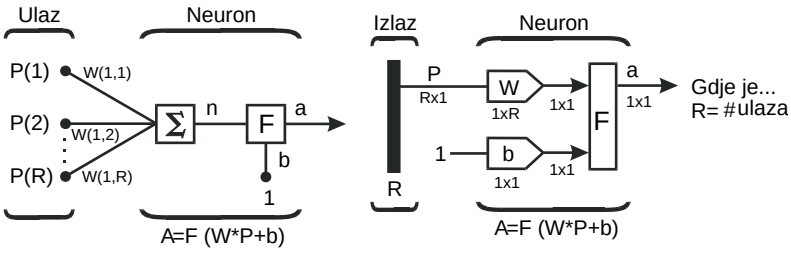
Na slici 3.4 prikazan je model umjetnog neurona (engl. *neuron, node, unit*) koji oponaša osnovne funkcije biološkog neurona. Tijelo biološkog neurona zamjenjuje sumator, ulogu dendrita preuzimaju ulazi u sumator ($x_1,x_2…x_n$), izlaz iz sumatora je akson umjetnog neurona, a uloga praga osjetljivosti bioloških neurona preslikava se na tzv. *aktivacijske funkcije* $f(net)_j$. Funkcije sinaptičke veze biološkog neurona s njegovom okolinom preslikavaju se na *težinske faktore* ili *težine* (w_{ij}), preko kojih se i ostvaruje veza umjetnog neurona s njegovom okolinom. Težinski faktori ili težine povezuju izlaze iz okoline neurona, odnosno izlaze drugih neurona s ulazima sumatora. Težinski faktor ili težina može biti nula, pozitivan ili negativan broj. Kada je težina jednaka nuli, to znači

da ne postoji veza neurona sa okolinom. Izlaz iz sumatora producira izlaz umjetnog neurona y_j.



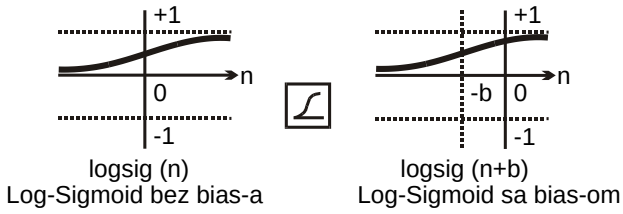
Slika 3.4. Model umjetnog neurona [58]

Jednostavni neuron sa *R* ulaza prikazan je na slici 3.5. prema notaciji iz [9]. Svakom ulazu od *P(1)* do *P(R)* dodijeljena je odgovarajuća težina ili težinski faktor *w(i,j)*. Suma produkta pojedinačnih iznosa ulaza i težina sa iznosom *bias* vektora *b* predstavlja ulaz u aktivacijsku funkciju *F* koja generira izlaz *a*. Aktivacijska funkcija može biti bilo koja diferencijabilna funkcija.



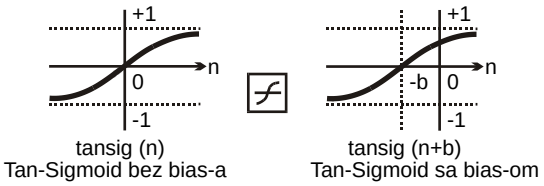
Slika 3.5. Prikaz backpropagation modela neurona [58]

Kod višeslojnih neuronskih mreža najčešće se koriste aktivacijske funkcije *log-sigmoid* tipa, slika 3.6.



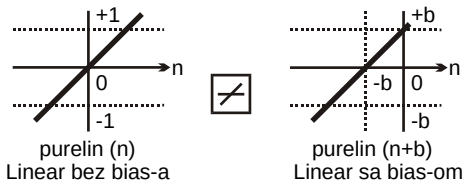
Slika 3.6. Aktivacijske funkcije log-sigmoid tipa [58]

Aktivacijske funkcije generiraju izlaze iz mreže koji imaju vrijednosti između 0 i 1 i to za ulaze u mrežu čiji se iznosi kreću iz '-' beskonačnosti do '+' beskonačnosti. Kao što je pokazano na slici 3.6. *log-sigmoid* aktivacijske funkcije mogu biti sa ili bez *bias-a*. Unošenjem *bias* vrijednosti u aktivacijsku funkciju postiže se viši stupanj slobode učenja mreže, te se povećava vjerojatnost mreže za pronalaženje prihvatljivog rješenja uz istovremeno smanjenje broja epoha potrebnih za učenje (treniranje) mreže.



Slika 3.7. Aktivacijske funkcije tan-sigmoid tipa [58]

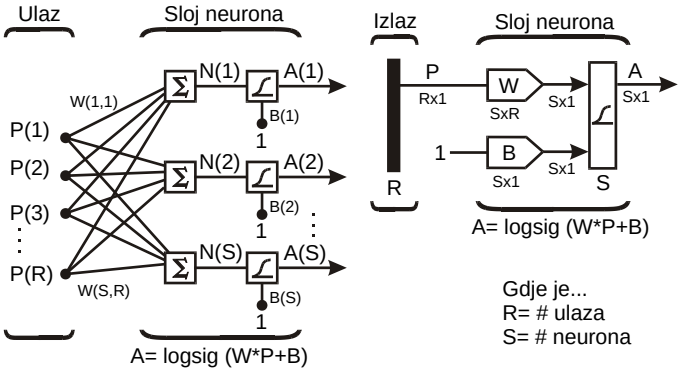
Alternativno, kod neuronskih mreža sa *povratnim prostiranjem pogreške* (engl. *backpropagation,* u daljnjem tekstu *BP*) mogu se upotrijebiti i aktivacijske funkcije *tan-sigmoid* tipa, prikazanih na slici 3.7. ili *lin* tipa prikazanih na slici 3.8., također sa *bias*-om ili bez *bias*-a.



Slika 3.8. Aktivacijske funkcije lin tipa [58]

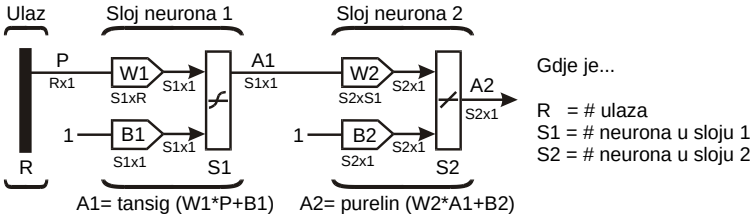
Prema [58], ako se u posljednjem sloju *BP* neuronske mreže nalazi aktivacijska funkcija *sigmoidalnog* tipa, tada su izlazi iz mreže ograničeni na vrijednosti u području 0 do +1. Ako pak je *linernog* tipa, tada izlazi iz mreže mogu poprimiti bilo koju vrijednost. Osim što su diferencijabilne ove aktivacijske funkcije su i monotono rastuće, što znači da iznos funkcije povećava proporcionalnu povećanju nezavisnih varijabli. Ovo su najčešća tri tipa aktivacijskih funkcija koje se upotrebljavaju kod *BP* neuronskih mreža, ali kao što je već spomenuto mogu se primijeniti bilo koje diferencijabilne funkcije.

Na slici 3.9. prikazana je jednoslojna unaprijedna backpropagation mreža sa *R* ulaza sa aktivacijskom funkcijom sigmoidalnog tipa. Unaprijedne *BP* mreže najčešće imaju jedan ili više skrivenih slojeva sa sigmoidalnim aktivacijskim funkcijama nakon čega obično slijedi izlazni sloj linearnog tipa. Više slojeva sa nelinearnim aktivacijskim funkcijama omogućava da mreža nauči nelinearne i linearne interakcije između ulaznih i izlaznih vektora. Linearni izlazni sloj omogućava da izlaz iz mreže poprimi vrijednost izvan intervala –1 do +1. S druge strane, ako želimo da je izlaz iz mreže u granicama 0 i 1 tada izlazni sloj mreže treba bit sastavljen od sigmoidalnih aktivacijskih funkcija.



Slika 3.9. Shema jednoslojne unaprijedne BP mreže [58]

Na slici 3.10. prikazana je shema uobičajene *unaprijedne BP* mreže sa linearnim izlaznim slojem čija je najčešća primjena pri aproksimaciji funkcija.

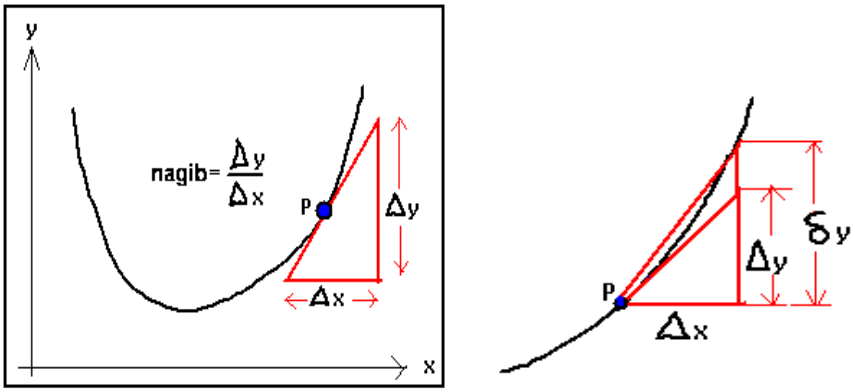


Slika 3.10. Shema dvoslojne unaprijedne backpropagation mreže [58]

Na prethodnim slikama '#' predstavlja skraćenicu za riječ '*broj'.* Oznaka *R* predstavlja broj ulaza, *B* je bias vektor, *F* je oznaka za aktivacijsku funkciju, *W* je vektor težina, *S* je broj neurona u pojedinom sloju i *A* (engl. *actual*) je vektor izračunatih izlaznih vrijednosti iz mreže.

Algoritam učenja povratne mreže

Većinom se algoritmi učenja *BP* mreža temelje na primjeni *algoritma najstrmijeg pada* (engl. *gradient or steepest descent*) za pronalaženje minimuma funkcije. Pretpostavimo da je y=f(x) čiji točan oblik nije poznat, te se želi pronaći x-koordinata gdje funkcija ima minimum, slika 3.11. Tada se izraz *nagib=Δy/Δx* koji prolazi bilo kojom točkom te funkcije naziva gradijent (nagib) tangente.



Slika 3.11. Gradijent ili nagib tangente funkcije y=f(x) [58]

Ako je Δx mali tada je i Δy mali tako da imamo:

$$\delta y \approx \Delta y = \frac{\Delta y}{\Delta x}\Delta x\,, \tag{3.27}$$

$$\delta y \approx nagib \times \Delta x\,. \tag{3.28}$$

Ako stavimo da je

$$\Delta x = -\alpha \times nagib\,, \tag{3.29}$$

gdje je $\alpha > 0$ te dovoljno mali da je $\delta_y \approx \Delta y$ tada je

$$\delta y \approx -\alpha (nagib)^2\,. \tag{3.30}$$

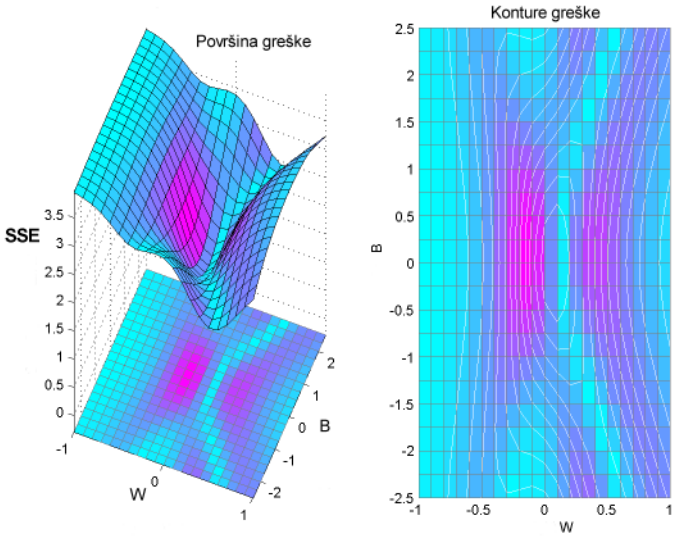
Kada je $\delta y < 0$, nagib teži nuli te se na taj način približavamo točki gdje funkcija ima minimum. Pri učenju neuronske mreže primjenom algoritma najstrmijeg pada izračunava se nagib ili gradijent *G* funkcije E=f(w) koju nazivamo pogreška (engl. *loss, objective, error*). *E* je zapravo funkcija težina w_{ij} za pojedini p_i ulaz u mrežu, slika 3.12.



Slika 3.12. Funkcije greške E=f(w) [58]

E je definirana kao **suma kvadrata pogreške** - ***SSE*** (engl. *sum squared error*) između izlaznih (t_i) i ulaznih (y_i) vektora podataka:

$$E = \frac{1}{2}\sum_{i=1}^{n}(t_i - y_i)^2 \equiv SSE\,, \tag{3.31}$$



Slika 3.13. Prikaz SSE površine kao funkcije od W i B [58]

Na slici 3.13. je prikazan primjer trodimenzionalnog i dvodimenzionalnog grafa *SSE* funkcije u zavisnosti od vektora težina *W* i bias vektora *B*. Pretpostavimo da želimo

naučiti višeslojnu *BP* mrežu tako da bude aproksimator nepoznate funkcije i to na temelju parova vektora (*x,t*). Vektor *x* predstavlja uzorke ulaznih podataka dok vektor *t* predstavlja uzorke izlaznih podataka. Kao što se može zaključiti iz slike 3.13, izračunavanje gradijenta *G,* odnosi se na svaki vektor težine w_{ij} unutar mreže, te nam govori o tome kako promjena težine utječe na ukupnu grešku E. U nastavku je prikazano izračunavanje gradijenta jediničnog uzorka za slučaj linearne a zatim i nelinearne *BP* neuronske mreže.

Izraz (3.31) također možemo napisati u obliku:

$$E = \sum_p E^p , \qquad E^p = \frac{1}{2}\sum_o (t_o^p - y_o^p)^2 , \tag{3.32}$$

gdje *p* predstavlja jednu točku iz skupa za učenje, a *o* područje (engl. *range*) skupa podataka za učenje, tada je izraz za gradijent *G*:

$$G = \frac{\partial E}{\partial w_{ij}} = \frac{\partial}{\partial w_{ij}} \sum_p E^p = \sum_p \frac{\partial E^p}{\partial w_{ij}} . \tag{3.33}$$

Zbog jednostavnosti izraza u nastavku će se izostaviti oznaka *p*. Ako gradijent *G* uz pomoć *lančanog pravila* (engl. *chain rule*) rastavimo na dva faktora dobivamo:

$$\frac{\partial E}{\partial w_{oi}} = \frac{\partial E}{\partial y_o} \frac{\partial y_o}{\partial w_{oi}} . \tag{3.34}$$

Prvi faktor dobije se derivacijom

$$\frac{\partial E}{\partial y_o} = -(t_o - y_o) . \tag{3.35}$$

Ako stavimo da je,

$$y_o = \sum_j w_{oj} y_j , \tag{3.36}$$

tada je drugi faktor:

$$\frac{\partial y_o}{\partial w_{oi}} = \frac{\partial}{\partial w_{oi}} \sum_j w_{oj} y_j = y_i . \tag{3.37}$$

Povezujući izraze (3.34 – 3.37) dobivamo konačan izraz za izračun gradijenta *G* za linearnu mrežu:

$$G = \frac{\partial E}{\partial w_{oi}} = -(t_o - y_o) y_i . \tag{3.38}$$

Ako uvedemo da je

$$net_i = \sum_j w_{ij} y_j \tag{3.39}$$

i

$$\delta_j = -\frac{\partial E}{\partial net_j}, \tag{3.40}$$

te da je

$$\Delta w_{ij} = -\frac{\partial E}{\partial w_{ij}}. \tag{3.41}$$

Tada izraz za gradijent prema (3.33) rastavljen na dva faktora pomoću lančanog pravila, poprima sljedeći oblik:

$$\Delta w_{ij} = -\frac{\partial E}{\partial net_i}\frac{\partial net_i}{\partial w_{ij}}. \tag{3.42}$$

Prvi faktor predstavlja grešku a drugi faktor je:

$$\frac{\partial net_i}{\partial w_{ij}} = \frac{\partial}{\partial w_{ij}} \sum_{k \in A_i} w_{ik} y_k = y_j. \tag{3.43}$$

Tada je konačan izraz za izračun gradijenta težine w_{ij} za nelinearnu *BP* mrežu:

$$\Delta w_{ij} = \delta_i y_j. \tag{3.44}$$

Za izračunavanje gradijenta prema izrazu (3.44) potrebno je poznavati iznose aktivnosti i greške svih neurona u mreži.

Aktivnost y_i i-tog neurona mreže izračunava se prema izrazu:

$$y_i = f_i(\sum_j w_{ij} y_j). \tag{3.45}$$

Pogreška *E* svih neurona mreže izračunava se prema izrazu,

$$E = \frac{1}{2}\sum_o (t_o - y_o)^2, \tag{3.46}$$

a pogreška izlaznog neurona prema izrazu

$$\delta_o = t_o - y_o. \tag{3.47}$$

Za izračun greške neurona u skrivenom sloju primjenjuje se algoritam pod nazivom *algoritam povratnog prostiranja pogreške* (engl. *error backpropagation*).

Tada je pogreška δ_j , *j*-tog neurona skrivenog sloja jednaka:

$$\delta_j = -\sum_j \frac{\partial E}{\partial net_i}\frac{\partial net_i}{\partial y_j}\frac{\partial y_j}{\partial net_j} \tag{3.48}$$

Prvi faktor unutar izraza (3.48) predstavlja grešku, a drugi faktor je iznos težine w_{ij}:

$$\frac{\partial net_i}{\partial y_j} = \frac{\partial}{\partial y_j} \sum_{k \in A_i} w_{ik} y_k = w_{ij} . \tag{3.49}$$

Treći faktor je derivacija aktivacijske funkcije *j*-tog neurona skrivenog sloja:

$$\frac{\partial y_j}{\partial net_i} = \frac{\partial f_j(net_j)}{\partial net_j} = f_j^{'}(net_j) . \tag{3.50}$$

Konačni izraz za izračun greške *j*-tog neurona skrivenog sloja je:

$$\delta_j = f_j^{i}(net_j) \sum_{i \in P_j} \delta_i w_{ij} . \tag{3.51}$$

Kod unaprijednih višeslojnih *BP* mreža svaki je neuron iz promatranog sloja spojen sa svakim neuronom iz sljedećeg sloja. Prikladnosti radi, često se *BP* algoritam iskazuje u matričnoj formi te se ulazi u mrežu, *bias*-i (*b*), težine (*w*), aktivacijske funkcije, i pogreške prikazuju u vektorskoj formi za svaki pojedini sloj (*L*). Tada se *BP* algoritam sažeto može prikazati kako slijedi:

1. Inicijalizacija ulaznog sloja:

$$\vec{y}_o = \vec{x} . \tag{3.52}$$

2. Rasprostiranje aktivnosti unaprijed: l=1,2,…,L

$$\vec{y_l} = f_l(W_l \vec{y}_{l-1} + \vec{b_l}) , \tag{3.53}$$

gdje je *l*-ulazni sloj, *L*-izlazni sloj, a *b*-bias vektor težina

3. Izračunavanje greške izlaznog sloja

$$\vec{\delta}_L = \vec{t} - \vec{y}_L . \tag{3.54}$$

4. Povratno prostiranje greške: od L-1, L-2,…, l

$$\vec{\delta}_l = (W_{l+1}^T \delta_{l+1}) f_l^{'}(\vec{net_l}) , \tag{3.55}$$

gdje je *T* označava transponiranu matricu.

5. Promjena težina i bias-a:

$$\Delta W_l = \vec{\delta_l} y_{l-1}^T , \qquad \Delta \vec{b_l} = \vec{\delta}_l . \tag{3.56}$$

Modifikacije povratnog algoritma

Danas u primjeni nalazimo mnoge modifikacije prethodno opisanog osnovnog algoritma *povratnog prostiranja pogreške.* Međusobno se ti algoritmi razlikuju prema brzini konvergencije prilikom traženja rješenja, a razlike mogu biti čak do nekoliko stotina puta. Razlikujemo dvije grupe algoritama kod unaprijednih statičkih *BP* mreža. Tzv. *spore* i *brze* algoritme. U spore algoritme npr. spadaju:

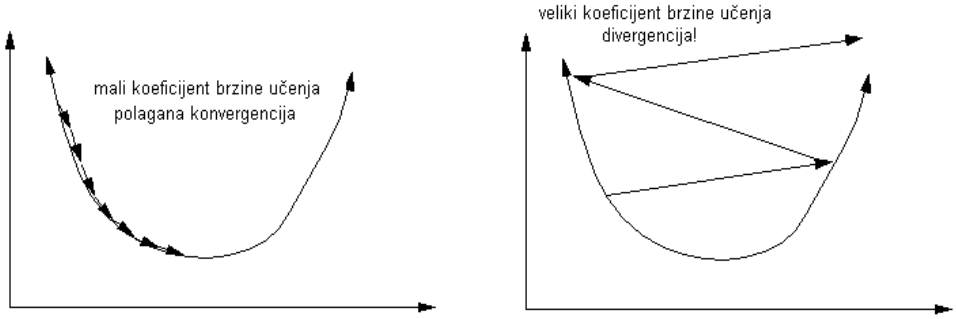
- *BP*-algoritam sa konstantnim koeficijentom brzine učenja (engl. *learning rate*) - μ,
- *BP*-algoritam sa konstantnim koeficijentom momentuma (engl. *momentum*) - α

U brze npr. spadaju:
- *BP*-algoritam sa varijabilnim ili adaptivnim (engl. *adaptive*) koeficijentom brzine učenja,
- *Levenberg-Marquardt* algoritam

Kod *brzih* algoritma primjenjuju se tzv. *heurističke tehnike* koje su proizišle iz analize uspješnosti učenja standardnih '*stepeest descent*' algoritama. Kod *Levenberg-Marquard* algoritma primjenjuju se klasične metode numeričke optimizacije. U nastavku su ukratko opisane nabrojane modifikacije algoritama.

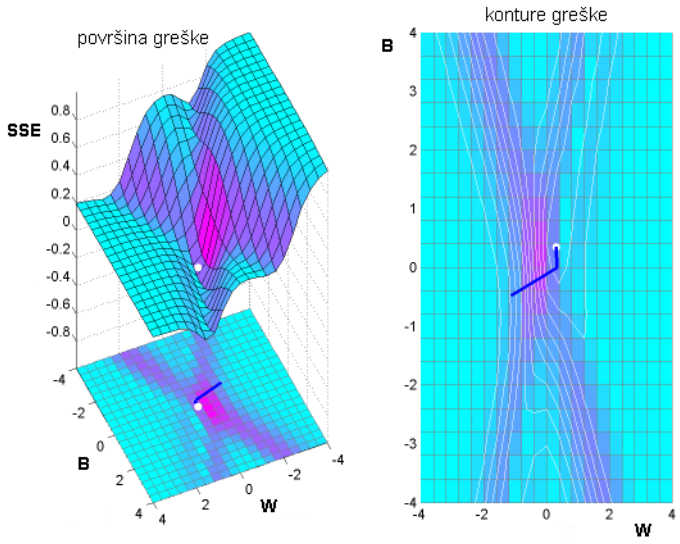
Povratni algoritam s konstantnim koeficijentom brzine učenja - μ

U osnovnoj verziji *BP* algoritma, vektora težina i *bias* vektor pomiču se u smjeru negativnog gradijenta funkcije uspješnosti. Kod ove modifikacije algoritma dodan je tzv. *koeficijent brzine učenja* (engl. *learning rate*) - μ, kojim određujemo iznos promjene vektora težina za svaki korak iteracije. Na slici 3.14. prikazano je značenje parametra učenja. Za slučaj malih vrijednosti koeficijenta, algoritmu će trebati puno vremena za konvergenciju, dok će za slučaj velikih vrijednosti algoritam postati nestabilan te će divergirati.



Slika 3.14. Značenje koeficijenta brzine učenja [58]

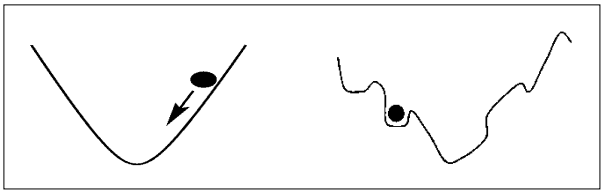
Na slici 3.15. prikazan je izgled površine greške (*SSE)* sa putanjom konvergencije uz primijenjeni koeficijent brzine učenja μ=2. Sa slike je očito da je μ je pravilno odabran jer nije došlo do divergencije. Za učenje su početni (inicijalni) iznosi vektora težina i bias vektora odabrani generatorom slučajnih brojeva. Koeficijent brzine učenja određuje učitelj, a njegova se vrijednost najčešće kreće između 10^{-3} i 10, prema [58].



Slika 3.15. Primjer površine greške sa putanjom konvergencije [58]

Povratni algoritam s konstantnim koeficijentom momentuma - α

Kod algoritma najstrmijeg pada traženje minimuma započinje u nekoj točki površine funkcije greške, pokušavajući se gibati u smjeru globalnog minimuma (engl. *global minimum*). Ako je funkcija površine greške slična onoj na slici 3.15., tada je ta situacija pojednostavljeno prikazana lijevom skicom na slici 3.16. Vidi se da će nas svaki korak u silaznom smjeru te funkcije dovesti bliže globalnom minimumu. Kod realnih problema površina funkcije greške puno je složenija, te više nalikuje funkciji prikazanoj desnom skicom na slici 3.16. Kao što se vidi ovdje postoji više lokalnih minimuma (engl. *local minima*), a u jednom od njih može završiti crna kuglica te je napredovanje u smjeru globalnog minimuma završeno. Kao što se vidi, daljnje napredovanje kuglice moguće je njenim uspinjanjem naviše prije spuštanja prema globalnom minimumu.



Slika 3.16. Problem lokalnog i globalnog minimuma [58]

Zbog izbjegavanja lokalnog minimuma, te da bi se ubrzao proces učenja, u algoritam najstrmijeg pada uvodi se *koeficijent količine gibanja* ili *momentum* (engl. *momentum*) - α te je tada modificirani izraz BP-algoritma sljedeći:

$$\Delta w_{ij}(n) = \mu_i \delta_i y_j + \alpha \Delta w_{ij}(n-1), \tag{3.57}$$

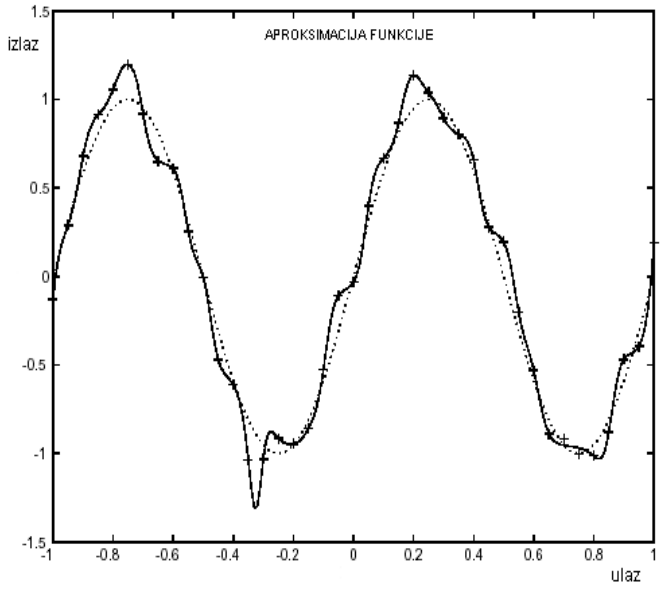
gdje *n* označava trenutnu promjenu težina, a *(n-1)* prethodnu promjenu težina u procesu učenja. Vrijednost koeficijenta momentuma također određuje učitelj, a njegova vrijednost obično s kreće u intervalu: $0 < \alpha < 1$. Ako je iznos momentuma 0 promjena težina temelji se na iznosu promjene gradijenta, a ako je iznos 1 promjena je jednaka iznosu prethodne promjene težina, a promjena gradijenta se izostavlja. U pravilu bi trebalo smanjivati koeficijent brzine učenja ako se koeficijent momentuma povećava, a u protivnom se može dogoditi da se sa primjenom velikog koeficijenta brzine učenja i velikog koeficijenta momentuma 'preskoči' globalni minimum.

Algoritam s varijabilnim ili adaptivnim koeficijentom brzine učenja - μ

Kod standardnog algoritma najstrmijeg pada, koeficijent brzine učenja je konstantan. Učinak algoritma vrlo je osjetljiv na pravilan izbor koeficijenta brzine učenja. Ako je izabran previsok koeficijent, algoritam može početi oscilirati te postati nestabilan, a ako je prenizak biti će potrebno više vremena za konvergenciju. Određivanje iznosa koeficijenta učenja prije procesa učenja nije od velikog praktičnog značaja, zbog toga što tako određeni koeficijent vjerojatno nije optimalni u cijelom području funkcije površine greške. Naime, kod složenih problema funkcija površine greške može imati puno lokalnih minimuma kod kojih može doći do divergencije tj. nestabilnosti algoritma. Ovi se problemi mogu izbjeći a time i povećati učinak standardnog algoritma, ako se dozvoli promjena (adaptacija) koeficijenta brzine u toku samog procesa učenja. Najčešći algoritam za adaptaciju koeficijenta brzine učenja je tzv. *bold driver* algoritam. Algoritam radi na sljedeći način: nakon svake epohe učenja uspoređuje se pogreška mreže *E(n)* sa iznosom greške u prethodnoj epohi *E(n-1)*. Ako se pogreška smanjila, povećava se koeficijent - μ u određenom postotnom iznosu (1% do 5%), a ako se pogreška povećala (u iznosu npr. 10^{-10}), poništava se zadnja promjena težina, a koeficijent - μ se smanjuje za 50%. Primjenom opisanih pravila dobiva se optimalan koeficijent brzine učenja za *'lokalni teren'* na funkciji površine greške. Varijabilni ili adaptivni koeficijent brzine učenja također se kombinira sa koeficijentom momentuma.

Metode za poboljšanje aproksimacije

Jedan od problema koji se javlja kod učenja ili treninga neuronske mreže je problem tzv. *overfitting*-a. Naime, pogreška na ulazno-izlaznom skupu podataka je nakon učenja vrlo mala, a kad se na ulaz mreže dovedu novi podaci na kojima mreža nije učila (trenirala), pogreška je jako velika. Kaže se da mreža nema dovoljno *'znanja'* za generalizaciju ili aproksimaciju novih situacija. Na slici 3.17. prikazan je primjer *overfitting*-a pri aproksimaciji sinus funkcije (prikazane točkicama) neuronskom mrežom (prikazane punom linijom) gdje se vidi da mreža ne generalizira dobro nove ulazne podatke (prikazane znakom '+').



Slika 3.17. Problem overfittinga mreže [58]

Jednostavni način za poboljšanje aproksimacije je taj da imamo dovoljan skup ulazno-izlaznih podataka za učenje mreže. Prema nekim literaturnim izvorima, minimalan tj. dovoljan broj skupova je *10* broj_ulaznih_varijabli*. U pravilu, što je veći broj skupova, to će biti bolja aproksimacija novih podataka koji nisu bili u skupu za učenje. Sljedeći način je pravilan izbor veličine mreže s obzirom na broj slojeva i broj neurona u svakom od njih. Ovdje je stvar puno složenija te se jedino eksperimentiranjem može doći do pravilnog izbora broja slojeva i broja neurona. I treći način predstavljaju metode koje su ugrađene u pojedine algoritme. Te metode su: *regularizacija* (engl. *regularization*) i *rano zaustavljanje* (engl. *early Stopping)*.

Metoda regularizacije

Najčešća funkcija uspješnosti (engl. *performance function*) za učenje unaprijednih neuronskih mreža, kao što je već spomenuto, je suma kvadrata pogreške ili *srednja suma kvadrata pogreške - MSE* (engl. *mean square error*):

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(e_i)^2 = \frac{1}{N}\sum_{i=1}^{N}(t_i - a_i)^2 \quad . \tag{3.58}$$

gdje je t_i – ciljni ili izlazni vektor, a a_i –procijenjeni vektor (engl. *actual*) od strane naučene neuronske mreže. Aproksimaciju ili generalizaciju mreže moguće je unaprijediti ako u funkciju uspješnosti proširimo za dva člana: prvi član je tzv. *performance ratio-*γ, a drugi član je *suma kvadrata pogreške težina i biasa - msw.* Izraz (3.58) tada izgleda:

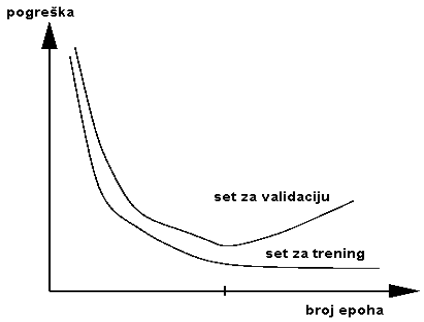
$$msereg = \gamma mse + (1-\gamma)msw\,, \tag{3.59}$$

$$msw = \frac{1}{n}\sum_{j=1}^{n} w_j^2\,. \tag{3.60}$$

Problem kod primjene metode regularizacije je taj što je teško odrediti optimalan iznos za *performance ratio-*γ. Ako ovaj parametar postavimo na veliku vrijednost postoji opasnost od pojave *overfitting-a.* Ako je pak premali, vjerojatno je da niti trening skup podataka neće biti dobro aproksimiran, a time naravno niti novi skupovi koji nisu bili u procesu učenja.

Metoda ranog zaustavljanja

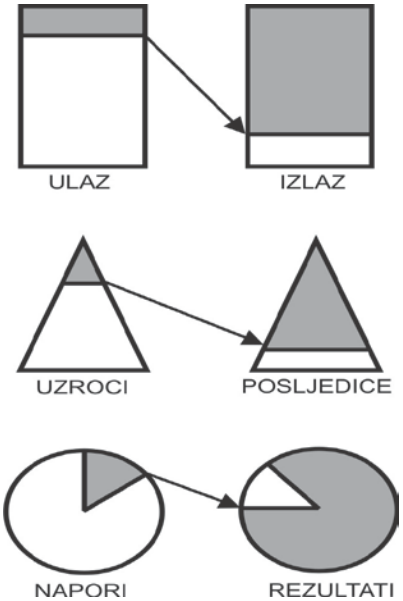
Kod ove tehnike za poboljšanje aproksimacije ulazni skup podataka dijeli se na dva dijela. Prvi skup je trening skup, pomoću kojeg se izračunava gradijent te se nakon toga mijenjaju vektori težina i bias vektori. Drugi skup je skup za validaciju (engl. *validation skup*) te se preko njega 'nadgleda' pogreška u procesu učenja. Pogreška validacije se u početnoj fazi učenja smanjuje isto kao i pogreška na trening skupu podataka. Kada se u procesu učenja počinje događati '*overfitting*', pogreška validacijskog skupa počinje rasti. Ako se pogreška validacije značajno poveća za određeni broj iteracija, zaustavlja se proces učenja te se za daljnje proračune (aproksimacije za nove skupove) uzimaju vektori težina i bias vektora kod kojih je pogreška validacije bila minimalna. Na slici 3.18. dan je shematski prikaz *early stoping* metode.



Slika 3.18. Shematski prikaz early stopping metode [58]

3.2.5. Pareto analiza

Primjena Paretovog načela, odnosno načela 80/20 prikazanog na slici 3.19., ima veliki značaj u kontekstu upravljanja kvalitetom s obzirom na svoju univerzalnost, odnosno primjena istog nije ograničena samo na područje ekonomije [59]. Paretovo načelo predstavlja jedan od sedam osnovnih (menadžerskih) alata u kontekstu upravljanja kvalitetom [60].



Slika 3.19. – Načelo 80/20 [59]

Na temelju Paretovog načela proizlazi Pareto analiza koja predstavlja metodu za klasificiranje uzroka problema prema stupnju njihovog značenja, odnosno postupak odabiranja prioritetnih problema za provođenje korektivnih radnji.

3.2.5.1. Pareto analiza u funkciji izrade prediktivnog modela tehničke ispravnosti vozila temeljem tehničkih pregleda

U svim (tadašnjim) stanicama za tehnički pregled vozila (146 stanica), tijekom 2009. godine, na teritoriju Republike Hrvatske obavljeno je ukupno 1.781.052 redovnih tehničkih pregleda svih kategorija vozila [61].

Od prethodno navedenog broja, na 373.706 vozila (20,98 %) je tijekom tehničkog pregleda utvrđena jedna ili više tehničkih neispravnosti što je za posljedicu imalo ne-ovjeravanje valjanosti prometne dozvole predmetnih vozila, odnosno njihov "ne-prolaz/pad" na tehničkom pregledu [61].
Predmet istraživanja ovog doktorskog rada su osobni automobili [22], odnosno motorna vozila koja imaju najmanje četiri kotača i koja se koristi za prijevoz putnika, s ne više od osam sjedala u koja nije uključeno sjedalo za vozača (M1 kategorija vozila) [2].

Izrada prediktivnog modela tehničke ispravnosti vozila temeljem tehničkih pregleda osobnih automobila otvara mogućnost izrade prediktivnih modela tehničke ispravnosti vozila bilo koje kategorije, s obzirom da se radi o univerzalnom načelu, odnosno univerzalnoj primijenjenoj metodologiji.

U svim stanicama za tehnički pregled vozila na teritoriju Republike Hrvatske, tijekom 2009. godine, obavljeno je ukupno 1.354.239 redovnih tehničkih pregleda osobnih automobila. Na 307.372 osobnih automobila je tijekom tehničkog pregleda utvrđena jedna ili više tehničkih neispravnosti što je za posljedicu imalo ne-ovjeravanje valjanosti prometne dozvole predmetnih vozila, odnosno njihov "ne-prolaz/pad" na tehničkom pregledu [62]. U sklopu provedenih redovnih tehničkih pregleda osobnih automobila tijekom 2009. godine pregledane su 143 marke istih (u rasponu od 1 do 229.258 osobnih automobila po pojedinoj marki), odnosno 609 tipova osobnih automobila (u rasponu od 1 do 115.535 osobnih automobila po pojedinom tipu) predmetnih trgovačkih marki [62]. S obzirom na veliki ukupni broj osobnih automobila pregledanih tijekom 2009. godine u sklopu redovnih tehničkih pregleda (1.354.239 osobnih automobila), odnosno s obzirom na veliki broj trgovačkih marki (143) i tipova osobnih automobila (609), bilo je potrebno reducirati broj trgovačkih marki i tipova osobnih

automobila koji su obrađeni u sklopu ovog istraživanja, radi preglednosti dijagramskog prikaza koji predstavlja rezultat Pareto analize.

Tablica 3.1. Trgovačke marke osobnih automobila sukladno kriteriju 1

Trgovačka marka osobnih automobila	Broj osobnih automobila na kojima je obavljen redovni TP
VOLKSWAGEN	229.258
OPEL	187.249
RENAULT	139.717
FIAT	121.252
FORD	71.839
ŠKODA	68.666
PEUGEOT	67.036
CITROEN	59.495
MERCEDES	46.482
AUDI	44.261
HYUNDAI	35.912
MAZDA	30.672
TOYOTA	28.394
SEAT	28.380
BMW	24.146
ZASTAVA	21.273
DAEWOO	19.685
KIA	18.436
HONDA	13.446
NISSAN	13.124
CHEVROLET	11.500
ALFA ROMEO	11.419
SUZUKI	10.232
UKUPNO:	**1.301.874**

Reduciranje broja trgovačkih marki i tipova osobnih automobila nije ugrozilo signifikantnost uzorka koji je obrađen u sklopu ovog istraživanja, u smislu njegovog udjela u ukupnom skupu, s obzirom da su za potrebe reduciranja definirana dva kriterija koja su opisana u nastavku teksta.

“Kriterij 1 – marke osobnih automobila” – one trgovačke marke koje su zastupljene s više od 10.000 pregledanih osobnih vozila u sklopu redovnog tehničkog pregleda tijekom 2009. godine. Kao što je prikazano u tablici 3.1. radi se o 23 trgovačke marke sa ukupnim brojem od 1.301.874 pregledanih osobnih vozila, što predstavlja 96,1 %

od ukupnog broja pregledanih osobnih automobila, čime je potvrđena signifikantnost uzorka u odnosu na ukupni skup.

“Kriterij 2 – tipovi osobnih automobila” – oni tipovi osobnih automobila, u okviru kriterija 1, kod kojih je redovnim tehničkim pregledom tijekom 2009. godine utvrđeno više od 150 tehnički neispravnih osobnih automobila (kojima nije ovjerena valjanost prometne dozvole, odnosno koji su „pali” na tehničkom pregledu) po pojedinom tipu.

Tablica 3.2. Tipovi osobnih automobila sukladno kriteriju 2

Tip osobnog automobila	Broj osobnih automobila na kojima je obavljen redovni TP	Broj tehnički neispravnih osobnih automobila koji su „pali“ na TP-u	Tip osobnog automobila	Broj osobnih automobila na kojima je obavljen redovni TP	Broj tehnički neispravnih osobnih automobila koji su „pali“ na TP-u
GOLF	115.535	38.788	SERIJA 3	6.721	872
PUNTO	60.134	14.710	BERLINGO	5.289	850
ASTRA	65.534	13.451	KANGOO	4.907	799
CORSA	49.415	9.717	SWIFT	4.441	787
VECTRA	30.285	9.428	A6	5.977	726
CLIO	48.780	9.238	156	3.451	715
PASSAT	42.120	9.185	YARIS	9.447	685
POLO	38.593	7.503	THALIA	3.583	665
FIESTA	20.671	6.045	A3	4.535	614
MEGANE	25.574	4.085	STILO	4.091	586
FABIA	23.454	3.400	SCENIC	5.360	579
OCTAVIA	21.849	3.126	MICRA	3.250	527
206	19.930	3.039	KLASA C	4.854	512
TWINGO	15.672	2.971	KLASA E	5.003	430
ACCENT	15.413	2.782	GETZ	5.093	419
SAXO	9.677	2.446	207	5.257	377
323	10.499	2.304	SERIJA 5	3.628	372
MONDEO	8.260	2.233	C4	4.278	326
IBIZA	10.977	2.105	C5	3.377	325
XSARA	12.702	2.067	ACCORD	3.025	321
A4	12.152	1.998	ZAFIRA	3.203	288
LAGUNA	9.712	1.849	6	4.488	246
CORDOBA	8.854	1.846	KLASA A	2.272	227
FOCUS	12.515	1.813	SPARK	3.027	200
CIVIC	6.959	1.502	407	3.172	171
307	12.953	1.412	AVEO	3.301	170
C3	11.403	1.209	3	4.093	157
COROLLA	7.870	995	RAV 4	3.186	150
			UKUPNO:	**849.801**	**174.343**

Kao što je prikazano u tablici 3.2. radi se o 56 tipova sa ukupnim brojem od 849.801 pregledanih osobnih vozila, te sa ukupnim brojem od 174.343 osobnih automobila koji su „pali“ na tehničkom pregledu s obzirom na utvrđene tehničke neispravnosti, što predstavlja 56,7 % od ukupnog broja osobnih automobila (307.372) koji su „pali“ na tehničkom pregledu u 2009. godini, čime je potvrđena signifikantnost uzorka u odnosu na ukupni skup.

