UNIVERZITET U BEOGRADU
MATEMATIČKI FAKULTET

Dejana Jeremić

INTERPRETACIJA NELINEARNIH STATISTIČKIH MODELA SA PRIMENOM U OSIGURANJU

master rad

Beograd, 2025.

Mentor:

dr Bojana Milošević, vanredni profesor
Univerzitet u Beogradu, Matematički fakultet

Članovi komisije:

dr Marko Obradović, docent
Univerzitet u Beogradu, Matematički fakultet

dr Marija Cuparić, docent
Univerzitet u Beogradu, Matematički fakultet

Datum odbrane: _____

Porodici

Naslov master rada: Interpretacija nelinearnih statističkih modela sa primenom u osiguranju

Rezime: Nelinearni statistički modeli su zbog svoje složene arhitekture često teško razumljivi. Sa intenzivnim razvojem veštačke inteligencije, postaje sve važnije imati uvid u način na koji ovi modeli funkcionišu, jer bez toga ne možemo očekivati dalji napredak i usavršavanje modela. Tek kada znamo šta se tačno događa u modelu, možemo shvatiti sve njegove mogućnosti i ograničenja, što je ključno za njegov dalji razvoj. U ovom radu koristićemo nelinearne modele statističkog učenja, kao što su slučajne šume i neuronske mreže, na podacima iz osiguranja. Metode za interpretaciju nelinearnih modela statističkog učenja koriste se u svim sferama poslovanja, i kao takve imaju veoma široku primenu. Cilj ovog rada je prikazati različite metode za interpretaciju ovih vrsta modela i iskoristiti dobijeno za izgradnju najboljeg mogućeg modela koji treba da predvidi broj budućih šteta koje osiguravajuća kompanija može da očekuje. Neke od metoda koje ćemo prikazati su grafici parcijalne zavisnosti (eng. Partial Dependence Plots, PDP), metod Individualnog uslovnog očekivanja (eng. Individual Conditional Expectation, ICE), metoda Akumuliranih lokalnih efekata (eng. Accumulated Local Effects, ALE), sa akcentom na najrasprostranjeniju Šeplijevu metodu (eng. Shapley Additive exPlanations, SHAP). Biće prikazana i objašnjena i matematička pozadina ovih modela i metoda interpretacije.

Ključne reči: model, interpretacija, osiguranje

Sadržaj

1 Uvod **1**

2 Metode interpretacije nelinearnih statističkih modela **3**

2.1 Grafici parcijalne zavisnosti 3

2.2 Metoda individualnog uslovnog očekivanja 5

2.3 Metoda akumuliranih lokalnih efekata 8

2.4 Šeplijeva metoda 12

2.5 Poređenje Šeplijeve metode sa PDP, ICE i ALE metodama 18

3 Primer u praksi **20**

3.1 Baza podataka 21

3.2 Linearan model 24

3.3 Uopšten linearan model 27

3.4 Model slučajnih šuma 30

3.5 Neuronska mreža 40

3.6 Poređenje interpretacije modela: Uopšteni linearan model, model slučajnih šuma i neuronska mreža 47

3.7 Izrada najoptimalnijeg modela 48

4 Zaključak **75**

Bibliografija **77**

Glava 1

Uvod

Savremena nauka i industrija se sve više oslanjaju na složene statističke modele za donošenje odluka i predviđanja. Nelinearni statistički modeli, kao što su slučajne šume i neuronske mreže, predstavljaju glavni alat u ovoj oblasti. Iako se ovi modeli ističu po svojoj sposobnosti da analiziraju velike skupove podataka i daju tačna predviđanja, jedan od glavnih izazova ostaje njihova interpretacija. Razumevanje kako modeli funkcionišu nije samo akademski izazov, već je značajno i u praksi, u kontekstu primene u različitim sektorima, uključujući i osiguranje.

U domenu osiguranja, potreba za tačnim i objašnjivim modelima je od suštinskog značaja. Osiguravajuća društva svakodnevno obrađuju ogromne količine podataka kako bi identifikovala rizike, utvrdila cene premija i predvidela buduće štete. U ovom kontekstu, upotreba nelinearnih modela omogućava dublje i preciznije analize, ali takođe postavlja pitanje njihove interpretacije.