Granične vrijednosti iz oba kriterija (trgovačke marke koje su zastupljene s više od 10.000 pregledanih osobnih vozila i tipovi osobnih automobile kod kojih je utvrđeno više od 150 tehnički neispravnih osobnih automobile) Hrvatski autoklub koristi u kontekstu stručnog nadzora nad radom STP, u sklopu provođenja temeljnih (engl. *Default*) pretraga [62].

Tipovi osobnih automobila iz tablice 3.2. predmet su provedene Pareto analize i to u skladu sa “kriterijem 2 – tipovi osobnih automobila”, odnosno u pogledu tehničke neispravnosti utvrđene u sklopu redovnih tehničkih pregleda provedenih tijekom 2009. godine.

U skladu s prethodno iznijetim očekivani rezultat (u obliku dijagramskog prikaza) Pareto analize jest detektiranje manjeg broja tipova osobnih automobila (utjecajnih čimbenika), koji su u određenom kontekstu – na primjer, u pogledu tehničke ispravnosti – od najvećeg značaja.

U pogledu istraživanja koje je predmet ovog doktorskog rada navedene utjecajne čimbenike predstavljaju tipovi osobnih automobila – tablica 3.3.

Prema [59] i [60], a sukladno s načelom *80/20* koje proizlazi iz Pareto analize, za graničnu postotnu vrijednost svih tipova osobnih automobile iz tablice 3.3. određena je granična vrijednost od 80 %.

Tablica 3.3. – Tipovi osobnih automobila - utjecajni čimbenici

Tip osobnog automobila	Broj tehnički neispravnih automobila koji su „pali“ na TP-u	Postotni udjel broja tehnički neispravnih automobila koji su „pali“ na TP-u [%]	Kumulativni postotni udjel broja tehnički neispravnih automobila koji su „pali“ na TP-u [%]
GOLF	38.788	22,25	22,25
PUNTO	14.710	8,44	30,69
ASTRA	13.451	7,72	38,40
CORSA	9.717	5,57	43,97
VECTRA	9.428	5,41	49,38
CLIO	9.238	5,30	54,68
PASSAT	9.185	5,27	59,95
POLO	7.503	4,30	64,25
FIESTA	6.045	3,47	67,72
MEGANE	4.085	2,34	70,06
FABIA	3.400	1,95	72,01
OCTAVIA	3.126	1,79	73,81
206	3.039	1,74	75,55
TWINGO	2.971	1,70	77,25
ACCENT	2.782	1,60	78,85
SAXO	2.446	1,40	80,25
323	2.304	1,32	81,57
MONDEO	2.233	1,28	82,85
IBIZA	2.105	1,21	84,06
XSARA	2.067	1,19	85,25
A4	1.998	1,15	86,39
LAGUNA	1.849	1,06	87,45
CORDOBA	1.846	1,06	88,51
FOCUS	1.813	1,04	89,55
CIVIC	1.502	0,86	90,41
307	1.412	0,81	91,22
C3	1.209	0,69	91,92
COROLLA	995	0,57	92,49
SERIJA 3	872	0,50	92,99
BERLINGO	850	0,49	93,48
KANGOO	799	0,46	93,93
SWIFT	787	0,45	94,39
A6	726	0,42	94,80
156	715	0,41	95,21
YARIS	685	0,39	95,61
THALIA	665	0,38	95,99
A3	614	0,35	96,34
STILO	586	0,34	96,67
SCENIC	579	0,33	97,01

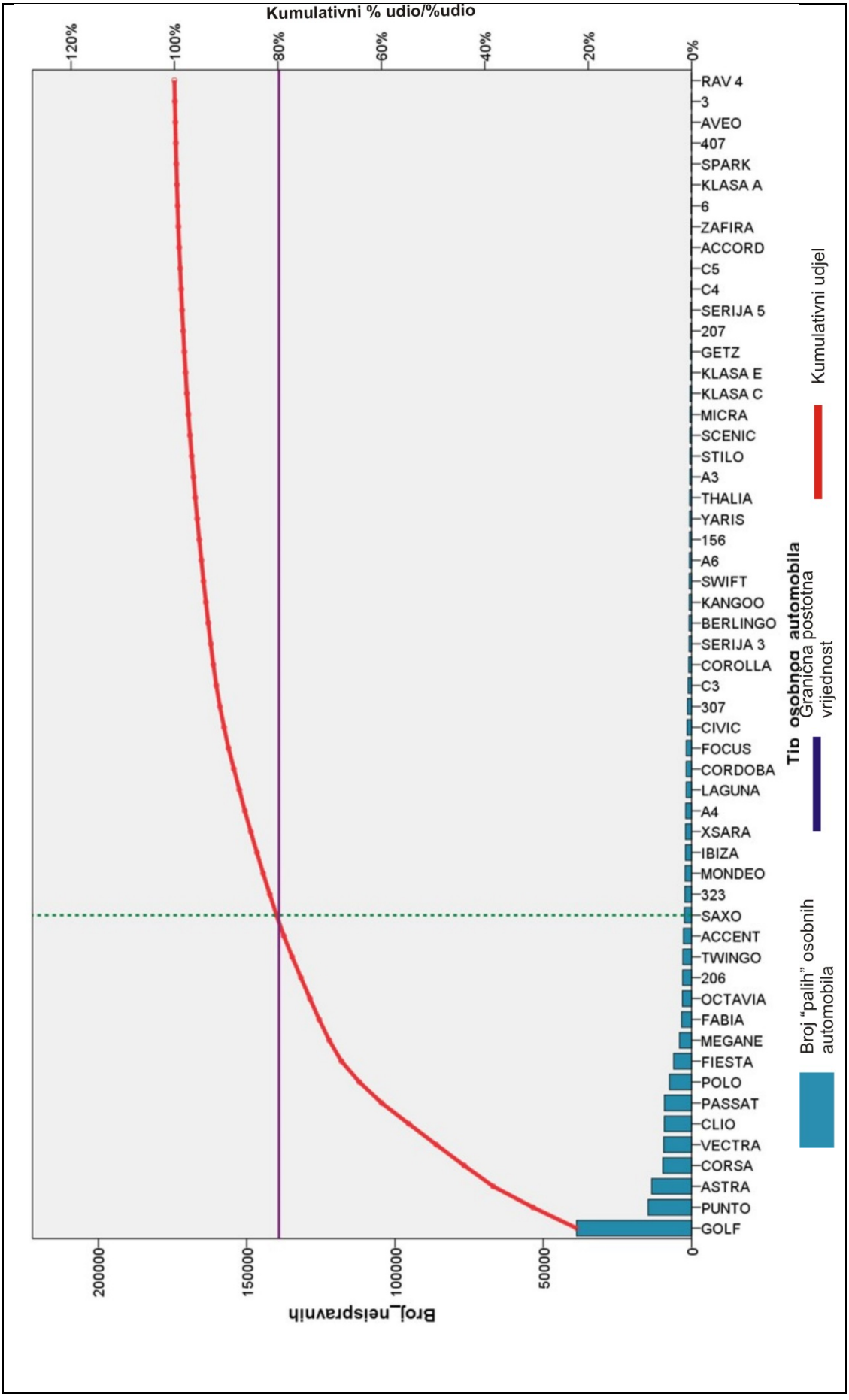
Tablica 3.3.– Utjecajni čimbenici (nastavak s prethodne stranice)

Tip osobnog automobila	Broj tehnički neispravnih automobila koji su „pali“ na TP-u	Postotni udjel broja tehnički neispravnih automobila koji su „pali“ na TP-u [%]	Kumulativni postotni udjel broja tehnički neispravnih automobila koji su „pali“ na TP-u [%]
MICRA	527	0,30	97,31
KLASA C	512	0,29	97,60
KLASA E	430	0,25	97,85
GETZ	419	0,24	98,09
207	377	0,22	98,31
SERIJA 5	372	0,21	98,52
C4	326	0,19	98,71
C5	325	0,19	98,89
ACCORD	321	0,18	99,08
ZAFIRA	288	0,17	99,24
6	246	0,14	99,38
KLASA A	227	0,13	99,51
SPARK	200	0,11	99,63
407	171	0,10	99,73
AVEO	170	0,10	99,82
3	157	0,09	99,91
RAV 4	150	0,09	100,00
UKUPNO:	**174.343**		

Provode se sljedeći koraci Pareto analize:

- za svaki tip osobnog automobila – utjecajni čimbenik – određuje se udjel u ukupnom iznosu/učinku i izračunava postotni udjel,
- utvrđuje se redoslijed tipova osobnih automobila - utjecajnih čimbenika - u skladu s pripadajućim iznosima/učincima,
- za svaki tip osobnog automobila – utjecajni čimbenik – zbrajaju se pripadajući iznosi/učinci, u skladu s redoslijedom,
- zbroj iznosa/učinaka se grafički prikazuje redoslijedom utjecajnih čimbenika.

U tablici 3.3. navedeno je koliki su iznosi/učinci pojedinih tipova osobnih automobila u kontekstu utvrđene tehničke neispravnosti u sklopu redovnih tehničkih pregleda. Dijagramski prikaz koji proizlazi iz rezultata Pareto analize prikazan je na slici 3.20.



Slika 3.20. – Dijagramski prikaz rezultata Pareto analize

Iz grafičkog prikaza sa slike 3.20. proizlazi 16 utjecajnih čimbenika – tipova osobnih automobila (VW Golf, Fiat Punto, Opel Astra, Opel Corsa, Opel Vectra, Renault Clio, VW Passat, VW Polo, Ford Fiesta, Renault Megane, Škoda Fabia, Škoda Octavia, Peugeot 206, Renault Twingo, Hyundai Accent i Citroen Saxo) čine 80 % tehnički neispravnih osobnih automobila u sklopu redovnog tehničkog pregleda, tijekom 2009. godine (osobni automobili koji su „pali" na tehničkom pregledu).

Iz rezultata Pareto analize vidljivo je koji tipovi osobnih automobila dominiraju u kontekstu utvrđene tehničke neispravnosti u sklopu redovnog tehničkog pregleda, odnosno na koje tipove osobnih automobila treba obratiti pozornost u smislu izrade prediktivnog modela tehničke ispravnosti temeljenog na rezultatima tehničkog pregleda.

Za potrebe ovog rada obrađena se prva tri utjecajna čimbenika (kritična) - tipa osobnih automobila (VW Golf, Fiat Punto i Opel Astra). Razlog takvoj odluci leži u činjenici da se radi o analogiji izrade prediktivnih modela za svaki od predmetnih tipova osobnih automobila (svih 16 tipova), a na primjeru od tri kritična tipa moguće je uočiti razlike u rezultatima temeljem kojih se mogu proizvesti adekvatni zaključci.

Obrazloženje činjenice da su u ovom radu definirani utjecajni čimbenici – 16 tipova osobnih automobila s obzirom na **apsolutni iznos (broj)** osobnih automobila predmetnih tipova koji su "pali" na tehničkom pregledu zbog utvrđenih tehničkih neispravnosti, a ne s obzirom na njihov relativni udio, proizlazi iz prioritetnog utjecajnog kriterija iz aspekta čimbenika "Vozilo" kao jednog od tri ključna čimbenika (uz "Vozača i "Cestu") koji direktno utječu na sigurnost cestovnog prometa, kao što je objašnjeno u poglavlju 1.

Naime, taj prioritetni utjecajni kriterij predstavljaju kritični tipovi automobila kojih je kao tehnički neispravnih u apsolutnom iznosu (odnosno u ukupnom broju) najviše u prometu na hrvatskim cestama, a ne oni kojih je u relativnom udjelu najviše.

Logika na kojoj se temelji prethodno navedeni prioritetni utjecajni kriterij najbolje se oslikava na sljedećem ilustrativnom primjeru u sklopu kojeg su zamišljena dva tipa automobila:

- *tip automobila X* - koji podrazumijeva ukupni skup od 100.000 automobila od kojih je **50.000 tehnički neispravnih** u prometu na cestama (odnosno postotni udio od “samo” 50% tehnički neispravnih), i

- *tip automobila Y* - koji podrazumijeva ukupni skup od 10 automobila od kojih je **svih 10 tehnički neispravnih** u prometu na cestama (odnosno postotni udio od čak 100% tehnički neispravnih).

Iz prethodno navedenog ilustrativnog primjera proizlazi neosporna činjenica da je po pitanju (negativnog) utjecaja na sigurnost prometa na cestama, iz aspekta čimbenika - “Vozilo”, **značajno kritičniji** zamišljeni *tip automobila X* s obzirom da isti podrazumijeva čak 50.000 tehnički neispravnih automobila u prometu na cestama, nego *tip automobila Y* s obzirom da isti podrazumijeva samo 10 tehnički neispravnih automobila u prometu na cestama.

4. MODELIRANJE PREDIKTIVNIH MODELA

Pod modeliranjem prediktivnih modela podrazumijeva se razrada metodologije otkrivanja znanja (idejna razrada algoritma u smislu dijagrama toka prediktivnih modela) definiranje varijabli i parametara, definiranje metrike za usporedbu preferencijalnog i alternativnog prediktivnog modela te njihova primjena i usporedba.

4.1. RAZRADA METODOLOGIJE OTKRIVANJA ZNANJA

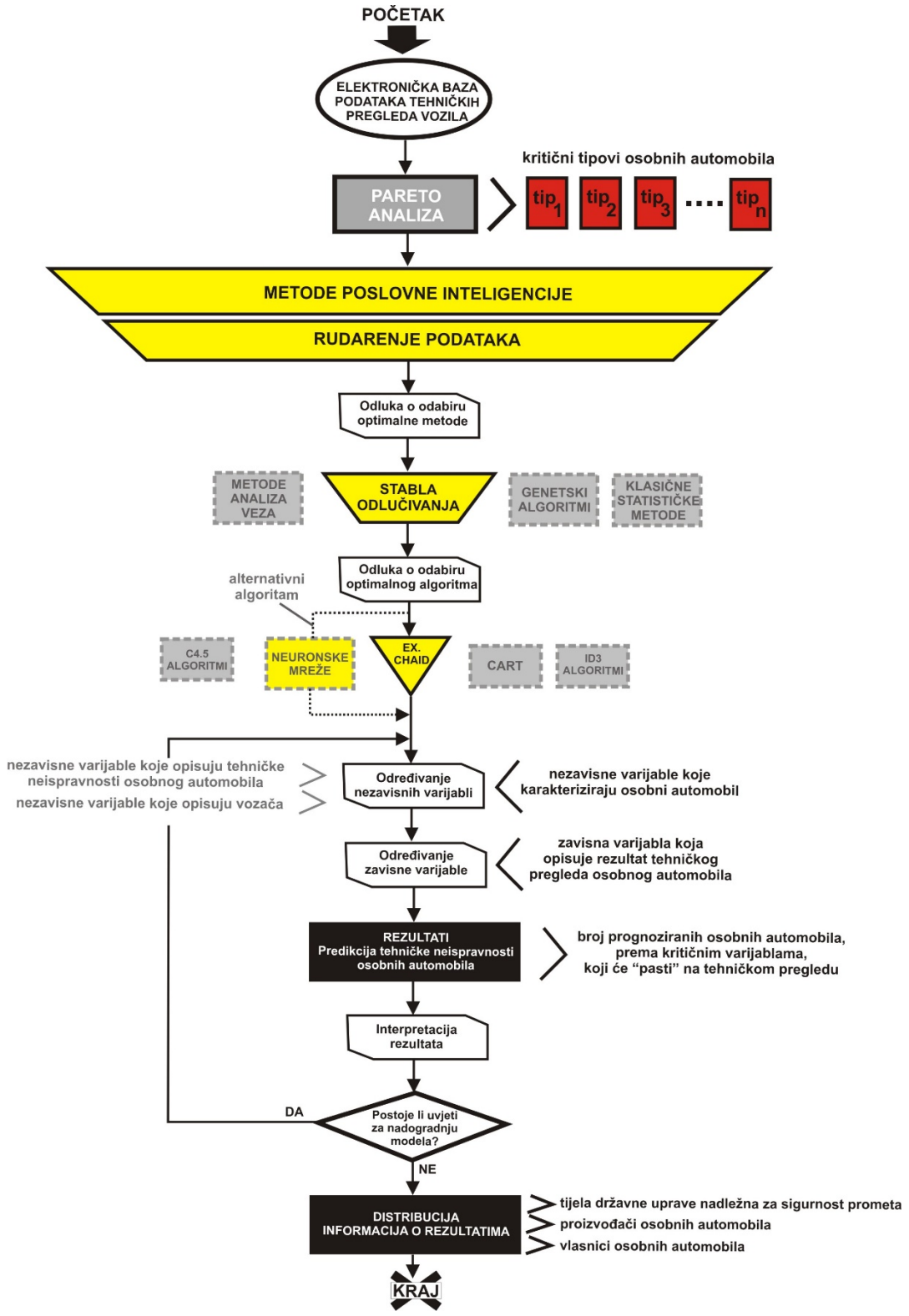
U idejnom smislu prediktivni modeli tehničke ispravnosti vozila/osobnih automobila sastoje se od mehanizma za segmentaciju ulaznih podataka u kontekstu filtriranja kritičnih tipova vozila/osobnih automobila (Pareto analiza) glede aspekta tehničke neispravnosti, te mehanizma za obradu segmentiranih ulaznih podataka u kontekstu dobivanja prognostičkih informacija o broju tehnički neispravnih vozila/osobnih automobila koji će "pasti" na tehničkom pregledu. Esencijalni mehanizam prediktivnih modela čini obrnuta piramida prikazana na slici 4.1.



Slika 4.1. - Esencijalni mehanizam prediktivnih modela

Algoritam (dijagram toka) prediktivnih modela prikazan je na slici 4.2.

Podaci o vozilima (osobnim automobilima) koji su ekstrahirani iz elektroničke baze podataka tehničkih pregleda vozila (osobnih automobila) obrađuju se metodom Pareto analize u svrhu filtriranja kritičnih tipova vozila/osobnih automobila glede aspekta tehničke neispravnosti.



Slika 4.2. – Novo-razvijeni algoritam prediktivnih modela tehničke ispravnosti osobnih automobila

Na temelju metoda poslovne inteligencije, primjenom rudarenja podataka, u skladu s utvrđenim kriterijima, prema [28], [35] i [36] glede preglednosti vizualne interpretacije, brzine obrade velike količine podataka, sposobnosti adaptiranja promjenama u ponašanju podataka, te kompatibilnosti s platformama SQL baza podataka, donosi se odluka o optimalnoj metodi - stablu odlučivanja.

U okviru metode stabla odlučivanja u skladu s utvrđenim kriterijima (koji su praktički isti kao i kriteriji za odabir optimalne metode rudarenja podataka) donosi se odluka o primjeni algoritma Exhaustive CHAID i algoritma perceptronskih neuronskih mreža.

Prema [28], [34] i [39] prednosti Exhaustive CHAID algoritma u odnosu na druge algoritme su: preglednost vizualne interpretacije, sposobnost obrade velike skupine podataka (ovaj kriterij se obično smatra ograničavajućim čimbenikom no u kontekstu ovog istraživanja, koje se temelji na izrazito velikom skupu podataka, isti predstavlja dodatnu korist), brzina obrade, sposobnost adaptiranja promjenama u ponašanju podataka, sposobnost integriranja i kompatibilnost s platformama SQL baza podataka, te mogućnost definiranja zadane vrijednosti razine štete kod pogrešne klasifikacije prediktivnih kategorija zavisne varijable.

Sukladno navedenom, prediktivni model temeljen na algoritmu Exhaustive CHAID u kontekstu ovog istraživanja smatra se preferencijalnim dok se kao alternativa istom koristi algoritam perceptronskih neuronskih mreža. Nakon usporedbe rezultata donosi se ocjena o prihvaćanju odnosno odbacivanju pojedinog prediktivnog modela.

Kao ulazni podaci u algoritme definirane se nezavisne (opće) varijable koje opisuju vozilo (osobni automobil), te je definirana zavisna varijabla koja opisuje rezultat tehničkog pregleda vozila (osobnog automobila) - "pad" na tehničkom pregledu uslijed utvrđenih tehničkih neispravnosti.

Nakon kompetentnog interpretiranja dobivenih rezultata, koji podrazumijevaju otkrivanje unaprijed poznatih i nepoznatih zakonitosti među promatranim varijablama, čime se na temelju postojećeg stanja predviđa buduće stanje, utvrđuje se je li postoje uvjeti za nadogradnju prediktivnog modela u smislu definiranja dodatnih nezavisnih

varijabli koje opisuju tehničke neispravnosti na vozilu (osobnom automobilu), te koje opisuju vozača u kontekstu načina vožnje i održavanja vozila (osobnog automobila).

U konačnici, provodi se distribucija visokovrijednih informacija o prognozi tehničke ispravnosti vozila prema tijelima državne uprave nadležnim za područje sigurnosti prometa na cestama, proizvođačima vozila (osobnih automobila) i vlasnicima/korisnicima vozila (osobnih automobila).

4.2. DEFINIRANJE VARIJABLI I PARAMETARA

Na temelju dosadašnjih razmatranja te provedene Pareto analize (skup 80%), odabrana su tri najzastupljenija tipa osobnih automobila – VW Golf, Fiat Punto i Opel Astra, unutar grupe 80% tehnički neispravnih osobnih automobila.

Za potrebe obavljanja poslova koje proizlaze iz predmetne javne ovlasti – stručnog nadzora nad radom stanica za tehnički pregled vozila – Hrvatski autoklub je objedinio i strukturirao elektroničku SQL bazu podataka tehničkih pregleda vozila [62] koja se generira putem svih stanica za tehnički pregled vozila u Republici Hrvatskoj u sklopu centraliziranog informatičkog sustava i jedinstvenog programskog rješenja za obradu podataka o tehničkim pregledima i registraciji vozila [63].

Za potrebe ovog istraživanja iz [62] su ekstrahirani podaci za 2010. godinu za najzastupljenije tipove osobnih automobila – VW Golf, Fiat Punto i Opel Astra, u formatu .xls (Microsoft Office Excel 2007), a koji je kompatibilan sa formatom statističkog paketa *IBM SPSS Statistics 22* [64]. U tablici 4.1. dan je primjer dijela ekstrahiranih podataka za skupinu osobnih automobila VW Golf.

Tablica 4.1.– Primjer dijela ekstrahiranih podataka za VW Golf

„snaga“	„ccm“	„starost_god“	„index_stp“	„vr_motora“	„vr_kocnica“	„index_km“	„tp_prolaz“
37	1093	32	1	1	1	5	0
40	1588	19	2	3	1	4	1
50	1896	11	3	3	2	4	1
55	1781	18	3	2	1	5	1
74	1595	14	3	2	2	4	0
77	1896	3	2	3	2	1	1
...	...	...	...	...	...	...	...

Postupak obrade predmetnih podataka putem [64] je identičan za sva tri navedena tipa osobnih automobila.

Grupe podataka za svaki tip osobnih automobila koji su bili predmet istraživanja, odnosno koji su obrađeni predmetnim algoritmima podrazumijevaju sljedeće nezavisne varijable nastavno na sliku 4.2.: snagu motora izraženu u kW, radni obujam motora izražen u cm^3, starost osobnog automobila (broj godina osobnog automobila proteklih od godine njegove proizvodnje do 2010. godine), indeks stanice za tehnički pregled vozila u kojoj je obavljen tehnički pregled, vrstu motora s obzirom na pogonsko gorivo, vrstu ugrađenog kočionog sustava, indeks prijeđenog puta (kilometara) i status osobnog automobila u kontekstu "prolaza" ili "ne-prolaza/pada" na tehničkim pregledu (odnosno da li su tijekom tehničkog pregleda konstatirane tehničke neispravnosti na osobnom automobilu ili ne). Nezavisne varijable koje opisuju tehničke neispravnosti osobnog automobila, te nezavisne varijable koje opisuju vozača potencijalno se mogu koristiti u daljnjim istraživanjima u skladu s objašnjenjima danim u točkama 4.8. i 5. ovog rada.

Prethodno navedene grupe podataka predstavljaju varijable. Varijable s obzirom na svoju strukturu mogu biti nominalne, redne (ordinalne) i metričke.

Varijable se smatraju nominalnim kada njihove vrijednosti predstavljaju kategorije bez unutarnjeg rangiranja (npr. vrste motora osobnog automobila s obzirom na ispušni sustav). Varijable se smatraju rednim (ordinalnim) kada njihove vrijednosti predstavljaju kategorije s unutarnjim rangiranjem (npr. razine zadovoljstva pruženom uslugom, u rasponu od izrazito zadovoljavajuća do izrazito nezadovoljavajuća). Metričke varijable su varijable kod kojih vrijednosti predstavljaju kategorije u metričkom poretku (npr. starost izražena u godinama).

Varijable s obzirom na svoju uvjetovanost mogu biti nezavisne i zavisne.

Pod nezavisnim varijablama se podrazumijevaju varijable koje se mjere neovisno o ishodu istraživanja, te koje prethode zavisnim varijablama odnosno iste uvjetuju. Pod zavisnim varijablama se podrazumijevaju varijable koje su posljedica djelovanja nezavisnih varijabli, te koje predstavljaju mjeru ishoda istraživanja.

Nezavisne varijable “snaga”, “ccm”, “starost_god” definirane su kao metričke varijable, nezavisne varijable “index_stp”, “vr_motora”, “vr_kocnica” definirane su kao nominalne varijable, a nezavisna varijabla “index_km” definirana je kao redna (ordinalna) varijabla. Nominalna varijabla “tp_prolaz” predstavlja zavisnu varijablu.

Za nezavisnu nominalnu varijablu “index_stp”, koja predstavlja varijablu veličine sredine u kojoj se nalazi pojedina stanica za tehnički pregled vozila, definirane su sljedeće oznake vrijednosti:

- oznaka 1 predstavlja stanicu za tehnički pregled vozila koja se nalazi u velikoj urbanoj sredini (Zagreb, Split, Osijek, Rijeka),
- oznaka 2 predstavlja stanicu za tehnički pregled vozila koja se nalazi u urbanoj sredini (Zadar, Pula, Slavonski Brod, Dubrovnik, Varaždin, Karlovac, Šibenik), te
- oznaka 3 predstavlja stanicu za tehnički pregled vozila koja se nalazi u maloj sredini (ostale sredine).

U kontekstu definiranja nominalne varijable “index_stp” polazi se od pretpostavke da se osobni automobili koji su obavili tehnički pregled u određenoj sredini dominantno koriste u smislu vožnje upravo unutar te sredine (prema podacima *HAK*-a [62], više od 95 % osobnih automobila obavlja tehnički pregled unutar sredine svog registarskog područja – npr., više od 95 % osobnih automobila zagrebačkih registarskih tablica obavlja tehnički pregled u nekoj od STP iz područja Grada Zagreba i Zagrebačke županije).

Za nezavisnu nominalnu varijablu “vr_motora”, koja predstavlja varijablu vrste motora osobnog automobila s obzirom na njegovo pogonsko gorivo odnosno vrstu ispušnog sustava, definirane su sljedeće oznake vrijednosti:

- oznaka 1 predstavlja benzinski motor bez ugrađenog katalizatora ili benzinski motor s ugrađenim nereguliranim katalizatorom (“BEZ KAT”),
- -oznaka 2 predstavlja benzinski (Ottov) motor s ugrađenim reguliranim katalizatorom (“REG KAT”), te
- oznaka 3 dizelski motor (“DIZEL”).

Za nezavisnu nominalnu varijablu “vr_kocnica”, koja predstavlja varijablu vrste ugrađenog kočnog sustava u osobni automobil, definirane su sljedeće oznake vrijednosti:

- oznaka 1 predstavlja klasični (ne-elektronički upravljani) kočni sustav (“klasicni”), te
- oznaka 2 kočni sustav koji je elektronički upravljan (“ECS”).

Za nezavisnu rednu (ordinalnu) varijablu “index_km”, koja predstavlja varijablu prijeđenog puta (kilometara) osobnog automobila, definirane su sljedeće oznake vrijednosti:

- oznaka 1 predstavlja prijeđeni put manji od 50.000 km,
- oznaka 2 predstavlja prijeđeni put veći od 50.000 km i manju od 100.000 km,
- oznaka 3 predstavlja prijeđeni put veći od 100.000 km i manju od 150.000 km,
- oznaka 4 predstavlja prijeđeni put veći od 150.000 km i manju od 200.000 km, te
- oznaka 5 predstavlja prijeđeni put veći od 200.000 km.

Za zavisnu nominalnu varijablu “tp_prolaz”, koja predstavlja varijablu statusa osobnog automobila u kontekstu utvrđene tehničke neispravnosti, definirane su sljedeće oznake vrijednosti:

- oznaka 0 predstavlja utvrđenu tehničku neispravnost, odnosno “ne-prolaz” osobnog automobila na tehničkom pregledu (”pad”), te
- oznaka 1 predstavlja utvrđenu tehničku ispravnost, odnosno “prolaz” osobnog automobila na tehničkom pregledu (”prolaz”).

Zavisna nominalna varijabla “tp_prolaz” predstavlja zavisnu (ciljanu) varijablu (“Target”). S obzirom da je sigurnost prometa na cestama, između ostalog, ugrožena tehničkom neispravnošću vozila, upravo se kategorija “ne-prolaza/pada” zavisne varijable odabire kao interesna (ciljana) kategorija. Drugim riječima, navedena interesna kategorija smatra se kategorijom od primarnog interesa za istraživanje.

U tablici broj 4.2. dan je pregled svih varijabli korištenih u algoritmima.

Tablica 4.2.– Pregled varijabli korištenih u algoritmima

Naziv varijable	Vrsta varijable s obzirom na uvjetovanost	Opis varijable	Oznaka varijable	Vrsta varijable s obzirom na strukturu
„snaga“	nezavisna	snaga motora	u stvarnom iznosu	metrička
„ccm“	nezavisna	radni obujam motora	u stvarnom iznosu	metrička
„starost_god“	nezavisna	broj godina osobnog automobila proteklih od godine njegove proizvodnje do 2010. godine	u stvarnom iznosu	metrička
„index_stp“	nezavisna	stanica za tehnički pregled vozila u kojem je obavljen tehnički pregled na osobnom automobilu	oznaka 1: velika urbana sredina oznaka 2: urbana sredina oznaka 3: mala sredina	nominalna
„vr_motora“	nezavisna	vrsta motora osobnog automobila s obzirom na njegovo pogonsko gorivo odnosno vrstu ispušnog sustava	oznaka 1: BEZ KAT oznaka 2: REG KAT oznaka 3: DIZEL	nominalna
„vr_kocnica“	nezavisna	vrsta ugrađenog kočnog sustava u osobni automobil	oznaka 1: klasični kočni sustav oznaka 2: elektronički upravljan kočni sustav	nominalna
„index_km“	nezavisna	prijeđeni put (broj kilometara) osobnog automobila	oznaka 1: <50.000 oznaka 2: >50.000<100.000 oznaka 3: >100.000<150.000 oznaka 4: >150.000<200.000 oznaka 5: >200.000	ordinalna
„tp_prolaz“	zavisna (ciljana)	status osobnog automobila u kontekstu utvrđene tehničke neispravnosti	oznaka 0: „ne prolaz“ („pad“) oznaka 1: „prolaz“	nominalna

U kontekstu određivanja validacijske metode glede točnosti predikcije stabla odlučivanja (“Decision Tree: Validation”), koja se automatski provodi u okviru statističkog paketa, određena je metoda dijeljenja uzorka (“Split sample validation”) u omjeru 15 % za trening uzorak i 85 % za testni uzorak za osobne automobile VW Golf i Opel Astra, te u omjeru 70 % za trening uzorak i 30 % za tetni uzorak.

U okviru definiranja kriterija za stablo odlučivanja (”Decison Tree: Criteria”), određene su granice rasta stabla (“Growth Limits”), pomoću kojih se ograničava broj

razina/stupnjeva stabla te se kontrolira minimalan broj slučajeva u "roditeljskim" i "dječjim" čvorovima ("parent and child nodes"). "Roditeljski" čvor podrazumijeva čvor iz kojeg proizlazi daljnje grananje "dječjih" čvorova, dok završni čvor ("list") podrazumijeva čvor iz kojeg nije moguće daljnje grananje čvorova.

Odabrana je dubina stabla odlučivanja ("Maximum Tree Depth") – tri razine ispod početnog/korijenskog čvora ("root node"). Prema [52], [53] i [57] riječ je o dubini stabla (treća razina) na kojoj u većini slučajeva dolazi do zaustavljanja rekurzivnog dijeljenja zadane skupine podataka u završnom čvoru ("listu"). Dubina stabla odlučivanja od tri razine ispod početnog/korijenskog čvora najveća je dubina koja se može koristiti u slučaju primjene Exhaustive CHAID algoritma i ista se u pravilu koristi redovno, a posebno kada se radi o obradi velikih skupina podataka (slučajeva), što je svakako slučaj u ovome radu.

Nakon provedenog eksperimentiranja, definiran je minimalan broj slučajeva i to po 100 slučajeva za "roditeljski" i za "dječji" čvor. Iz navedenog proizlazi da čvorovi koji ne zadovoljavaju ove kriterije neće biti dijeljeni na daljnje čvorove. Broj minimalnih slučajeva po čvoru direktno proporcionalno utječe na broj čvorova u stablu odlučivanja, na način da manji broj slučajeva po čvoru znači veći broj čvorova.

U okviru Pearson metode (odnosno Pearson-ovog χ^2 testa), u pogledu definiranja kriterija za stablo odlučivanja glede vrijednosti značajnosti testa ("Significance Level") određene su, standardne vrijednosti (*default*-ne vrijednosti) za dijeljene čvorove ("Splitting Nodes") – 0,05, te za kategorije udruživanja ("Merging Categories") – 0,05., (radi se o uobičajenoj graničnoj vrijednosti kod χ^2 testa pri kojoj se potvrđuje ili odbacuje nul-hipoteza, prema [43], [47], [48], [53]). Riječ je o zadovoljavajućim vrijednostima značajnosti testa koje se uobičajeno koriste u tehničkom znanstvenom području.

U pogledu prilagođavanja (korigiranja) *p* - vrijednosti testa (razina značajnosti) provodi se Bonferroni metoda, s obzirom da se radi o standardnoj korektivnoj metodi za rješavanje problema višestruke usporedbe (problem pogrešnog odbacivanja nul hipoteza u kontekstu višestruke usporedbe).

U pogledu definiranja opcija za stablo odlučivanja (“Decision Tree: Options”), definirano je da se eventualne nedostajuće vrijednosti (“Missing Values”) pojedinih nezavisnih nominalnih varijabli tretiraju zasebno, upravo kao nedostajuće vrijednosti, a ne da se tretiraju kao da se radi o stvarnim (postojećim) vrijednostima.

U pogledu definiranja opcija za stablo odlučivanja definirana je penalizacija u smislu predviđene razine negativnih posljedica kod pogrešne klasifikacije.

Na taj način se u predmetnu metodu integrira informacija o pripadajućoj razini negativnih posljedica (odnosno o pripadajućoj penalizaciji) kod pogrešne klasifikacije. Kao što je, na primjer, razina penalizacije kod pogrešnog klasificiranja pojedinca s visokim rizikom od srčane bolesti u klasu niskog rizika mnogo veća nego obrnuto (kod pogrešnog klasificiranja pojedinca s niskim rizikom od srčane bolesti u klasu visokog rizika), tako je u pogledu sigurnosti cestovnog prometa razina penalizacije kod pogrešnog klasificiranja vozila koje je tehnički neispravno u klasu tehnički ispravnih vozila mnogo veća od razine penalizacije kod pogrešnog klasificiranja vozila koje je tehnički ispravno u klasu tehnički neispravnih vozila.

Sukladno navedenom, za potrebe radnji koje se provode radi poboljšanja modela (tzv. korektivne radnje), što je objašnjeno u točkama 4.4.1.2., 4.4.2.2. i 4.4.3.2., razina penalizacije kod pogrešnog klasificiranja osobnog automobila koji je tehnički neispravan u klasu tehnički ispravnih osobnih automobila, definirana je s vrijednošću 2 (kod osobnih automobila VW Golf), odnosno s vrijednošću 3 (kod osobnih automobila Fiat Punto i Opel Astra) nakon provedenog eksperimentiranja, dok je razina penalizacije kod pogrešnog klasificiranja osobnog automobila koji je tehnički ispravan u klasu tehnički neispravnih osobnih automobila definirana s vrijednošću 1.

Kod definiranja inicijalnog prediktivnog modela, prije provođenja radnji radi poboljšanja modela, penalizacija nije predviđena, odnosno vrijednost kod pogrešnog klasificiranja osobnog automobila koji je tehnički neispravan u klasu tehnički ispravnih osobnih automobile izjednačena je s vrijednošću kod pogrešnog klasificiranja osobnog automobila koji je tehnički ispravan u klasu tehnički neispravnih osobnih automobila (obje iznose vrijednost 1).

Prikaz definiranih parametara Exhaustive CHAID algoritma (preferencijalnog) dan je u tablici 4.3.

Tablica 4.3.– Pregled definiranih parametara Exhaustive CHAID algoritma u poboljšanom prediktivnom modelu

Parametar	Opis	Vrijednost	Način odabira
Validacijska metoda	metoda dijeljenja uzorka	15 % za trening uzorak, 85 % za testni uzorak VW Golf i Opel Astra, 70 % za trening uzorak, 30 % za testni uzorak Fiat Punto	- nakon provedenog eksperimentiranja
Dubina stabla odlučivanja	broj razina ispod početnog (korijenskog) čvora	tri razine	- preporuka statističkog paketa (*default*) - primjena u prijašnjim istraživanjima
Minimalan broj slučajeva po čvorovima		- „roditeljski“ čvor: 100 slučajeva - „dječji“ čvor: 100 slučajeva	- preporuka statističkog paketa (*default*)
Značajnost testa	razina signifikantnosti	- dijeljenje čvorova (*splitting*): 0,05 - integriranje čvorova (*merging*): 0,05	- preporuka statističkog paketa (*default*) - primjena u tehničkom znanstvenom području
Prilagođavanje (korigiranje) *p* - vrijednosti testa	Bonferroni metoda	-	zadana metoda
Penalizacija kod pogrešne klasifikacije (radnje koje se provode radi poboljšanja modela)	Razina negativnih posljedica	- razina penalizacije kod pogrešnog klasificiranja osobnog automobila koji je tehnički neispravan u klasu tehnički ispravnih osobnih automobila: 2 odnosno 3 - razina penalizacije kod pogrešnog klasificiranja osobnog automobila koji je tehnički ispravan u klasu tehnički neispravnih osobnih automobila: 1	- nakon provedenog eksperimentiranja

Pomoću primijenjenog statističkog paketa [64] definiran je prikaz rizika koji podrazumijeva proračun rizika te standardne devijacije/pogreške, odnosno procjenu predviđene točnosti stabla. Za kategorijske zavisne varijable proračun rizika podrazumijeva omjer pogrešno klasificiranih slučajeva nakon prilagođavanja prethodnih vjerojatnosti i penalizacije u smislu predviđenih negativnih posljedica kod pogrešne klasifikacije. Za metričke zavisne varijable proračun rizika je u okviru varijance čvora.

Prikaz klasifikacije u sklopu kontingencijske tablice podrazumijeva pregled broja točno i pogrešno klasificiranih slučajeva za svaku kategoriju zavisne varijable.

Za kategorijske zavisne varijable s definiranim ciljanim kategorijama izlazni podaci dani su u obliku tablice iz koje je vidljiva postotna dobit, postotak odziva i postotni indeks po pojedinom čvoru ili postotnom razredu.

U pogledu vizualnog opisivanja značajki čvorova stabla odlučivanja definiran je prikaz dijagrama dobiti i dijagrama indeksa. Dobit predstavlja postotak svih slučajeva u okviru ciljane kategorije svakog pojedinog čvora, prema izračunu: (ciljana kategorija po pojedinom čvoru/ ciljana kategorija ukupno) x 100. Dijagram dobiti predstavlja linijski dijagram kumulativne postotne dobiti, prema izračunu: (kumulativan postotak ciljane kategorije/ ciljana kategorija ukupno) x 100. Indeks predstavlja omjer postotnog odziva čvora za ciljanu kategoriju u usporedbi s postotnim odzivom ukupne ciljne kategorije za cijeli uzorak. Dijagram indeksa predstavlja linijski dijagram kumulativne postotne vrijednosti indeksa. Kumulativni postotni indeks izračunava se prema: (kumulativni postotni odziv / ukupni postotni odziv) x 100.

4.3. METRIKA ZA USPOREDBU PREDIKTIVNIH MODELA

Kod procesa strojnog učenja kao mjerilo kvalitete binarne klasifikacije (koja podrazumijeva dva stanja) koristi se tzv. Matthews koeficijent korelacije (u daljnjem tekstu - MCC), prema [65] i [66]. MCC uzima u obzir točne i netočne pozitivne i negativne statuse te se smatra izbalansiranom mjerom koja se može koristiti čak i u slučajevima kada su razredi izrazito različitih veličina. U suštini, MCC predstavlja korelacijski koeficijent između opaženih i prediktivnih binarnih klasifikacija čija je vrijednost između -1 i +1, a razrada vrijednosti u smislu jačine korelacije je sljedeća: od -0,70 do -1 izrazito jaka negativna korelacija, od -0,40 do -0,69 jaka negativna korelacija, od -0,30 do -0,39 umjerena negativna korelacija, od -0,20 do -0,29 slaba negativna korelacija, od -0,01 do -0,19 te od +0,01 do +0,19 bez korelacije, od +0,20 do +0,29 slaba pozitivna korelacija, od +0,30 do +0,39 umjerena pozitivna korelacija, od +0,40 do +0,69 jaka pozitivna korelacija, te od +0,70 do +1 izrazito jaka pozitivna korelacija. MCC se temelji na *hi*-kvadratu za 2x2 kontingencijsku tablicu

$$|MCC| = \sqrt{\frac{\chi^2}{n}} \tag{4.1}$$

gdje je *n* ukupni broj opažanja.

Iako nema savršenog načina kojim bi se matrica konfuzije točnih i netočnih pozitivnih i negativnih statusa opisala pojedinim brojem, MCC se općenito smatra mjerom koja je tome najbliža. Druge mjere, kao što je udio točnih predikcija (odnosno točnost), nisu upotrebljive u slučajevima kada su dva razreda značajno različitih veličina.

MCC se može direktno proračunati temeljem matrice konfuzije prema

$$|MCC| = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \tag{4.2}$$

gdje je *TP* broj stvarno pozitivnih statusa, *TN* broj stvarno negativnih statusa, *FP* broj pogrešnih pozitivnih statusa i *FN* broj pogrešnih negativnih statusa.