Značaj interpretacije nelinearnih modela

Nelinearni modeli su postali standard u mnogim oblastima zbog svoje moći da uoče složene odnose među podacima. Međutim, njihova složena struktura često otežava razumevanje toga na koji način se donose odluke. U praksi, ovo može dovesti do nepoverenja u rezultate modela, posebno u visoko regulisanim industrijama poput osiguranja, gde je transparentnost presudna.

Metode interpretacije kao što su grafici parcijalne zavisnosti (eng. Partial Dependence Plots, PDP), metod individualnog uslovnog očekivanja (eng. Individual Conditional Expectation, ICE), metoda akumuliranih lokalnih efekata (eng. Accumulated Local Effects, ALE) i Šeplijeva metoda (eng. Shapley Additive exPlanations,

SHAP), razvijene su kako bi odgovorile na ovaj izazov. Ove metode omogućavaju stručnjacima da bolje razumeju ponašanje modela, i pružaju uvid u to kako pojedinačne promenljive utiču na predviđanja. Osim što poboljšavaju transparentnost, ove tehnike pružaju i mogućnost dalje optimizacije modela, otkrivajući mogućnosti za poboljšanje.

Primena u osiguranju

Industrija osiguranja je specifična po tome što se oslanja na složene matematičke i statističke metode za analizu rizika i upravljanje kapitalom. Uz pomoć interpretativnih metoda, osiguravajuća društva mogu bolje razumeti ne samo kako modeli funkcionišu, već i kako da ih unaprede. Interpretacija modela može pomoći u identifikovanju ključnih faktora koji utiču na učestalost i veličinu šteta, čime se omogućava preciznije određivanje cena premija i strategija za upravljanje rizikom.

Takođe, razumevanje modela je od suštinskog značaja za izgradnju poverenja između osiguravajućih kompanija i njihovih klijenata. Ukoliko se može jasno objasniti kako je određena premija ili zašto je određena šteta odbijena, to ne samo da poboljšava reputaciju kompanije, već i omogućava bolju regulatornu usaglašenost.

Struktura rada

Ovaj rad ima za cilj da predstavi ključne metode interpretacije nelinearnih modela i njihovu primenu u industriji osiguranja. U prvom delu rada, biće obrađena matematička pozadina pomenutih metoda interpretacija - PDP, ICE, ALE i Šeplijeva metode.

Nakon toga, ove metode će biti primenjene na stvarnim podacima iz osiguranja, u cilju dobijanja što boljeg uvida u podatke i kako bismo razvili najbolji mogući model za predviđanje broja šteta.

U završnom delu rada, biće prikazani rezultati i zaključci.

Glava 2

Metode interpretacije nelinearnih statističkih modela

2.1 Grafici parcijalne zavisnosti

Pretpostavimo da naš model f predviđa izlaz na osnovu vektora nezavisnih promenljivih

$$\mathbf{X} = (X_1, X_2, \ldots, X_p).$$

Želimo da ispitamo kako promena određene nezavisne promenljive ili skupa nezavisnih promenljivih $\mathbf{X}_S$ utiče na izlaz $f(\mathbf{X})$, pri čemu ostale nezavisne promenljive označavamo sa $\mathbf{X}_C$ (gde je $S \cup C = \{1, 2, \ldots, p\}$).

Definicija PDP

Funkcija parcijalne zavisnosti za skup nezavisnih promenljivih $\mathbf{X}_S$ definiše se kao uslovno očekivanje modela f preko raspodele preostalih nezavisnih promenljivih $\mathbf{X}_C$:

$$PD_{\mathbf{X}_S}(\mathbf{x}_S) = \mathbb{E}_{X_C}\left[f(\mathbf{x}_S, \mathbf{X}_C)\right] = \int f(\mathbf{x}_S, \mathbf{x}_C)\, p(\mathbf{x}_C)\, d\mathbf{x}_C,$$

gde $\mathbf{X}_C$ označava slučajne promenljive iz komplementarnog skupa nezavisnih promenljivih, a $p(\mathbf{x}_C)$ njihovu marginalnu raspodelu.