Prema [66] izraz 4.2 originalno glasi

$$N = TN + TP + FN + FP \tag{4.3}$$

$$S = \frac{TP+FN}{N} \tag{4.4}$$

$$P = \frac{TP+FP}{N} \tag{4.5}$$

$$MCC = \frac{\frac{TP}{N} - S \times P}{\sqrt{PS(1-S)(1-P)}} \tag{4.6}$$

Matrica konfuzije

U području procesa strojnog učenja matrica konfuzije poznata je i pod imenom kontingencijska tablica, odnosno matrica greške. Matrica konfuzije predstavlja specifični tablični prikaz koji na pregledan način prikazuje performanse algoritma. Svaki stupac matrice konfuzije predstavlja primjere u prediktivnom razredu, dok svaki red predstavlja primjere u predmetnom (aktualnom) razredu. Naziv matrice konfuzije proizlazi iz činjenice da je iz iste lako vidljivo da li sustav pogrešno zamjenjuje dva razreda jedan za drugi, te na taj način stvara konfuziju u rezultatima.

Neka se pretpostavi da određena pojava ima *P* pozitivnih statusa i *N* negativnih statusa. Tada četiri rezultata mogu biti formulirana u obliku 2x2 matrice konfuzije odnosno kontingencijske tablice, kako je prikazano u tablici 4.4.

Tablica 4.4. – Primjer matrice konfuzije (kontingencijske tablice)

		Stanje statusa			
	Ukupna populacija	Pozitivni	Negativni	Prevalencija = $= \frac{\sum pozitivni}{\sum ukupna\ populacija}$	
Rezultat testa	Pozitivan rezultat testa	**Stvarno pozitivni**	Pogrešno pozitivni	Pozitivna predikcija (PPV) = $= \frac{\sum stvarno\ pozitivni}{\sum pozitivan\ rezultat\ testa}$	Stopa otkrivanja pogrešnih (FDR) = $= \frac{\sum pogrešno\ pozitivni}{\sum pozitivan\ rezultat\ testa}$
	Negativan rezultat testa	Pogrešno negativni	**Stvarno negativni**	Stopa pogrešnih izostavljanja (FOR) = $= \frac{\sum pogrešno\ negativni}{\sum negativan\ rezultat\ testa}$	Negativna predikcija (NPV) = $= \frac{\sum stvarno\ negativni}{\sum negativan\ rezultat\ testa}$
	Točnost (ACC) = $= \frac{\sum stvarno\ pozitivni + \sum stvarno\ negativni}{\sum ukupna\ populacija}$	Stopa stvarno pozitivnih (TPR) = $= \frac{\sum stvarno\ pozitivni}{\sum pozitivni}$	Stopa pogrešno pozitivnih (FPR) = $= \frac{\sum pogrešno\ pozitivni}{\sum negativni}$	Omjer vjerojatnosti za pozitivne $(LR+) = \frac{TPR}{FPR}$	Omjer dijagnostičke vjerojatnosti $(DOR) = \frac{LR+}{LR-}$
		Stopa spogrešno negativnih (FNR) = $= \frac{\sum spogrešno\ negativni}{\sum pozitivni}$	Stopa stvarno negativnih (TNR) = $= \frac{\sum stvarno\ negativni}{\sum negativni}$	Omjer vjerojatnosti za negativne $(LR-) = \frac{FNR}{TNR}$	

Derivacije iz tablice 4.4. mogu biti izražene i na način prikazan u tablici 4.5.

Tablica 4.5. – Derivacije matrice konfuzije

Stopa stvarno pozitivnih	*TPR = TP/P = TP/(TP + FN)*
Stopa stvarno negativnih	*TNR = TN/N = TN/(FP + TN)*
Pozitivna predikcija	*PPV = TP/(TP + FP)*
Negativna predikcija	*NPV = TN/(TN + FN)*
Stopa pogrešno pozitivnih	*FPR = FP/N = FP/(FP + TN)*
Stopa pogrešno negativnih	*FNR = FN/P = FN/(FN + TP)*
Stopa otkrivanja pogrešnih	*FDR = FP/(FP + TP) = 1 - PPV*
Točnost	*ACC = (TP + TN)/(P + N)*

4.4. IZRADA I USPOREDBA PREDIKTIVNIH MODELA

4.4.1. Prediktivni modeli za osobne automobile VW Golf

U tablici 4.6. (postavke i zbirni rezultati modela) sadržane su informacije o postavkama i rezultatima dobivenim iz modela temeljenog na Exhaustive CHAID algoritmu. Usporedni rezultati sva tri obrađena tipa osobnih automobila (kritične nezavisne varijable u kontekstu "pada" na tehničkom pregledu) dani su u tablici 6.1.

Tablica 4.6. – Postavke i zbirni rezultati modela za VW Golf (Exhaustive CHAID)

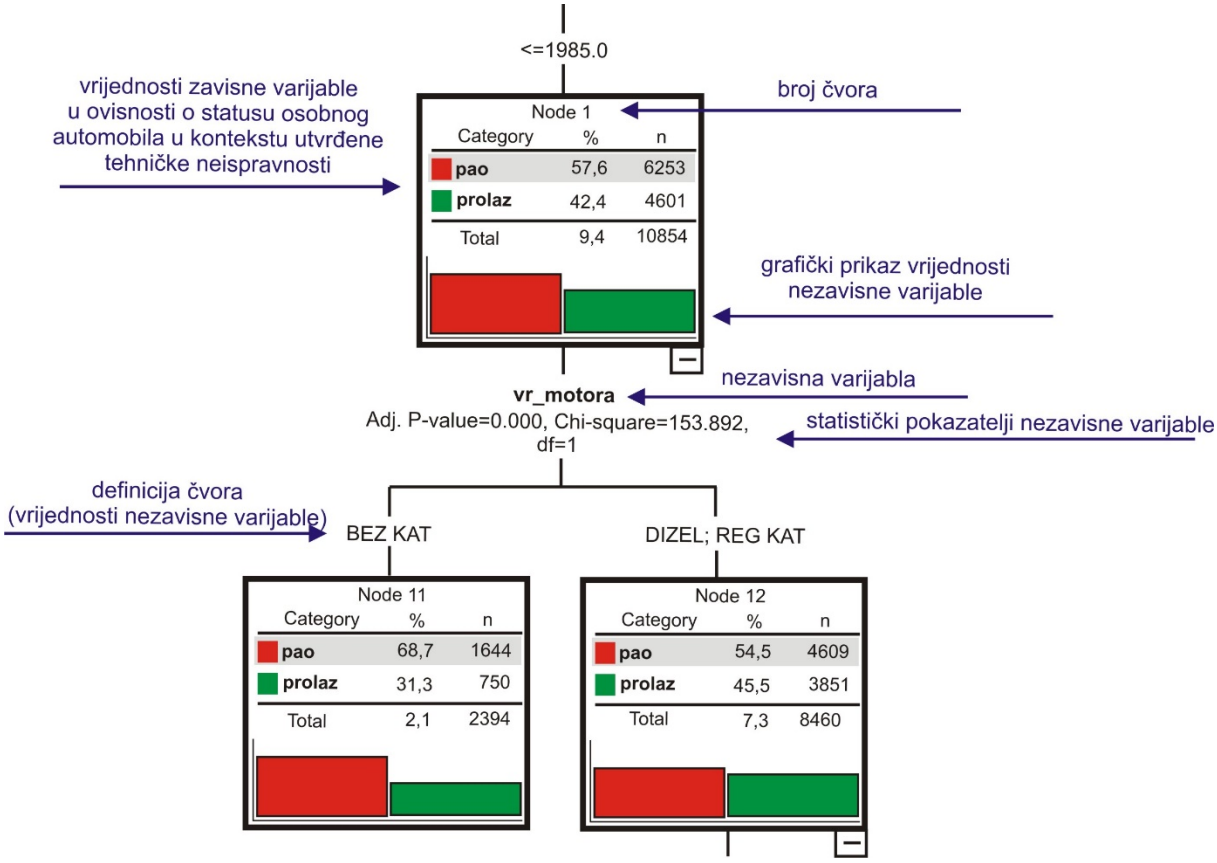
Specifikacije	Metoda	Exhaustive CHAID
	Zavisna varijabla	tp_prolaz
	Nezavisne varijable	snaga, ccm, index_stp, vr_motora, vr_kocnica, index_km, starost_god
	Validacija	Metoda dijeljenja uzorka
	Najveća dubina stabla (broj razina)	3
	Najmanji broj slučajeva u "roditeljskom" čvoru	100
	Najmanji broj slučajeva u "dječjem" čvoru	100
Rezultati	Uključene nezavisne varijable	starost_god, index_stp, index_km, snaga, ccm
	Broj čvorova	48
	Broj završnih čvorova	32
	Dubina (broj razina)	3

Informacije o postavkama modela odnose se na varijable, vrstu validacije stabla odlučivanja, dubinu stabla i najmanji broj slučajeva po čvorovima. Informacije o rezultatima odnose se na nezavisne varijable koje su uključene u izradu modela, broj ukupnih čvorova, broj završnih čvorova te dubina (broj razina) stabla.

Kao što je prikazano u tablici 4.6. od početno specificiranih sedam nezavisnih varijabli pet nezavisnih varijabli je u konačnici uključeno u završni model. Nezavisne varijable "vr_kocnica" (vrsta kočionog sustava ugrađenog na osobni automobil) i "vr_motora" (vrsta motora osobnog automobila s obzirom na njegovo pogonsko gorivo odnosno

vrstu ispušnog sustava), nisu značajno pridonijele modelu te su iste od strane algoritma automatski isključene iz završnog modela.

Sadržaj segmenata stabla odlučivanja, odnosno čvorova, u okviru grafičkog prikaza stabla odlučivanja dan je na slici 4.3.



Slika 4.3. – Prikaz sadržaja čvorova u okviru stabla odlučivanja

Kao što je prikazano na slici 4.4., na kojoj je dano stablo odlučivanja trening uzorka prediktivnog modela tehničke ispravnosti osobnih automobila VW Golf, radi se o trideset završnih čvorova, u okviru četrdeset sedam ukupnih čvorova, unutar tri razine modela. Iz predmetnog stabla odlučivanja proizlazi da je nezavisna varijabla "starost_god" (broj godina osobnog automobila proteklih od godine njegove proizvodnje do 2010. godine) varijabla s najvećim prediktivnim potencijalom (prva razina dubine stabla) u kontekstu tehničke neispravnosti osobnih automobila VW Golf. Drugim riječima, predmetna varijabla ima najveći potencijal u podjeli osobnih automobila u dvije skupine – osobni automobili bez utvrđenih tehničkih neispravnosti na tehničkom pregledu koji za posljedicu imaju "prolaz" na istom, odnosno osobni

automobili kod kojih je tijekom tehničkog pregleda utvrđena jedna ili više tehničkih neispravnosti te koji za posljedicu imaju “pad” na istom.

Nezavisna varijabla “ starost_god ” dijeli uzorak na deset segmenata u okviru kojih su izražene različite godine „starosti“ osobnih automobila, te gdje su zastupljene značajne razlike u postotku tehnički neispravnih osobnih automobila : osobni automobili „stari“ 5 godina i manje - 3,7 % tehnički neispravnih, osobni automobili koji su „stari“ 6 do 9 godina, uključujući 9 godina - 10,1 % tehnički neispravnih, osobni automobili koji su „stari“ 10 do 12 godina, uključujući 12 godina - 20,5 % tehnički neispravnih, osobni automobili „stari“ 13 godina - 27,7 % tehnički neispravnih, osobni automobili koji su „stari“ 14 i 15 godina - 22,6 % tehnički neispravnih, osobni automobili „stari“ 16 i 17 godina - 31,4 % tehnički neispravnih, osobni automobili „stari“ 18 i 19 godina - 42,9 % tehnički neispravnih, osobni automobili „stari“ 20 i 21 godinu - 54,4 % tehnički neispravnih, osobni automobili koji su „stari“ 22 do 24 godine, uključujući 24 godine – 63,8 % tehnički neispravnih i osobni automobili „stari“ od 25 i više godina - 57,2 % tehnički neispravnih.

U pogledu nezavisne varijable “ starost_god ”, a u okviru prve razine dubine stabla, čvorovi pod brojevima 1, 4, 5, 6, 7, 9 i 10 predstavlja “roditeljski” čvor, dok su čvorovi pod brojevima 2, 3 i 8 završne čvorove.

U okviru druge razine dubine stabla vidljivo je da nezavisna varijabla “index_stp” (veličina sredine u kojoj se nalazi pojedina stanica za tehnički pregled) dijeli uzorak na dva segmenta kod osobnih automobila „starih“ 5 godina i manje: osobni automobili kod kojih je tehnički pregled obavljen u stanici za tehnički pregled vozila koja se nalazi u velikoj urbanoj sredini - 5,4 % tehnički neispravnih i osobni automobili kod kojih je tehnički pregled obavljen u stanici za tehnički pregled vozila koja se nalazi u urbanoj sredini ili maloj sredini - 2,0 % tehnički neispravnih. Objašnjenje glede presumpcija u svezi nezavisne varijable “index_stp” dano je u okviru točke 6. ovog rada.

U pogledu nezavisne varijable “ index_stp ”, a u okviru druge razine dubine stabla, čvor pod brojem 12 predstavlja “roditeljski” čvor, dok je čvor pod brojem 11 završni čvor.

Druga nezavisna varijabla druge razine dubine stabla “index_km” (prijeđeni put osobnih automobila) dijeli uzorak na dva segmenta kod osobnih automobila „starih“ 13 godina: osobni automobili koji su prešli do 150.000 km - 6,5 % tehnički neispravnih i osobni automobili koji su prešli više od 150.000 km - 32,7 % tehnički neispravnih.

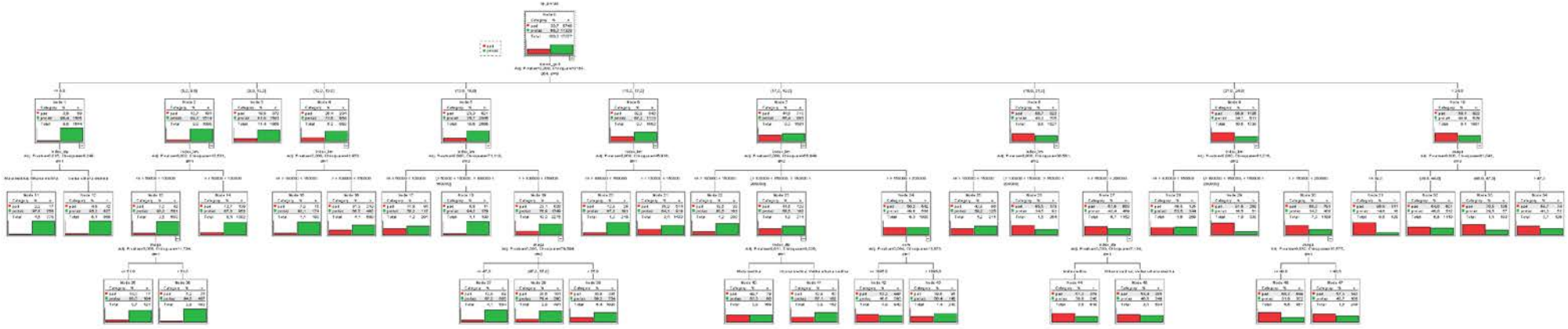
Navedena nezavisna varijabla još u četiri slučaja dijeli uzorak u dva, tri, odnosno četiri segmenta. Kod osobnih automobila „starih“ 14 i 15 godina: osobni automobili koji su prešli do 100.000 km - 40,6 % tehnički neispravnih, osobni automobili koji su prešli više od 100.000 km i manje od 150.000 km - 8,9 % tehnički neispravnih, osobni automobili koji su prešli više od 150.000 km i manje od 200.000 km - 19,5 % tehnički neispravnih, osobni automobili koji su prešli više od 200.000 km - 24,3 % tehnički neispravnih, kod osobnih automobila „starih“ 16 i 17 godina: osobni automobili koji su prešli do 150.000 km - 11,0 % tehnički neispravnih, osobni automobili koji su prešli više od 150.000 km - 34,5 % tehnički, kod osobnih automobila „starih“ 18 i 19 godina: osobni automobili koji su prešli do 150.000 km - 23,0 % tehnički neispravnih, osobni automobili koji su prešli više od 150.000 km - 45,6 % tehnički neispravnih i kod osobnih automobila koji su „stari“ 22 do 24 godine, uključujući 24 godine: osobni automobili koji su prešli do 150.000 km - 51,1 % tehnički neispravnih, osobni automobili koji su prešli više od 150.000 km i manje od 200.000 km - 75,3 % tehnički neispravnih i osobni automobili koji su prešli više od 200.000 - 64,0 % tehnički neispravnih.

U pogledu nezavisne varijable “index_km”, a u okviru druge razine dubine stabla, čvorovi pod brojem 14, 16, 17, 18, 20, 22 i 25 predstavljaju “roditeljske” čvorove, dok su čvorovi pod brojem 13, 15, 19, 21, 23 i 24 završni čvorovi.

Treća nezavisna varijabla druge razine dubine stabla “snaga” (snaga motora) dijeli uzorak na dva segmenta kod osobnih automobila „starih“ 25 i više godina: osobni automobili kod kojih je snaga motora manja ili jednaka 39 kW - 88,3 % tehnički neispravnih i osobni automobili kod kojih je snaga motora veća od 39 kW - 54,4 % tehnički neispravnih.

U pogledu nezavisne varijable “ snaga ”, a u okviru druge razine dubine stabla, čvor pod brojem 27 predstavlja “roditeljski” čvor, dok je čvor pod brojem 26 završni čvor.

.



Slika 4.4. - Stablo odlučivanja trening uzorka prediktivnog modela za VW Golf

U okviru treće (završne) razine dubine stabla vidljivo je da nezavisna varijabla “index_km” dijeli uzorak na dva segmenta kod osobnih automobila „starih“ 5 godina i manje i kod kojih je tehnički pregled obavljen u stanici za tehnički pregled vozila koja se nalazi u urbanoj sredini ili maloj sredini: osobni automobili koji su prešli do 50.000 km - 0,2 % tehnički neispravnih i osobni automobili koji su prešli više od 50.000 km - 4,0 % tehnički neispravnih.

U pogledu nezavisne varijable “index_km”, a u okviru treće (završne) razine dubine stabla, čvorovi pod brojem 28 i 29 su završni čvorovi.

Druga nezavisna varijabla treće (završne) razine dubine stabla “vr_kocnica” (vrsta ugrađenog kočnog sustava u osobni automobil) dijeli uzorak na dva segmenta kod osobnih automobila koji su „stari“ 13 godina, te koji su prešli više od 150.000 km: osobni automobili s ugrađenim klasičnim kočnim sustavom - 36,8 % tehnički neispravnih i osobni automobili s ugrađenim ECS kočnim sustavom - 29,6 % tehnički neispravnih.

U pogledu nezavisne varijable “vr_kocnica”, a u okviru treće (završne) razine dubine stabla, čvorovi pod brojem 30 i 31 su završni čvorovi.

Treća nezavisna varijabla treće (završne) razine dubine stabla “vr_motora” (vrsta motora osobnog automobila s obzirom na njegovo pogonsko gorivo odnosno vrstu ispušnog sustava) dijeli uzorak na dva segmenta kod osobnih automobila koji su „stari“ 14 i 15 godina i koji su prešli više od 100.000 km i manje od 150.000 km: osobni automobili s dizelskim motorom (DIZEL) - 12,8 % tehnički neispravnih i osobni automobili s benzinskim (Ottovim) motorom i ugrađenim reguliranim katalizatorom (REG KAT) - 5,2 % tehnički neispravnih.

Navedena nezavisna varijabla još u jednom slučaju dijeli uzorak u dva segmenta. Kod osobnih automobila „starih“ 16 i 17 godina, a koji su prešli više od 150.000 km: osobni automobili s benzinskim (Ottovim) motorom i ugrađenim reguliranim katalizatorom (REG KAT), odnosno bez ugrađenog katalizatora (BEZ KAT) - 40,7 % tehnički neispravnih i osobni automobili s dizelskim motorom (DIZEL) - 31,5 %

U pogledu nezavisne varijable “vr_motora”, a u okviru treće (završne) razine dubine stabla, čvorovi pod brojem 32, 33, 39 i 40 su završni čvorovi.

Četvrta nezavisna varijabla treće (završne) razine dubine stabla “snaga” (snaga motora) dijeli uzorak na dva segmenta kod osobnih automobila koji su „stari“ 14 i 15 godina, te koji su prešli više od 150.000 km i manje od 200.000 km: osobni automobili s motorom snage manje ili jednake 47 kW - 9,5 % tehnički neispravnih i osobni automobili s motorom snage veće od 47 kW - 24,0 % tehnički neispravnih.

Navedena nezavisna varijabla još u jednom slučaju dijeli uzorak u tri segmenta. Kod osobnih automobila koji su „stari“ 14 i 15 godina, te koji su prešli više od 200.000 km: osobni automobili s motorom snage manje ili jednake 47 kW - 20,3 % tehnički neispravnih, osobni automobili s motorom snage veće od 47 kW a manje ili jednake 55 kW - 13,7 % tehnički neispravnih i osobni automobili s motorom snage veće od 55 kW - 30,4 % tehnički neispravnih

U pogledu nezavisne varijable “snaga”, a u okviru treće (završne) razine dubine stabla, čvorovi pod brojem 34, 35, 36, 37 i 38 su završni čvorovi.

Peta nezavisna varijabla treće (završne) razine dubine stabla “ccm” (radni obujam motora) dijeli uzorak na dva segmenta kod osobnih automobila koji su „stari“ 18 i 19 godina, te koji su prešli više od 150.000 km: osobni automobili s motorom radnog obujma manjeg ili jednakog 1780 cm^3 - 49,1 % tehnički neispravnih i osobni automobili s motorom obujma većeg od 1780 cm^3 - 39,2 % tehnički neispravnih.

U pogledu nezavisne varijable “ccm”, a u okviru treće (završne) razine dubine stabla, čvorovi pod brojem 41 i 42 su završni čvorovi.

Šesta nezavisna varijabla treće (završne) razine dubine stabla “index_stp” (veličina sredine u kojoj se nalazi pojedina stanica za tehnički pregled) dijeli uzorak na dva segmenta kod osobnih automobila koji su „stari“ 22 do 24 godine, uključujući 24 godine, te koji su prešli više od 200.000 km: osobni automobili kod kojih je tehnički pregled obavljen u stanici za tehnički pregled vozila koja se nalazi u velikoj urbanoj sredini i u urbanoj sredini - 58,5 % tehnički neispravnih i osobni automobili kod kojih je

tehnički pregled obavljen u stanici za tehnički pregled vozila koja se nalazi u maloj sredini - 68,5 % tehnički neispravnih.

Navedena nezavisna varijabla još u jednom slučaju dijeli uzorak u dva segmenta. Kod osobnih automobila „starih“ 25 i više godina, a čija je snaga motora veća od 39 kW: osobni automobili kod kojih je tehnički pregled obavljen u stanici za tehnički pregled vozila koja se nalazi u velikoj urbanoj sredini - 47,7 % tehnički neispravnih i osobni automobili kod kojih je tehnički pregled obavljen u stanici za tehnički pregled vozila koja se nalazi u urbanoj sredini i u maloj sredini - 56,5 % tehnički neispravnih.

U pogledu nezavisne varijable “index_stp”, a u okviru treće (završne) razine dubine stabla, čvorovi pod brojem 43, 44, 45 i 46 su završni čvorovi.

U tablici 4.7. (tablica dobiti po završnim čvorovima) sadržani su podaci o završnim čvorovima, odnosno o čvorovima kod kojih je prestalo grananje stabla u daljnje čvorove, za trening i testni uzorak. Završni čvorovi, sami po sebi, predstavljaju najbolji prediktivni potencijal predmetnog modela. U predmetnoj tablici sadržani su broj i postotak slučajeva/osobnih automobila po pojedinom završnom čvoru. U okviru stupca “Dobit” također su sadržani broj i postotak slučajeva/osobnih automobila po pojedinom završnom čvoru u okviru ciljane kategorije (broj i postotak slučajeva/osobnih automobila s utvrđenim tehničkim neispravnostima tijekom tehničkog pregleda što je rezultiralo “padom”).

Tablica 4.7. - Dobit po završnim čvorovima za VW Golf

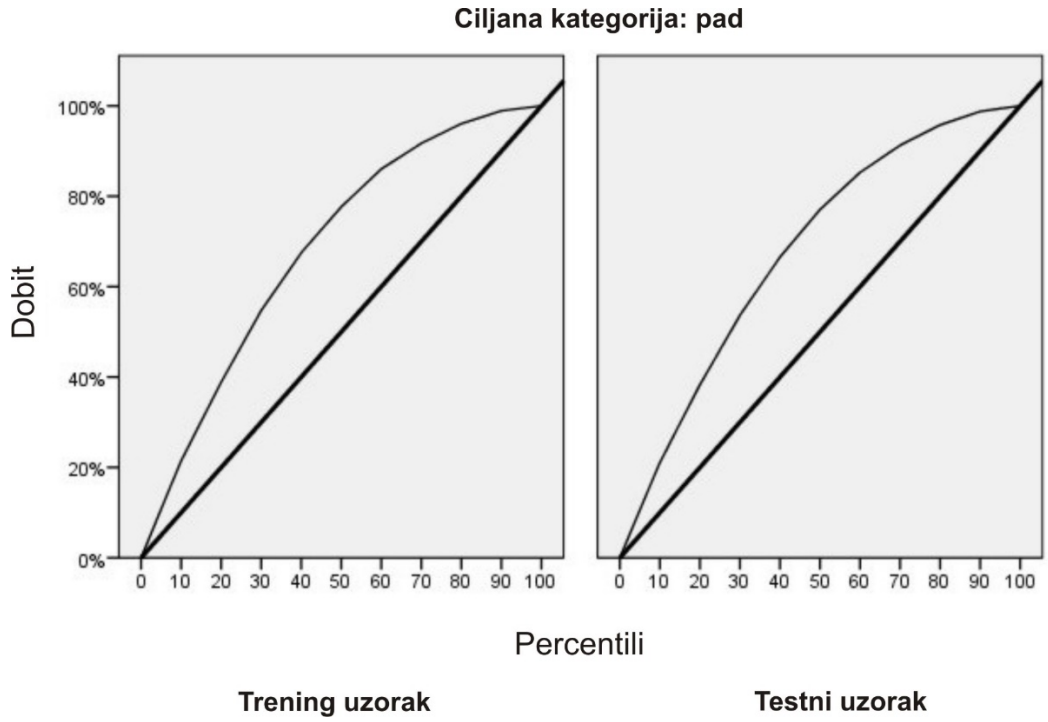
Uzorak	Završni čvor	Završni čvor		Dobit		Odziv	Indeks
		Ukupni broj VW Golf osobnih automobila po završnom čvoru	Postotak u odnosu na ukupni skup VW Golf osobnih automobila	Broj "palih" VW Golf osobnih automobila po završnom čvoru	Postotak u odnosu na ukupni skup "palih" VW Golf osobnih automobila		
	I	II	III	IV	V	VI	VII
Trening	31	129	0,8 %	111	1,9 %	86,1 %	255,6 %
	29	330	1,9 %	269	4,7 %	81,5 %	242,2 %
	33	193	1,1 %	136	2,4 %	70,5 %	209,4 %
	46	951	5,6 %	649	11,3 %	68,2 %	202,7 %
	26	264	1,5 %	173	3,0 %	65,5 %	194,7 %
	44	618	3,6 %	378	6,6 %	61,2 %	181,7 %
	34	126	0,7 %	74	1,3 %	58,7 %	174,5 %
	47	248	1,5 %	142	2,5 %	57,3 %	170,1 %
	32	1.113	6,5 %	601	10,5 %	54,0 %	160,4 %
	45	534	3,1 %	285	5,0 %	53,4 %	158,6 %
	42	840	4,9 %	447	7,8 %	53,2 %	158,1 %
	40	159	0,9 %	79	1,4 %	49,7 %	147,6 %
	28	269	1,6 %	125	2,2 %	46,5 %	138,1 %
	17	201	1,2 %	84	1,5 %	41,8 %	124,2 %
	25	211	1,2 %	86	1,5 %	40,8 %	121,1 %
	43	240	1,4 %	95	1,7 %	39,6 %	117,6 %
	21	1.433	8,4 %	514	8,9 %	35,9 %	106,6 %
	41	152	0,9 %	50	0,9 %	32,9 %	97,7 %
	16	699	4,1 %	219	3,8 %	31,3 %	93,1 %
	39	1.090	6,4 %	336	5,8 %	30,8 %	91,6 %
	38	491	2,9 %	101	1,8 %	20,6 %	61,1 %
	22	200	1,2 %	39	0,7 %	19,5 %	57,9 %
	3	1.955	11,4 %	372	6,5 %	19,0 %	56,5 %
	35	121	0,7 %	17	0,3 %	14,0 %	41,7 %
	37	694	4,1 %	89	1,5 %	12,8 %	38,1 %
	20	219	1,3 %	28	0,5 %	12,8 %	38,0 %
	14	1.092	6,4 %	139	2,4 %	12,7 %	37,8 %
	15	189	1,1 %	15	0,3 %	7,9 %	23,6 %
	18	190	1,1 %	11	0,2 %	5,8 %	17,2 %
	36	482	2,8 %	25	0,4 %	5,2 %	15,4 %
	12	869	5,1 %	42	0,7 %	4,8 %	14,4 %
	11	775	4,5 %	17	0,3 %	2,2 %	6,5 %
Test	31	736	0,8 %	603	1,9 %	81,9 %	248,4 %
	29	1.822	1,9 %	1.461	4,5 %	80,2 %	243,1 %
	33	1.155	1,2 %	736	2,3 %	63,7 %	193,2 %
	46	5.497	5,6 %	3.656	11,3 %	66,5 %	201,7 %
	26	1.507	1,5 %	903	2,8 %	59,9 %	181,7 %
	44	3.593	3,7 %	2.096	6,5 %	58,3 %	176,9 %
	34	789	0,8 %	467	1,4 %	59,2 %	179,5 %
	47	1.440	1,5 %	833	2,6 %	57,8 %	175,4 %
	32	6.381	6,5 %	3.371	10,5 %	52,8 %	160,2 %
	45	2.982	3,1 %	1.565	4,9 %	52,5 %	159,1 %
	42	4.991	5,1 %	2.336	7,3 %	46,8 %	141,9 %
	40	914	0,9 %	414	1,3 %	45,3 %	137,3 %
	28	1.703	1,7 %	905	2,8 %	53,1 %	116,1 %
	17	1.069	1,1 %	429	1,3 %	40,1 %	121,7 %
	25	1.112	1,1 %	495	1,5 %	44,5 %	135,0 %
	43	1.390	1,4 %	453	1,4 %	32,6 %	98,8 %
	21	8.039	8,2 %	2.880	8,9 %	35,8 %	108,6 %
	41	894	0,9 %	353	1,1 %	39,5 %	119,7 %
	16	4.110	4,2 %	1.264	3,9 %	30,8 %	93,2 %
	39	6.254	6,4 %	1.820	5,6 %	29,1 %	88,2 %
	38	2.963	3,0 %	567	1,8 %	19,1 %	58,0 %
	22	1.041	1,1 %	204	0,6 %	19,6 %	59,4 %
	3	10.675	10,9 %	2.125	6,6 %	19,9 %	60,4 %
	35	728	0,7 %	63	0,2 %	8,7 %	26,2 %
	37	3.735	3,8 %	569	1,8 %	15,2 %	46,2 %
	20	1.422	1,5 %	141	0,4 %	9,9 %	30,1 %
	14	6.083	6,2 %	729	2,3 %	12,0 %	36,3 %
	15	1.040	1,1 %	66	0,2 %	6,3 %	19,2 %

	18	1.246	1,3 %	97	0,3 %	7,8 %	23,6 %
	36	2.921	3,0 %	248	0,8 %	8,5 %	25,7 %
	12	4.948	5,1 %	220	0,7 %	4,4 %	13,5 %
	11	4.501	4,6 %	147	0,5 %	3,3 %	9,9 %

Za kategorijsku zavisnu varijablu, stupac “Odziv” prikazuje unutarnju strukturu završnih čvorova, odnosno u istom su sadržani postotci slučajeva/osobnih automobila u pojedinom završnom čvoru u okviru predmetne ciljane kategorije (osobni automobili koji su “pali” na tehničkom pregledu”), a u odnosu na ukupni broj slučajeva/osobnih automobila po pojedinom završnom čvoru (prema: IV x 100/II).

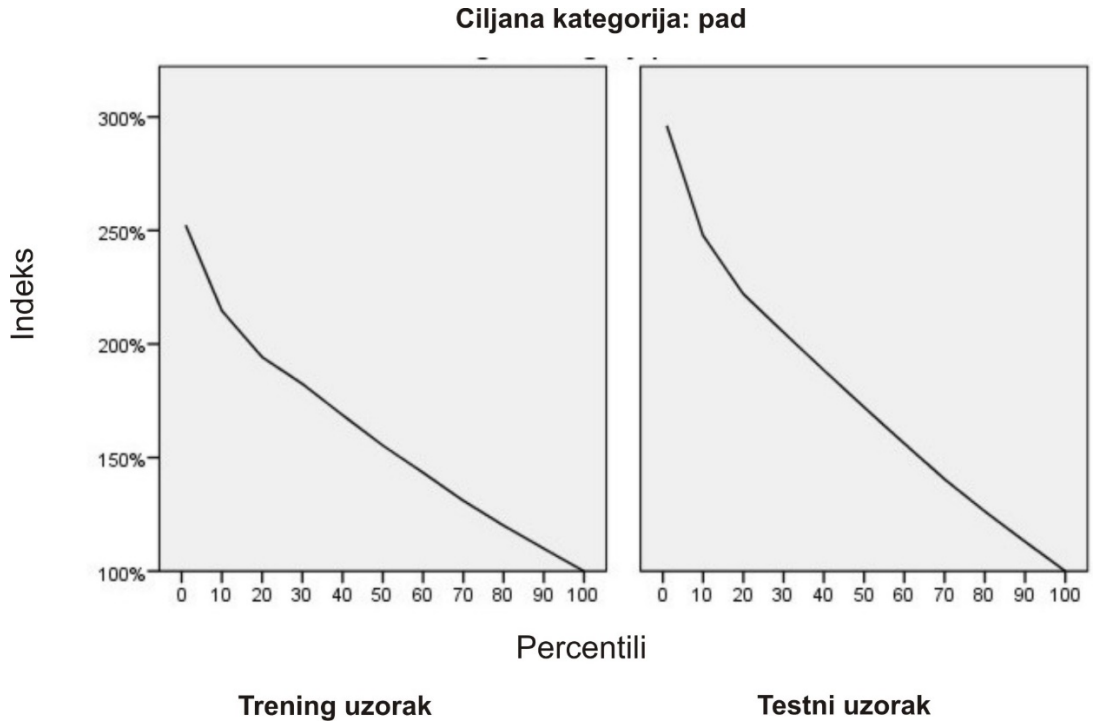
Za kategorijsku zavisnu varijablu indeks podrazumijeva omjer odziva za ciljanu kategoriju u odnosu na odziv za ukupni skup. Drugim riječima, pomoću indeksa se dobiva informacija o tome koliko se značajno postotak promatrane ciljane kategorije za predmetni završni čvor razlikuje u odnosu na očekivani postotak ciljane kategorije (za ukupni skup) koji je izražen u korijenskom čvoru, te za osobne automobile VW Golf, u trening uzorku, iznosi 33,7 % (na primjer, aproksimativno, u završnom čvoru broj 31 postotna vrijednost indeksa iznosi 86,1 % x 100/33,7 % = 255,6 %). Vrijednost indeksa koja prelazi 100 % ukazuje da se u predmetnom završnom čvoru nalazi veći postotni iznos ciljane kategorije u usporedbi s očekivanim postotnim iznosom ciljane kategorije za ukupni skup.

Na slici 4.5. dan je grafički prikaz dobiti za ciljanu kategoriju “pad” nezavisne varijable “tp_prolaz”, za trening i testni uzorak. Na apscisi se nalaze percentili dok se na ordinati nalaze vrijednosti dobiti izražene u postotcima.



Slika 4.5. - Dijagram dobiti za ciljanu kategoriju "pad" za VW Golf

Na slici 4.6. dan je grafički prikaz indeksa za ciljanu kategoriju "pad" nezavisne varijable "tp_prolaz", za trening i testni uzorak. Na apscisi se nalaze percentili dok se na ordinati nalaze vrijednosti indeksa izražene u postotcima.



Slika 4.6. - Dijagram indeksa za ciljanu kategoriju "pad" za VW Golf

Temeljem podataka iz tablica 4.8. i 4.9. donosi se ocjena o prediktivnom potencijalu predmetnog modela.

Tablica 4.8.- Rizik za VW Golf

Uzorak	Procijenjena vrijednost	Standardna devijacija
Trening	0,267	0,003
Testni	0,277	0,001

Iz podataka sadržanih u tablici rizika prikazan je rizik predikcije koji predstavlja postotak pogrešne klasifikacije kategorija. U pogledu provedenog istraživanja proizlazi da, u slučaju kada su poznate karakteristike osobnih automobila VW Golf u smislu predmetnih sedam nezavisnih (prediktivnih) varijabli, rizik pogrešne klasifikacije osobnih automobila VW Golf glede utvrđivanja tehničke neispravnosti tijekom tehničkog pregleda (odnosno "pada" na tehničkom pregledu) jest 26,7 % (primjenom metode dijeljenja uzorka) za trening uzorak, odnosno 27,7 % za testni uzorak.

Podaci iz kontingencijske tablice potvrđuju podatke iz tablice rizika u smislu da predmetni model, u ukupno 73,3 % slučajeva (osobnih automobila VW Golf) trening uzorka točno klasificira kategoriju, odnosno u ukupno 72,3 % slučajeva testnog uzorka.

Tablica 4.9. - Klasifikacija za VW Golf (Exhaustive CHAID)

Uzorak	Promatrane kategorije	Prediktivno		
		„pad“	„prolaz“	Postotak točnosti
Trening	„pad“	**3.265**	**2.483**	**56,8 %**
	„prolaz“	**2.081**	**9.248**	81,6 %
	Ukupni postotak	31,3 %	68,7 %	**73,3 %**
Testni	„pad“	**18.027**	**14.189**	**56,0 %**
	„prolaz“	**12.866**	**52.599**	80,3 %
	Ukupni postotak	31,6 %	68,4 %	**72,3 %**

Nadalje, u kontekstu analize učinkovitosti prediktivnog modela, prema izrazu 4.2. izračunat je MCC za trening i testni uzorak, a rezultati su prikazani u tablici 4.10.

Tablica 4.10. – Prikaz vrijednosti MCC za VW Golf (Exhaustive CHAID)

Uzorak	Kriterij	Vrijednost
Trening	pogrešno pozitivni	2.483
	pogrešno negativni	2.081
	stvarno pozitivni	9.248
	stvarno negativni	3.265
	MCC	**0,37**
Testni	pogrešno pozitivni	14.189
	pogrešno negativni	12.866
	stvarno pozitivni	52.599
	stvarno negativni	18.027
	MCC	**0,35**

Kao što je prikazano u tablici 4.11., nakon primjene Exhaustive CHAID algoritma na bazi podataka osobnih automobila VW Golf izraženoj u vidu sedam nezavisnih varijabli te jedne zavisne varijable, u okviru statističkog paketa [64] generirane su četiri nove varijable u predmetnoj bazi podataka.

Varijabla “Broj čvora” predstavlja broj završnog čvora pojedinog slučaja/osobnog automobila.

Varijabla “Prediktivna vrijednost ” predstavlja prediktivnu vrijednost zavisne varijable za svaki pojedini slučaj/osobni automobil. S obzirom da u okviru zavisne varijable postoje dvije prediktivne kategorije – “0” (“pad”) i “1” (“prolaz”) – prediktivna vrijednost “0” podrazumijeva da se predviđa tehnička neispravnost predmetnog osobnog automobila, odnosno njegov “pad” na tehničkom pregledu.

Varijabla “Prediktivna vjerojatnost” predstavlja vjerojatnost da će pojedini slučaj/osobni automobil potpasti pod određenu kategoriju zavisne varijable. S obzirom na činjenicu da postoje samo dvije kategorije varijabla “ Prediktivna vjerojatnost 1” predstavlja vjerojatnost da će predmetni slučaj/osobni automobil potpasti pod kategoriju “pad”, dok varijabla “ Prediktivna vjerojatnost 2” predstavlja vjerojatnost da će predmetni slučaj/osobni automobil potpasti pod kategoriju “prolaz”.

Tablica 4.11. – Izvod dijela prediktivnih vrijednosti za VW Golf

Redni broj	„snaga“	„ccm“	„starost_god“	„index_stp“	„vr_motora“	„vr_kocnica“	„index_km“	„tp_prolaz“	Broj čvora	Prediktivna vrijednost	Prediktivna vjerojatnost 1	Prediktivna vjerojatnost 2
1.	37	1093	33	3	1	1	5	1	11	0	0,69	0,31
2.	37	1093	32	1	1	1	5	0	11	0	0,69	0,31
3.	37	1093	31	1	1	1	5	0	11	0	0,69	0,31
4.	37	1093	31	1	1	1	4	0	11	0	0,69	0,31
5.	37	1093	30	1	1	1	3	1	11	0	0,69	0,31
6.	37	1093	30	3	3	1	5	1	35	0	0,56	0,44
7.	37	1093	34	1	3	1	4	0	14	0	0,83	0,17
8.	37	1093	34	3	1	1	5	1	11	0	0,69	0,31
9.	37	1093	34	1	1	1	5	0	11	0	0,69	0,31
...	...	...	...	...	...	...	...	...	...	...	...	...

Drugim riječima, varijabla “Prediktivna vjerojatnost” predstavlja opseg slučajeva/osobnih automobila u pojedinoj kategoriji zavisne varijable za završni čvor u okviru kojeg se nalazi pojedini slučaj/osobni automobil. U primjeru iz tablice 4.11., u čvoru 35, 56 % slučajeva/osobnih automobila potpada pod kategoriju “pad” (“Prediktivna vrijednost” - 0), a 44 % slučajeva/osobnih automobila potpada pod kategoriju “prolaz” (“Prediktivna vrijednost ” - 1), što predstavlja prediktivne vjerojatnosti od 0,56 i 0,44.

4.4.1.1. Analiza prediktivnog potencijala modela temeljenog na Exhaustive CHAID algoritmu

Procjena prediktivnog potencijala modela proizlazi iz analize podataka sadržanih u tablici rizika (tablica 4.8.) i kontingencijskoj tablici (tablica 4.9.).

Iz kontingencijske tablice (tablica 4.9.) proizlazi podatak o 56,8 % točno predviđenih slučajeva/osobnih automobila u kategoriji “pad" u okviru trening uzorka te 56,0 % u okviru testnog uzorka, odnosno čak 43,2 % pogrešno klasificiranih slučajeva (pogrešno klasificiranih tehnički neispravnih osobnih automobila u kategoriju tehničkih ispravnih osobnih automobila koji su “prošli” tehnički pregled) u okviru trening uzorka te 44,0 % u okviru testnog uzorka. Ovakav oblik pogrešne klasifikacije ima mnogo veći negativni značaj u kontekstu sigurnosti cestovnog prometa, od obrnutog slučaja – kod

pogrešno klasificiranih tehnički ispravnih osobnih automobila u kategoriju tehničkih neispravnih osobnih automobila koji su “pali” na tehničkom pregledu.