Aproksimacija u praksi

U praksi se integral ne računa analitički, već se aproksimira na osnovu dostupnih podataka. Pretpostavimo da imamo uzorak od n posmatranja, tj. da za svako $i =$

$1, 2, \ldots, n$ znamo realizaciju $\mathbf{x}_C^{(i)}$. Tada aproksimacija PDP funkcije izgleda ovako:

$$\widehat{PD}_{\mathbf{X}_S}(\mathbf{x}_S) = \frac{1}{n}\sum_{i=1}^{n} f(\mathbf{x}_S, \mathbf{x}_C^{(i)}), \tag{2.1}$$

gde je $\mathbf{x}_C^{(i)}$ vrednost nezavisne promenljive iz komplementarnog skupa $\mathbf{x}_C$ za i-to posmatranje.

Poseban slučaj - Jedna nezavisna promenljiva

Kada nas zanima uticaj samo jedne nezavisne promenljive X_j (tj. $S = \{j\}$), formula postaje:

$$PD_{X_j}(x_j) = \mathbb{E}_{X_C}\big[f(x_j, X_C)\big] = \int f(x_j, \mathbf{x}_C)\, p(\mathbf{x}_C)\, d\mathbf{x}_C,$$

a ocena:

$$\widehat{PD}_{X_j}(x_j) = \frac{1}{n}\sum_{i=1}^{n} f(x_j, \mathbf{x}_C^{(i)}).$$

Procena važnosti nezavisnih promenljivih

PDP se takođe koristi za procenu važnosti nezavisnih promenljivih u modelu. Merimo koliko promena u vektoru nezavisnih promenljivih $\mathbf{X}_S$ utiče na izlaz modela u poređenju sa drugim nezavisnim promenljivama. Ovo je posebno korisno za identifikovanje ključnih promenljivih koje imaju najveći uticaj na predikciju.

Za numeričke promenljive, važnost nezavisnih promenljivih zasnovana na grafikonima delimične zavisnosti se meri kao standardno odstupanje svake od K jedinstvenih vrednosti vektora $\mathbf{X}_S$ od prosečne krive:

$$I(\mathbf{x}_S) = \sqrt{\frac{1}{K-1}\sum_{k=1}^{K}\left(\widehat{PD}_{\mathbf{X}_S}(\mathbf{x}_S^{(k)}) - \frac{1}{K}\sum_{k=1}^{K}\widehat{PD}_{\mathbf{X}_S}(\mathbf{x}_S^{(k)})\right)^2},$$

gde:

- $\mathbf{x}_S^{(k)}$ je K-ta jedinstvena vrednost vektora nezavisnih promenljivih $\mathbf{X}_S$,
- $\widehat{PD}_{\mathbf{X}_S}(\mathbf{x}_S^{(k)})$ je očekivana vrednost izlaza modela za ulaznu vrednost $\mathbf{x}_S^{(k)}$.

Za kategoričke promenljive, važnost nezavisnih promenljivih zasnovana na grafikonima delimične zavisnosti se računa kao razlika između maksimalne i minimalne vrednosti predikcije za različite vrednosti promenljivih:

$$I(\mathbf{x}_S) = \frac{\max_k \left(\widehat{PD}_{\mathbf{x}_S}(\mathbf{x}_S^{(k)})\right) - \min_k \left(\widehat{PD}_{\mathbf{x}_S}(\mathbf{x}_S^{(k)})\right)}{4}.$$

Ova mera nam pomaže da odredimo koliko promena u vektoru nezavisnih promenljivih $\mathbf{X}_S$ utiče na izlaz modela. Promenljive koje izazivaju veće promene u predikcijama modela imaju veću važnost.

Nedostaci metode

Matematički, PDP predstavlja srednju vrednost funkcije modela f preko raspodele ostalih nezavisnih promenljivih, čime se dobija funkcija koja nam pruža uvid u uticaj odabranih promenljivih ili skupa promenljivih. Ovo je posebno korisno u interpretaciji kompleksnih, nelinearnih modela, ali je važno napomenuti da se pretpostavlja da promenljiva x_j nije prejako korelisana sa ostalim promenljivama, jer u suprotnom prosek može sakriti značajne interakcije.