Potrebno je naglasiti da u literaturi koja opisuje CHAID algoritme (bilo da se radi o originalnoj verziji, odnosno poboljšanoj inačici), uključujući istraživanja citirana u ovom radu, ne postoji striktno određenje granične vrijednosti postotnog iznosa točno predviđenih slučajeva ciljane kategorije zavisne varijable (niti granične vrijednosti iznosa procijenjene vrijednosti rizika predikcije), koja bi se smatrala minimalno zadovoljavajućim u kontekstu prediktivnog potencijala modela.

Proizlazi da su navedene granične vrijednosti stvar su pojedinačne procjene onoga tko provodi istraživanje, u ovisnosti glede specifičnosti pojedinog prediktivnog modela i zadanih prioriteta. Iz analize dosadašnjih istraživanja vidljivo je da su ti iznosi značajno različiti (za točnost predviđenih slučajeva ciljane kategorije zavisne varijable): 44,69 % [51], 61,80 % [54], 60,00 % [67], 64,53 % [68], 91,6 % [52].

U skladu s prethodno navedenim te uvažavajući specifičnost istraživanja opisanog u ovom radu, u pogledu prediktivnog modela tehničke ispravnosti vozila (osobnih automobila) postotni iznos od 70 % točno predviđenih slučajeva/osobnih automobila u kategoriji “pad" smatra se minimalno zadovoljavajućim u kontekstu prediktivnog potencijala modela.

U skladu s prethodno navedenim proizlazi da iznos od 56,8 % (56,0 %) točno predviđenih slučajeva/osobnih automobila u kategoriji “pad", odnosno čak 43,2 % (44,0 %) pogrešno klasificiranih slučajeva nije zadovoljavajući s obzirom da se radi o ciljanoj kategoriji zavisne varijable, odnosno model nema zadovoljavajući prediktivni potencijal, te je potrebno izvršiti radnju radi poboljšanja (korektivne radnje) koja će dovesti do povećanja prediktivnog potencijala modela.

Vrijednosti korelacijskog koeficijenta MCC za oba uzorka (trening i testni uzorak) su srednje pozitivni +0,37 i +0,35, što podrazumijeva dobru predikciju već sada (teoretska vrijednost +1 podrazumijeva savršenu predikciju, vrijednost 0 podrazumijeva predikciju koja nije bolja od nasumične predikcije, teoretska vrijednost -1 podrazumijeva potpuno razmimoilaženje između predikcije i opažanja).

4.4.1.2. Poboljšanje prediktivnog potencijala modela temeljenog na Exhaustive CHAID algoritmu

U cilju povećanja prediktivnog potencijala modela potrebno je izvršiti radnju za poboljšanje na način da se redefiniraju zadane vrijednosti penalizacije, odnosno razina negativnih posljedica, kod pogrešne klasifikacije prediktivnih kategorija zavisne varijable.

Za razliku od prvotno definirane penalizacije, odnosno razina negativnih posljedica, kada su obje prediktivne kategorije zavisne varijable vrednovane jednakim vrijednostima (vrijednost 1 za kategoriju “prolaz” i vrijednost 1 za kategoriju “pad”), korektivna radnja podrazumijeva različito vrednovanje penalizacije, odnosno razina negativnih posljedica prediktivnih kategorija zavisne varijable

Različito vrednovanje penalizacije, odnosno razina negativnih posljedica, prediktivnih kategorija zavisne varijable mora biti u skladu s logikom da je penalizacija, odnosno razina negativnih posljedica kod pogrešnog klasificiranja vozila koje je tehnički neispravno u klasu tehnički ispravnih vozila mnogo veći od penalizacije, odnosno razine negativnih posljedica kod pogrešnog klasificiranja vozila koje je tehnički ispravno u klasu tehnički neispravnih vozila.

Sukladno navedenom, za potrebe istraživanja koje je predmet ovog doktorskog rada, a kao što je prethodno navedeno, korektivna radnja u predmetnom kontekstu podrazumijeva vrednovanje penalizacije, odnosno razine negativnih posljedica kod pogrešnog klasificiranja osobnog automobila koji je tehnički neispravan u klasu tehnički ispravnih osobnih automobila s cjelobrojnim vrijednostima većim od jedan (počev od vrijednosti 2), naspram vrijednosti 1 za penalizaciju, odnosno razinu negativnih posljedica kod pogrešnog klasificiranja osobnog automobila koji je tehnički ispravan u klasu tehnički neispravnih osobnih automobila. Prva cjelobrojna vrijednost veća od 1 kod koje se postiže postotni iznos od 70 % točno predviđenih slučajeva/osobnih automobila u kategoriji “pad" smatra se optimalnom vrijednošću penalizacije.
Prethodno navedeno vrednovanje penalizacije, u kontekstu osobnih automobila VW Golf, izvršeno je u omjeru 2:1 (dvostruko veća vrijednost negativnih posljedica kod

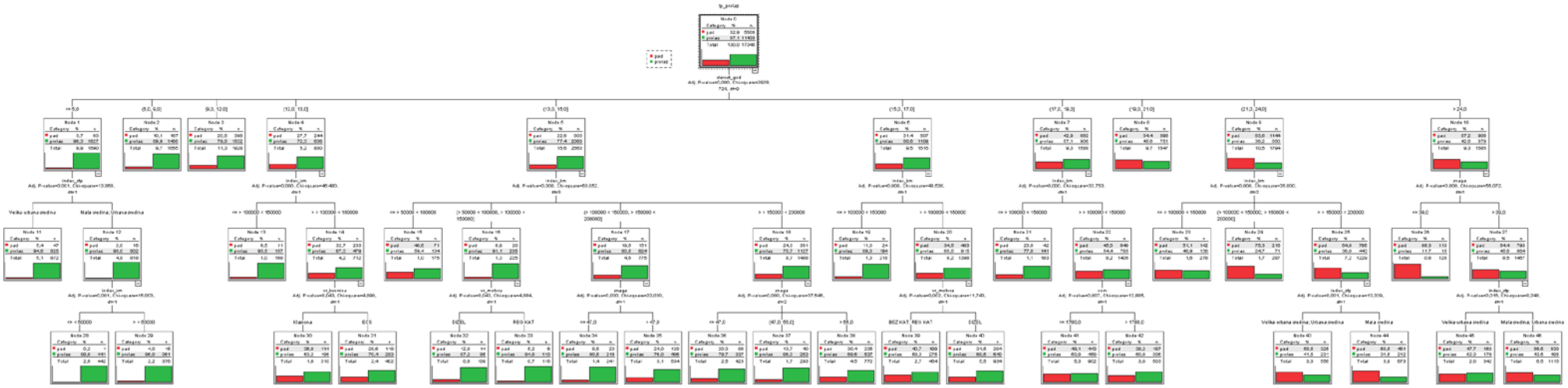
pogrešnog klasificiranja osobnog automobila koji je tehnički neispravan u klasu tehnički ispravnih osobnih automobila, nego u obrnutom slučaju) nakon provedenog eksperimentiranja, iz razloga što je upravo pri predmetnom omjeru postignut minimalni određeni postotni iznos od 70 % točno predviđenih slučajeva/osobnih automobila u kategoriji “pad" (što je vidljivo iz tablica 4.15., 4.26., 4.38. i 6.1. ovog rada), čime se smatra da je postignut zadovoljavajući prediktivni potencijal modela (u skladu s točkom 4.4.1.1. ovog rada).

Drugim riječima, penalizacija pogrešnog klasificiranja osobnog automobila VW Golf koji je tehnički neispravan u klasu tehnički ispravnih osobnih automobila je dva puta veća od penalizacije pogrešnog klasificiranja osobnog automobila koji je tehnički ispravan u klasu tehnički neispravnih osobnih automobila.

4.4.1.3. Analiza nakon provedenog poboljšanja modela temeljenog na Exhaustive CHAID algoritmu

Provođenjem predmetne radnje vidljive su promjene u stablu odlučivanja prediktivnog modela tehničke ispravnosti osobnih automobila VW Golf. Predmetne promjene ne odnose se na distribuciju slučajeva/osobnih automobila po pojedinim čvorovima, koja je ostala ista, već se odnose na dominantne prediktivne kategorije zavisne varijable, kao što je prikazano na slici 4.7.

S obzirom na dvostruko veće vrednovanje penalizacije pogrešnog klasificiranja osobnog automobila koji je tehnički neispravan u klasu tehnički ispravnih osobnih automobila u odnosu na penalizaciju pogrešnog klasificiranja osobnog automobila koji je tehnički ispravan u klasu tehnički neispravnih osobnih automobila, u svim čvorovima koji su prije provođenja korektivne radnje imali podjednaku distribuciju obje prediktivne kategorije, ili nešto veću distribuciju kategorije “prolaz”, nakon provođenja korektivne radnje u istima će dominirati kategorija “pad”.



Slika 4.7. - Stablo odlučivanja trening uzorka prediktivnog modela za VW Golf nakon provođenja radnje za poboljšanje modela

Promjene dominantne prediktivne kategorije, iz kategorije “prolaz” u kategoriju “pad” zabilježene su u čvorovima broj 20 i 22, te u završnim čvorovima broj 15, 39, 41, 42 i 45.

Kao što je vidljivo iz tablice 4.12. nakon provođenja radnje za poboljšanje prediktivnog modela došlo je do promjene rezultata u odnosnu na rezultate prije provođenja radnje za poboljšanje prediktivnog modela (tablica 4.6.) i to u broju čvorova, broju završnih čvorova te u pogledu nezavisnih varijabli koje su pridonijele modelu (u model je uključeno svih sedam nezavisnih varijabli).

Tablica 4.12. – Zbirni rezultati modela za VW Golf nakon provođenja radnje za poboljšanje (Exhaustive CHAID)

Specifikacije	Metoda	Exhaustive CHAID
	Zavisna varijabla	tp_prolaz
	Nezavisne varijable	snaga, ccm, index_stp, vr_motora, vr_kocnica, index_km, starost_god
	Validacija	Metoda dijeljenja uzorka
	Najveća dubina stabla (broj razina)	3
	Najmanji broj slučajeva u “roditeljskom” čvoru	100
	Najmanji broj slučajeva u “dječjem” čvoru	100
Rezultati	Uključene nezavisne varijable	**starost_god, index_stp, index_km, vr_kocnica, vr_motora, snaga, ccm**
	Broj čvorova	**47**
	Broj završnih čvorova	**30**
	Dubina (broj razina)	3

Nadalje, provođenjem radnje za poboljšanje modela došlo je do promjena u tablici rizika (tablica 4.13.), te kontingencijskoj tablici (tablica 4.14.).

Tablica 4.13. - Rizik za VW Golf nakon provođenja radnje za poboljšanje

Uzorak	Procijenjena vrijednost	Standardna devijacija
Trening	0,394	0,005
Testni	0,397	0,002

“Nusprodukt” provođenja radnje za poboljšanje modela jest očekivano povećanje rizika predikcije, u okviru trening uzorka, s 26,7% prije provođenja korektivne radnje, na 39,4 % poslije provođenja korektivne radnje, odnosno u okviru testnog uzorka s 27,7% na 39,7 %

Međutim, provođenje radnje za poboljšanje modela dovelo je do željenog efekta, odnosno značajnog povećanja prediktivnog potencijala modela u kontekstu ciljane prediktivne kategorije zavisne varijable “pad” koji proizlazi iz većeg postotka točno klasificiranih slučajeva/osobnih automobila glede utvrđivanja tehničke neispravnosti tijekom tehničkog pregleda (odnosno “pada” na tehničkom pregledu), i to u okviru trening uzorka s 56,8 %, prije provođenja radnje za poboljšanje modela, na 70,6 % nakon provođenja radnje za poboljšanje modela, odnosno u okviru testnog uzorka s 56,0 % na 70,3 %, što je predstavlja rezultat koji zadovoljava, sukladno točci 4.4.1.1.

Tablica 4.14. - Klasifikacija za VW Golf nakon provođenja radnje za poboljšanje (Exhaustive CHAID)

Uzorak	**Promatrane kategorije**	**Prediktivno**		
		„pad“	„prolaz“	Postotak točnosti
Trening	„pad“	**3.960**	**1.648**	**70,6 %**
	„prolaz“	**3.420**	**8.018**	70,1 %
	Ukupni postotak	43,3 %	56,7 %	**70,3 %**
Testni	„pad“	**22.759**	**9.597**	**70,3 %**
	„prolaz“	**19.596**	**45.760**	70,0 %
	Ukupni postotak	43,3 %	56,7 %	**70,1 %**

Povećanje postotka točno klasificiranih slučajeva/osobnih automobila kod kojih je utvrđena tehnička neispravnost tijekom tehničkog pregleda (što za posljedicu ima “pad” osobnih automobila na tehničkom pregledu) za direktnu posljedicu ima smanjeni postotak točno klasificiranih slučajeva/osobnih automobila kod kojih nije utvrđena tehnička neispravnost tijekom tehničkog pregleda (kategorija “prolaz”), s 81,6 % u okviru trening uzorka na 70,1 %, te u okviru testnog uzorka s 80,3 % na 70,0 %(odnosno pad ukupnog postotka točno klasificiranih slučajeva/osobnih automobila u okviru trening uzorka s 73,3 % na 70,3 %, te u okviru testnog uzorka s 72,3 % na 70,1 %), međutim sukladno prethodno iznesenoj argumentaciji radi se o prediktivnoj kategoriji sekundarnog značaja, odnosno radi se o prediktivnoj kategoriji koja ima manji značaj za ovo istraživanje s obzirom da ista nema negativan utjecaj na sigurnost prometa na cestama (osobni automobili kod kojih nije utvrđena tehnička neispravnost

pogrešno su klasificirani kao osobni automobili koji su “pali” na tehničkom pregledu zbog utvrđene tehničke neispravnosti).

Nadalje, u kontekstu analize učinkovitosti prediktivnog modela, prema izrazu 4.2. nakon izvršene radnje za poboljšanje modela izračunat je MCC za trening i testni uzorak, a rezultati su prikazani u tablici 4.15.

Tablica 4.15. – Prikaz vrijednosti MCC za VW Golf nakon provođenja radnje za poboljšanje (Exhaustive CHAID)

Uzorak	Kriterij	Vrijednost
Trening	pogrešno pozitivni	1.648
	pogrešno negativni	3.420
	stvarno pozitivni	8.018
	stvarno negativni	3.960
	MCC	**0,38**
Testni	pogrešno pozitivni	9.597
	pogrešno negativni	19.596
	stvarno pozitivni	45.760
	stvarno negativni	22.759
	MCC	**0,37**

Vrijednosti korelacijskog koeficijenta MCC za oba uzorka (trening i testni uzorak), nakon izvršene radnje za poboljšanje modela, su i dalje srednje pozitivne +0,38 i +0,37 (uz minimalno povećanje), što potvrđuje dobru predikciju.

U konačnici, prediktivni model tehničke ispravnosti za osobne automobile VW Golf a koji se temelji na metodi stabla odlučivanja i algoritmu Exhaustive CHAID, može se izraziti i u obliku poslovnih pravila, kako slijedi u tablici 4.16.:

Tablica 4.16. – Prikaz poslovnih pravila za poboljšani Exhaustive CHAID prediktivni model VW Golf

/* Čvor 11 */.
IF (starost_god NOT MISSING AND (starost_god <= 5)) AND (index_stp != "Mala sredina" AND index_stp != "Urbana sredina") THEN **Prediction = 1, Probability = 0.946101**

/* Čvor 28 */.
IF (starost_god NOT MISSING AND (starost_god <= 5)) AND (index_stp = "Mala sredina" OR index_stp = "Urbana sredina") AND (index_km IS MISSING OR (index_km <= "< 50000")) THEN **Prediction = 1, Probability = 0.997738**

/* Čvor 29 */.
IF (starost_god NOT MISSING AND (starost_god <= 5)) AND (index_stp = "Mala sredina" OR index_stp = "Urbana sredina") AND (index_km NOT MISSING AND (index_km > "< 50000")) THEN **Prediction = 1, Probability = 0.960106**

/* Čvor 2 */.
IF (starost_god NOT MISSING AND (starost_god > 5 AND starost_god <= 9)) THEN **Prediction = 1, Probability = 0.899094**

/* Čvor 3 */.
IF (starost_god NOT MISSING AND (starost_god > 9 AND starost_god <= 12)) THEN **Prediction = 1, Probability = 0.794606**

/* Čvor 13 */.
IF (starost_god NOT MISSING AND (starost_god > 12 AND starost_god <= 13)) AND (index_km NOT MISSING AND (index_km <= "> 100000 < 150000")) THEN **Prediction = 1, Probability = 0.934524**

/* Čvor 30 */.
IF (starost_god NOT MISSING AND (starost_god > 12 AND starost_god <= 13)) AND (index_km IS MISSING OR (index_km > "> 100000 < 150000")) AND (vr_kocnica = "Klasicna") THEN **Prediction = 0, Probability = 0.367742**

/* Čvor 31 */.
IF (starost_god NOT MISSING AND (starost_god > 12 AND starost_god <= 13)) AND (index_km IS MISSING OR (index_km > "> 100000 < 150000")) AND (vr_kocnica != "Klasicna") THEN **Prediction = 1, Probability = 0.703980**

/* Čvor 15 */.
IF (starost_god IS MISSING OR (starost_god > 13 AND starost_god <= 15)) AND (index_km NOT MISSING AND (index_km <= "> 50000 < 100000")) THEN **Prediction = 0, Probability = 0.405714**

/* Čvor 32 */.
IF (starost_god IS MISSING OR (starost_god > 13 AND starost_god <= 15)) AND (index_km NOT MISSING AND (index_km > "> 50000 < 100000" AND index_km <= "> 100000 < 150000")) AND (vr_motora = "DIZEL") THEN **Prediction = 1, Probability = 0.871560**

/* Čvor 33 */.
IF (starost_god IS MISSING OR (starost_god > 13 AND starost_god <= 15)) AND (index_km NOT MISSING AND (index_km > "> 50000 < 100000" AND index_km <= "> 100000 < 150000")) AND (vr_motora != "DIZEL") THEN **Prediction = 1, Probability = 0.948276**

/* Čvor 34 */.
IF (starost_god IS MISSING OR (starost_god > 13 AND starost_god <= 15)) AND (index_km NOT MISSING AND (index_km > "> 100000 < 150000" AND index_km <= "> 150000 < 200000")) AND (snaga NOT MISSING AND (snaga <= 47)) THEN **Prediction = 1, Probability = 0.904564**

/* Čvor 35 */.
IF (starost_god IS MISSING OR (starost_god > 13 AND starost_god <= 15)) AND (index_km NOT MISSING AND (index_km > "> 100000 < 150000" AND index_km <= "> 150000 < 200000")) AND (snaga IS MISSING OR (snaga > 47)) THEN **Prediction = 1, Probability = 0.760300**

/* Čvor 36 */.
IF (starost_god IS MISSING OR (starost_god > 13 AND starost_god <= 15)) AND (index_km IS MISSING OR (index_km > "> 150000 < 200000")) AND (snaga NOT MISSING AND (snaga <= 47)) THEN **Prediction = 1, Probability = 0.796690**

/* Čvor 37 */.
IF (starost_god IS MISSING OR (starost_god > 13 AND starost_god <= 15)) AND (index_km IS MISSING OR (index_km > "> 150000 < 200000")) AND (snaga NOT MISSING AND (snaga > 47 AND snaga <= 55)) THEN **Prediction = 1, Probability = 0.863481**

/* Čvor 38 */.
IF (starost_god IS MISSING OR (starost_god > 13 AND starost_god <= 15)) AND (index_km IS MISSING OR (index_km > "> 150000 < 200000")) AND (snaga IS MISSING OR (snaga > 55)) THEN **Prediction = 1, Probability = 0.695596**

/* Čvor 19 */.
IF (starost_god NOT MISSING AND (starost_god > 15 AND starost_god <= 17)) AND (index_km NOT MISSING AND (index_km <= "> 100000 < 150000")) THEN **Prediction = 1, Probability = 0.889908**

/* Čvor 39 */.
IF (starost_god NOT MISSING AND (starost_god > 15 AND starost_god <= 17)) AND (index_km IS MISSING OR (index_km > "> 100000 < 150000")) AND (vr_motora = "BEZ KAT" OR vr_motora = "REG KAT") THEN **Prediction = 0, Probability = 0.407328**

/* Čvor 40 */.
IF (starost_god NOT MISSING AND (starost_god > 15 AND starost_god <= 17)) AND (index_km IS MISSING OR (index_km > "> 100000 < 150000")) AND (vr_motora != "BEZ KAT" AND vr_motora != "REG KAT") THEN **Prediction = 1, Probability = 0.685225**

/* Čvor 21 */.
IF (starost_god NOT MISSING AND (starost_god > 17 AND starost_god <= 19)) AND (index_km NOT MISSING AND (index_km <= "> 100000 < 150000")) THEN **Prediction = 1, Probability = 0.770492**

/* Čvor 41 */.
IF (starost_god NOT MISSING AND (starost_god > 17 AND starost_god <= 19)) AND (index_km IS MISSING OR (index_km > "> 100000 < 150000")) AND (ccm IS MISSING OR (ccm <= 1780)) THEN **Prediction = 0, Probability = 0.491131**

/* Čvor 42 */.
IF (starost_god NOT MISSING AND (starost_god > 17 AND starost_god <= 19)) AND (index_km IS MISSING OR (index_km > "> 100000 < 150000")) AND (ccm NOT MISSING AND (ccm > 1780)) THEN **Prediction = 0, Probability = 0.391650**

/* Čvor 8 */.
IF (starost_god NOT MISSING AND (starost_god > 19 AND starost_god <= 21)) THEN **Prediction = 0, Probability = 0.544019**

/* Čvor 23 */.
IF (starost_god NOT MISSING AND (starost_god > 21 AND starost_god <= 24)) AND (index_km NOT MISSING AND (index_km <= "> 100000 < 150000")) THEN **Prediction = 0, Probability = 0.510791**

/* Čvor 24 */.
IF (starost_god NOT MISSING AND (starost_god > 21 AND starost_god <= 24)) AND (index_km NOT MISSING AND (index_km > "> 100000 < 150000" AND index_km <= "> 150000 < 200000"))
THEN **Prediction = 0, Probability = 0.752613**

/* Čvor 43 */.
IF (starost_god NOT MISSING AND (starost_god > 21 AND starost_god <= 24)) AND (index_km IS MISSING OR (index_km > "> 150000 < 200000")) AND (index_stp = "Velika urbana sredina" OR index_stp = "Urbana sredina") THEN **Prediction = 0, Probability = 0.584532**

/* Čvor 44 */.
IF (starost_god NOT MISSING AND (starost_god > 21 AND starost_god <= 24)) AND (index_km IS MISSING OR (index_km > "> 150000 < 200000")) AND (index_stp != "Velika urbana sredina" AND index_stp != "Urbana sredina") THEN **Prediction = 0, Probability = 0.684993**

/* Čvor 26 */.
IF (starost_god NOT MISSING AND (starost_god > 24)) AND (snaga NOT MISSING AND (snaga <= 39)) THEN **Prediction = 0, Probability = 0.882813**

/* Čvor 45 */.
IF (starost_god NOT MISSING AND (starost_god > 24)) AND (snaga IS MISSING OR (snaga > 39)) AND (index_stp = "Velika urbana sredina") THEN **Prediction = 0, Probability = 0.476608**

/* Čvor 46 */.
IF (starost_god NOT MISSING AND (starost_god > 24)) AND (snaga IS MISSING OR (snaga > 39)) AND (index_stp != "Velika urbana sredina") THEN **Prediction = 0, Probability = 0.565022**

4.4.1.4. Alternativni prediktivni model temeljen na algoritmu perceptronskih neuronskih mreža

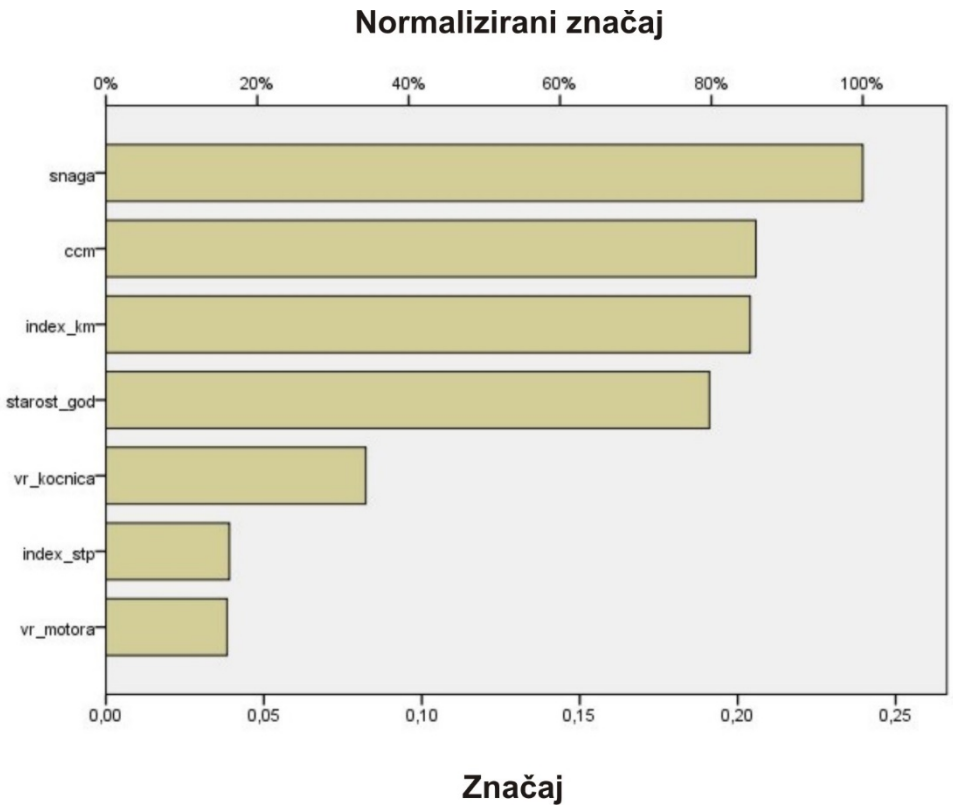
Kao alternativa preferencijalnom prediktivnom modelu koji se temelji na Exhaustive CHAID algoritmu korišten je algoritam perceptronske neuronske mreže. U tablici 4.17. dane su parametri strukture neuronske mreže za osobne automobile VW Golf.

Tablica 4.17. – Parametri strukture neuronske mreže za VW Golf

Ulazni sloj	Varijable	1	„snaga“
		2	„ccm“
		3	„index_stp“
		4	„vr_motora“
		5	„vr_kocnica“
		6	„index_km“
		7	„starost_god“
	Broj neurona		87
	Broj skrivenih slojeva		1
Skriveni sloj(evi)	Broj neurona u skrivenim slojevima 1		8
	Aktivacijska funkcija		Hipebolička tangenta
Izlazni sloj	Zavisne varijable 1		„tp_prolaz“
	Broj neurona		2
	Aktivacijska funkcija		softmax
	Funkcija pogreške		križna entropija

Kao što je vidljivo iz slike 4.8. procjena značaja u smislu prediktivnog potencijala pojedinih nezavisnih varijabli za osobne automobile VW Golf proizašlih iz algoritma neuronske mreže značajno se razlikuje od procjene njihovog značaja proizašle iz Exhaustive CHAID algoritma.

Proizlazi da je najznačajnija nezavisna varijabla „snaga“ koju prema razini značajnosti slijede nezavisne varijable „ccm“, „index_km“ te „starost_god“, dok nezavisne varijable „vr_kocnica“, „index_stp“ i „vr_motora“ imaju najmanji značaj.



Slika 4.8. – Prikaz nezavisnih varijabli prema procjeni razine značaja za VW Golf (neuronska mreža)

Iz kontingencijske tablice (tablica 4.18.) proizlazi podatak o 37,2 % točno predviđenih slučajeva/osobnih automobila u kategoriji “pad" u okviru trening uzorka te 38,2 % u okviru testnog uzorka, odnosno čak 62,8 % pogrešno klasificiranih slučajeva (pogrešno klasificiranih tehnički neispravnih osobnih automobila u kategoriju tehničkih ispravnih osobnih automobila koji su “prošli” tehnički pregled) u okviru trening uzorka te 61,8 % u okviru testnog uzorka.

Tablica 4.18. - Klasifikacija za osobne automobile VW Golf (neuronska mreža)

Uzorak	**Promatrane kategorije**	**Prediktivno**		
		„pad“	„prolaz“	Postotak točnosti
Trening	„pad“	**3.193**	**4.662**	**37,2 %**
	„prolaz“	**1.696**	**25.823**	93,8 %
	Ukupni postotak	12,8 %	87,2 %	81,8 %
Testni	„pad“	**2.270**	**3.679**	**38,2 %**
	„prolaz“	**1.293**	**21.084**	94,2 %
	Ukupni postotak	12,6 %	87,4	82,4 **%**
Validacijski	„pad“	585	896	39,5 **%**
	„prolaz“	314	5.118	94,2 %
	Ukupni postotak	13,0 %	87,0	82,5 %

S obzirom da su predmetni rezultati algoritma neuronske mreže u pogledu točno predviđenih slučajeva/osobnih automobila u kategoriji “pad" – 37,2 % u okviru trening uzorka, odnosno 38,2 % u okviru testnog uzorka - značajno ispod minimalne vrijednosti od 70 % koja je definirana kao zadovoljavajuća, odnosno da su predmetni rezultati značajno ispod vrijednosti postignutih uporabom Exhaustive CHAID algoritma, alternativni prediktivni model se odbacuje.

Nadalje, u kontekstu analize učinkovitosti prediktivnog modela temeljenog na neuronskoj mreži, izračunat je MCC za trening i testni uzorak, a rezultati su prikazani u tablici 4.19.

Tablica 4.19. – Prikaz vrijednosti MCC za VW Golf (neuronska mreža)

Uzorak	Kriterij	Vrijednost
Trening	pogrešno pozitivni	4.662
	pogrešno negativni	1.696
	stvarno pozitivni	25.823
	stvarno negativni	3.193
	MCC	0,40
Testni	pogrešno pozitivni	3.679
	pogrešno negativni	1.293
	stvarno pozitivni	21.084
	stvarno negativni	2.270
	MCC	0,39

Vrijednosti korelacijskog koeficijenta MCC za oba uzorka (trening i testni uzorak) su na razini srednjih pozitivnih vrijednosti +0,40 i +0,39 što daje obilježje dobre predikcije.

4.4.2. Prediktivni modeli za osobne automobile Fiat Punto

U tablici 4.20. (postavke i zbirni rezultat modela) sadržane su informacije o postavkama i rezultatima dobivenim iz modela temeljenog na Exhaustive CHAID algoritmu. Usporedni rezultati sva tri obrađena tipa osobnih automobila (kritične nezavisne varijable u kontekstu “pada” na tehničkom pregledu dani su u tablici 6.1.

Informacije o postavkama modela odnose se na varijable, vrstu validacije stabla odlučivanja, dubinu stabla i najmanji broj slučajeva po čvorovima.

Informacije o rezultatima odnose se na nezavisne varijable koje su uključene u izradu modela, broj ukupnih čvorova, broj završnih čvorova te dubina (broj razina) stabla.

Kao što je prikazano u tablici 4.20. od početno specificiranih sedam nezavisnih varijabli njih šest je u konačnici uključeno u završni model. Nezavisna varijabla "vr_kocnica" (vrsta kočionog sustava ugrađenog na osobni automobil) nije značajno pridonijela modelu te je ista od strane algoritma automatski isključena iz završnog modela.

Tablica 4.20. – Postavke i zbirni rezultati modela za Fiat Punto (Exhaustive CHAID)

Specifikacije	Metoda	Exhaustive CHAID
	Zavisna varijabla	tp_prolaz
	Nezavisne varijable	snaga, ccm, index_stp, vr_motora, vr_kocnica, index_km, starost_god
	Validacija	Metoda dijeljenja uzorka
	Najveća dubina stabla (broj razina)	3
	Najmanji broj slučajeva u "roditeljskom" čvoru	100
	Najmanji broj slučajeva u "dječjem" čvoru	100
Rezultati	Uključene nezavisne varijable	index_km, starost_god, ccm, index_stp, vr_motora, snaga
	Broj čvorova	63
	Broj završnih čvorova	42
	Dubina (broj razina)	3

Kao što je prikazano na stablu odlučivanja trening uzorka prediktivnog modela tehničke ispravnosti osobnih automobila Fiat Punto (slika 4.9.), radi se o četrdeset dva završna čvora, u okviru šezdeset tri čvora, unutar tri razine modela.

Iz predmetnog stabla odlučivanja proizlazi da je nezavisna varijabla "indeks_km" (prijeđeni put osobnih automobila) varijabla s najvećim prediktivnim potencijalom (prva razina dubine stabla) u kontekstu tehničke neispravnosti osobnih automobila Fiat Punto.

Nezavisna varijabla "indeks_km" dijeli uzorak na pet segmenata u okviru kojih su izraženi različiti prijeđeni putevi osobnih automobila, te gdje su zastupljene značajne razlike u postotku tehnički neispravnih osobnih automobila : osobni automobili koji su prešli do 50.000 km - 7,0 % tehnički neispravnih, osobni automobili koji su prešli od

50.000 do 100.000 km - 15,6 % tehnički neispravnih, osobni automobili koji su prešli od 100.000 do 150.000 km - 25,1 % tehnički neispravnih, osobni automobili koji su prešli od 150.000 do 200.000 km - 39,9 % tehnički neispravnih, osobni automobili koji su prešli više od 200.000 km - 68,1 % tehnički neispravnih.

U pogledu nezavisne varijable “indeks_km”, a u okviru prve razine dubine stabla, nema završnih čvorova, odnosno svi čvorovi (od broja 1 do broja 5) su “roditeljski” čvorovi.

U okviru druge razine dubine stabla prikazano je da nezavisna varijabla “starost_god” dijeli uzorak na četiri segmenta u okviru kojih je izražena različita „starost“ osobnih automobila, kod osobnih automobila koji su prešli do 50.000 km: osobni automobili „stari“ 6 godina i manje – 4,8 % tehnički neispravnih, osobni automobili „stari“ 7 do 9 godina, uključujući 9 godina – 8,6 % tehnički neispravnih, osobni automobili „stari“ 10 i 11 godina – 14,3 % tehnički neispravnih i osobni automobili „stariji“ od 11 godina – 27,5 % tehnički neispravnih.

Navedena nezavisna varijabla još u tri slučaja dijeli uzorak na pet (jedanput) i šest (dvaput) segmenata. Kod osobnih automobila koji su prešli između 50.000 i 100.000 km: osobni automobili „stari“ 8 godina i manje – 9,7 % tehnički neispravnih, osobni automobili „stari“ 9 godina – 12,7 % tehnički neispravnih, osobni automobili „stari“ 10 i 11 godina – 8,3 % tehnički neispravnih, osobni automobili „stari“ 12 godina – 38,5 % tehnički neispravnih, osobni automobili „stari“ 13 i 14 godina – 33,2 % tehnički neispravnih i osobni automobili „stariji“ od 14 godina – 40,1 % tehnički neispravnih. Kod osobnih automobila koji su prešli između 100.000 i 150.000 km: osobni automobili „stari“ 4 godine i manje – 52,3 % tehnički neispravnih, osobni automobili „stari“ 5 i 6 godina – 37,1 % tehnički neispravnih, osobni automobili „stari“ 7 do 9 godina, uključujući 9 godina – 19,0 % tehnički neispravnih, osobni automobili „stari“ 10 i 11 godina – 24,4 % tehnički neispravnih, osobni automobili „stari“ 12 do 14 godina, uključujući 14 godina – 28,5 % tehnički neispravnih i osobni automobili „stariji“ od 14 godina – 24,5 % tehnički neispravnih. Kod osobnih automobila koji su prešli između 150.000 i 200.000 km: osobni automobili „stari“ 7 godine i manje – 47,3 % tehnički neispravnih, osobni automobili „stari“ 8 i 9 godina – 28,8 % tehnički neispravnih, osobni automobili „stari“ 10 do 12 godina, uključujući 12 godina – 37,8 % tehnički neispravnih,

osobni automobili „stari“ 13 i 14 godina – 43,5 % tehnički neispravnih, i osobni automobili „stariji“ od 14 godina – 39,7 % tehnički neispravnih

U pogledu nezavisne varijable “starost_god”, a u okviru druge razine dubine stabla, čvorovi pod brojevima 6, 7, 9, 10, 13, 14, 15, 18, 19, 20, 21, 24, 25 i 26 predstavljaju “roditeljske” čvorove, dok su čvorovi pod brojem 8, 11, 12, 16, 17, 22 i 23 završni čvorovi.

U okviru druge razine dubine stabla prikazano je da nezavisna varijabla “ccm” dijeli uzorak na dva segmenta u pogledu radnog obujma motora, kod osobnih automobila koji su prešli više od 200.000 km: osobni automobili s motorom radnog obujma manjeg ili jednakog 1.242 cm^3 – 71,2 % tehnički neispravnih i osobni automobili s motorom radnog obujma većeg od 1.242 cm^3 – 42,2 % tehnički neispravnih.

U pogledu nezavisne varijable “ccm”, a u okviru druge razine dubine stabla, čvor pod brojem 27 predstavlja “roditeljski” čvor, dok je čvor pod brojem 28 završni čvor.

U okviru treće (završne) razine dubine stabla prikazano je da nezavisna varijabla “ccm” dijeli uzorak na dva segmenta kod osobnih automobila koji su prešli do 50.000 km, te koji su „stari“ 6 godina i manje: osobni automobili s motorom radnog obujma manjeg ili jednakog 1.242 cm^3 – 5,5 % tehnički neispravnih i osobni automobili s motorom radnog obujma većeg od 1.242 cm^3 – 4,0 % tehnički neispravnih.

Navedena nezavisna varijabla još u četiri slučaja dijeli uzorak na dva (dvaput) i tri (dvaput) segmenata. Kod osobnih automobila koji su prešli između 50.000 i 100.000 km, te koji su „stari“ 12 godina: osobni automobili s motorom radnog obujma manjeg ili jednakog 1108 cm^3 – 24,6 % tehnički neispravnih i osobni automobili s motorom radnog obujma većeg od 1.108 cm^3 – 75,6 % tehnički neispravnih. Kod osobnih automobila koji su prešli između 100.000 i 150.000 km, te koji su „stari“ više od 14 godina: osobni automobili s motorom radnog obujma manjeg ili jednakog 1.108 cm^3 – 19,1 % tehnički neispravnih, osobni automobili s motorom radnog obujma između 1.108 i 1.242 cm^3 – 30,0 % tehnički neispravnih i osobni automobili s motorom radnog obujma većeg od 1.242 cm^3 – 41,8 % tehnički neispravnih. Kod osobnih automobila koji su prešli između 150.000 i 200.000 km, te koji su „stari“ 10 do 12 godina, uključujući

12 godina: osobni automobili s motorom radnog obujma manjeg ili jednakog 1.108 cm^3 – 32,8 % tehnički neispravnih i osobni automobili s motorom radnog obujma većeg od 1108 cm^3 – 41,8 % tehnički neispravnih. Kod osobnih automobila koji su prešli između 150.000 i 200.000 km, te koji su „stariji“ od 14 godina: osobni automobili s motorom radnog obujma manjeg ili jednakog 1.108 cm^3 – 41,6 % tehnički neispravnih, osobni automobili s motorom radnog obujma između 1.108 i 1.242 cm^3 – 34,6 % tehnički neispravnih i osobni automobili s motorom radnog obujma većeg od 1.242 cm^3 – 43,1 % tehnički neispravnih.

U pogledu nezavisne varijable “ccm”, a u okviru treće (završne) razine dubine stabla, čvorovi pod brojem 29, 30, 37, 38, 51, 52, 53, 54 i 55 su završni čvorovi.

Druga nezavisna varijabla treće (završne) razine dubine stabla “index_stp” dijeli uzorak na dva segmenta kod osobnih automobila koji su prešli do 50.000 km, te koji su „stari“ 7 do 9 godina, uključujući 9 godina: osobni automobili kod kojih je tehnički pregled obavljen u stanici za tehnički pregled vozila koja se nalazi u velikoj urbanoj sredini i maloj sredini - 9,5 % tehnički neispravnih i osobni automobili kod kojih je tehnički pregled obavljen u stanici za tehnički pregled vozila koja se nalazi u urbanoj sredini - 5,6 % tehnički neispravnih.

Navedena nezavisna varijabla još u pet slučajeva dijeli uzorak na dva (četiri puta) i tri (jedanput) segmenta. Kod osobnih automobila koji su prešli do 50.000 km, te koji su „stari“ više od 11 godina: osobni automobili kod kojih je tehnički pregled obavljen u stanici za tehnički pregled vozila koja se nalazi u maloj sredini - 36,5 % tehnički neispravnih i osobni automobili kod kojih je tehnički pregled obavljen u stanici za tehnički pregled vozila koja se nalazi u velikoj urbanoj sredini i urbanoj sredini - 20,8 % tehnički neispravnih. Kod osobnih automobila koji su prešli između 50.000 i 100.000 km, te koji su „stari“ više od 14 godina: osobni automobili kod kojih je tehnički pregled obavljen u stanici za tehnički pregled vozila koja se nalazi u maloj sredini i urbanoj sredini - 22,1 % tehnički neispravnih i osobni automobili kod kojih je tehnički pregled obavljen u stanici za tehnički pregled vozila koja se nalazi u velikoj urbanoj sredini - 57,5 % tehnički neispravnih. Kod osobnih automobila koji su prešli između 100.000 i 150.000 km, te koji su „stari“ 10 i 11 godina: osobni automobili kod kojih je tehnički pregled obavljen u stanici za tehnički pregled vozila koja se nalazi u maloj sredini - 31,4 % tehnički neispravnih i osobni automobili kod kojih je tehnički pregled obavljen

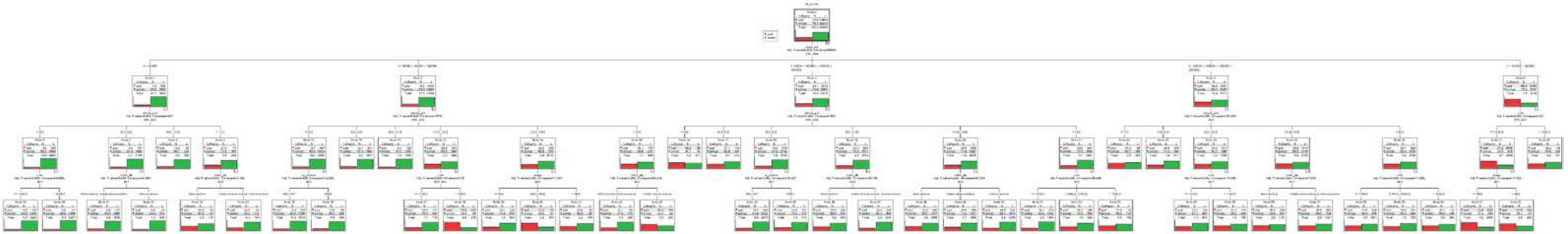
u stanici za tehnički pregled vozila koja se nalazi u velikoj urbanoj sredini i urbanoj sredini - 19,8 % tehnički neispravnih. Kod osobnih automobila koji su prešli između 100.000 i 150.000 km, te koji su „stari“ 12 do 14 godina, uključujući 14 godina: osobni automobili kod kojih je tehnički pregled obavljen u stanici za tehnički pregled vozila koja se nalazi u maloj sredini - 33,3 % tehnički neispravnih, osobni automobili kod kojih je tehnički pregled obavljen u stanici za tehnički pregled vozila koja se nalazi u velikoj urbanoj sredini - 28,8 % tehnički neispravnih i osobni automobili kod kojih je tehnički pregled obavljen u stanici za tehnički pregled vozila koja se nalazi u urbanoj sredini - 19,9 % tehnički neispravnih. Kod osobnih automobila koji su prešli između 150.000 i 200.000 km, te koji su „stari“ 13 i 14 godina: osobni automobili kod kojih je tehnički pregled obavljen u stanici za tehnički pregled vozila koja se nalazi u maloj sredini - 46,4 % tehnički neispravnih i osobni automobili kod kojih je tehnički pregled obavljen u stanici za tehnički pregled vozila koja se nalazi u velikoj urbanoj sredini i urbanoj sredini - 40,8 % tehnički neispravnih.

U pogledu nezavisne varijable “index_stp”, a u okviru treće (završne) razine dubine stabla, čvorovi pod brojem 31, 32, 33, 34, 42, 43, 46, 47, 48, 49, 50, 56 i 57 su završni čvorovi.

Treća nezavisna varijabla treće (završne) razine dubine stabla “vr_motora” dijeli uzorak na dva segmenta kod osobnih automobila koji su prešli između 50.000 i 100.000 km, te koja su „stara“ 8 godina i manje: osobni automobili s benzinskim (Ottovim) motorom i ugrađenim reguliranim katalizatorom (REG KAT) - 9,2 % tehnički neispravnih i osobni automobili s dizelskim motorom (DIZEL) - 14,8 % tehnički neispravnih.

Navedena nezavisna varijabla još u jednom slučaju dijeli uzorak na dva segmenta. Kod osobnih automobila koji su prešli između 100.000 i 150.000 km, te koji su „stari“ 7 do 9 godina, uključujući 9 godina: osobni automobili s benzinskim (Ottovim) motorom i ugrađenim reguliranim katalizatorom (REG KAT) - 18,2 % tehnički neispravnih i osobni automobili s dizelskim motorom (DIZEL) - 24,8 % tehnički neispravnih.

U pogledu nezavisne varijable “vr_motora”, a u okviru treće (završne) razine dubine stabla, čvorovi pod brojem 35, 36, 44 i 45 su završni čvorovi.



Slika 4.9. - Stablo odlučivanja trening uzorka prediktivnog modela za Fiat Punto

Četvrta nezavisna varijabla treće (završne) razine dubine stabla “snaga” dijeli uzorak na tri segmenta kod osobnih automobila koji su prešli između 50.000 i 100.000 km, te koja su „stara“ 13 i 14 godina: osobni automobili kod kojih snaga motora nije veća od 43 kW - 28,1 % tehnički neispravnih, osobni automobili kod kojih je snaga motora 44 kW - 71,2 % tehnički neispravnih i osobni automobili kod kojih je snaga motora veća od 44 kW - 34,6 % tehnički neispravnih.

Navedena nezavisna varijabla još u jednom slučaju dijeli uzorak na dva segmenta. Kod osobnih automobila koji su prešli više od 200.000 km, te koji imaju motor radnog obujma manjeg ili jednakog 1.242 cm^3: osobni automobili kod kojih snaga motora nije veća od 44 kW - 72,6 % tehnički neispravnih i osobni automobili kod kojih je snaga motora veća od 44 kW - 61,9 % tehnički neispravnih.

U pogledu nezavisne varijable “snaga”, a u okviru treće (završne) razine dubine stabla, čvorovi pod brojem 39, 40, 41, 61 i 62 su završni čvorovi.

U tablici 4.21. (dobit po završnim čvorovima) sadržani su podaci o završnim čvorovima, odnosno o čvorovima kod kojih je prestalo grananje stabla u daljnje čvorove. Završni čvorovi, sami po sebi, predstavljaju najbolji prediktivni potencijal predmetnog modela.

U predmetnoj tablici sadržani su broj i postotak slučajeva/osobnih automobila po pojedinom završnom čvoru. U okviru stupca “Dobit” također su sadržani broj i postotak slučajeva/osobnih automobila po pojedinom završnom čvoru u okviru ciljane kategorije (broj i postotak slučajeva/osobnih automobila s utvrđenim tehničkim neispravnostima tijekom tehničkog pregleda što je rezultiralo “padom”).

Tablica 4.21. - Dobit po završnim čvorovima za osobne automobile Fiat Punto

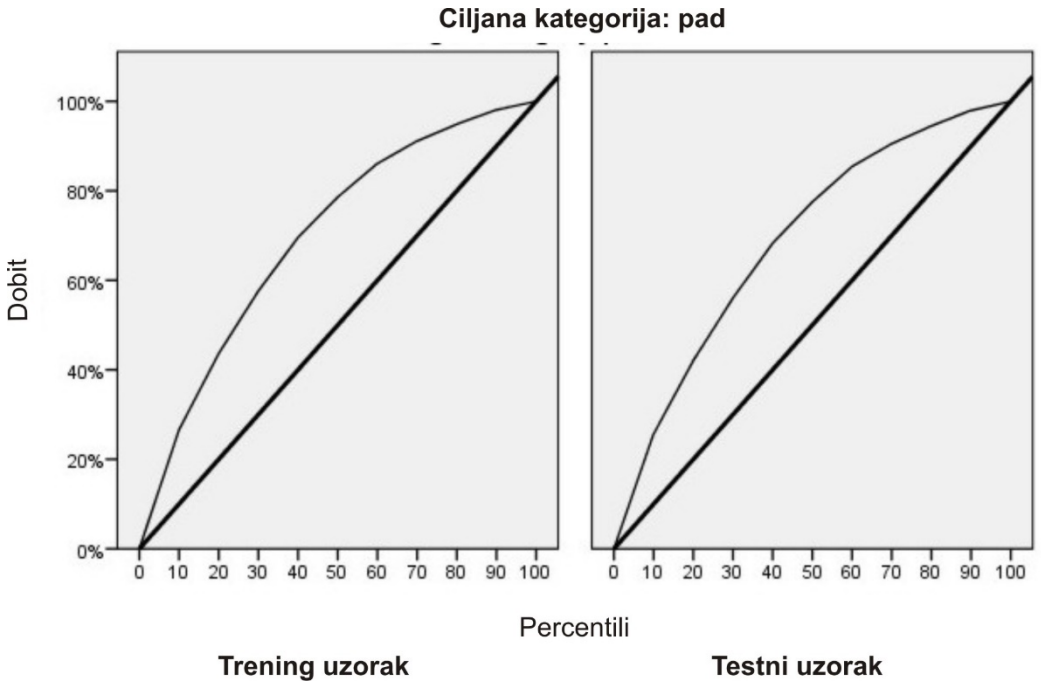
Uzorak	Završni čvor	Završni čvor		Dobit		Odziv	Indeks
		Ukupni broj Fiat Punto osobnih automobila po završnom čvoru	Postotak u odnosu na ukupni skup Fiat Punto osobnih automobila	Broj "palih" Fiat Punto osobnih automobila po završnom čvoru	Postotak u odnosu na ukupni skup "palih" Fiat Punto osobnih automobila		
	I	II	III	IV	V	VI	VII
Trening	38	270	0,6 %	204	1,9 %	75,6 %	303,4 %
	61	2.508	5,8 %	1.820	17,0 %	72,6 %	291,4 %
	40	104	0,2 %	74	0,7 %	71,2 %	285,7 %
	62	360	0,8 %	223	2,1 %	61,9 %	248,7 %
	43	233	0,5 %	134	1,3 %	57,5 %	230,9 %
	16	111	0,3 %	58	0,5 %	52,3 %	209,8 %
	22	201	0,5 %	95	0,9 %	47,3 %	189,8 %
	56	1.133	2,6 %	526	4,9 %	46,4 %	186,4 %
	60	260	0,6 %	112	1,0 %	43,1 %	173,0 %
	28	351	0,8 %	148	1,4 %	42,2 %	169,3 %
	55	756	1,8 %	316	3,0 %	41,8 %	167,8 %
	53	189	0,4 %	79	0,7 %	41,8 %	167,8 %
	58	1.501	3,5 %	625	5,8 %	41,6 %	167,2 %
	57	1.197	2,8 %	488	4,6 %	40,8 %	163,7 %
	17	313	0,7 %	116	1,1 %	37,1 %	148,8 %
	33	115	0,3 %	42	0,4 %	36,5 %	146,6 %
	41	104	0,2 %	36	0,3 %	34,6 %	139,0 %
	59	755	1,8 %	261	2,4 %	34,6 %	138,8 %
	48	2.065	4,8 %	688	6,4 %	33,3 %	133,8 %
	54	603	1,4 %	198	1,9 %	32,8 %	131,8 %
	46	834	1,9 %	262	2,5 %	31,4 %	126,1 %
	52	828	,1,9 %	248	2,3 %	30,0 %	120,3 %
	23	611	1,4 %	176	1,6 %	28,8 %	115,7 %
	49	1.596	3,7 %	459	4,3 %	28,8 %	115,5 %
	39	804	1,9 %	226	2,1 %	28,1 %	112,9 %
	45	412	1,0 %	102	1,0 %	24,8 %	99,4 %
	37	719	1,7 %	177	1,7 %	24,6 %	98,8 %
	42	226	0,5 %	50	0,5 %	22,1 %	88,8 %
	34	154	0,4 %	32	0,3 %	20,8 %	83,4 %
	50	1.215	2,8 %	242	2,3 %	19,9 %	80,0 %
	47	1.245	2,9 %	246	2,3 %	19,8 %	79,3 %
	51	1.434	3,3 %	274	2,6 %	19,1 %	76,7 %
	44	2.974	6,9 %	542	5,1 %	18,2 %	73,2 %
	36	398	0,9 %	59	0,6 %	14,8 %	59,5 %
	8	266	0,6 %	38	0,4 %	14,3 %	57,4 %
	11	1.821	4,2 %	231	2,2 %	12,7 %	50,9 %
	31	1.645	3,8 %	157	1,5 %	9,5 %	38,3 %
	35	5.610	13,1 %	525	4,9 %	9,4 %	37,6 %
	12	1.555	3,6 %	129	1,2 %	8,3 %	33,3 %
	32	540	1,3 %	30	0,3 %	5,6 %	22,3 %
	29	2.623	6,1 %	145	1,4 %	5,5 %	22,2 %
	30	2.257	5,3 %	91	0,9 %	4,0 %	16,2 %
Test	38	121	0,7 %	82	1,8 %	67,8 %	267,5 %
	61	1.041	5,6 %	752	16,1 %	72,2 %	285,1 %
	40	51	0,3 %	32	0,7 %	62,7 %	247,7 %
	62	143	0,8 %	85	1,8 %	59,4 %	234,6 %
	43	100	0,5 %	52	1,1 %	52,0 %	205,2 %
	16	49	0,3 %	29	0,6 %	59,2 %	233,6 %
	22	93	0,5 %	42	0,9 %	45,2 %	178,3 %
	56	459	2,5 %	233	5,0 %	50,0 %	200,4 %
	60	106	0,6 %	38	0,8 %	35,0 %	141,5 %
	28	151	0,8 %	62	1,3 %	41,1 %	162,1 %
	55	332	1,8 %	135	2,9 %	40,7 %	160,5 %
	53	96	0,5 %	37	0,8 %	38,5 %	152,1 %
	58	662	3,6 %	272	5,8 %	41,1 %	162,2 %
	57	515	2,8 %	209	4,5 %	40,6 %	160,2 %
	17	116	0,6 %	43	0,9 %	37,1 %	146,3 %
	33	46	0,2 %	14	0,3 %	30,4 %	120,1 %
	41	43	0,2 %	15	0,3 %	34,9 %	137,7 %
	59	380	2,1 %	144	3,1 %	37,9 %	149,6 %

	48	881	4,8 %	286	6,1 %	32,5 %	128,1 %
	54	233	1,3 %	86	1,8 %	36,9 %	145,7 %
	46	382	2,1 %	111	2,4 %	29,1 %	114,7 %
	52	392	2,1 %	138	3,0 %	35,2 %	139,0 %
	23	270	1,5 %	93	2,0 %	34,4 %	136,0 %
	49	649	3,5 %	181	3,9 %	27,9 %	110,1 %
	39	353	1,9 %	103	2,2 %	29,2 %	115,2 %
	45	192	1,0 %	48	1,0 %	25,0 %	98,7 %
	37	286	1,6 %	81	1,7 %	28,3 %	111,8 %
	42	79	0,4 %	20	0,4 %	25,3 %	99,9 %
	34	68	0,4 %	15	0,3 %	22,1 %	87,1 %
	50	517	2,8 %	99	2,1 %	19,1 %	75,6 %
	47	596	3,2 %	120	2,6 %	20,1 %	79,5 %
	51	583	3,2 %	136	2,9 %	23,3 %	92,1 %
	44	1.283	7,0 %	234	5,0 %	18,2 %	72,0 %
	36	193	1,0 %	24	0,5 %	12,4 %	49,1 %
	8	105	0,6 %	12	0,3 %	11,4 %	45,1 %
	11	749	4,1 %	107	2,3 %	14,3 %	56,4 %
	31	742	4,0 %	72	1,5 %	9,7 %	38,3 %
	35	2.421	13,1 %	245	5,2 %	10,1 %	39,9 %
	12	633	3,4 %	57	1,2 %	9,0 %	35,5 %
	32	234	1,3 %	15	0,3 %	6,4 %	25,3 %
	29	1.164	6,3 %	70	1,5 %	6,0 %	23,7 %
	30	920	5,0 %	40	0,9 %	4,3 %	17,2 %

Za kategorijsku zavisnu varijablu, stupac “Odziv” prikazuje unutarnju strukturu završnih čvorova, odnosno u istom su sadržani postotci slučajeva/osobnih automobila u pojedinom završnom čvoru u okviru predmetne ciljane kategorije (osobni automobili koji su “pali” na tehničkom pregledu”), a u odnosu na ukupni broj slučajeva/osobnih automobila po pojedinom završnom čvoru (prema: IV x 100/II).

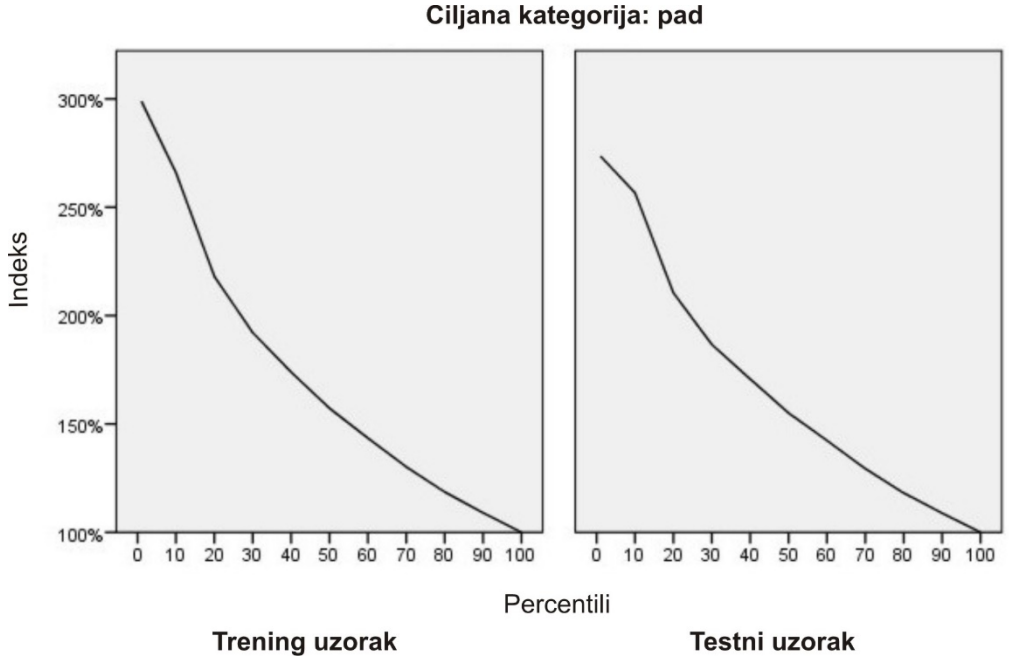
Za kategorijsku zavisnu varijablu indeks podrazumijeva omjer odziva za ciljanu kategoriju u odnosu na odziv za ukupni skup. Pomoću indeksa se dobiva informacija o tome koliko se značajno postotak promatrane ciljane kategorije za predmetni završni čvor razlikuje u odnosu na očekivani postotak ciljane kategorije (za ukupni skup) koji je izražen u korijenskom čvoru, te za osobne automobile Fiat Punto iznosi 24,90 % (na primjer, aproksimativno, u završnom čvoru broj 38 postotna vrijednost indeksa iznosi 75,6 % x 100/24,90 % = 303,4 %). Vrijednost indeksa koja prelazi 100 % ukazuje da se u predmetnom završnom čvoru nalazi veći postotni iznos ciljane kategorije u usporedbi s očekivanim postotnim iznosom ciljane kategorije za ukupni skup.

Na slici 4.10. dan je grafički prikaz dobiti za ciljanu kategoriju “pad” nezavisne varijable “tp_prolaz”, za trening i testni uzorak. Na apscisi se nalaze percentili dok se na ordinati nalaze vrijednosti dobiti izražene u postotcima.



Slika 4.10. - Dijagram dobiti za ciljanu kategoriju "pad" za Fiat Punto

Na slici 4.11. dan je grafički prikaz indeksa za ciljanu kategoriju "pad" nezavisne varijable "tp_prolaz", za trening i testni uzorak. Na apscisi se nalaze percentili dok se na ordinati nalaze vrijednosti indeksa izražene u postotcima.



Slika 4.11. – Dijagram indeksa za ciljanu kategoriju "pad" za Fiat Punto

Temeljem podataka iz tablice rizika (tablica 4.22.) i klasifikacijske tablice (tablica 4.23.) donosi se ocjena o prediktivnom potencijalu predmetnog modela.

Tablica 4.22. - Rizik za Fiat Punto

Uzorak	Procijenjena vrijednost	Standardna devijacija
Trening	0,215	0,002
Testni	0,223	0,003

Iz podataka sadržanih u tablici rizika prikazan je rizik predikcije koji predstavlja postotak pogrešne klasifikacije kategorija. U pogledu provedenog istraživanja proizlazi da, u slučaju kada su poznate karakteristike osobnih automobila Fiat Punto u smislu predmetnih šest nezavisnih (prediktivnih) varijabli, rizik pogrešne klasifikacije osobnih automobila Fiat Punto glede utvrđivanja tehničke neispravnosti tijekom tehničkog pregleda (odnosno "pada" na tehničkom pregledu) jest 21,5 % (primjenom metode dijeljenja uzorka) za trening uzorak, odnosno 22,3 % za testni uzorak.

Podaci iz kontingencijske tablice potvrđuju podatke iz tablice rizika u smislu da predmetni model, u ukupno 78,5 % slučajeva (osobnih automobila VW Golf) trening uzorka točno klasificira kategoriju, odnosno u ukupno 77,7 % slučajeva testnog uzorka.

Tablica 4.23. - Klasifikacija za Fiat Punto (Exhaustive CHAID)

Uzorak	Promatrane kategorije	Prediktivno		
		„pad“	„prolaz“	Postotak točnosti
Trening	„pad“	**2.513**	**8.171**	**23,5 %**
	„prolaz“	**1.073**	**31.139**	96,7 %
	Ukupni postotak	8,4 %	91,6 %	**78,5 %**
Testni	„pad“	**1.032**	**3.637**	**22,1 %**
	„prolaz“	**473**	**13.287**	96,6 %
	Ukupni postotak	8,2 %	91,8 %	**77,7 %**

Nadalje, u kontekstu analize učinkovitosti prediktivnog modela, prema izrazu 4.2. izračunat je MCC za trening i testni uzorak, a rezultati su prikazani u tablici 4.24.

Tablica 4.24. – Prikaz vrijednosti MCC za Fiat Punto (Exhaustive CHAID)

Uzorak	Kriterij	Vrijednost
Trening	pogrešno pozitivni	8.171
	pogrešno negativni	1.073
	stvarno pozitivni	31.139
	stvarno negativni	2.513
	MCC	**0,31**
Testni	pogrešno pozitivni	3.637
	pogrešno negativni	473
	stvarno pozitivni	13.287
	stvarno negativni	1.032
	MCC	**0,29**

Kao što je prikazano u tablici 4.25., nakon primjene Exhaustive CHAID metode na bazi podataka osobnih automobila Fiat Punto izraženoj u vidu sedam nezavisnih varijabli te jedne zavisne varijable, generirane su četiri nove varijable u predmetnoj bazi podataka, po istom principu kao kod osobnog automobila VW Golf (točka 4.4.1. ovog rada).

Tablica 4.25. – Izvod dijela prediktivnih vrijednosti za Fiat Punto

Redni broj	„snaga“	„ccm“	„starost_god“	„index_stp“	„vr_motora“	„vr_kocnica“	„index_km“	„tp_prolaz“	Broj čvora	Prediktivna vrijednost	Prediktivna vjerojatnost 1	Prediktivna vjerojatnost 2
1.	40	1108	17	3	2	1	5	0	19	0	0,72	0,28
2.	40	1108	17	1	2	1	4	0	17	1	0,35	0,65
3.	40	1108	17	2	2	1	4	0	18	1	0,37	0,63
4.	40	1108	17	2	2	1	4	0	18	1	0,37	0,63
5.	40	1108	17	3	2	1	4	0	30	1	0,44	0,56
6.	40	1108	17	2	2	1	5	0	19	0	0,72	0,28
7.	43	1242	17	1	2	1	4	0	17	1	0,35	0,65
8.	43	1242	17	2	2	1	4	0	18	1	0,37	0,63
9.	43	1242	17	3	2	1	4	0	30	1	0,44	0,56
...	...	...	...	...	...	...	...	...	...	...	...	...

Varijabla “Prediktivna vrijednost” predstavlja opseg slučajeva/osobnih automobila u pojedinoj kategoriji zavisne varijable za završni čvor u okviru kojeg se nalazi pojedini slučaj/osobni automobil. U primjeru iz tablice 4.25., u čvoru 17, 35 % slučajeva/osobnih automobila potpada pod kategoriju “pad” (“Prediktivna vrijednost” - 0), a 65 % slučajeva/osobnih automobila potpada pod kategoriju “prolaz” (“Prediktivna vrijednost ” - 1), što predstavlja prediktivne vjerojatnosti od 0,35 i 0,65.

4.4.2.1. Analiza prediktivnog potencijala modela temeljenog na Exhaustive CHAID algoritmu

Procjena prediktivnog potencijala modela proizlazi iz analize podataka sadržanih u tablici rizika (tablica 4.22.) i kontingencijskoj tablici (tablica 4.23.).

Iz kontingencijske tablice (tablica 4.23.) proizlazi podatak o samo 23,5 % točno predviđenih slučajeva/osobnih automobila u kategoriji "pad" u okviru trening uzorka te samo 22,1 % u okviru testnog uzorka, odnosno čak 76,5 % pogrešno klasificiranih slučajeva (pogrešno klasificiranih tehnički neispravnih osobnih automobila u kategoriju tehničkih ispravnih osobnih automobila koji su "prošli" tehnički pregled) u okviru trening uzorka te 77,9 % u okviru testnog uzorka. Ovakav oblik pogrešne klasifikacije ima mnogo veći negativni značaj u kontekstu sigurnosti cestovnog prometa, od obrnutog slučaja – kod pogrešno klasificiranih tehnički ispravnih osobnih automobila u kategoriju tehničkih neispravnih osobnih automobila koji su "pali" na tehničkom pregledu.

U skladu s objašnjenjem iznesenim u točci 4.4.1.1. ovog rada postotni iznos od 70,0 % točno predviđenih slučajeva/osobnih automobila u kategoriji "pad" (uz definiranu toleranciju pogreške) smatra se minimalno zadovoljavajućim u kontekstu prediktivnog potencijala modela.

Temeljem prethodno navedenog proizlazi da iznos od 23,5 % (22,1 %) točno predviđenih slučajeva/osobnih automobila u kategoriji "pad", odnosno čak 76,5 % (77,9 %) pogrešno klasificiranih slučajeva nije zadovoljavajući s obzirom da se radi o ciljanoj kategoriji zavisne varijable, odnosno model nema zadovoljavajući prediktivni potencijal, te je potrebno izvršiti radnju radi poboljšanja (korektivne radnje) koja će dovesti do povećanja prediktivnog potencijala modela.

Vrijednosti korelacijskog koeficijenta MCC za oba uzorka (trening i testni uzorak) su niže srednje pozitivni +0,31 i +0,29, što podrazumijeva zadovoljavajuću predikciju već sada (teoretska vrijednost +1 podrazumijeva savršenu predikciju, vrijednost 0 podrazumijeva predikciju koja nije bolja od nasumične predikcije, teoretska vrijednost -1 podrazumijeva potpuno razmimoilaženje između predikcije i opažanja).

4.4.2.2. Poboljšanje prediktivnog potencijala modela temeljenog na Exhaustive CHAID algoritmu

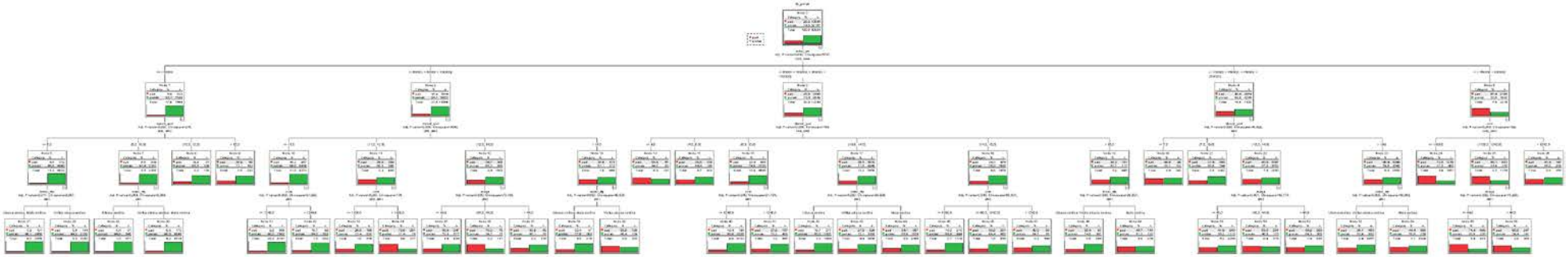
Metodologija provođenja radnje radi poboljšanja prediktivnog potencijala modela za osobne automobile Fiat Punto identična je primijenjenoj metodologiji u slučaju prediktivnog modela za osobne automobile VW Golf, kako je izneseno u točci 4.4.1.2.

Međutim, s obzirom da penalizacija negativnih posljedica, kod pogrešnog klasificiranja osobnih automobila Fiat Punto koji su tehnički neispravni u klasu tehnički ispravnih osobnih automobila, s vrijednošću 2 (kao u slučaju penalizacije kod osobnih automobila VW Golf) još uvijek nije pokazala zadovoljavajuće rezultate, ista je provedena s vrijednošću 3.

Prethodno navedeno vrednovanje penalizacije, u kontekstu osobnih automobila Fiat Punto, izvršeno je u omjeru 3:1 (trostruko veća vrijednost negativnih posljedica kod pogrešnog klasificiranja osobnog automobila koji je tehnički neispravan u klasu tehnički ispravnih osobnih automobila, nego u obrnutom slučaju) nakon provedenog eksperimentiranja, iz razloga što je upravo pri predmetnom omjeru postignut minimalni određeni postotni iznos od 70 % točno predviđenih slučajeva/osobnih automobila u kategoriji “pad" čime se smatra da je postignut zadovoljavajući prediktivni potencijal modela (u skladu s točkom 4.4.1.1. ovog rada).

4.4.2.3. Analiza nakon provedenog poboljšanja modela temeljenog na Exhaustive CHAID algoritmu

Provođenjem predmetne radnje vidljive su promjene u stablu odlučivanja prediktivnog modela tehničke ispravnosti osobnih automobila Fiat Punto. Predmetne promjene ne odnose se na distribuciju slučajeva/osobnih automobila po pojedinim čvorovima, koja je ostala ista, već se odnose na dominantne prediktivne kategorije zavisne varijable, kao što je prikazano na slici 4.12.



Slika 4.12. - Stablo odlučivanja trening uzorka prediktivnog modela za Fiat Punto nakon provođenja radnje za poboljšanje modela

S obzirom na trostruko veće vrednovanje penalizacije pogrešnog klasificiranja osobnog automobila koji je tehnički neispravan u klasu tehnički ispravnih osobnih automobila u odnosu na penalizaciju pogrešnog klasificiranja osobnog automobila koji je tehnički ispravan u klasu tehnički neispravnih osobnih automobila, u svim čvorovima koji su prije provođenja korektivne radnje imali podjednaku distribuciju obje prediktivne kategorije, ili nešto veću distribuciju kategorije "prolaz", nakon provođenja korektivne radnje u istima će dominirati kategorija "pad".

Promjene dominantne prediktivne kategorije, iz kategorije "prolaz" u kategoriju "pad" zabilježene su u čvorovima broj 3, 4, 11, 12, 13, 16, 17, 19, 22, i 23 te u završnim čvorovima broj 9, 15, 20, 21, 26, 33, 35, 37, 41, 43, 44, 46, 47, 48, 49, 50, 52, 53 i 54.

Kao što je vidljivo iz tablice 4.26. nakon provođenja radnje za poboljšanje prediktivnog modela došlo je do promjene rezultata u odnosnu na rezultate prije provođenja radnje za poboljšanje prediktivnog modela (tablica 4.20.) i to u broju čvorova, broju završnih čvorova te u pogledu nezavisnih varijabli koje su pridonijele modelu (u model nisu uključene nezavisne varijable „vr_kocnica" i „vr_motora").

Tablica 4.26. – Zbirni rezultati modela za Fiat Punto nakon provođenja radnje za poboljšanje (Exhaustive CHAID)

Specifikacije	Metoda	Exhaustive CHAID
	Zavisna varijabla	tp_prolaz
	Nezavisne varijable	snaga, ccm, index_stp, vr_motora, vr_kocnica, index_km, starost_god
	Validacija	Metoda dijeljenja uzorka
	Najveća dubina stabla (broj razina)	3
	Najmanji broj slučajeva u "roditeljskom" čvoru	100
	Najmanji broj slučajeva u "dječjem" čvoru	100
Rezultati	Uključene nezavisne varijable	**index_km, starost_god, index_stp, ccm, snaga**
	Broj čvorova	**57**
	Broj završnih čvorova	**38**
	Dubina (broj razina)	3

Provođenjem korektivne radnje došlo je do promjena u tablici rizika (tablica 4.27.), te klasifikacijskoj tablici (tablica 4.28.).

Tablica 4.27. - Rizik za Fiat Punto nakon provođenja radnje za poboljšanje

Uzorak	Procijenjena vrijednost	Standardna devijacija
Trening	0,455	0,004
Testni	0,454	0,006

"Nusprodukt" provođenja radnje za poboljšanje modela jest očekivano povećanje rizika predikcije, u okviru trening uzorka, s 21,5 % prije provođenja korektivne radnje, na 45,5 % poslije provođenja korektivne radnje, odnosno u okviru testnog uzorka s 22,3% na 45,5 %

Međutim, provođenje radnje za poboljšanje modela dovelo je do željenog efekta, odnosno značajnog povećanja prediktivnog potencijala modela u kontekstu ciljane prediktivne kategorije zavisne varijable "pad" koji proizlazi iz većeg postotka točno klasificiranih slučajeva/osobnih automobila glede utvrđivanja tehničke neispravnosti tijekom tehničkog pregleda (odnosno "pada" na tehničkom pregledu), i to u okviru trening uzorka s 23,5 %, prije provođenja radnje za poboljšanje modela, na 74,6 % nakon provođenja radnje za poboljšanje modela, odnosno u okviru testnog uzorka s 22,1 % na 73,7 %, što je predstavlja rezultat koji zadovoljava, sukladno točci 4.4.1.1.

Tablica 4.28. - Klasifikacija za Fiat Punto nakon provođenja radnje za poboljšanje (Exhaustive CHAID)

Uzorak	Promatrane kategorije	Prediktivno		
		„pad“	„prolaz“	Postotak točnosti
Trening	„pad“	**8.087**	**2.757**	**74,6 %**
	„prolaz“	**11.282**	**20.875**	64,9 %
	Ukupni postotak	45,0 %	55,0 %	67,4 %
Testni	„pad“	**3.322**	**1.187**	**73,7 %**
	„prolaz“	**4.757**	**9.058**	65,6 %
	Ukupni postotak	44,1 %	55,9 %	67,6 %

Povećanje postotka točno klasificiranih slučajeva/osobnih automobila kod kojih je utvrđena tehnička neispravnost tijekom tehničkog pregleda (što za posljedicu ima "pad" osobnih automobila na tehničkom pregledu) za direktnu posljedicu ima smanjeni postotak točno klasificiranih slučajeva/osobnih automobila kod kojih nije utvrđena tehnička neispravnost tijekom tehničkog pregleda (kategorija "prolaz"), s 96,7 % u okviru trening uzorka na 64,9 %, te u okviru testnog uzorka s 96,6 % na 65,6 %

(odnosno pad ukupnog postotka točno klasificiranih slučajeva/osobnih automobila u okviru trening uzorka s 78,5 % na 67,4 %, te u okviru testnog uzorka s 77,7 % na 67,6 %), međutim sukladno prethodno iznesenoj argumentaciji radi se o prediktivnoj kategoriji sekundarnog značaja, odnosno radi se o prediktivnoj kategoriji koja ima manji značaj za ovo istraživanje s obzirom da ista nema negativan utjecaj na sigurnost prometa na cestama (osobni automobili kod kojih nije utvrđena tehnička neispravnost pogrešno su klasificirani kao osobni automobili koji su "pali" na tehničkom pregledu zbog utvrđene tehničke neispravnosti).

Nadalje, u kontekstu analize učinkovitosti prediktivnog modela, prema izrazu 4.2. nakon izvršene radnje za poboljšanje modela izračunat je MCC za trening i testni uzorak, a rezultati su prikazani u tablici 4.29.

Tablica 4.29. – Prikaz vrijednosti MCC za Fiat Punto nakon provođenja radnje za poboljšanje (Exhaustive CHAID)

Uzorak	Kriterij	Vrijednost
Trening	pogrešno pozitivni	2.757
	pogrešno negativni	11.282
	stvarno pozitivni	20.875
	stvarno negativni	8.087
	MCC	**0,36**
Testni	pogrešno pozitivni	1.187
	pogrešno negativni	4.757
	stvarno pozitivni	9.058
	stvarno negativni	3.322
	MCC	**0,36**

Vrijednosti korelacijskog koeficijenta MCC za oba uzorka (trening i testni uzorak), nakon izvršene radnje za poboljšanje modela, su povišene na razinu srednjih pozitivnih vrijednosti +0,36, što potvrđuje dobru predikciju (odnosno bolju predikciju nego prije izvršene radnje za poboljšanje modela).

U konačnici, prediktivni model tehničke ispravnosti za osobne automobile Fiat Punto a koji se temelji na metodi stabla odlučivanja i algoritmu Exhaustive CHAID, može se izraziti i u obliku poslovnih pravila, kako slijedi u tablici 4.30.:

Tablica 4.30. – Prikaz poslovnih pravila za poboljšani Exhaustive CHAID prediktivni model Fiat Punto

/* Čvor 27 */.
IF (index_km NOT MISSING AND (index_km <= "< 50000")) AND (starost_god IS MISSING OR (starost_god <= 6)) AND (index_stp != "Velika urbana sredina") THEN **Prediction = 1**, **Probability = 0.960531**

/* Čvor 28 */.
IF (index_km NOT MISSING AND (index_km <= "< 50000")) AND (starost_god IS MISSING OR (starost_god <= 6)) AND (index_stp = "Velika urbana sredina") THEN **Prediction = 1** **Probability = 0.942112**

/* Čvor 29 */.
IF (index_km NOT MISSING AND (index_km <= "< 50000")) AND (starost_god NOT MISSING AND (starost_god > 6 AND starost_god <= 10)) AND (index_stp = "Urbana sredina") THEN **Prediction = 1**, **Probability = 0.945709**

/* Čvor 30 */.
IF (index_km NOT MISSING AND (index_km <= "< 50000")) AND (starost_god NOT MISSING AND (starost_god > 6 AND starost_god <= 10)) AND (index_stp != "Urbana sredina") THEN **Prediction = 1**, **Probability = 0.904631**
/* Čvor 8 */.
IF (index_km NOT MISSING AND (index_km <= "< 50000")) AND (starost_god NOT MISSING AND (starost_god > 10 AND starost_god <= 12)) THEN **Prediction = 1**, **Probability = 0.801471**

/* Čvor 9 */.
IF (index_km NOT MISSING AND (index_km <= "< 50000")) AND (starost_god NOT MISSING AND (starost_god > 12)) THEN **Prediction = 0**, **Probability = 0.269058**

/* Čvor 31 */.
IF (index_km NOT MISSING AND (index_km > "< 50000" AND index_km <= "> 50000 < 100000")) AND (starost_god IS MISSING OR (starost_god <= 11)) AND (ccm IS MISSING OR (ccm <= 1248)) THEN **Prediction = 1**, **Probability = 0.901842**

/* Čvor 32 */.
IF (index_km NOT MISSING AND (index_km > "< 50000" AND index_km <= "> 50000 < 100000")) AND (starost_god IS MISSING OR (starost_god <= 11)) AND (ccm NOT MISSING AND (ccm > 1248)) THEN **Prediction = 1**, **Probability = 0.843354**

/* Čvor 33 */.
IF (index_km NOT MISSING AND (index_km > "< 50000" AND index_km <= "> 50000 < 100000")) AND (starost_god NOT MISSING AND (starost_god > 11 AND starost_god <= 12)) AND (ccm IS MISSING OR (ccm <= 1108)) **THEN Prediction = 0, Probability = 0.265537**

/* Čvor 34 */.
IF (index_km NOT MISSING AND (index_km > "< 50000" AND index_km <= "> 50000 < 100000")) AND (starost_god NOT MISSING AND (starost_god > 11 AND starost_god <= 12)) AND (ccm NOT MISSING AND (ccm > 1108)) THEN **Prediction = 0**, **Probability = 0.725632**

/* Čvor 35 */.
IF (index_km NOT MISSING AND (index_km > "< 50000" AND index_km <= "> 50000 < 100000")) AND (starost_god NOT MISSING AND (starost_god > 12 AND starost_god <= 14)) AND (snaga IS MISSING OR (snaga <= 43)) THEN **Prediction = 0**, **Probability = 0.299757**

/* Čvor 36 */.
IF (index_km NOT MISSING AND (index_km > "< 50000" AND index_km <= "> 50000 < 100000")) AND (starost_god NOT MISSING AND (starost_god > 12 AND starost_god <= 14)) AND (snaga NOT MISSING AND (snaga > 43 AND snaga <= 44)) THEN **Prediction = 0**, **Probability = 0.702703**

/* Čvor 37 */.
IF (index_km NOT MISSING AND (index_km > "< 50000" AND index_km <= "> 50000 < 100000")) AND (starost_god NOT MISSING AND (starost_god > 12 AND starost_god <= 14)) AND (snaga NOT MISSING AND (snaga > 44)) THEN **Prediction = 0**, **Probability = 0.338983**

/* Čvor 38 */.
IF (index_km NOT MISSING AND (index_km > "< 50000" AND index_km <= "> 50000 < 100000")) AND (starost_god NOT MISSING AND (starost_god > 14)) AND (index_stp = "Urbana sredina" OR index_stp = "Mala sredina") THEN **Prediction = 1**, **Probability = 0.776190**

/* Čvor 39 */.
IF (index_km NOT MISSING AND (index_km > "< 50000" AND index_km <= "> 50000 < 100000")) AND (starost_god NOT MISSING AND (starost_god > 14)) AND (index_stp != "Urbana sredina" AND index_stp != "Mala sredina") THEN **Prediction = 0**, **Probability = 0.536170**

/* Čvor 14 */.
IF (index_km IS MISSING OR (index_km > "> 50000 < 100000" AND index_km <= "> 100000 < 150000")) AND (starost_god NOT MISSING AND (starost_god <= 4)) THEN **Prediction = 0**, **Probability = 0.558333**

/* Čvor 15 */.
IF (index_km IS MISSING OR (index_km > "> 50000 < 100000" AND index_km <= "> 100000 < 150000")) AND (starost_god NOT MISSING AND (starost_god > 4 AND starost_god <= 6)) THEN **Prediction = 0**, **Probability = 0.356436**

/* Čvor 40 */.
IF (index_km IS MISSING OR (index_km > "> 50000 < 100000" AND index_km <= "> 100000 < 150000")) AND (starost_god NOT MISSING AND (starost_god > 6 AND starost_god <= 10)) AND (ccm IS MISSING OR (ccm <= 1248)) THEN **Prediction = 1**, **Probability = 0.806452**

/* Čvor 41 */.
IF (index_km IS MISSING OR (index_km > "> 50000 < 100000" AND index_km <= "> 100000 < 150000")) AND (starost_god NOT MISSING AND (starost_god > 6 AND starost_god <= 10)) AND (ccm NOT MISSING AND (ccm > 1248)) THEN **Prediction = 0**, **Probability = 0.277876**

/* Čvor 42 */.
IF (index_km IS MISSING OR (index_km > "> 50000 < 100000" AND index_km <= "> 100000 < 150000")) AND (starost_god IS MISSING OR (starost_god > 10 AND starost_god <= 14)) AND (index_stp = "Urbana sredina") THEN **Prediction = 1**, **Probability = 0.802707**

/* Čvor 43 */.
IF (index_km IS MISSING OR (index_km > "> 50000 < 100000" AND index_km <= "> 100000 < 150000")) AND (starost_god IS MISSING OR (starost_god > 10 AND starost_god <= 14)) AND (index_stp = "Velika urbana sredina") THEN **Prediction = 0**, **Probability = 0.278897**

/* Čvor 44 */.
IF (index_km IS MISSING OR (index_km > "> 50000 < 100000" AND index_km <= "> 100000 < 150000")) AND (starost_god IS MISSING OR (starost_god > 10 AND starost_god <= 14)) AND (index_stp != "Urbana sredina" AND index_stp != "Velika urbana sredina") THEN **Prediction = 0**, **Probability = 0.341082**

/* Čvor 45 */.
IF (index_km IS MISSING OR (index_km > "> 50000 < 100000" AND index_km <= "> 100000 < 150000")) AND (starost_god NOT MISSING AND (starost_god > 14 AND starost_god <=

15)) AND (ccm IS MISSING OR (ccm <= 1108)) THEN **Prediction = 1**, **Probability = 0.817642**

/* Čvor 46 */.
IF (index_km IS MISSING OR (index_km > "> 50000 < 100000" AND index_km <= "> 100000 < 150000")) AND (starost_god NOT MISSING AND (starost_god > 14 AND starost_god <= 15)) AND (ccm NOT MISSING AND (ccm > 1108 AND ccm <= 1242)) THEN **Prediction = 0**, **Probability = 0.306402**

/* Čvor 47 */.
IF (index_km IS MISSING OR (index_km > "> 50000 < 100000" AND index_km <= "> 100000 < 150000")) AND (starost_god NOT MISSING AND (starost_god > 14 AND starost_god <= 15)) AND (ccm NOT MISSING AND (ccm > 1242)) THEN **Prediction = 0**, **Probability = 0.402778**

/* Čvor 48 */.
IF (index_km IS MISSING OR (index_km > "> 50000 < 100000" AND index_km <= "> 100000 < 150000")) AND (starost_god NOT MISSING AND (starost_god > 15)) AND (index_stp != "Mala sredina") THEN **Prediction = 0**, **Probability = 0.253788**
/* Čvor 49 */.
IF (index_km IS MISSING OR (index_km > "> 50000 < 100000" AND index_km <= "> 100000 < 150000")) AND (starost_god NOT MISSING AND (starost_god > 15)) AND (index_stp = "Mala sredina") THEN **Prediction = 0**, **Probability = 0.487179**

/* Čvor 20 */.
IF (index_km NOT MISSING AND (index_km > "> 100000 < 150000" AND index_km <= "> 150000 < 200000")) AND (starost_god NOT MISSING AND (starost_god <= 7)) THEN **Prediction = 0**, **Probability = 0.468750**

/* Čvor 21 */.
IF (index_km NOT MISSING AND (index_km > "> 100000 < 150000" AND index_km <= "> 150000 < 200000")) AND (starost_god NOT MISSING AND (starost_god > 7 AND starost_god <= 10)) THEN **Prediction = 0**, **Probability = 0.314103**

/* Čvor 50 */.
IF (index_km NOT MISSING AND (index_km > "> 100000 < 150000" AND index_km <= "> 150000 < 200000")) AND (starost_god IS MISSING OR (starost_god > 10 AND starost_god <= 14)) AND (snaga IS MISSING OR (snaga <= 40)) THEN**Prediction = 0**, **Probability = 0.417996**

/* Čvor 51 */.
IF (index_km NOT MISSING AND (index_km > "> 100000 < 150000" AND index_km <= "> 150000 < 200000")) AND (starost_god IS MISSING OR (starost_god > 10 AND starost_god <= 14)) AND (snaga NOT MISSING AND (snaga > 40 AND snaga <= 44)) THEN **Prediction = 0**, **Probability = 0.534574**

/* Čvor 52 */.
IF (index_km NOT MISSING AND (index_km > "> 100000 < 150000" AND index_km <= "> 150000 < 200000")) AND (starost_god IS MISSING OR (starost_god > 10 AND starost_god <= 14)) AND (snaga NOT MISSING AND (snaga > 44)) THEN **Prediction = 0**, **Probability = 0.409385**

/* Čvor 53 */.
IF (index_km NOT MISSING AND (index_km > "> 100000 < 150000" AND index_km <= "> 150000 < 200000")) AND (starost_god NOT MISSING AND (starost_god > 14)) AND (index_stp = "Urbana sredina" OR index_stp = "Velika urbana sredina") THEN **Prediction = 0**, **Probability = 0.361177**

/* Čvor 54 */.
IF (index_km NOT MISSING AND (index_km > "> 100000 < 150000" AND index_km <= "> 150000 < 200000")) AND (starost_god NOT MISSING AND (starost_god > 14)) AND

(index_stp != "Urbana sredina" AND index_stp != "Velika urbana sredina") THEN **Prediction = 0**, **Probability = 0.443943**

/* Čvor 24 */.
IF (index_km NOT MISSING AND (index_km > "> 150000 < 200000")) AND (ccm IS MISSING OR (ccm <= 1108)) THEN **Prediction = 0**, **Probability = 0.724569**

/* Čvor 55 */.
IF (index_km NOT MISSING AND (index_km > "> 150000 < 200000")) AND (ccm NOT MISSING AND (ccm > 1108 AND ccm <= 1242)) AND (snaga IS MISSING OR (snaga <= 44)) THEN **Prediction = 0**, **Probability = 0.713764**

/* Čvor 56 */.
IF (index_km NOT MISSING AND (index_km > "> 150000 < 200000")) AND (ccm NOT MISSING AND (ccm > 1108 AND ccm <= 1242)) AND (snaga NOT MISSING AND (snaga > 44)) THEN **Prediction = 0**, **Probability = 0.606145**

/* Čvor 26 */.
IF (index_km NOT MISSING AND (index_km > "> 150000 < 200000")) AND (ccm NOT MISSING AND (ccm > 1242)) THEN **Prediction = 0**, **Probability = 0.414085**

4.4.2.4. Alternativni prediktivni model temeljen na algoritmu perceptronskih neuronskih mreža

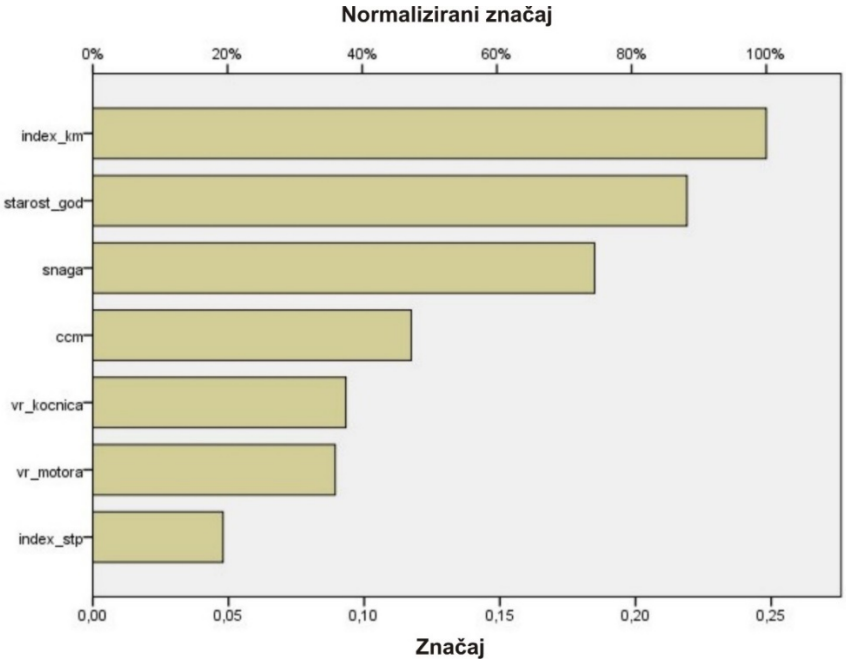
Kao alternativa preferencijalnom prediktivnom modelu koji se temelji na Exhaustive CHAID algoritmu korišten je algoritam perceptronske neuronske mreže. U tablici 4.31. dane su parametri strukture neuronske mreže za osobne automobile Fiat Punto.

Tablica 4.31. – Parametri strukture neuronske mreže za Fiat Punto

Ulazni sloj	Varijable	1	„snaga“
		2	„ccm“
		3	„index_stp“
		4	„vr_motora“
		5	„vr_kocnica“
		6	„index_km“
		7	„starost_god“
	Broj neurona		63
	Broj skrivenih slojeva		1
Skriveni sloj(evi)	Broj neurona u skrivenim slojevima 1		7
	Aktivacijska funkcija		Hiperbolička tangenta
Izlazni sloj	Zavisne varijable 1		„tp_prolaz“
	Broj neurona		2
	Aktivacijska funkcija		softmax
	Funkcija pogreške		križna entropija

Kao što je vidljivo iz slike 4.13. procjena značaja u smislu prediktivnog potencijala pojedinih nezavisnih varijabli za osobne automobile Fiat Punto proizašlih iz algoritma neuronske mreže identična je procjeni dvije nezavisne varijable s najvećim značajem proizašloj iz Exhaustive CHAID algoritma.

Proizlazi da je najznačajnija nezavisna varijabla „index_km“ koju prema razini značajnosti slijede nezavisne varijable „starost_god“, „snaga“ te „ccm“, dok nezavisne varijable „vr_kocnica“, „vr_motora“ i „index_stp“ imaju najmanji značaj.



Slika 4.13. – Prikaz nezavisnih varijabli prema procjeni razine značaja za Fiat Punto (neuronska mreža)

Iz kontingencijske tablice (tablica 4.32.) proizlazi podatak o 42,2 % točno predviđenih slučajeva/osobnih automobila u kategoriji “pad" u okviru trening uzorka te 41,6 % u okviru testnog uzorka, odnosno čak 57,8 % pogrešno klasificiranih slučajeva (pogrešno klasificiranih tehnički neispravnih osobnih automobila u kategoriju tehničkih ispravnih osobnih automobila koji su “prošli” tehnički pregled) u okviru trening uzorka te 58,4 % u okviru testnog uzorka.

Tablica 4.32. - Klasifikacija za Fiat Punto (neuronska mreža)

Uzorak	**Promatrane kategorije**	**Prediktivno**		
		„pad“	„prolaz“	Postotak točnosti
Trening	„pad“	**4.055**	**5.565**	**42,2 %**
	„prolaz“	**4.125**	**25.158**	85,9 %
	Ukupni postotak	21,0 %	79,0 %	75,1 %
Testni	„pad“	**1.781**	**2.499**	**41,6 %**
	„prolaz“	**1.779**	**10.859**	85,9 %
	Ukupni postotak	21,0 %	79,0 %	**74,7 %**
Validacijski	„pad“	607	845	41,8 %
	„prolaz“	590	3.459	85,4 %
	Ukupni postotak	21,8 %	78,2 %	73,9 %

S obzirom da su predmetni rezultati algoritma neuronske mreže u pogledu točno predviđenih slučajeva/osobnih automobila u kategoriji “pad" – 42,2 % u okviru trening uzorka, odnosno 41,6 % u okviru testnog uzorka - značajno ispod minimalne vrijednosti od 70 % koja je definirana kao zadovoljavajuća, odnosno da su predmetni rezultati značajno ispod vrijednosti postignutih uporabom Exhaustive CHAID algoritma, alternativni prediktivni model se odbacuje.

Nadalje, u kontekstu analize učinkovitosti prediktivnog modela temeljenog na neuronskoj mreži, izračunat je MCC za trening i testni uzorak, a rezultati su prikazani u tablici 4.33.

Tablica 4.33. – Prikaz vrijednosti MCC za Fiat Punto (neuronska mreža)

Uzorak	Kriterij	Vrijednost
Trening	pogrešno pozitivni	5.565
	pogrešno negativni	4.125
	stvarno pozitivni	25.158
	stvarno negativni	4.055
	MCC	0,30
Testni	pogrešno pozitivni	2.499
	pogrešno negativni	1.779
	stvarno pozitivni	10.859
	stvarno negativni	1.781
	MCC	0,29

Vrijednosti korelacijskog koeficijenta MCC za oba uzorka (trening i testni uzorak) su na razini nižih srednjih pozitivnih vrijednosti +0,30 i +0,29 što daje obilježje zadovoljavajuće predikcije.

4.4.3. Prediktivni modeli za osobne automobile Opel Astra

U tablici 4.34. (postavke i zbirni rezultat modela) sadržane su informacije o postavkama i rezultatima dobivenim iz modela temeljenog na Exhaustive CHAID algoritmu. Usporedni rezultati sva tri obrađena tipa osobnih automobila (kritične nezavisne varijable u kontekstu “pada” na tehničkom pregledu dani su u tablici 6.1.

Informacije o postavkama modela odnose se na varijable, vrstu validacije stabla odlučivanja, dubinu stabla i najmanji broj slučajeva po čvorovima.

Informacije o rezultatima odnose se na nezavisne varijable koje su uključene u izradu modela, broj ukupnih čvorova, broj završnih čvorova te dubina (broj razina) stabla. Kao što je prikazano u tablici 4.34. od početno specificiranih sedam nezavisnih varijabli njih pet je u konačnici uključeno u završni model.

Tablica 4.34. – Postavke i zbirni rezultati modela za Opel Astra (Exhaustive CHAID)

Specifikacije	Metoda	Exhaustive CHAID
	Zavisna varijabla	tp_prolaz
	Nezavisne varijable	snaga, ccm, index_stp, vr_motora, vr_kocnica, index_km, starost_god
	Validacija	Metoda dijeljenja uzorka
	Najveća dubina stabla (broj razina)	3
	Najmanji broj slučajeva u "roditeljskom" čvoru	100
	Najmanji broj slučajeva u "dječjem" čvoru	100
Rezultati	Uključene nezavisne varijable	index_km, starost_god, index_stp, snaga, vr_motora
	Broj čvorova	31
	Broj završnih čvorova	21
	Dubina (broj razina)	3

Nezavisne varijable "vr_kocnica" (vrsta kočionog sustava ugrađenog na osobni automobil) i „ccm" (radni obujam motora) nisu značajno pridonijele modelu te su iste od strane algoritma automatski isključene iz završnog modela.

Kao što je prikazano na stablu odlučivanja trening uzorka prediktivnog modela tehničke ispravnosti osobnih automobila Opel Astra (slika 4.14.), radi se o dvadeset jednom završnom čvoru, u okviru trideset jednog čvora, unutar tri razine modela.

Iz predmetnog stabla odlučivanja proizlazi da je, kao i kod osobnih automobila Fiat Punto, nezavisna varijabla "indeks_km" (prijeđeni put osobnih automobila) varijabla s najvećim prediktivnim potencijalom (prva razina dubine stabla) u kontekstu tehničke neispravnosti osobnih automobila Opel Astra.

Nezavisna varijabla “indeks_km” dijeli uzorak na pet segmenata u okviru kojih su izraženi različiti prijeđeni putevi osobnih automobila, te gdje su zastupljene značajne razlike u postotku tehnički neispravnih osobnih automobila : osobni automobili koji su prešli do 50.000 km - 5,0 % tehnički neispravnih, osobni automobili koji su prešli od 50.000 do 100.000 km - 8,7 % tehnički neispravnih, osobni automobili koji su prešli od 100.000 do 150.000 km - 20,1 % tehnički neispravnih, osobni automobili koji su prešli od 150.000 do 200.000 km - 40,6 % tehnički neispravnih, osobni automobili koji su prešli više od 200.000 km - 50,3 % tehnički neispravnih.

U pogledu nezavisne varijable “indeks_km”, a u okviru prve razine dubine stabla, nema završnih čvorova, odnosno svi čvorovi (od broja 1 do broja 5) su “roditeljski” čvorovi.

U okviru druge razine dubine stabla prikazano je da nezavisna varijabla “starost_god” dijeli uzorak na dva segmenta u okviru kojih je izražena različita „starost“ osobnih automobila, kod osobnih automobila koji su prešli do 50.000 km: osobni automobili „stari“ 2 godine i manje – 2,2 % tehnički neispravnih i osobni automobili „stariji“ od 2 godine – 6,2 % tehnički neispravnih.

Navedena nezavisna varijabla još u tri slučaja dijeli uzorak na tri, četiri i šest segmenata. Kod osobnih automobila koji su prešli između 100.000 i 150.000 km: osobni automobili „stari“ 6 godina i manje – 12,6 % tehnički neispravnih, osobni automobili „stari“ 7 i 8 godina – 30,2 % tehnički neispravnih, osobni automobili „stari“ 9 do 11 godina, uključujući 11 godina – 19,7 % tehnički neispravnih, osobni automobili „stari“ 12 i 13 godina – 5,7 % tehnički neispravnih, osobni automobili „stari“ 14 i 15 godina – 15,2 % tehnički neispravnih i osobni automobili „stariji“ od 15 godina – 52,8 % tehnički neispravnih. Kod osobnih automobila koji su prešli između 150.000 i 200.000 km: osobni automobili „stari“ 8 godina i manje – 32,4 % tehnički neispravnih, osobni automobili „stari“ 9 do 11 godina, uključujući 11 godina – 21,7 % tehnički neispravnih, osobni automobili „stari“ 12 i 13 godina – 33,6 % tehnički neispravnih i osobni automobili „stariji“ od 13 godina – 55,2 % tehnički neispravnih. Kod osobnih automobila koji su prešli više od 200.000 km: osobni automobili „stari“ 11 godina i manje – 25,1 % tehnički neispravnih, osobni automobili „stari“ 12 do 15 godina, uključujući 15 godina – 39,5 % tehnički neispravnih i osobni automobili „stariji“ od 15 godina – 74,3 % tehnički neispravnih.

U pogledu nezavisne varijable “starost_god”, a u okviru druge razine dubine stabla, čvorovi pod brojevima 6, 7, 11 i 18 predstavljaju “roditeljske” čvorove, dok su čvorovi pod brojem 10, 12, 13, 14, 15, 16, 17, 19, 20, 21 i 22 završni čvorovi.
U okviru druge razine dubine stabla prikazano je da nezavisna varijabla “snaga” dijeli uzorak na dva segmenta u pogledu snage motora, kod osobnih automobila koji su prešli više između 50.000 i 100.000 km: osobni automobili s motorom snage manje ili jednake 50 kW – 15,6 % tehnički neispravnih i osobni automobili s motorom snage veće od 50 kW – 7,8 % tehnički neispravnih.

U pogledu nezavisne varijable “snaga”, a u okviru druge razine dubine stabla, čvorovi pod brojevima 8 i 9 predstavljaju završne čvorove.

U okviru treće (završne) razine dubine stabla prikazano je da nezavisna varijabla “index_stp” dijeli uzorak na dva segmenta kod osobnih automobila koji su prešli do 50.000 km, te koji su „stari“ 2 godine i manje: osobni automobili kod kojih je tehnički pregled obavljen u stanici za tehnički pregled vozila koja se nalazi u urbanoj sredini i maloj sredini - 0,8 % tehnički neispravnih i osobni automobili kod kojih je tehnički pregled obavljen u stanici za tehnički pregled vozila koja se nalazi u velikoj urbanoj sredini - 3,7 % tehnički neispravnih.

Navedena nezavisna varijabla još u jednom slučaju dijeli uzorak na dva segmenta. Kod osobnih automobila koji su prešli do 50.000 km, te koji su „stari“ više od 2 godine: osobni automobili kod kojih je tehnički pregled obavljen u stanici za tehnički pregled vozila koja se nalazi u maloj sredini - 4,3 % tehnički neispravnih i osobni automobili kod kojih je tehnički pregled obavljen u stanici za tehnički pregled vozila koja se nalazi u velikoj urbanoj sredini i urbanoj sredini - 7,3 % tehnički neispravnih.

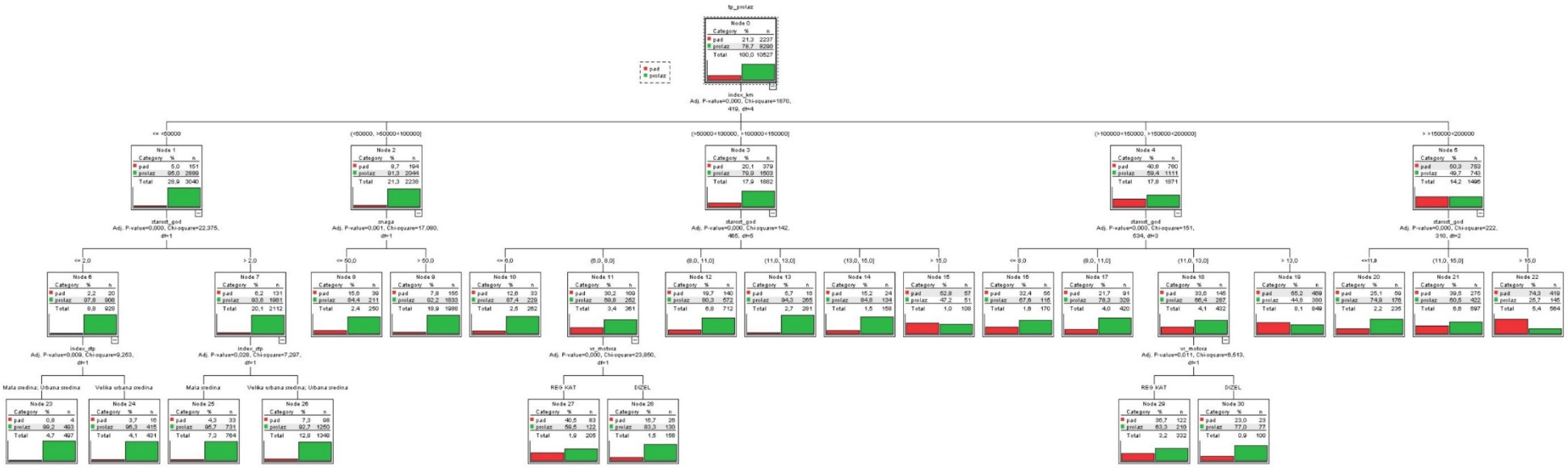
U pogledu nezavisne varijable “index_stp”, a u okviru treće (završne) razine dubine stabla, čvorovi pod brojem 23, 24, 25 i 26 su završni čvorovi.

Treća nezavisna varijabla treće (završne) razine dubine stabla “vr_motora” dijeli uzorak na dva segmenta kod osobnih automobila koji su prešli između 100.000 i 150.000 km, te koji su „stari“ 7 i 8 godina: osobni automobili s benzinskim (Ottovim) motorom i ugrađenim reguliranim katalizatorom (REG KAT) - 40,5 % tehnički

neispravnih i osobni automobili s dizelskim motorom (DIZEL) - 16,7 % tehnički neispravnih.

Navedena nezavisna varijabla još u jednom slučaju dijeli uzorak na dva segmenta. Kod osobnih automobila koji su prešli između 150.000 i 200.000 km, te koji su „stari“ 12 i 13 godina: osobni automobili s benzinskim (Ottovim) motorom i ugrađenim reguliranim katalizatorom (REG KAT) - 36,7 % tehnički neispravnih i osobni automobili s dizelskim motorom (DIZEL) - 23,0 % tehnički neispravnih.

U pogledu nezavisne varijable “vr_motora”, a u okviru treće (završne) razine dubine stabla, čvorovi pod brojem 27, 28, 29 i 30 su završni čvorovi.



Slika 4.14. - Stablo odlučivanja trening uzorka prediktivnog modela za Opel Astra

U tablici 4.35. (dobit po završnim čvorovima) sadržani su podaci o završnim čvorovima, odnosno o čvorovima kod kojih je prestalo grananje stabla u daljnje čvorove. Završni čvorovi, sami po sebi, predstavljaju najbolji prediktivni potencijal predmetnog modela.

Tablica 4.35. - Dobit po završnim čvorovima za osobne automobile Opel Astra

Uzorak	Završni čvor	Završni čvor		Dobit		Odziv	Indeks
		Ukupni broj Opel Astra osobnih automobila po završnom čvoru	Postotak u odnosu na ukupni skup Opel Astra osobnih automobila	Broj "palih" Opel Astra osobnih automobila po završnom čvoru	Postotak u odnosu na ukupni skup "palih" Opel Astra osobnih automobila		
	I	II	III	IV	V	VI	VII
Trening	22	564	5,4 %	419	18,7 %	74,4 %	349,6 %
	19	849	8,1 %	469	21,0 %	55,2 %	260,0 %
	15	108	1,0 %	57	2,5 %	52,8 %	248,4 %
	27	205	1,9 %	83	3,7 %	40,5 %	190,5 %
	21	697	6,6 %	275	12,3 %	39,5 %	185,7 %
	29	332	3,2 %	122	5,5 %	36,7 %	172,9 %
	16	170	1,6 %	55	2,5 %	32,4 %	152,2 %
	20	235	2,2 %	59	2,6 %	25,1 %	118,1 %
	30	100	0,9 %	23	1,0 %	23,0 %	108,2 %
	17	420	4,0 %	91	4,1 %	21,7 %	102,0 %
	12	712	.6,8 %	140	6,3 %	19,7 %	92,5 %
	28	156	1,5 %	26	1,2 %	16,7 %	78,4 %
	8	250	2,4 %	39	1,7 %	15,6 %	73,4 %
	14	158	1,5 %	24	1,1 %	15,2 %	71,5 %
	10	262	2,5 %	33	1,5 %	12,6 %	59,3 %
	9	1.988	18,9 %	155	6,9 %	7,8 %	36,7 %
	26	1.348	12,8 %	98	4,4 %	7,3 %	34,2 %
	13	281	2,7 %	16	0,7 %	5,7 %	26,8 %
	25	764	7,3 %	33	1,5 %	4,3 %	20,3 %
	24	431	4,1 %	16	0,7 %	3,7 %	17,5 %
	23	497	4,7 %	4	0,2 %	0,8 %	3,8 %
Test	22	3.396	5,7 %	2.492	19,7 %	73,4 %	346,9 %
	19	4.592	7,7 %	2.458	19,5 %	53,5 %	253,0 %
	15	548	0,9 %	319	2,5 %	58,2 %	275,2 %
	27	1.171	2,0 %	485	3,8 %	41,4 %	195,8 %
	21	3.702	6,2 %	1.437	11,4 %	38,8 %	183,5 %
	29	1.717	2,9 %	606	4,8 %	35,3 %	166,8 %
	16	993	1,7 %	303	2,4 %	30,5 %	144,2 %
	20	1.219	2,0 %	322	2,6 %	26,4 %	124,9 %
	30	664	1,1 %	221	1,8 %	33,3 %	157,3 %
	17	2.516	4,2 %	669	5,3 %	26,6 %	125,7 %
	12	3.948	6,6 %	733	5,8 %	18,6 %	87,8 %
	28	915	1,5 %	158	1,3 %	17,3 %	81,6 %
	8	1.557	2,6 %	184	1,5 %	11,8 %	55,9 %
	14	911	1,5 %	199	1,6 %	21,8 %	103,3 %
	10	1.340	2,2 %	151	1,2 %	11,3 %	53,3 %
	9	11.108	18,6 %	920	7,3 %	8,3 %	39,2 %
	26	7.853	13,2 %	450	3,6 %	5,7 %	27,1 %
	13	1.601	2,7 %	158	1,3 %	9,9 %	46,7 %
	25	4.657	7,8 %	241	1,9 %	5,2 %	24,5 %
	24	2.473	4,1 %	83	0,7 %	3,4 %	15,9 %
	23	2.786	4,7 %	33	0,3 %	1,2 %	5,6 %

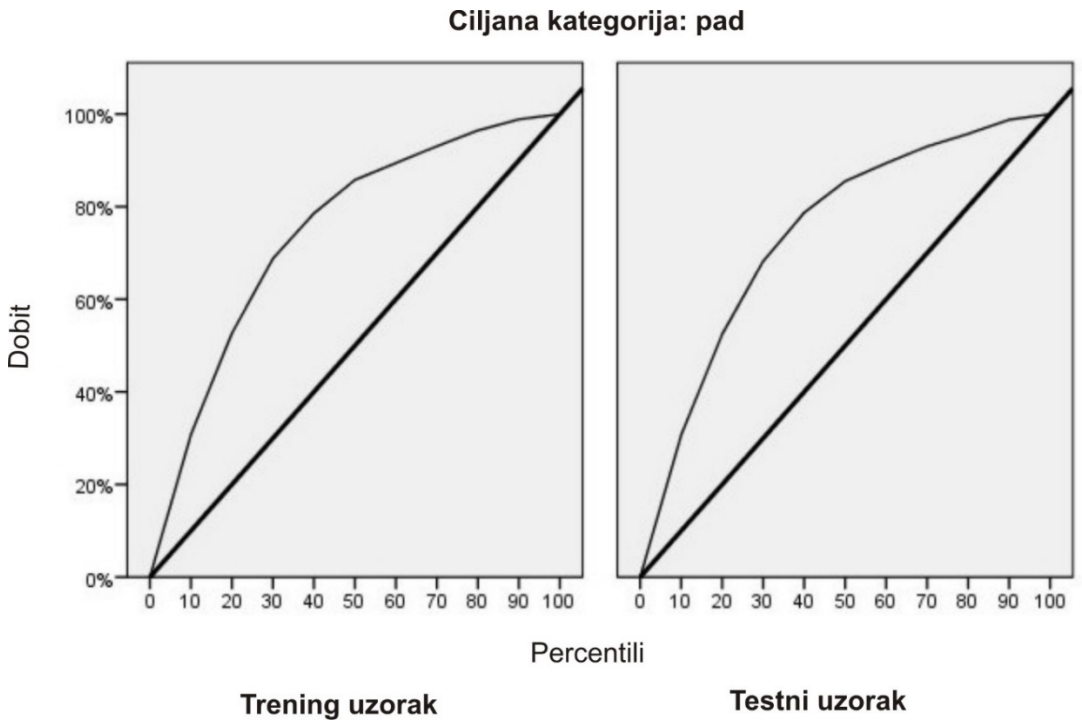
U predmetnoj tablici sadržani su broj i postotak slučajeva/osobnih automobila po pojedinom završnom čvoru. U okviru stupca "Dobit" također su sadržani broj i postotak slučajeva/osobnih automobila po pojedinom završnom čvoru u okviru ciljane kategorije

(broj i postotak slučajeva/osobnih automobila s utvrđenim tehničkim neispravnostima tijekom tehničkog pregleda što je rezultiralo “padom”).

Za kategorijsku zavisnu varijablu, stupac “Odziv” prikazuje unutarnju strukturu završnih čvorova, odnosno u istom su sadržani postotci slučajeva/osobnih automobila u pojedinom završnom čvoru u okviru predmetne ciljane kategorije (osobni automobili koji su “pali” na tehničkom pregledu”), a u odnosu na ukupni broj slučajeva/osobnih automobila po pojedinom završnom čvoru (prema: IV x 100/II).

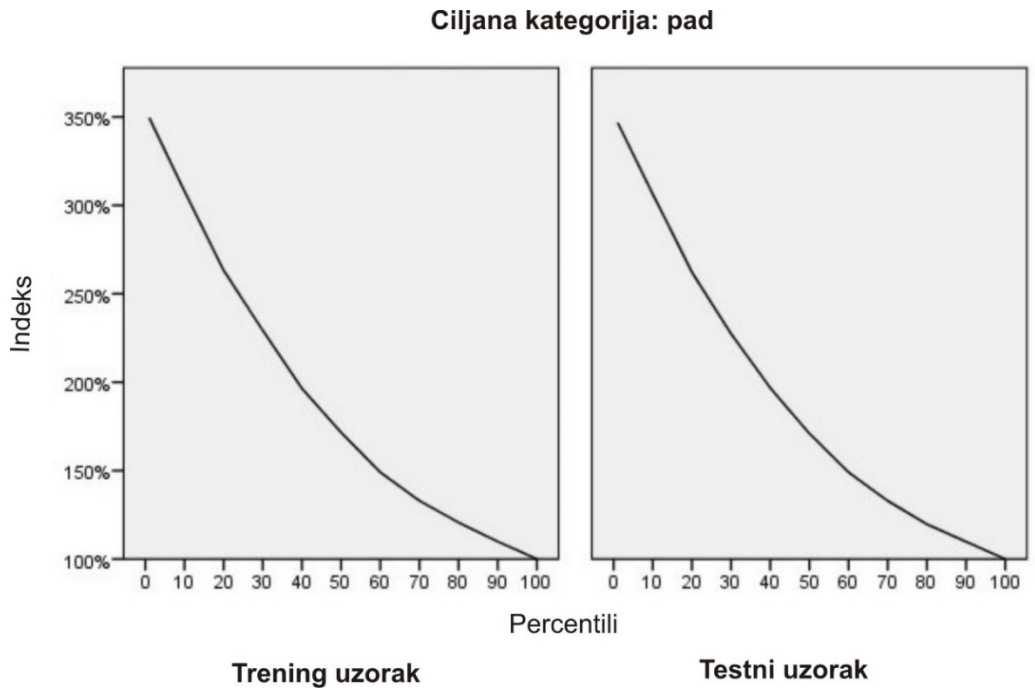
Za kategorijsku zavisnu varijablu indeks podrazumijeva omjer odziva za ciljanu kategoriju u odnosu na odziv za ukupni skup. Pomoću indeksa se dobiva informacija o tome koliko se značajno postotak promatrane ciljane kategorije za predmetni završni čvor razlikuje u odnosu na očekivani postotak ciljane kategorije (za ukupni skup) koji je izražen u korijenskom čvoru, te za osobne automobile Opel Astra iznosi 21,30 % (na primjer, aproksimativno, u završnom čvoru broj 22 postotna vrijednost indeksa iznosi 74,4 % x 100/21,30 % = 349,4 %). Vrijednost indeksa koja prelazi 100 % ukazuje da se u predmetnom završnom čvoru nalazi veći postotni iznos ciljane kategorije u usporedbi s očekivanim postotnim iznosom ciljane kategorije za ukupni skup.

Na slici 4.15. dan je grafički prikaz dobiti za ciljanu kategoriju “pad” nezavisne varijable “tp_prolaz”, za trening i testni uzorak. Na apscisi se nalaze percentili dok se na ordinati nalaze vrijednosti dobiti izražene u postotcima.



Slika 4.15. - Dijagram dobiti za ciljanu kategoriju "pad" za Opel Astra

Na slici 4.16. dan je grafički prikaz indeksa za ciljanu kategoriju "pad" nezavisne varijable "tp_prolaz", za trening i testni uzorak. Na apscisi se nalaze percentili dok se na ordinati nalaze vrijednosti indeksa izražene u postotcima.



Slika 4.16. - Dijagram indeksa za ciljanu kategoriju "pad" za Opel Astra

Temeljem podataka tablice rizika (tablica 4.36.) i klasifikacijske tablice (tablica 4.37.) donosi se ocjena o prediktivnom potencijalu predmetnog modela.

Tablica 4.36. - Rizik za Opel Astra

Uzorak	Procijenjena vrijednost	Standardna devijacija
Trening	0,177	0,004
Testni	0,178	0,002

Iz podataka sadržanih u tablici rizika prikazan je rizik predikcije koji predstavlja postotak pogrešne klasifikacije kategorija. U pogledu provedenog istraživanja proizlazi da, u slučaju kada su poznate karakteristike osobnih automobila Opel Astra u smislu predmetnih šest nezavisnih (prediktivnih) varijabli, rizik pogrešne klasifikacije osobnih automobila Opel Astra glede utvrđivanja tehničke neispravnosti tijekom tehničkog pregleda (odnosno "pada" na tehničkom pregledu) jest 17,7 % (primjenom metode dijeljenja uzorka) za trening uzorak, odnosno 17,8 % za testni uzorak.

Podaci iz kontingencijske tablice 4.37. potvrđuju podatke iz tablice rizika u smislu da predmetni model, u ukupno 82,3 % slučajeva (osobnih automobila Opel Astra) trening uzorka točno klasificira kategoriju, odnosno u ukupno 82,2 % slučajeva testnog uzorka.

Tablica 4.37. - Klasifikacija za Opel Astra (Exhaustive CHAID)

Uzorak	Promatrane kategorije	Prediktivno		
		„pad“	„prolaz“	Postotak točnosti
Trening	„pad“	**945**	**1.292**	**42,2 %**
	„prolaz“	**576**	**7.714**	93,1 %
	Ukupni postotak	14,4 %	85,6 %	**82,3 %**
Testni	„pad“	**5.269**	**7.353**	**41,7 %**
	„prolaz“	**3.267**	**43.778**	93,1 %
	Ukupni postotak	14,3 %	85,7 %	**82,2 %**

Nadalje, u kontekstu analize učinkovitosti prediktivnog modela, prema izrazu 4.2. izračunat je MCC za trening i testni uzorak, a rezultati su prikazani u tablici 4.38.

Tablica 4.38. – Prikaz vrijednosti MCC za Opel Astra (Exhaustive CHAID)

Uzorak	Kriterij	Vrijednost
Trening	pogrešno pozitivni	1.292
	pogrešno negativni	576
	stvarno pozitivni	7.714
	stvarno negativni	945
	MCC	**0,40**
Testni	pogrešno pozitivni	7.353
	pogrešno negativni	3.267
	stvarno pozitivni	43.778
	stvarno negativni	5.269
	MCC	**0,40**

Kao što je prikazano u tablici 4.39., nakon primjene Exhaustive CHAID algoritma na bazi podataka osobnih automobila Opel Astra izraženoj u vidu sedam nezavisnih varijabli te jedne zavisne varijable, generirane su četiri nove varijable u predmetnoj bazi podataka po istom principu kao kod osobnog automobila VW Golf (točka 4.4.1. ovog rada) i osobnog automobila Fiat Punto (točka 4.4.2. ovog rada).

Tablica 4.39. – Izvod dijela prediktivnih vrijednosti za Opel Astra

Redni broj	„snaga“	„ccm“	„starost_god“	„index_stp“	„vr_motora“	„vr_kocnica“	„index_km“	„tp_prolaz“	Broj čvora	Prediktivna vrijednost	Prediktivna vjerojatnost 1	Prediktivna vjerojatnost 2
1.	48	1199	9	1	2	1	3	0	19	1	0,22	0,78
2.	48	1199	11	3	2	1	2	0	16	1	0,11	0,89
3.	48	1199	12	1	2	1	4	0	15	1	0,31	0,69
4.	48	1199	11	3	2	1	1	0	16	1	0,11	0,89
5.	48	1199	11	3	2	1	3	0	16	1	0,11	0,89
6.	48	1199	11	1	2	1	5	0	17	1	0,27	0,73
7.	48	1199	11	3	2	1	5	0	17	1	0,27	0,73
8.	48	1199	11	3	2	1	5	0	17	1	0,27	0,73
9.	48	1199	10	2	2	1	3	0	19	1	0,22	0,78
...	...	...	...	...	...	...	...	...	...	...	...	...

Varijabla “Prediktivna vrijednost” predstavlja opseg slučajeva/osobnih automobila u pojedinoj kategoriji zavisne varijable za završni čvor u okviru kojeg se nalazi pojedini slučaj/osobni automobil. U primjeru iz tablice 4.39., u čvoru 15, 31 % slučajeva/osobnih automobile potpada pod kategoriju “pad” (“Prediktivna vrijednost ” - 0), a 69 % slučajeva/osobnih automobila potpada pod kategoriju “prolaz” (“Prediktivna vrijednost ” - 1), što predstavlja prediktivne vjerojatnosti od 0,31 i 0,69.

4.4.3.1. Analiza prediktivnog potencijala modela temeljenog na Exhaustive CHAID algoritmu

Procjena prediktivnog potencijala modela proizlazi iz analize podataka sadržanih u tablici rizika (tablica 4.36.) i kontingencijskoj tablici (tablica 4.37.).

Iz kontingencijske tablice (tablica 4.37.) proizlazi podatak o samo 42,2 % točno predviđenih slučajeva/osobnih automobila u kategoriji “pad" u okviru trening uzorka te samo 41,7 % u okviru testnog uzorka, odnosno čak 57,8 % pogrešno klasificiranih slučajeva (pogrešno klasificiranih tehnički neispravnih osobnih automobila u kategoriju tehničkih ispravnih osobnih automobila koji su “prošli” tehnički pregled) u okviru trening uzorka te 58,3 % u okviru testnog uzorka. Ovakav oblik pogrešne klasifikacije ima mnogo veći negativni značaj u kontekstu sigurnosti cestovnog prometa, od obrnutog slučaja – kod pogrešno klasificiranih tehnički ispravnih osobnih automobila u kategoriju tehničkih neispravnih osobnih automobila koji su “pali” na tehničkom pregledu.

U skladu s objašnjenjem iznesenim u točci 4.4.1.1. ovog rada postotni iznos od 70,0 % točno predviđenih slučajeva/osobnih automobila u kategoriji “pad" (uz definiranu toleranciju pogreške) smatra se minimalno zadovoljavajućim u kontekstu prediktivnog potencijala modela.

Temeljem prethodno navedenog proizlazi da iznos od 42,2 % (41,7 %) točno predviđenih slučajeva/osobnih automobila u kategoriji “pad", odnosno čak 57,8 % (58,3 %) pogrešno klasificiranih slučajeva nije zadovoljavajući s obzirom da se radi o ciljanoj kategoriji zavisne varijable, odnosno model nema zadovoljavajući prediktivni potencijal, te je potrebno izvršiti radnju radi poboljšanja (korektivne radnje) koja će dovesti do povećanja prediktivnog potencijala modela.

Vrijednost korelacijskog koeficijenta MCC je identičan za oba uzorka (trening i testni uzorak) i isti je srednje pozitivan +0,40, što podrazumijeva dobru predikciju već sada (teoretska vrijednost +1 podrazumijeva savršenu predikciju, vrijednost 0 podrazumijeva predikciju koja nije bolja od nasumične predikcije, teoretska vrijednost -1 podrazumijeva potpuno razmimoilaženje između predikcije i opažanja).

4.4.3.2. Poboljšanje prediktivnog potencijala modela temeljenog na Exhaustive CHAID algoritmu

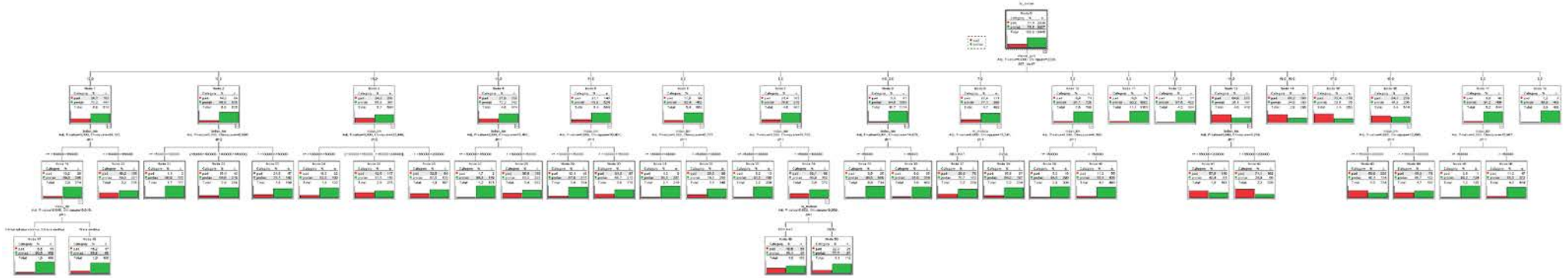
Metodologija provođenja radnje radi poboljšanja prediktivnog potencijala modela za osobne automobile Opel Astra identična je primijenjenoj metodologiji u slučaju prediktivnog modela za osobne automobile VW Golf i Fiat Punto, kako je izneseno u točkama 4.4.1.2. i 4.4.2.2. ovog rada.

Kao i u slučaju osobnih automobila Fiat Punto, s obzirom da penalizacija negativnih posljedica, kod pogrešnog klasificiranja osobnih automobila Opel Astra koji su tehnički neispravni u klasu tehnički ispravnih osobnih automobila, s vrijednošću 2 (kao u slučaju penalizacije kod osobnih automobila VW Golf) još uvijek nije pokazala zadovoljavajuće rezultate, ista je provedena s vrijednošću 3.

Prethodno navedeno vrednovanje penalizacije, u kontekstu osobnih automobila Opel Astra, izvršeno je u omjeru 3:1 (trostruko veća vrijednost negativnih posljedica kod pogrešnog klasificiranja osobnog automobila koji je tehnički neispravan u klasu tehnički ispravnih osobnih automobila, nego u obrnutom slučaju) nakon provedenog eksperimentiranja, iz razloga što je upravo pri predmetnom omjeru postignut minimalni određeni postotni iznos od 70 % točno predviđenih slučajeva/osobnih automobila u kategoriji “pad" čime se smatra da je postignut zadovoljavajući prediktivni potencijal modela (u skladu s točkom 4.4.1.1. ovog rada).

4.3.3.3. Analiza nakon provedenog poboljšanja modela temeljenog na Exhaustive CHAID algoritmu

Provođenjem predmetne radnje vidljive su promjene u stablu odlučivanja prediktivnog modela tehničke ispravnosti osobnih automobila Opel Astra. Predmetne promjene ne odnose se na distribuciju slučajeva/osobnih automobila po pojedinim čvorovima, koja je ostala ista, već se odnose na dominantne prediktivne kategorije zavisne varijable, kao što je prikazano na slici 4.17.



Slika 4.17. - Stablo odlučivanja trening uzorka prediktivnog modela za Opel Astra nakon provođenja radnje za poboljšanje modela

S obzirom na trostruko veće vrednovanje penalizacije pogrešnog klasificiranja osobnog automobila koji je tehnički neispravan u klasu tehnički ispravnih osobnih automobila u odnosu na penalizaciju pogrešnog klasificiranja osobnog automobila koji je tehnički ispravan u klasu tehnički neispravnih osobnih automobila, u svim čvorovima koji su prije provođenja korektivne radnje imali podjednaku distribuciju obje prediktivne kategorije, ili nešto veću distribuciju kategorije "prolaz", nakon provođenja korektivne radnje u istima će dominirati kategorija "pad".

Promjene dominantne prediktivne kategorije, iz kategorije "prolaz" u kategoriju "pad" zabilježene su u čvorovima broj 1, 3, 4 i 34 te u završnim čvorovima broj 20, 25, 26, 28, 30, 32, 37 i 49.

Kao što je vidljivo iz tablice 4.40. nakon provođenja radnje za poboljšanje prediktivnog modela došlo je do promjene rezultata u odnosnu na rezultate prije provođenja radnje za poboljšanje prediktivnog modela (tablica 4.34.) i to u broju čvorova, broju završnih čvorova te u pogledu nezavisnih varijabli koje su pridonijele modelu (u model nisu uključene nezavisne varijable „vr_kocnica", „snaga" i „ccm").

Tablica 4.40. – Zbirni rezultati modela za Opel Astra nakon provođenja radnje za poboljšanje (Exhaustive CHAID)

Specifikacije	Metoda	Exhaustive CHAID
	Zavisna varijabla	tp_prolaz
	Nezavisne varijable	snaga, ccm, index_stp, vr_motora, vr_kocnica, index_km, starost_god
	Validacija	Metoda dijeljenja uzorka
	Najveća dubina stabla (broj razina)	3
	Najmanji broj slučajeva u "roditeljskom" čvoru	100
	Najmanji broj slučajeva u "dječjem" čvoru	100
Rezultati	Uključene nezavisne varijable	**starost_god, index_km, , index_stp, vr_motora**
	Broj čvorova	**51**
	Broj završnih čvorova	**35**
	Dubina (broj razina)	3

Provođenjem korektivne radnje došlo je do promjena u tablici rizika (tablica 4.41.), te klasifikacijskoj tablici (tablica 4.42.).

Tablica 4.41. - Rizik za Opel Astra nakon izvršene radnje za poboljšanje

Uzorak	Procijenjena vrijednost	Standardna devijacija
Trening	0,350	0,007
Testni	0,354	0,003

"Nusprodukt" provođenja radnje za poboljšanje modela jest očekivano povećanje rizika predikcije, u okviru trening uzorka, s 17,7 % prije provođenja korektivne radnje, na 35,0 % poslije provođenja korektivne radnje, odnosno u okviru testnog uzorka s 17,8% na 35,4 %

Međutim, kao i u prethodnim slučajevima, provođenje radnje za poboljšanje modela dovelo je do željenog efekta, odnosno značajnog povećanja prediktivnog potencijala modela u kontekstu ciljane prediktivne kategorije zavisne varijable "pad" koji proizlazi iz većeg postotka točno klasificiranih slučajeva/osobnih automobila glede utvrđivanja tehničke neispravnosti tijekom tehničkog pregleda (odnosno "pada" na tehničkom pregledu), i to u okviru trening uzorka s 42,2 %, prije provođenja radnje za poboljšanje modela, na 75,7 % nakon provođenja radnje za poboljšanje modela, odnosno u okviru testnog uzorka s 41,7 % na 75,7 %, što je predstavlja rezultat koji zadovoljava, sukladno točci 4.4.1.1.

Tablica 4.42. - Klasifikacija za Fiat Punto nakonizvršene radnje za poboljšanje (Exhaustive CHAID)

Uzorak	Promatrane kategorije	Prediktivno		
		„pad"	„prolaz"	Postotak točnosti
Trening	„pad"	**1.695**	**543**	**75,7 %**
	„prolaz"	**2.023**	**6.184**	75,4 %
	Ukupni postotak	35,6 %	64,4 %	75,4 %
Testni	„pad"	**9.557**	**3.064**	**75,7 %**
	„prolaz"	**11.974**	**35.154**	74,6 %
	Ukupni postotak	36,0 %	64,0 %	74,8 %

Povećanje postotka točno klasificiranih slučajeva/osobnih automobila kod kojih je utvrđena tehnička neispravnost tijekom tehničkog pregleda (što za posljedicu ima "pad" osobnih automobila na tehničkom pregledu) za direktnu posljedicu ima smanjeni postotak točno klasificiranih slučajeva/osobnih automobila kod kojih nije utvrđena tehnička neispravnost tijekom tehničkog pregleda (kategorija "prolaz"), s 93,1 % u okviru trening uzorka na 75,4 %, te u okviru testnog uzorka s 93,1 % na 74,6 %

(odnosno pad ukupnog postotka točno klasificiranih slučajeva/osobnih automobila u okviru trening uzorka s 82,3 % na 75,4 %, te u okviru testnog uzorka s 82,2 % na 74,8 %), međutim sukladno prethodno iznesenoj argumentaciji radi se o prediktivnoj kategoriji sekundarnog značaja, odnosno radi se o prediktivnoj kategoriji koja ima manji značaj za ovo istraživanje s obzirom da ista nema negativan utjecaj na sigurnost prometa na cestama (osobni automobili kod kojih nije utvrđena tehnička neispravnost pogrešno su klasificirani kao osobni automobili koji su "pali" na tehničkom pregledu zbog utvrđene tehničke neispravnosti).

Nadalje, u kontekstu analize učinkovitosti prediktivnog modela, prema izrazu 4.2. nakon izvršene radnje za poboljšanje modela izračunat je MCC za trening i testni uzorak, a rezultati su prikazani u tablici 4.43.

Tablica 4.43. – Prikaz vrijednosti MCC za Opel Astra nakon provođenja radnje za poboljšanje (Exhaustive CHAID)

Uzorak	Kriterij	Vrijednost
Trening	pogrešno pozitivni	543
	pogrešno negativni	2.023
	stvarno pozitivni	6.184
	stvarno negativni	1.695
	MCC	**0,45**
Testni	pogrešno pozitivni	3.064
	pogrešno negativni	11.974
	stvarno pozitivni	35.154
	stvarno negativni	9.557
	MCC	**0,44**

Vrijednosti korelacijskog koeficijenta MCC za oba uzorka (trening i testni uzorak), nakon izvršene radnje za poboljšanje modela, su djelomično povećane te su ostale na razini srednjih pozitivnih vrijednosti +0,45 i +0,44 što potvrđuje dobru predikciju (odnosno bolju predikciju nego prije izvršene radnje za poboljšanje modela).

U konačnici, prediktivni model tehničke ispravnosti za osobne automobile Opel Astra a koji se temelji na metodi stabla odlučivanja i algoritmu Exhaustive CHAID, može se izraziti i u obliku poslovnih pravila, kako slijedi u tablici 4.44.:

Tablica 4.44. – Prikaz poslovnih pravila za poboljšani Exhaustive CHAID prediktivni model Opel Astra

/* Čvor 47 */.
IF (starost_god = 12) AND (index_km NOT MISSING AND (index_km <= ">100000<150000")) AND (index_stp != "Mala sredina") THEN **Prediction = 1**, **Probability = 0.934911**

/* Čvor 48 */.
IF (starost_god = 12) AND (index_km NOT MISSING AND (index_km <= ">100000<150000")) AND (index_stp = "Mala sredina") THEN **Prediction = 1**, **Probability = 0.838095**

/* Čvor 20 */.
IF (starost_god = 12) AND (index_km IS MISSING OR (index_km > ">100000<150000")) THEN **Prediction = 0**, **Probability = 0.401786**

/* Čvor 21 */.
IF (starost_god = 10) AND (index_km NOT MISSING AND (index_km <= ">50000<100000")) THEN **Prediction = 1**, **Probability = 0.988701**

/* Čvor 22 */.
IF (starost_god = 10) AND (index_km IS MISSING OR (index_km > ">50000<100000" AND index_km <= ">100000<150000")) THEN **Prediction = 1**, **Probability = 0.845560**
/* Čvor 23 */.
IF (starost_god = 10) AND (index_km NOT MISSING AND (index_km > ">100000<150000")) THEN **Prediction = 1**, **Probability = 0.751323**

/* Čvor 24 */.
IF (starost_god = 14) AND (index_km NOT MISSING AND (index_km <= ">100000<150000")) THEN **Prediction = 1**, **Probability = 0.819672**

/* Čvor 25 */.
IF (starost_god = 14) AND (index_km IS MISSING OR (index_km > ">100000<150000" AND index_km <= ">150000<200000")) THEN **Prediction = 0**, **Probability = 0.425455**

/* Čvor 26 */.
IF (starost_god = 14) AND (index_km NOT MISSING AND (index_km > ">150000<200000")) THEN **Prediction = 0**, **Probability = 0.324873**

/* Čvor 27 */.
IF (starost_god = 13) AND (index_km NOT MISSING AND (index_km <= ">100000<150000")) THEN **Prediction = 1**, **Probability = 0.983471**

/* Čvor 28 */.
IF (starost_god = 13) AND (index_km IS MISSING OR (index_km > ">100000<150000")) THEN **Prediction = 0**, **Probability = 0.368272**

/* Čvor 29 */.
IF (starost_god = 11) AND (index_km IS MISSING OR (index_km <= ">100000<150000")) THEN **Prediction = 1**, **Probability = 0.878531**

/* Čvor 30 */.
IF (starost_god = 11) AND (index_km NOT MISSING AND (index_km > ">100000<150000")) THEN **Prediction = 0**, **Probability = 0.312903**

/* Čvor 31 */.
IF (starost_god = 9) AND (index_km NOT MISSING AND (index_km <= ">50000<100000")) THEN **Prediction = 1**, **Probability = 0.958333**

/* Čvor 32 */.

IF (starost_god = 9) AND (index_km IS MISSING OR (index_km > ">50000<100000")) THEN **Prediction = 0**, **Probability = 0.260116**

/* Čvor 33 */.
IF (starost_god = 8) AND (index_km NOT MISSING AND (index_km <= ">50000<100000")) THEN **Prediction = 1**, **Probability = 0.937799**

/* Čvor 49 */.
IF (starost_god = 8) AND (index_km IS MISSING OR (index_km > ">50000<100000")) AND (vr_motora != "DIZEL") THEN **Prediction = 0**, **Probability = 0.406250**

/* Čvor 50 */.
IF (starost_god = 8) AND (index_km IS MISSING OR (index_km > ">50000<100000")) AND (vr_motora = "DIZEL") THEN **Prediction = 1**, **Probability = 0.776786**

/* Čvor 35 */.
IF (starost_god = 4 OR starost_god = 2) AND (index_km IS MISSING OR (index_km <= "<50000")) THEN **Prediction = 1**, **Probability = 0.964986**

/* Čvor 36 */.
IF (starost_god = 4 OR starost_god = 2) AND (index_km NOT MISSING AND (index_km > "<50000")) THEN **Prediction = 1**, **Probability = 0.910000**

/* Čvor 37 */.
IF (starost_god = 7) AND (vr_motora != "DIZEL") THEN **Prediction = 0**, **Probability = 0.293436**

/* Čvor 38 */.
IF (starost_god = 7) AND (vr_motora = "DIZEL") THEN **Prediction = 1**, **Probability = 0.841880**

/* Čvor 39 */.
IF (starost_god = 5) AND (index_km NOT MISSING AND (index_km <= "<50000")) THEN **Prediction = 1**, **Probability = 0.947712**

/* Čvor 40 */.
IF (starost_god = 5) AND (index_km IS MISSING OR (index_km > "<50000")) THEN **Prediction = 1**, **Probability = 0.888438**

/* Čvor 11 */.
IF (starost_god != 12 AND starost_god != 10 AND starost_god != 14 AND starost_god != 13 AND starost_god != 11 AND starost_god != 9 AND starost_god != 8 AND starost_god != 4 AND starost_god != 7 AND starost_god != 5 AND starost_god != 2 AND starost_god != 1 AND starost_god != 16 AND starost_god != 18 AND starost_god != 17 AND starost_god != 15 AND starost_god != 19 AND starost_god != 6 AND starost_god != 0) THEN **Prediction = 1**, **Probability = 0.931955**

/* Čvor 12 */.
IF (starost_god = 1) THEN **Prediction = 1**, **Probability = 0.974654**

/* Čvor 41 */.
IF (starost_god = 16) AND (index_km NOT MISSING AND (index_km <= ">150000<200000")) THEN **Prediction = 0**, **Probability = 0.575916**

/* Čvor 42 */.
IF (starost_god = 16) AND (index_km IS MISSING OR (index_km > ">150000<200000")) THEN **Prediction = 0**, **Probability = 0.710526**

/* Čvor 14 */.
IF (starost_god = 18 OR starost_god = 19) THEN **Prediction = 0**, **Probability = 0.652027**

/* Čvor 15 */.

IF (starost_god = 17) THEN **Prediction = 0**, **Probability = 0.703557**

/* Čvor 43 */.
IF (starost_god = 15) AND (index_km IS MISSING OR (index_km <= ">150000<200000")) THEN **Prediction = 0**, **Probability = 0.598802**

/* Čvor 44 */.
IF (starost_god = 15) AND (index_km NOT MISSING AND (index_km > ">150000<200000")) THEN **Prediction = 0**, **Probability = 0.433333**

/* Čvor 45 */.
IF (starost_god = 6) AND (index_km NOT MISSING AND (index_km <= "<50000")) THEN **Prediction = 1**, **Probability = 0.992000**

/* Čvor 46 */.
IF (starost_god = 6) AND (index_km IS MISSING OR (index_km > "<50000")) THEN **Prediction = 1**, **Probability = 0.887828**

/* Čvor 18 */.
IF (starost_god = 0) THEN **Prediction = 1**, **Probability = 0.987745**

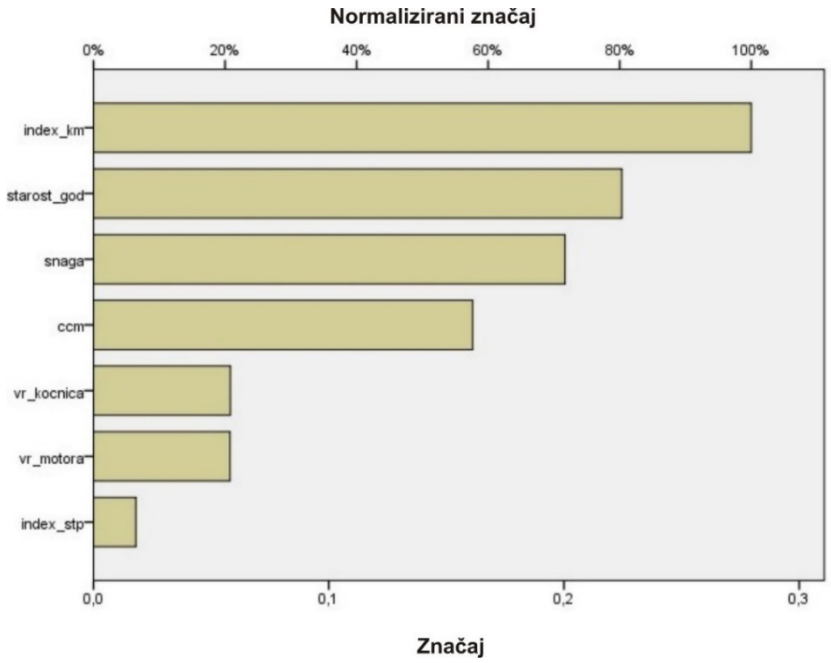
4.4.3.4. Alternativni prediktivni model temeljen na algoritmu perceptronskih neuronskih mreža

Kao alternativa preferencijalnom prediktivnom modelu koji se temelji na Exhaustive CHAID algoritmu korišten je algoritam perceptronske neuronske mreže. U tablici 4.45. dane su parametri strukture neuronske mreže za osobne automobile Opel Astra.

Tablica 4.45. – Parametri strukture neuronske mreže za Opel Astra

Ulazni sloj	Varijable	1	„snaga“
		2	„ccm“
		3	„index_stp“
		4	„vr_motora“
		5	„vr_kocnica“
		6	„index_km“
		7	„starost_god“
	Broj sneurona		87
	Broj skrivenih slojeva		1
Skriveni sloj(evi)	Broj neurona u skrivenim slojevima 1		8
	Aktivacijska funkcija		Hiperbolička tangenta
Izlazni sloj	Zavisne varijable 1		„tp_prolaz“
	Broj neurona		2
	Aktivacijska funkcija		softmax
	Funkcija pogreške		križna entropija

Kao što je vidljivo iz slike 4.18. procjena značaja u smislu prediktivnog potencijala pojedinih nezavisnih varijabli za osobne automobile Opel Astra proizašlih iz algoritma neuronske mreže razlikuje se od procjene njihovog značaja proizašle iz Exhaustive CHAID algoritma.



Slika 4.18. – Prikaz nezavisnih varijabli prema procjeni razine značaja za Opel Astra (neuronska mreža)

Proizlazi da je najznačajnija nezavisna varijabla „index_km“, koju prema razini značajnosti slijede nezavisne varijable „starost_god“, „snaga“ te „ccm“, dok nezavisne varijable „vr_kocnica“, „vr_motora“ i „index_stp“ imaju najmanji značaj.

Iz kontingencijske tablice (tablica 4.46.) proizlazi podatak o 37,2 % točno predviđenih slučajeva/osobnih automobila u kategoriji “pad" u okviru trening uzorka te 38,2 % u okviru testnog uzorka, odnosno čak 62,8 % pogrešno klasificiranih slučajeva (pogrešno klasificiranih tehnički neispravnih osobnih automobila u kategoriju tehničkih ispravnih osobnih automobila koji su “prošli” tehnički pregled) u okviru trening uzorka te 61,8 % u okviru testnog uzorka.

Tablica 4.46. - Klasifikacija za Opel Astra (neuronska mreža)

Uzorak	**Promatrane kategorije**	**Prediktivno**		
		„pad“	„prolaz“	Postotak točnosti
Trening	„pad“	**2.766**	**4.662**	**37,2 %**
	„prolaz“	**1.696**	**25.823**	93,8 %
	Ukupni postotak	12,8 %	87,2 %	81,8 %
Testni	„pad“	**2.270**	**3.679**	**38,2 %**
	„prolaz“	**1.293**	**21.084**	94,2 %
	Ukupni postotak	12,6 %	87,4 %	82,4 %
Validacijski	„pad“	585	896	39,5 %
	„prolaz“	314	5.118	94,2 %
	Ukupni postotak	13,0 %	87,0 %	82, 5%

S obzirom da su predmetni rezultati algoritma neuronske mreže u pogledu točno predviđenih slučajeva/osobnih automobila u kategoriji “pad" – 37,2 % u okviru trening uzorka, odnosno 38,2 % u okviru testnog uzorka - značajno ispod minimalne vrijednosti od 70 % koja je definirana kao zadovoljavajuća, odnosno da su predmetni rezultati značajno ispod vrijednosti postignutih uporabom Exhaustive CHAID algoritma, alternativni prediktivni model se odbacuje.

Nadalje, u kontekstu analize učinkovitosti prediktivnog modela temeljenog na neuronskoj mreži, izračunat je MCC za trening i testni uzorak, a rezultati su prikazani u tablici 4.47.

Tablica 4.47. – Prikaz vrijednosti MCC za Opel Astra (neuronska mreža)

Uzorak	Kriterij	Vrijednost
Trening	pogrešno pozitivni	4.662
	pogrešno negativni	1.696
	stvarno pozitivni	25.823
	stvarno negativni	2.766
	MCC	0,37
Testni	pogrešno pozitivni	3.679
	pogrešno negativni	1.293
	stvarno pozitivni	21.084
	stvarno negativni	2.270
	MCC	0,39

Vrijednosti korelacijskog koeficijenta MCC za oba uzorka (trening i testni uzorak) su na razini srednjih pozitivnih vrijednosti +0,37 i +0,39 što daje obilježje dobre predikcije.

5. VALIDACIJA MODELA

U skladu s točkom 4.2. ovog rada u okviru Exhaustive CHAID algoritma u kontekstu validacije prediktivnog modela korištena je metoda dijeljenja uzorka (Split sample validation) koja se uobičajeno koristi u slučajevima kada postoje velike skupine podataka koji se obrađuju. U okviru ove metode validacije prediktivni model se generira na temelju trening uzorka te se testira pomoću tzv. potvrđujućeg uzorka (holdout sample).

U okviru tablice 5.1. dana je usporedba rezultata primijenjenih prediktivnih modela, odnosno usporedba njihove uspješnosti u kontekstu prediktivnog potencijala. Sve zabilježene vrijednosti postotaka točnosti predikcije tehnički neispravnih osobnih automobila (odnosno predviđenih slučajeva/osobnih automobila u kategoriji “pad"), nakon izvršene radnje za poboljšanje prediktivnog modela, u okviru Exhaustive CHAID algoritma, su iznad minimalne vrijednosti od 70 % koja je definirana kao zadovoljavajuća, čime je dokazan zadovoljavajući prediktivni potencijal predmetnog prediktivnog modela.

Istovremeno, utvrđena je visoka razina podudarnosti (praktički se radi o identičnim vrijednostima) postotaka točnosti predikcije trening uzorka i testnog uzorka što dokazuje izbalansiranost ukupnog skupa podataka te stabilnost prediktvinog modela.

Tablica 5.1. – Usporedba uspješnosti prediktivnih modela

Osobni automobil	Uzorak	Postotak točnosti predikcije Exhaustive CHAID algoritma	MCC (E.CHAID)	Postotak točnosti predikcije neuronske mreže	MCC (neuronska mreža)
VW Golf	Trening	**70,6 %**	**0,38**	**37,2 %**	**0,40**
	Test	**70,3 %**	**0,37**	**38,2 %**	**0,39**
Fiat Punto	Trening	**74,6 %**	**0,36**	**42,2 %**	**0,30**
	Test	**73,7 %**	**0,36**	**41,6 %**	**0,29**
Opel Astra	Trening	**75,7 %**	**0,45**	**37,2 %**	**0,37**
	Test	**75,7 %**	**0,44**	**38,2 %**	**0,39**

Također, zabilježene vrijednosti MCC korelacijskog koeficijenta na razini srednjih pozitivnih vrijednosti pokazuju dobru predikciju prediktivnog modela koji se temelji na Exhaustive CHAID algoritmu.

Za razliku od prediktivnog modela koji se temelji na Exhaustive CHAID algoritmu, prediktivni model koji se temelji na algoritmu neuronske mreže ne pokazuje dobar prediktivni potencijal s obzirom da su zabilježene vrijednosti postotaka točnosti predikcije tehnički neispravnih osobnih automobila (odnosno predviđenih slučajeva/osobnih automobila u kategoriji “pad") značajno ispod minimalne vrijednosti od 70 % koja je definirana kao granično zadovoljavajuća i predstavlja primarni kriterij za ocjenu uspješnosti prediktivnih modela, dok sekundarni kriterij predstavlja vrijednost korelacijskog koeficijenta MCC. Upravo navedena činjenica predstavlja razlog zbog čega se prediktivni model koji se temelji na algoritmu neuronske mreže odbacuje, iako isti, s obzirom na sekundarni kriterij (MMC za neuronsku mrežu), zadovoljava.

Što se tiče eventualnog provođenja dodatne validacije prediktivnog modela na način da se prognozirani rezultati tehničkih pregleda predmetnih tipova osobnih automobila (na temelju podataka iz 2010. godine), u kontekstu tehničke neispravnosti, uspoređuju sa stvarnim podacima iz prve sljedeće godine (2011.), procijenjeno je da takva dodatna validacija nije potrebna.

Razlog takvoj procjeni leži u činjenici da su promjene u strukturi ukupnog skupa vozila odnosno osobnih automobila (u smislu fluktuacije) na godišnjoj razini male, te da značajno ne utječu na strukturu rezultata koji proizlaze iz predmetnog prediktivnog modela. Promjene u strukturi ukupnog skupa vozila odnosno osobnih automobila podrazumijevaju realnu situaciju kada zbog čimbenika na koje nije moguće utjecati, u konkretnom slučaju, dolazi do priliva novih (ne nužno u smislu godine proizvodnje nego i u smislu pristupanja tehničkom pregledu radi registracije nakon određenog razdoblja neregistriranosti) automobila, odnosno dolazi do odliva određenog broja automobila (što iz razloga dotrajalosti, odnosno “pada” na tehničkom pregledu, što iz drugih razloga). Međutim, predmetne promjene u veličini osnovnog skupa i podskupova u pravilu nisu velike, osim ako se ne radi o nekim okolnostima obično uzrokovanih vanjskim čimbenicima (npr. promjene u zakonskoj i podzakonskoj legislativi). Stoga opisane situacije u pravilu nemaju presudan utjecaj na eventualni značajan poremećaj u strukturi podataka koji bi doveo do kompromitiranja razrađenog prediktivnog modela.

Prethodno iznesenu tezu potvrđuju podaci iz tablice 5.2., prema [69] i [70], u kojoj je dana usporedba skupova podataka u kontekstu tehničkih pregleda vozila iz 2010.

godine (godina iz koje su se podaci o rezultatima tehničkih pregleda osobnih automobila koristili za izradu prediktivnog modela), te iz 2011. godine.

Tablica 5.2. – Usporedba skupova podataka

REDOVNI TEHNIČKI PREGLEDI VOZILA	**2010. GODINA** [broj vozila/automobila]	**2011. GODINA** [broj vozila/automobila]	**RAZLIKA**	
			broj	postotak
I	II	III	IV	
Vozila svih kategorija koja su pristupila tehničkom pregledu	**1.822.105**	**1.848.172**	**+ 26.067**	**+ 1,4 %**
Vozila svih kategorija koja su "pala" na tehničkom pregledu	**380.820**	**375.918**	**- 4.902**	**-1,3 %**
Osobni automobili koji su pristupili tehničkom pregledu	**1.390.055**	**1.404.792**	**- 14.737**	**- 1,1 %**
Osobni automobili koji su "pali" na tehničkom pregledu	**311.789**	**310.663**	**- 1,126**	**- 0,4 %**

6. INTERPRETACIJA REZULTATA

U tablici 6.1. dana je usporedba rezultata istraživanja u smislu usporedbe nezavisnih varijable s prediktivnim potencijalom glede “pada” osobnih automobila na tehničkom pregledu, odnosno glede utvrđivanja tehničke neispravnosti, te postotnih iznosa točnosti prediktivnih kategorija u okviru trening i testnog uzorka, nakon provedene radnje za poboljšanje prediktivnog modela temeljenog na Exhaustive CHAID algoritmu.

Kod sva tri tipa osobnih automobila dvije nezavisne varijabla s najvećim prediktivnim potencijalom (nezavisne varijable poredane su prema veličini prediktivnog potencijala, od najvećeg na vrhu do najmanjeg na dnu, po pojedinom tipu osobnog automobila) jesu „starost_god“ (broj godina osobnog automobila proteklih od godine njegove proizvodnje do 2010. godine) te „index_km“ (ukupno prijeđeni putevi (kilometri) osobnog automobila). Ovakav rezultat je očekivan s obzirom da je logično pretpostaviti da su starija vozila u pravilu izloženija tehničkim neispravnostima s obzirom da su dulje vremena u eksploataciji, odnosno više godina eksploatacije osobnih automobila u pravilu podrazumijeva i više prijeđenih kilometara.

Sljedeća nezavisna varijabla s najvećim prediktivnim potencijalom (kod osobnih automobila VW Golf i Opel Astra) jest „vr_motora“ (vrsta motora s obzirom na pogonsko gorivo i vrstu ispušnog sustava), te „index_stp“ (veličina sredine u kojoj se nalazi STP u kojoj je izvršen tehnički pregled osobnih automobila) čija specifičnost proizlazi iz činjenice da se ne radi o “egzaktnoj” varijabli koja se direktno ne odnosi na neku od značajki osobnih automobila (kao što su, na primjer, snaga, starost ili broj prijeđenih kilometara), već je riječ o varijabli koja implicira određene okolnosti glede eksploatacije osobnih automobila, a koje u konačnici dovode do povećane mogućnosti za stvaranje tehničkih neispravnosti na osobnom automobilu, utemeljene na dvjema logičkim presumpcijama.

Prva logička presumpcija odnosi se na različite režime vožnje osobnih automobila uvjetovane različitim rješenjima prometne infrastrukture, koja su pak uvjetovana različitim veličinama sredina.

Tablica 6.1. – Usporedba rezultata istraživanja između obrađenih osobnih automobila

Osobni automobili – predmet istraživanja	**Uključene prediktivne (nezavisne) varijable izložene prema visini prediktivnog potencijala**	**Isključene prediktivne (nezavisne) varijable**	**Prediktivne kategorije**	
			"PAD" Udjel (postotni iznos i broj automobila)	**"PROLAZ" Udjel (postotni iznos)**
VW Golf	„starost" osobnog automobila (starost_god) ukupno prijeđeni putevi (kilometri) osobnog automobila (index_km) vrsta motora s obzirom na pogonsko gorivo, odnosno vrstu ispušnog sustava (vr_motora) snaga motora osobnog automobila (snaga) vrsta kočionog sustava ugrađenog u osobni automobil (vr_kocnica) veličina sredine u kojoj se nalazi STP u kojoj je izvršen tehnički pregled osobnog automobila (index_stp) radni obujam motora automobila (ccm)	-	**70,6 % 3.960 automobila** **(70, 3 % 22.759 automobila**)	70,1 % (70,0 %)
Fiat Punto	ukupno prijeđeni putevi (kilometri) osobnog automobila (index_km) „starost" osobnog automobila (starost_god) veličina sredine u kojoj se nalazi STP u kojoj je izvršen tehnički pregled osobnog automobila (index_stp) radni obujam motora osobnog automobila (ccm) snaga motora osobnog automobila (snaga)	vrsta kočionog sustava ugrađenog u osobni automobil (vr_kocnica), vrsta motora s obzirom na pogonsko gorivo, odnosno vrstu ispušnog sustava (vr_motora)	**74,6 % 8.087 Automobila** **(73,7 % 3.322 automobila)**	64,9 % (65,6 %)
Opel Astra	„starost" osobnog automobila (starost_god) ukupno prijeđeni putevi (kilometri) osobnog automobila (index_km) vrsta motora s obzirom na pogonsko gorivo, odnosno vrstu ispušnog sustava (vr_motora) veličina sredine u kojoj se nalazi STP u kojoj je izvršen tehnički pregled osobnog automobila (index_stp)	vrsta kočionog sustava ugrađenog u osobni automobil (vr_kocnica), radni obujam motora osobnog automobila (ccm), snaga motora osobnog automobila (snaga)	**75,7 % 1.695 Automobila** **(ukupno 11.252 automobila)**	75,4 % (74,6 %)

Naime, nedvojbena je činjenica da veličina sredine u kojoj se nalazi STP u kojoj je obavljen tehnički pregled predmetnog osobnog automobila (skupina 1 – velike urbane sredine: Zagreb, Split, Osijek, Rijeka, skupina 2 – urbane sredine: Zadar, Pula, Slavonski Brod, Dubrovnik, Varaždin, Karlovac, Šibenik, skupina 3 – male sredine: ostale sredine) uvjetuje specifična rješenja prometne infrastrukture.

Ta specifična rješenja prometne infrastrukture podrazumijevaju da veća sredina u odnosu na manje sredine ima veći broj raskrižja sa vertikalnom svjetlosnom signalizacijom (semaforima), veći broj kružnih tokova, veći broj pješačkih prelaza, veći broj cestovnih traka rezerviranih za javni prijevoz, itd.

Nadalje, također je nedvojbena činjenica da takva rješenja prometne infrastrukture utječu na način vožnje (tzv. gradski i izvangradski režimi vožnje) u smislu češće izmjene stupnjeva prijenosa snage motora (veći eksploatacijski utjecaj na elemente prijenosa snage osobnog automobila), češćeg kočenja (veći eksploatacijski utjecaj na kočioni sustav osobnog automobila) te veće potrošnja goriva (veći eksploatacijski utjecaj na ispušni sustav osobnog automobila).

U konačnici, iz prethodno navedenih činjenica proizlazi presumpcija da režimi vožnje koji podrazumijevaju veći eksploatacijski utjecaj pojedinih uređaja/sklopova osobnog automobila za posljedicu mogu imati stvaranje većeg potencijala za nastanak tehničkih neispravnosti na osobnom automobilu, koji uvjetuju ne-prolazak (“pad”) osobnog automobila na tehničkom pregledu.

S druge pak strane, druga logička presumpcija u pogledu predmetne varijable odnosi se na očekivanu kvalitetu i intenzitet redovnog održavanja vozila u pojedinim sredinama, uzimajući u obzir ekonomski aspekt, odnosno financijski potencijal vlasnika/korisnika vozila. Naime, za očekivati je da u većim sredinama, koje su ekonomski razvijenije, vlasnici/korisnici vozila u prosjeku kvalitetnije i intenzivnije održavaju osobne automobile u kontekstu tehničke ispravnosti, s obzirom da isti u prosjeku imaju veći financijski potencijal nego vlasnici/korisnici osobnih automobila u manjim sredinama.

Glede oba slučaja, kako je prethodno navedeno, polazi se od pretpostavke da se osobni automobili koji su obavili tehnički pregled u određenoj sredini dominantno

koriste u smislu vožnje upravo unutar te sredine (prema podacima *HAK*-a [62], više od 95 % osobnih automobila obavlja tehnički pregled unutar sredine svog registarskog područja – npr., više od 95 % osobnih automobila zagrebačkih registarskih tablica obavlja tehnički pregled u nekoj od STP iz područja Grada Zagreba i Zagrebačke županije).

Sljedeće nezavisne varijable s određenim prediktivnim potencijalom jesu „vr_kocnica“ (vrsta kočionog sustava ugrađenog u osobni automobil) (kod osobnih automobila VW Golf), te „ccm“ (radni obujam motora osobnih automobila) i „snaga“ (snaga motora osobnih automobila). Logična je povezanost radnog obujma i snage motora, odnosno podjednak prediktivni potencijal s obzirom da se radi o varijablama koje su, u pravilu, direktno proporcionalne, odnosno veći radni obujam motora podrazumijeva i veću snagu osobnog automobila.

U tablici 6.2. prikazana je pojednostavljena usporedba kritičnih aspekata osobnih automobila, u kontekstu “pada” na tehničkom pregledu, između obrađenih osobnih automobila, koja proizlazi iz rezultata istraživanja iznijetim u točkama 4.4.1., 4.4.2. i 4.4.3. ovog rada.

Tablica 6.2. – Usporedba kritičnih nezavisnih varijabli u kontekstu "pada" osobnih automobila na tehničkom pregledu

	Trgovačka marka i tip osobnog automobila – predmet istraživanja					
	VW Golf		**Fiat Punto**		**Opel Astra**	
Kritične nezavisne varijable	načelno svi osobni automobili „stari" 20 i više godina (po svim nezavisnim varijablama)	**2.946 automobila** (17.174 automobila)	načelno svi osobni automobili koji su prešli više od 150.000 km (po svim prediktivnim varijablama)	**5.042 automobila** (2.097 automobila)	načelno svi osobni automobili „stari" 15 i više godina (po svim nezavisnim varijablama)	921 automobila (5.323 automobila
	osobni automobili „stari" 18 i 19 godina, koji su prešli više od 150.000 km, neovisno o snazi motora	**640 automobila** (3.587 automobila)	osobni automobili koji su prešli između 100.000 km i 150.000 km, te koji su „stari" 6 godina i manje	**175** automobila (71 automobil)	osobni automobili „stari" 11 do 14 godina, uključujući 14 godina, koji su prešli više od 150.000 km	**543** automobila (2.984 automobila
	osobni automobili „stari" 16 i 17 godina, koji su prešli više od 150.000 km, te koji imaju benzinski (Ottov) motor i ugrađeni regulirani katalizator (REG KAT), odnosno koji imaju benzinski (Ottov) motor bez katalizatora (BEZ KAT)	**189 automobila** (1.046 automobila)	osobni automobili koji su prešli između 100.000 km i 150.000 km, koji su „stari" 7 do 10 godina, uključujići 10 godina, te koji imaju motor radnog obujma većeg od 1.248 cm^3	**157** automobila (52 automobila)	osobni automobili „stari" 9 godina, koji su prešli više od 100.000 km	**90** automobila (482 automobila

Kritične nezavisne varijable	osobni automobili „stari“ 14 i 15 godina, koji su prešli manje od 100.000 km,	**71 automobil** (442 automobila)	osobni automobili koji su prešli između 100.000 km i 150.000 km, koji su „stari“ 11 do 14 godina, uključujući 14 godina, te kod kojih je tehnički pregled obavljen u stanici za tehnički pregled vozila koja se nalazi u velikoj urbanoj sredini, odnosno u malim sredinama	**1.333** automobila (562 automobila)	osobni automobili „stari“ 8 godina, koji su prešli više od 100.000 km, te koji imaju benzinski (Ottov) motor i ugrađeni regulirani katalizator (REG KAT)	**65** automobila (381 automobila)
	osobni automobili „stari“ 13 godina, koji su prešli više od 150.000 km, te koji imaju klasičan kočni sustav	**114 automobila** (510 automobila)	osobni automobili koji su prešli između 100.000 km i 150.000 km, koji su „stari“ 15 godina, te koji imaju motor radnog obujma većeg od 1.108 cm^3	**259** automobila (94 automobila)	osobni automobili „stari“ 7 godina, koji imaju benzinski (Ottov) motor i ugrađeni regulirani katalizator (REG KAT)	**76** automobila (387 automobila)
			osobni automobili koji su prešli između 100.000 km i 150.000 km, koji su „stariji“ od 15 godina, neovisno o sredini u kojoj je obavljen tehnički pregled	**181** automobil (87 automobila)		
			osobni automobili koji su prešli između 50.000 km i 100.000 km, koji su „stari“ 12 godina, neovisno o radnom obujmu motora	**389** automobila (155 automobila)		

			osobni automobili koji su prešli između 50.000 km i 100.000 km, koji su „stari“ 13 i 14 godina, neovisno o snazi motora	**365** automobila (121 automobil)		
			osobni automobili koji su prešli između 50.000 km i 100.000 km, koji su „stariji“ od 14 godina, te kod kojih je tehnički pregled obavljen u stanici za tehnički pregled vozila koja se nalazi u velikok urbanoj sredini	**126** automobila (60 automobila)		
			osobni automobili koji su prešli do 50.000 km, te koji su „stariji“ od 12 godina	**60** automobila (23 automobila)		
	VW GOLF 70,6 % (70,3 %)	**∑ 3.960 automobila (∑22.759 automobila)**	**FIAT PUNTO 74,6 % (73,7 %)**	**∑ 8.087 automobila (∑3.322 automobila)**	**OPEL ASTRA 75,7 % (75,7 %)**	**∑ 1.695 automobila (∑9.557 automobila)**

Iz tablica 6.1. i 6.2. proizlazi da prediktivni model primijenjen u ovom radu, u slučaju kada su poznate karakteristike osobnih automobila u smislu četiri (Opel Astra), pet (Fiat Punto), odnosno šest determiniranih prediktivnih (nezavisnih) varijabli (VW Golf), vrši točno predviđanje “pada” na tehničkom pregledu uslijed utvrđene tehničke neispravnosti u 70,6 % slučajeva za osobne automobile VW Golf, u 74,6 % slučajeva za osobne automobile Fiat Punto, te u 75,7 % slučajeva za osobne automobile Opel Astra.

6.1. OGRANIČENJA MODELA

Kao što je navedeno u točci 2.1. ovog rada u legislativi RH koja regulira područje tehničkih pregleda vozila propisano je da vozilo (osobni automobil), kod kojeg je tehnička neispravnost utvrđena na bilo kojem od 134 dijelova uređaja/sklopova, osim na parkirnim svijetlima, svijetlima registracijske pločice i žutim rotacijskim ili treptavim svijetlima (drugim riječima, praktički na svim dijelovima uređaja/sklopova koji se na vozilu/osobnom automobilu kontroliraju), za posljedicu ima "pad" na tehničkom pregledu.

Problem koji proizlazi iz prethodno navedenih činjenica jest u tome što predmetna legislativa ne definira dovoljno precizno pojam tehničke neispravnosti, odnosno ne predviđa kategorizaciju (stupnjevanje) tehničkih neispravnosti. Upravo u ovom aspektu legislativa RH koja regulira područje tehničkih pregleda vozila nije usklađena s legislativom Europske unije.

Nekoliko je negativnih posljedica aktualnog legislativnog rješenja. Nelogično je na isti način u represivnom smislu ("pad" vozila na tehničkom pregledu) tretirati sve potencijalno utvrđene tehničke nedostatke na vozilu (koji se općenito smatraju tehničkim neispravnostima) tijekom procesa tehničkog pregleda – nedostatke u kontekstu aktivne sigurnosti, nedostatke u kontekstu pasivne sigurnosti, nedostatke u kontekstu ekološke podobnosti, nedostatke u kontekstu općeg stanja – neovisno o razini stvarnog utjecaja tih tehničkih nedostataka na vozilu na sigurnost u prometu.

Pa tako, po analogiji ovakvog legislativnog rješenja, jednaku težinu u smislu kriterija za utvrđivanje pojedinog vozila tehnički neispravnim (što za posljedicu ima "pad" na tehničkom pregledu) imaju, na primjer, nedostatak u kontekstu pasivne sigurnosti - izolirano (minimalno) napuknuće stražnjeg stakla, te nedostatak u kontekstu aktivne sigurnosti - neadekvatan koeficijent kočenja, ili nedostatak u kontekstu općeg stanja - izolirano područje korozije na blatobranu kotača, te nedostatak u kontekstu aktivne sigurnosti – korozija i zračnost u ležajevima glavine kotača, ili nedostatak u kontekstu opreme – nepostojanje sigurnosnog trokuta, te nedostatak u kontekstu ekološke podobnosti – nepostojanje *lambda* sonde, itd.

Stoga je u sustav utvrđivanja tehničkih neispravnosti na vozilu ključno uvesti kriterij u smislu procjene razine stvarnog utjecaja određene tehničke neispravnosti na vozilu na sigurnost u prometu, odnosno uvesti kategorizaciju (stupnjevanje) tehničkih neispravnosti. Na taj način bi se izbjegla situacija da utvrđivanje određene tehničke neispravnosti na vozilu, koja realno uopće nema ili nema značajan stvarni utjecaj na sigurnost u prometu, za posljedicu ima "pad" osobnog automobila na tehničkom pregledu. Nadalje, na taj način bi se omogućio objektivan uvid u stvarnu strukturu utvrđenih tehničkih neispravnosti na vozilu jer bi nadzornici tehničke ispravnosti vozila (osoblje STP koje provodi kontrolu tehničke ispravnosti vozila) evidentirali sve stvarno utvrđene tehničkih neispravnosti na vozilu tijekom tehničkom pregleda, što trenutno često nije slučaj.

Naime, jedan od sustavnih problema u radu STP na koji *HAK* tijekom provođenja stručnih nadzora nad radom STP obraća posebnu pozornost jest činjenica da nadzornici tehničke ispravnosti vozila često hotimice izbjegavaju evidentirati sve utvrđene tehničke neispravnosti na vozilu svjesni činjenice da će bilo koja utvrđena tehnička neispravnost na vozilu (neovisno o razini njezinog stvarnog utjecaja na sigurnost u prometu) za posljedicu imati "pad" osobnog automobila na tehničkom pregledu.

Proizlazi da analiza statističkih podataka dobivenih u okviru aktualne metodologije tehničkih pregleda vozila trenutno ne može dovesti do adekvatnih zaključaka o stvarnom stanju ovog segmenta sigurnosti prometa na cestama. Također, hrvatski statistički podaci o, na primjer, prosječno zabilježenom postotku tehničke ispravnosti vozila u sustavu tehničkih pregleda vozila (koji je prema [69] za 2010. godinu 20,9 %, te prema [70] za 2011. godinu 20,3 %) nisu usporedivi sa identičnim statističkim podacima većine država članica Europske unije upravo iz razloga što su u prikupljanu i obradi istih korištene različite metodologije.

Iz navedenih razloga, u ovom trenutku, u prediktivni model nije bilo moguće uvrstiti varijable koje bi predstavljale same tehničke neispravnosti, a čije je utvrđivanje na osobnom automobilu tijekom tehničkog pregleda dovelo do "pada" osobnog automobila na tehničkom pregledu te je, u ovom trenutku, istraživanje ograničeno na

uvrštavanje u prediktivni model varijable koja predstavlja krajnji rezultat tehničkog pregleda (*tp_prolaz),* ne ulazeći u strukturu tehničkih neispravnosti.

Dodatno ograničenje u ovom istraživanju predstavlja evaluacija nezavisne (prediktivne) varijable *index_stp,* u smislu utjecaja na tehničku neispravnost osobnih automobila*.* Naime, kao što je to prethodno obrazloženo postoje nezanemarivi logički argumenti za presumpciju utjecaja predmetne varijable na tehničku neispravnost osobnih automobila, no ne i egzaktni podaci utemeljeni na istraživanju.

6.2. PRIJEDLOG DALJNJEG ISTRAŽIVANJA

Potencijal izgrađenog modela predikcije tehničke ispravnosti osobnih automobila temeljem rezultata tehničkih pregleda, koji je odabran na temelju rezultata istraživanja provedenog u ovom radu, jest velik. Da bi taj potencijal bio iskorišten do maksimuma neophodno je da dođe do stvaranja preduvjeta za otklanjanje čimbenika koji isti ograničavaju, o čemu je bilo riječi u prethodnoj točci.

U tom kontekstu, upravo uvođenjem nove metodologije utvrđivanja tehničkih neispravnosti, u skladu sa modelom kategorizacije (stupnjevanja) tehničkih neispravnosti primijenjenim u legislativi Europske unije, značajan broj nedostataka manjeg utjecaja na sigurnost vozila više ne bi bio tretiran kao razlog za “pad” vozila na tehničkom pregledu, te bi na taj način u velikoj mjeri nestao razlog za manipulaciju tehničkim pregledom od strane nadzornika tehničke ispravnosti vozila. Oni bi sukladno propisanoj proceduri predmetne nedostatke uredno evaluirali i evidentirali, vlasnik/korisnik vozila bi dobio informaciju o istima i preporuku o njihovom otklanjanju, bez represivne mjere u vidu “pada” vozila na tehničkom pregledu.

Na ovaj način bi se dobila mnogo realnija slika o stvarnom stanju tehničke ispravnosti vozila, u kontekstu tehničkih pregleda, u RH. Nadalje, na ovaj način bi hrvatski statistički podaci o rezultatima tehničkih pregleda vozila bili u potpunosti usporedivi sa identičnim statističkim podacima većine država članica Europske unije iz razloga što bi u prikupljanu i obradi istih bile korištene identične metodologije.

U konačnici, na ovaj način bi se otklonila postojeća ograničenja u kontekstu primjene modela predikcije tehničke ispravnosti osobnih automobila temeljem rezultata tehničkih pregleda, čije bi područje primjene bilo prošireno uvođenjem rednih/ordinalnih varijabli tehničke neispravnosti osobnih automobila, u smislu uređaja/sklopova na kojima je utvrđen nedostatak. Navedeno bi za posljedicu imalo unaprjeđenje postojećeg prediktivnog modela koji bi pokazao zakonitosti, odnosno korelacije između podataka koji proizlaze iz općih tehničkih značajki osobnih automobila (podaci o snazi i radnom obujmu motora osobnih automobila, vrsti pogonskog goriva i ispušnog sustava, vrsti kočionog sustava, prijeđenim kilometrima), podataka dobivenih kontrolom i ispitivanjem tehničke ispravnosti osobnih automobila tijekom tehničkog pregleda (podaci o stanju svih 17 cjelina, odnosno uređaja/sklopova na osobnom automobilu), te podataka o konačnom rezultatu tehničkog pregleda (podaci o "prolazu" ili "padu", te o pretežnom režimu vožnje).

Potrebno je naglasiti da su aktivnosti u smjeru usklađivanja metodologije tehničkih pregleda i kriterija za ocjenjivanje tehničke ispravnosti vozila s legislativom Europske Unije u tijeku. Naime, Vlada RH utvrdila je mjere u tom smjeru u okviru Nacionalnog programa sigurnosti cestovnog prometa Republike Hrvatske, u razdoblju 2011.-2020. godina [71]. Implementacija ovih mjera utvrđena je kao prioritetna, a odgovorni za njihovo provođenje su upravo *CVH* i *HAK* (autor ovog rada je dio radnog tima CVH/*HAK* za pripremu implementacije).

Nadalje, rezultati ovog istraživanja mogu se nadograditi s rezultatima projekta uvođenja obveznog tehničkog pregleda za vozila koja su sudjelovala i koja su oštećena u teškim prometnim nesrećama (prometne nesreće sa smrtnom posljedicom). Naime ovaj projekt je također utvrđen u okviru [71], uz implementaciju predviđenu u dugoročnom razdoblju, a odgovorni za njegovo provođenje su, osim *CVH* i *HAK*-a, i *Ministarstvo nadležno za promet*, *Ministarstvo unutarnjih poslova* i *Hrvatski ured za osiguranje*.

Temeljni cilj ovog projekta je da se iz podataka dobivenih iz obveznog tehničkog pregleda za vozila koja su sudjelovala u prometnim nesrećama u kojima je bilo poginulih osoba formira registar o broju prometnih nesreća u kojima je uzrok prometne nesreće bilo tehnički neispravno vozilo. Registar bi također trebao sadržavati i

rezultate tehničkog pregleda vozila kako bi se dobio kompletan uvid koji su sustavi na vozilu najčešći uzročnici prometnih nesreća u slučajevima kada je čimbenik vozilo odgovoran za nastanak prometne nesreće.

Model predikcije tehničke ispravnosti osobnih automobila temeljem rezultata tehničkih pregleda, koji je opisan i primijenjen u okviru istraživanja ovog rada, dodatno bi mogao biti unaprijeđen inkorporiranjem nezavisnih varijabli koje bi podrazumijevale podatke o samom vlasniku/korisniku, odnosno vozaču osobnog automobila, a pogotovo podatke iz kojih bi se mogle iskristalizirati dodatne informacije o njegovom uobičajenom načinu vožnje, te načinu održavanja vozila.

Informacije o vozačevom uobičajenom načinu upravljanja osobnim automobilom mogle bi biti dobivene iz podataka o godini kada je isti osposobljen za upravljanje osobnim automobilom te podatka o vozačevoj starosti (navedeni podaci impliciraju iskustvo u vožnji), podataka o počinjenim prekršajima iz konteksta upravljanja osobnim automobilom (potrebna uvezanost baze podataka tehničkih pregleda i baze podataka *MUP*-a o prekršajima vozača), podataka o stručnoj spremi i broju djece (navedeni podaci impliciraju razinu agresivnosti vožnje), podataka o potencijalnoj *Ecowill* vožnji (režim upravljanja osobnim automobilom sukladan europskim standardima za ekološki podobnu vožnju), itd.

Informacije o vozačevom načinu održavanja osobnog automobila mogle bi biti dobivene iz podataka koji se nalaze u servisnoj knjižici osobnog automobila, a i izostanak takvih podataka (odnosno nepostojanje servisne knjižice) također bi predstavljao podatak o načinu održavanja osobnog automobila.

Predmetni podaci bi se obvezno prikupljali tijekom tehničkog pregleda vozila te evidentirali u okviru baze podataka jedinstvenog informatičkog programskog rješenja za obradu podataka o tehničkim pregledima i registraciji vozila u stanicama za tehnički pregled vozila.

U konačnici, model predikcije tehničke ispravnosti osobnog automobila temeljem rezultata tehničkih pregleda, koji je odabran na temelju rezultata istraživanja provedenog u ovom radu, po analogiji primjenjiv je ne samo za osobne automobile, nego za sve vrste/kategorije vozila koja obvezno pristupaju tehničkom pregledu.

7. ZAKLJUČAK

Baza podataka koja proizlazi iz jedinstvenog programskog rješenja za obradu podataka o tehničkim pregledima i registraciji vozila u stanicama za tehnički pregled vozila, koju koriste *MUP*, *CVH* i *HAK*, svaki u okviru svojih ingerencija kako je to opisano u točci 2. ovog rada, objedinjuje tehničke podatke o cca 2.000.000 vozila koja su podvrgnuta tehničkom pregledu i registrirana u RH. Ako se uzme u obzir da je za svako pojedino vozilo vezano nekoliko desetaka podataka, jasno je o kojem redu veličine baze podataka se radi.

Do sada je ova baza podataka obrađivana isključivo osnovnim statističkim metodama, primarno u svrhu statističkih izvještajnih analiza proteklih razdoblja. Upravo iz tog razloga, a s obzirom na svoj golemi potencijal u smislu generiranja visoko vrijednih informacija koje mogu poslužiti za analiziranje *pro futuro*, odnosno u prediktivnom kontekstu, ova baza podataka idealna je da se na njoj primjeni pristup poslovne inteligencije, te kroz proces rudarenja podataka dođe do željenih informacija.

U okviru istraživanja koje je opisano u ovom radu dokazan je potencijal primjenjivosti prediktivnog modela tehničke neispravnosti osobnih automobila koji se temelji na metodi stabla odlučivanja i Exhaustive CHAID algoritmu, a na osnovi rezultata tehničkih pregleda, uzimajući u obzir sve subjekte koji bi trebali imati koristi od istog (navedeni u nastavku ovog zaključka), kako iz aspekta sigurnosti prometa na cestama tako i iz aspekta gospodarstva (prodaja vozila). Iako se predmetni model odnosi na predikciju tehničke neispravnosti osobnih automobila (isti je strukturiran temeljem obrade kritičnih trgovačkih marki i tipova osobnih automobila glede tehničke neispravnosti utvrđene tijekom tehničkih pregleda), s obzirom na univerzalnost načela na kojem se predmetni model zasniva, potencijal za njegovu primjenu nije ograničen samo na vrstu vozila u koju spadaju osobni automobili, već se analogan model može modelirati u kontekstu predikcije tehničke neispravnosti ostalih vrsta cestovnih vozila.

Povratne informacije proizašle iz prediktivnog modela koji je predmet ovog istraživanja podrazumijevaju spoznaju, u prognostičkom smislu (uz zadovoljavajući postotni iznos točnosti prediktivnih kategorija), o očekivanom broju tehnički neispravnih osobnih automobila/vozila na tehničkom pregledu, u ovisnosti o trgovačkoj marki/tipu osobnog

automobila/vozila, snazi i radnom obujmu njegovog motora, prijeđenim kilometrima, vrsti motora s obzirom na pogonsko gorivo (odnosno vrsti ispušnog sustava), vrsti ugrađenog kočnog sustava, te u ovisnosti o pretpostavljenom dominantnom režimu vožnje proizašlim iz eksploatacijskog okruženja u infrastrukturnom smislu.

U okviru ovog istraživanja takve informacije dobivene su za sve tri prediktivne kategorije osobnih automobila, koja su utvrđena kao kritična po pitanju tehničkih neispravnosti detektiranih tijekom tehničkog pregleda.

Prediktivni model tehničke ispravnosti osobnih automobila koji se temelji na Exhaustive CHAID algoritmu zadovoljio je prema oba kriterija (primarnom i sekundarnom) definirana u točki 5. ovog rada a koji podrazumijevaju: točnost predikcije tehnički neispravnih osobnih automobila (odnosno predviđenih slučajeva/osobnih automobila u kategoriji “pad") višu od granične minimalne vrijednosti od 70 % (primarni kriterij) i pozitivnu vrijednost korelacijskog koeficijenta MCC (sekundarni kriterij). Alternativni prediktivni model, temeljen na algoritmu perceptronske neuronske mreže, nije zadovoljio po pitanju primarnog kriterija te je isti odbačen.

Dodatni razlog za odabir prediktivnog modela koji se temelji na Exhaustive CHAID algoritmu proizlazi iz činjenice da isti podrazumijeva lakše razumijevanje strukture poslovnih pravila nego što je to slučaj u alternativnom modelu s obzirom na kompleksnost matrica težina slojeva neuronske mreže.

Kao što je prethodno opisano u točci 1. ovog rada glavne resurse u dosadašnjim istraživanjima i analizama glede podataka o utjecaju vozila (osobnih automobila) na sigurnost cestovnog prometa, iz aspekta tehničke ispravnosti, predstavljali su sljedeći izvori:

- proizvođači vozila, odnosno ovlašteni servisi - u pogledu podataka o tehničkim neispravnostima na vozilima detektiranim tijekom servisiranja i homologacijskih ispitivanja vozila,

- nadležna tijela državne uprave u čijoj ingerenciji je sigurnost cestovnog prometa (policija, sudstvo/državno odvjetništvo, itd.) - u pogledu podataka o prometnim nesrećama vozila,

- subjekti koji provode poslove tehničke pomoći na cesti - u pogledu podataka o tehničkim neispravnostima na vozilima detektiranim tijekom pružanja tehničke pomoći na cesti.

Predmetni podaci nedvojbeno su relevantni kako po pitanju vjerodostojnosti tako i po pitanju volumena.

Međutim, činjenica jest da podaci o tehničkoj neispravnosti vozila (osobnih automobila) prikupljeni kroz proces tehničkog pregleda vozila (osobnih automobila), a posebno u kontekstu predikcije tehničke ispravnosti, daju sasvim novu dimenziju ne samo u pogledu prevencije sigurnosti cestovnog prometa nego i u pogledu određenih gospodarskih aspekata.

Naime, podaci prikupljeni od strane ovlaštenih servisa, koliko god bili značajni u pogledu svog volumena ipak su ograničeni s obzirom da se samo manji dio ukupnog voznog parka servisira na taj način (a posebno po isteku jamstvenog roka vozila). Nadalje, podaci prikupljeni od strane nadležnih tijela državne uprave u čijoj ingerenciji je sigurnost cestovnog prometa također su ograničeni s obzirom na relativno mali broj vozila koja sudjeluju u prometnim nesrećama u odnosu na ukupni broj vozila u prometu. U konačnici, svakako su ograničeni i podaci prikupljeni i od strane subjekata koji provode poslove tehničke pomoći na cesti, s obzirom da su predmetne usluge ograničene samo na vozila članova koji po vlastitom izboru plaćaju usluge tehničke pomoći na cesti.

Posebna vrijednost predmetnih podataka prikupljenih kroz proces tehničkog pregleda vozila, za razliku od prethodno navedenih predmetnih podataka, proizlazi iz činjenice da isti nisu ograničeni s obzirom da kompletan vozni park pojedine države, odnosno sva registrirana vozila, obvezno moraju biti podvrgnuta procesu tehničkog pregleda i što je posebno bitno, na način koji podrazumijeva potpuno istu metodologiju i iste vrste tehničkih (kontrolnih) ispitivanja u skladu s ujednačenim kriterijima.

Koristi od informacija glede predikcije tehničke ispravnosti osobnih automobila, odnosno putem primjene poslovnih pravila za pojedini tip automobila, a koje proizlaze iz primjene ovog modela, primarno bi trebali imati sljedeći subjekti:

- tijela državne uprave u čijoj je nadležnosti sigurnost cestovnog prometa – na način da predmetne informacije ciljano koriste u kontekstu poboljšanja postojeće legislative koja se odnosi na područje sigurnosti prometa na cestama, te za fokusiranje preventivno-sigurnosnih akcija koje promiču tehničku ispravnost osobnih automobila kao funkciju sigurnosti prometa na cestama,

- proizvođači osobnih automobila - na način da predmetne informacije koriste u cilju modificiranja procesa projektiranja osobnih automobila/vozila u cilju postizanja njihovih boljih sigurnosnih i ekoloških karakteristika posebno u pogledu višegodišnje eksploatacije. Na ovaj način proizvođači osobnih automobila/vozila ostvarili bi korist kako iz aspekta veće sigurnosti prometa na cestama, tako i iz aspekta prodaje osobnih automobila u marketinško-promidžbenom kontekstu,

- vlasnici/korisnici osobnih automobila - na način da predmetne informacije koriste u cilju što boljeg održavanja osobnih automobila. Također, potencijalni vlasnici osobnih automobila predmetne informacije mogu koristiti kao jedan od bitnijih kriterija za odabir pri kupnji istih,

- prodavači rabljenih osobnih automobila - ključni aspekt za optimalno proračunavanje troškova jamstva rabljenih automobila, a time posljedično i za optimalno modeliranje rokova jamstva, predstavlja razina vjerodostojnosti, odnosno kvaliteta podataka o tehničkim neispravnostima na automobilima/vozilima, u funkciji starosti i razine eksploatacije automobila/vozila, a izvor kojih su do sada bili isključivo ovlašteni servisi [16]. Proizlazi logičan zaključak da će osiguravanje novog izvora predmetnih podataka adekvatnog volumena i velike vjerodostojnosti (kvalitete) - što nepobitno jest tehnički pregled automobila - za posljedicu imati stvaranje preduvjeta za još preciznije proračunavanje troškova jamstva, odnosno kvalitetnije modeliranje rokova jamstva,

- nadzornici tehničke ispravnosti osobnih automobila (djelatnici STP-a koji neposredno provode tehnički pregled automobila) - na način da unaprijed (prije početka tehničkog pregleda predmetnog automobila), putem programske aplikacije za provođenje tehničkih pregleda, imaju upozornu informaciju o predikciji tehničke ispravnosti za predmetni automobil, te da posljedično mogu obratiti posebnu pozornost na određene segmente predmetnog automobila.

Prostor za povećanje prediktivnog potencijala modela izrađenog u okviru istraživanja koje je opisano u ovom radu postoji u kontekstu uvođenja nezavisnih varijabli koje predstavljaju tehničke neispravnosti osobnih automobila (u smislu uređaja/sklopova na kojima je utvrđen tehnički nedostatak), nakon što su iste kategorizirane (stupnjevane) sukladno procjeni glede razine njihovog stvarnog utjecaja na sigurnost prometa. Nadalje, prostor za povećanje prediktivnog potencijala postoji i u kontekstu povezivanja baze podataka o rezultatima tehničkih pregleda osobnih automobila/vozila s bazom podataka o vlasnicima/korisnicima osobnih automobila, te pronalaženjem zakonitosti između te dvije kategorije podataka.

Također, taj prostor postoji i u kontekstu povezivanja baze podataka o rezultatima (redovnih) tehničkih pregleda osobnih automobila s bazom podataka o tehničkim pregledima osobnih automobila koja su sudjelovala u teškim prometnim nesrećama.

Zasebna obrada baze podataka koje sadrže podatke o (tehničkim) karakteristikama "vozila", odnosno "vozača" za posljedicu svakako imaju dobivanje korisnih informacija. Međutim, u takvim slučajevima, najčešće je riječ o informacijama koje potvrđuju određene pretpostavke.

Integriranjem podataka o (tehničkim) značajkama "vozila" i karakteristikama "vozača" u jedinstvenu bazu, te obrada iste metodama rudarenja podataka za posljedicu bi imalo prikupljanje posebno vrijednih (nepoznatih) informacija u pogledu otkrivanja novih znanja, odnosno otkrivanja korelacija između određenih podataka za koje se prvotno uopće ne pretpostavlja da postoje.

Korištenje takvih informacija od strane prethodno navedenih strana nedvojbeno bi dovelo do značajnog podizanja razine sigurnosti u prometu na cestama.

8. LITERATURA

[1] M. Hafner, L.Poljančić: *Statistički pokazatelji rezultata tehničkog pregleda vozila, Stručni bilten broj 129,* Centar za vozila Hrvatske, Zagreb, 2009.

[2] *Council Directive 2009/40/EC* (OJ L 141, 6.6.2009, p. 1), EU, 2009.

[3] *Council Directive 2007/46/EC* (OJ L 263, 9.10.2007, p. 1), EU, 2007.

[4] *Cars - Strong & weak points*, Bilprovningen, ISSN 1402-3814, Sweden, 2004.

[5] *ADAC Breakdown statistics in 2007*, ADAC, IN 25214, Germany, 2008.

[6] *ADAC Breakdown statistics in 2008*, ADAC, IN 25797, Germany, 2009.

[7] *ADAC Breakdown statistics in 2009*, ADAC, IN 26387, Germany, 2010.

[8] *ADAC Breakdown statistics in 2010*, ADAC, IN 27008, Germany, 2011.

[9] *Roadworthiness Package, Comission staff working paper - Impact Assessment*, European Commisssion, SWD (2012) 206 final, Brussels, 2012.

[10] *AUTOFORE - Study on the Future Options for Roadworthiness Enforcement in the European Union*, CITA, Brussels, 2008.

[11] *MOT Scheme Evidence-base*, Department for Transport, UK, 2008.

[12] *Der statistick Bericht*, Deutschland Amtliche Verkehrsstatistiken, Germany, 2009.

[13] *Internationale Strategien zur Unfallvermeidung - Technische Sicherheit im Strassenverkehr*, DEKRA Fachschrift 58/05, DEKRA Automobil GmbH, Germany, 2005.

[14] *Motorcycle Road safety Report*, DEKRA Automobil GmbH, Germany, 2010.

[15] O. van Schoor, J.L. van Niekerk, B. Grobbelaar: *Mechanical failures as a contributing cause to motor vehicle accidents - South Africa*, Accident Analysis and Prevention, Vol.33, Issue 6, pp.713 - 721, 2001.

[16] C. S. Aksezer: *Failure analysis and warranty modeling of used cars*, Engineering Failure Analysis, Vol.16, Issue 6, pp.1520 - 1526, 2011.

[17] W.R. Blischke, E.M. Scheuer: *Calculation of the cost of warranty policies as a function of estimated life distributions*, Naval Research Logistics Quarterly, Vol.22, Issue 4, pp.681 - 696, 1975.

[18] D.G. Nguyen, D.N.P. Murthy: *A general model for estimating warranty costs for repairable products*, IIE Trans, Vol. 16, Issue 4, pp.379 - 386, 1984.

[19] W.R. Blischke, D.N.P. Murthy: *Product warranty management - A taxonomy for warranty policies*, European Journal of Operational Research, Vol. 62, Issue 2, pp.127 - 148, 1992.

[20] G.N. Chattopadhyay, W. Yun: *Modeling and analysis of warranty cost for 2D-policies associated with sale of second-hand products*, International Journal of Reliabilty, Quality and Safety Engineering , Vol.7, Issue 1, pp.71 - 77, 2006.

[21] G.N. Chattopadhyay, D.N.P. Murthy: *Modeling and analysis of warranty cost for 2D-policies associated with sale of second-hand products*, International Journal of Reliabilty, Quality and Safety Engineering , Vol.7, Issue 1, pp.71 - 77, 2006.

[22] *Zakon o sigurnosti prometa na cestama*, Narodne novine (67/2008, 74/2011), Republika Hrvatska.

[23] *Pravilnik o tehničkim pregledima vozila*, Narodne novine (148/2008,36/2010), Republika Hrvatska, 2008.

[24] *Directive 2014/45/EU of the European Parliament and of the Council of 3 April 2014 on periodic roadworthiness tests for motor vehicles and their trailers and repealing Directive 2009/40/EC* (OJ L 127, 29.4.2014, p. 1), EU, 2014.

[25] B. Liautaud, M.Hammond: *E-Business Intelligence. Turning Information into Knowledge into Profit,* McGraw-Hill, New York, USA, 2002.

[26] M. Petrini, M.Pozzebon: *What role is "Business Intelligence" playing in developing countries? A picture of Brazilian companies,* Data Mining Applications for Empowering Knowledge Societies, IGI Global, pp. 237-257, 2008.

[27] C. M. Olszak, E. Ziemba: *Business Intelligence as a Key to Managament of an Enterprise,* Informing Science Institute, Santa Rosa, USA, 2003.

[28] B. de Ville: *Decision Trees for Business Intelligence and Data Mining: Using SAS Enterprise Miner,* SAS Publishing, SAS Institute, Cary, USA, 2006.

[29] M. Petrini, M.Pozzebon: *Managing sustainability with the support of business intelligence: Integrating socio-environmental indicators and organisatioanl context,* The Journal od Strategic Information Systems, Vol. 18, Issue 4, pp. 178-191, 2009.

[30] Ž. Panian, G. Klepac: *Poslovna inteligencija,* Masmedia, Zagreb, 2003.

[31] Ž. Panian i suradnici: *Poslovna inteligencija: studije slučajeva iz hrvatske prakse,* Narodne novine d.d., Zagreb, 2007.

[32] G. Klepac, L. Mršić: *Poslovna inteligencija kroz poslovne slučajeve,* Lider press d.d./TIM press d.o.o., Zagreb, 2006.

[33] S.H. Liao, P.H. Chu, P.Y. Hsiao: *Data mining techniques and applications - A decade review from 2000 to 2011*, Expert Systems with Applications, Vol.39, Issue 12, pp.11303 - 11311, 2012.

[34] T. Hill, P. Lewicki: *Statistics - Methods and Applications,* Statsoft, Tulsa, USA, 1980.

[35] M. J .A. Berry, G. S. Linoff: *Data Mining Techniques for Marketing, Sales and Customer Relationship Management – Second Edition,* Wiley Publishing Inc., Indianapolis, USA, 2004.

[36] M.Jadrić, Ž.Garača, M. Ćukušić: *Student dropout analysis with application of data mining methods*, Management, Vol.15, 1, pp.31 - 46, 2010.

[37] A.S.Koyuncugil, N. Ozgulbas: *Financial early warning system model and data mining application for risk detection,* Expert Systems with Applications, Vol. 39, Issue 6, pp. 6238-6253, 2012.

[38] S. Buć: *Model upravljanja rizicima kod javnih projekata, magistarski rad,* Sveučlilište u Zagrebu, Fakultet organizacije i informatike Varaždin, Varaždin, 2007.

[39] G. V. Kass: *An Exploratory Technique for Investigating Large Quantities of Categorical Data,* Journal of the Royal Statistical society (Applied Statistics), Vol. 29, Issue 2, pp. 119-127, South Africa, 1980.

[40] N. Sut, O. Simsek: *Comparison of regression tree data mining methods for prediction of mortality in head injury,* Expert Systems with Applications, Vol. 38, Issue 12, pp. 15534-15539, 2011.

[41] M. Ture, I. Kurt, T.A. Kurum, K. Ozdamar: *Comparing classification techniques for predicting essential hypertension,* Expert Systems with Applications, Vol. 29, Issue 3, pp. 583-588, 2005.

[42] R.L.Plackett: *Karl Perason and the Chi-squared Test,* International Statistical Review, 51, pp. 59-72, UK, 1983.

[43] B. Petz: *Osnovne statističke metode za nematematičare,* Naklada Slap, RH, 2007.

[44] I.Šošić: *Primjenjena statistika,* Školska knjiga, RH, 2004.

[45] D.Biggs, B.Ville, E.Suen: *A Method of Choosing Multiway Partitions for Classification and Decesion Trees,* Journal of Applied Statistics, 18, 1, 1991.

[46] L.A.Goodman: *Simple Models for the Analysis of Association in Cross-Classifications Having Ordered Categories,* Journal of the American Statistical Association, 74, 1979.

[47] A.Badea-Romero, J. Lenard: *Source of head injury for pedestrians and pedal cyclists: Striking vehicle or road?,* Accident Analysis & Prevention, Vol. 50, pp. 1140-1150, 2013.

[48] X. Yan, S. Richards, X. Su: *Using hierarchical tree-based regression model to predict train-vehicle crashes at passive highway-rail grade crossings,* Accident Analysis & Prevention, Vol. 42, Issue 1, pp. 64-74, 2010.

[49] A. Della Corte, C. Bancone, C.Quarto, G. Dialetto, F.E.Covino, M. Scardone, G. Caianiello, M. Cotrufo: *Predictors of ascending aortic dilatation with bicuspid aortic valve: a wide spectrum of disease expression,* European Journal of Cardio-thoracic Surgery, Volume 31, pp. 397-405, Italy, 2007.

[50] Lj. Preradović, V. Đajić, D. Račić, D. Tadić, Lj. Popović. T. Rabi Žikić, M. Vujinić, M. Žikić: *Primjena savremenih informacionih tehnologija za analizu podataka i predviđanje u medicini,* Journal of Association of Serbian Neurologists, god. XVII, br.1-2, UDK 61:007, Srbija, 2009.

[51] M. Ramaswami, R. Bhaskaran: *A CHAID Based Performance Prediction Model in Educational Data Mining,* IJCSI International Journal of Computer Science Issues, Vol.7, Issue 1, No.1, ISSN:1694-0814, pp. 10-18, 2010.

[52] J.Soldić Aleksić: *Prediktivni model segmentacije tržišta: primena modela logističke regresije i CHAID procedure,* Marketing, Godina 40, Broj 3, pp. 129-138, Srbija, 2009.

[53] P. Oom do Valle, P. Pintassilgo, A. Matias, F. Andre: *Tourist attitudes towards an accommodation tax earmarked for environmental protection: A survey in the Algarve,* Tourism Management, Vol. 33, Issue 6, pp. 1408-1416, 2012.

[54] A.I. Abdelrahman, D.H. Abdel Hady: *Classification of Insolvent Small Businesses in Egypt by Some Running Cost Variables: A Decision Tree Approach,* interstat.statjournals.net, 2010.

[55] J. Shury, M. Speed, D. Vivian, A. Kuechel, S. Nicholas: *Crime against retail and manufacturing premises: finding from the 2002 Commercial Victimisation Survey,* Supplementary web report 37/05, Research Development and Statistics Directorate, Home Office, UK Govermant, www.homeoffice.gov.uk/rds, UK, 2005.

[56] C.F. Chi, Y.Y.Lin: *Flow diagram analysis of electrical fatalities in construction industry,* Safety Science, Vol. 50, Issue 5, pp. 1205-1214, 2012.

[57] L. Kleppin, R. Pesch, W. Schröder: *CHAID Models on boundary conditions of metal accumulation in mosses collected in Germany in 1990, 1995 and 2000,* Atmospheric Environment, Vol. 42, Issue 21, pp. 5220-5231, 2008.

[58] D.Lisjak: Primjena metoda umjetne inteligencije pri izboru materijala, Doktorski rad, Fakultet strojarstva i brodogradnje Sveučilišta u Zagrebu, RH, 2004.

[59] R. Koch: *Living the 80/20 Way*, Nicholas Brealey Publishing, USA, 2004.

[60] N.R.Tague: *The Quality Toolbox. Second edition,* Quality Press, USA, 2004.

[61] *Izvješće o stanju tehničkih pregleda vozila u Republici Hrvatskoj u 2009. godini*, Centar za vozila Hrvatske, Republika Hrvatska, 2010.

[62] Hrvatski autoklub - *Elektronička SQL baza podataka tehničkih pregleda vozila* (za 2009. godinu)

[63] *Pravilnik o informatičkom sustavu i jedinstvenom programskom rješenju za obradu podataka o tehničkim pregledima i registraciji vozila u stanicama za tehnički pregled vozila*, Narodne novine (33/2009), Republika Hrvatska, 2009.

[64] *IBM SPSS Statistics 22 Software, SAD*.

[65] B. W. Matthews: *Comparison of the predicted and observed secondary structure of T4 phage lysozyme*. Biochimica et Biophysica Acta (BBA) - Protein Structure 405 (2), USA, 1975.

[66] D.M.W.Powers: *Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation*, Journal of Machine Learning Technologies 2(1), Australia, 2011.

[67] E. Antipov, E. Pokryshevskaya: *Applying CHAID for logistic regression diagnostics and classification accuracy improvement*, Journal of Targeting, Measurement and Analysis for Marketing, Vol. 18, pp. 108-117, 2010.

[68] S. Kusakci: *Segmentation of student market: A CHAID based analysis,* 6th international Student Conference, Izmir University of Economics, Turkey, 2010.

[69] *Izvješće o stanju i funkcioniranju sustava tehničkih pregleda i registracije vozila u RH u 2010. godini, iz aspekta stručnog nadzora nad radom stanica za tehnički pregled vozila,* Hrvatski autoklub, Zagreb, 2011.

[70] *Izvješće o stanju i funkcioniranju sustava tehničkih pregleda i registracije vozila u RH u 2011. godini, iz aspekta stručnog nadzora nad radom stanica za tehnički pregled vozila,* Hrvatski autoklub, Zagreb, 2012.

[71] *Nacionalni program sigurnosti cestovnog prometa Republike Hrvatske 2011.-2020. godina,* Vlada Republike Hrvatske, Zagreb, 2011.

PRILOG 1 Sklopovi (17 cjelina) i dijelova (134 pozicije) čije se postojanje i tehnička ispravnost obvezno kontroliraju tijekom redovnog tehničkog pregleda osobnog automobila

1. cjelina: Uređaj za upravljanje

pozicije: - kolo upravljača
- stup upravljača
- prijenosni mehanizam
- poluge i zglobovi upravljača
- pojačalo sile zakretanja upravljača
- amortizer upravljača
- graničnik kuta zakretanja upravljača.
- zakretno postolje priključnog vozila

2. cjelina: Uređaj za kočenje

pozicije: - radna kočnica
- pomoćna kočnica
- parkirna kočnica
- komanda radne kočnice
- komanda pomoćne kočnice
- zračna kočna instalacija - dijelovi
- hidraulična kočna instalacija - dijelovi
- mehanička kočna instalacija - dijelovi
- izvršni kočni elementi
- spojne glave za kočnicu prikolice.

3. cjelina: Uređaj za osvjetljavanje i svjetlosnu signalizaciju

pozicije: - kratko svijetlo
- dugo svijetlo
- prednje svijetlo za maglu
- reflektor za radove
- svijetlo za vožnju unatrag
- prednja pozicijska svijetla
- stražnja pozicijska svijetla

- stražnja svijetla za maglu
- parkirna svijetla
- gabaritna svijetla
- svijetla registarske tablice
- žuta rotacijska ili treptava svijetla
- plava/crvena rotacijska ili treptava svijetla
- katadiopteri
- stop svijetla
- pokazivači smjera
- četiri pokazivača pravca.

4. cjelina: Uređaji za osiguravanje vidljivosti

pozicije: - vjetrobran
- staklene površine
- brisači vjetrobrana
- perači vjetrobrana
- retrovizori.

5. cjelina: Samonosiva karoserija te šasija s kabinom i nadogradnjom

pozicije: - samonosiva karoserija
- šasija
- kabina
- nadogradnja.

6. cjelina: Elementi ovjesa, osovine, kotači

pozicije: - polužje ovjesa
- zglobovi ovjesa
- amortizeri
- opruge
- glavina kotača
- naplatci
- gume.

7. cjelina: Motor

pozicije: - oslonci motora
- zauljenost motora
- ispušni sustav
- usisni sustav
- sustav za paljenje
- sustav za napajanje gorivom
- razvodni mehanizam.

8. cjelina: Buka vozila

pozicije: - buka u mirovanju
- buka u pokretu.

9. cjelina: Elektro uređaji i elektro instalacije

pozicije: - elektropokretač
- generator
- akumulator
- kontakt brava
- električni vodovi

10. cjelina: Prijenosni mehanizam

pozicije: - spojka
- mjenjač
- vratila
- poluvratila
- diferencijal
- lanac, lančanici
- remen, remenice.

11. cjelina: Kontrolni i signalni uređaji

pozicije: - brzinomjer s putomjerom
- plava lampa za dugo svijetlo
- sirena
- tahograf

- ograničavač brzine
- signal pokazivača smjera
- ostali signalni uređaji.

12. cjelina: Ispitivanje ispušnih plinova motornih vozila (EKO test)

pozicije: - ispušni sustav
- usisni sustav
- sustav za paljenje
- sustav za napajanje gorivom
- razvodni mehanizam
- BEZ KAT
- REG KAT
- DIZEL.

13. cjelina: Uređaj za spajanje vučnog i priključnog vozila

pozicije: - mehanička spojka
- električni priključak spojke.

14. cjelina: Ostali uređaji i dijelovi vozila

pozicije: - unutrašnjost kabine, sjedala i prostor
- uređaj za ventilaciju kabine vjetrobr.
- vrata vozila
- pokretni prozori i krovovi
- brave
- izlazi u slučaju opasnosti
- blatobrani
- branici
- stražnja zaštita protiv podlijetanja
- bočna zaštita protiv podlijetanja
- sigurnosni pojasevi
- dodatne invalidne komande
- ograničavač brzine na mopedima.

15. cjelina: Oprema vozila

pozicije: - aparat za gašenje požara
- sigurnosni trokut
- kutija prve pomoći
- klinasti podmetači
- čekić za razbijanje stakla u nuždi
- rezervne žarulje
- rezervni kotač s opremom.

16. cjelina: Registracijske tablice i oznake

pozicije: - registracijske tablice
- ploče za “teška” vozila
- ploče za “duga” vozila
- ploče za “spora” vozila.

17. cjelina: Plinska instalacija

pozicije: - spremnik plina
- armatura spremnika plina
- priključak za punjenje
- priključak za pražnjenje
- višesmjerni ventil
- pokazivač količine plina
- pročistač plina
- isparivač plina
- regulator tlaka
- ventil plina
- ventil tekućeg goriva
- vodovi visokog tlaka (VT)
- vodovi niskog tlaka (NT)
- vodovi za sredstvo za grijanje
- električni uređaji i instalacije
- *lambda* sonda
- regulator količine plina
- elektronički uređaj za *lambda* kontrolu

- mješač plina
- brizgaljke plina.

ŽIVOTOPIS

Igor Šiško rođen je 15. veljače 1972. godine u Zagrebu gdje je pohađao osnovnu školu "Narodni heroj Anka Butorac", te srednju školu "Centar odgoja i usmjerenog obrazovanja Ruđer Bošković". Nakon završene srednje škole, 1990. godine upisuje Fakultet strojarstva i brodogradnje Sveučilišta u Zagrebu. Po završetku studija "Proizvodno strojarstvo" diplomira u mjesecu studenom 1996. godine. Nakon završetka vojnog roka, u veljači 1998. godine zapošljava se u Hrvatskom autoklubu, na radnom mjestu samostalnog referenta za homologaciju vozila. U studenom 2006. godine preuzima rukovođenje Sektorom za vozila u Hrvatskom autoklubu u okviru kojeg rukovodi procesom stručnog nadzora nad radom stanica za tehnički pregled vozila, procesom homologacije vozila te procesom preuzimanja, prikupljanja, čuvanja i javne prodaje vozila pod carinskim nadzorom. U prosincu 2005. godine završava školovanje za menadžera kvalitete prema zahtjevima EOQ (engl. *EOQ Quality Systems Manager*) harmonizirane sheme te je osposobljen za uspostavljanje i upravljanje sustava kvalitete prema normi ISO 9001. Od studenoga 2006. godine voditelj je projekta implementacije zahtjeva norme HRN EN ISO/IEC 17020:2005, koja propisuje zahtjeve za nadzorne ustanove, u Sektor za vozila – Odjel stručnog nadzora nad radom stanica za tehnički pregled vozila Hrvatskog autokluba. U svibnju 2008. godine navedena ustrojstvena cjelina Hrvatskog autokluba akreditirana je od strane Hrvatske akreditacijske agencije prema zahtjevima predmetne norme. U svom dosadašnjem radu autor je i ko-autor nekoliko stručnih i znanstvenih radova iz područja tehničkih pregleda i homologacije vozila, te je član nekoliko tuzemnih strukovnih tijela iz predmetnih područja. Također je član radnih tijela međunarodne organizacije ustanova koje obavljaju tehničke preglede vozila CITA –e (*Comité International de l'inspection Technique Automobile*). Od početka 2014. godine preuzima funkciju pomoćnika Glavnog tajnika Hrvatskog autokluba.

BIOGRAPHY

Igor Šiško was born on 15th of February 1972 in Zagreb, where he attended elementary school “People’s Hero Anka Butorac” and high school “Center of directed education Ruđer Bošković”. After high school he enrolled in the Faculty of Mechanical Engineering and Naval Architecture University of Zagreb in 1990. He graduated in November of 1996 in scope of studies of “Producted Mechanical Engineering”. After finishing military service in February of 1998 he started to work as a specialist for homologation (type approval) of vehicles in Croatian automobile. He took over the managing of Vehicles division of Croatian automobile club in November of 2006. The scope of his work has been managing the process of inspection (supervision) of vehicle roadworthiness tests stations, managing the process of homologation of vehicles and managing the process of Customs affairs (vehicles under Customs supervision). He completed education for EOQ Quality Systems Manager in December of 2005 in accordance with the EOQ Hamonized Scheme and has proven the knowledge and practice in establishing and managing of a quality system due to standard ISO 9001. Since November of 2006 till May of 2008 he was in charge for the project of implementing requirements for inspection bodies due to standard HRN EN ISO/IEC 17020:2005, in Vehicles division – Inspection of vehicles roadworthiness tests stations department of Croatian automobile club, which was accredited by Croatian Accreditation Agency in May of 2008. He published and co-published several experts and scientific papers from the field of roadworthiness tests of vehicles and homologation of vehicles and he is a member of several domestic experts bodies from mentioned fields. Also, he is a member of working bodies of CITA (International Motor Vehicle inspection Committe). Since the beginning of 2014 he took over the position of Assistent Secretary General of Croatian automobile club.