2.2 Metoda individualnog uslovnog očekivanja

Metoda Individualnog uslovnog očekivanja (eng. Individual Conditional Expectation, ICE), je tehnika za interpretaciju modela statističkog učenja koja omogućava vizualizaciju zavisnosti predikcija modela od pojedinačnih promenljivih, za svaki uzorak posebno. Za razliku od globalne metode kao što je PDP, koja prikazuju prosečan efekat promenljivih na predikciju, ICE se fokusira na lokalne efekte, pružajući detaljniji uvid u ponašanje modela na nivou pojedinačnih instanci.

ICE metoda se zasniva na analizi kako promena vrednosti određene promenljive ili skupa promenljivih utiče na predikciju modela za svaki uzorak posebno. Za svaku instancu u skupu podataka, vrednost odabranih promenljivih se sistematski menja unutar definisanog opsega, dok ostale promenljive ostaju nepromenjene. Za svaku od tih modifikovanih vrednosti, računa se predikcija modela, što rezultira krivom koja prikazuje odnos između vrednosti promenljive i predikcije za tu instancu. Kombinovanjem ovih krivih za sve instance, dobija se ICE plot koji omogućava vizualizaciju varijacija u efektima promenljivih na predikciju kroz različite uzorke.

Matematička formulacija

Neka je f funkcija koja predstavlja model statističkog učenja,

$$\mathbf{X} = (X_1, X_2, \ldots, X_p)$$

vektor ulaznih promenljivih, a $\mathbf{X}_S$ podskup promenljivih od interesa. Individualna uslovna očekivanja za instancu i definišu se kao:

$$ICE^{(i)}_{\mathbf{X}_S}(\mathbf{x}_S) = f(\mathbf{x}_S, \mathbf{x}^{(i)}_C), \tag{2.2}$$

gde je $\mathbf{x}_S$ vrednost promenljivih od interesa, a $\mathbf{x}^{(i)}_C$ vrednosti komplementarnog skupa promenljivih za instancu i, $\mathbf{X}_C$, (gde je $S \cup C = \{1, 2, \ldots, p\}$). Promenom $\mathbf{x}_S$ kroz željeni opseg vrednosti i zadržavanjem $\mathbf{x}^{(i)}_C$ konstantnim, dobijamo niz predikcija koje čine ICE krivu za datu instancu. Ponavljanjem ovog postupka za sve instance u skupu podataka, možemo konstruisati ICE plot koji prikazuje individualne efekte promenljivih $\mathbf{X}_S$ na predikciju modela.

Centralizovan ICE grafik

Jedan od izazova kod ICE grafikona je to što krive za različite instance mogu počinjati od različitih nivoa predikcije, što otežava njihovo poređenje. Jednostavno rešenje ovog problema je centriranje svih krivih u odnosu na određenu referentnu tačku promenljive. Na taj način, ICE grafik prikazuje samo razlike u predikciji u odnosu na tu zajedničku tačku. Ovaj pristup vodi ka stvaranju **centriranog ICE (c-ICE) grafika**. Najbolje je krive centrirati na donjoj granici promenljive. Svaka od krivih može se definisati kao:

$$ICE^{(i)}_{\mathbf{X}_S}(\mathbf{x}_S) = f(\mathbf{x}_S, \mathbf{x}^{(i)}_C) - f(a, \mathbf{x}^{(i)}_C),$$

gde je f prediktivni model, a a je tačka koju koristimo kao centar.

Centrirani ICE grafik omogućava lakše poređenje krivih za pojedinačne instance. Ovaj pristup je koristan kada želimo da analiziramo razliku u predikcijama u odnosu na fiksnu tačku opsega promenljive, umesto da se fokusiramo na apsolutnu promenu predikcije.

Poređenje sa PDP

PDP predstavlja prosečni efekat promena $\mathbf{X}_S$ preko svih instanci (formula 2.1). Ako su pojedinačne ICE linije slične, PDP funkcija verovatno dobro reprezentuje

uticaj promenljivih $\mathbf{X}_S$. Međutim, značajne razlike među ICE linijama ukazuju na postojanje interakcija između $\mathbf{X}_S$ i ostalih promenljivih.

Prednosti i ograničenja ICE metode

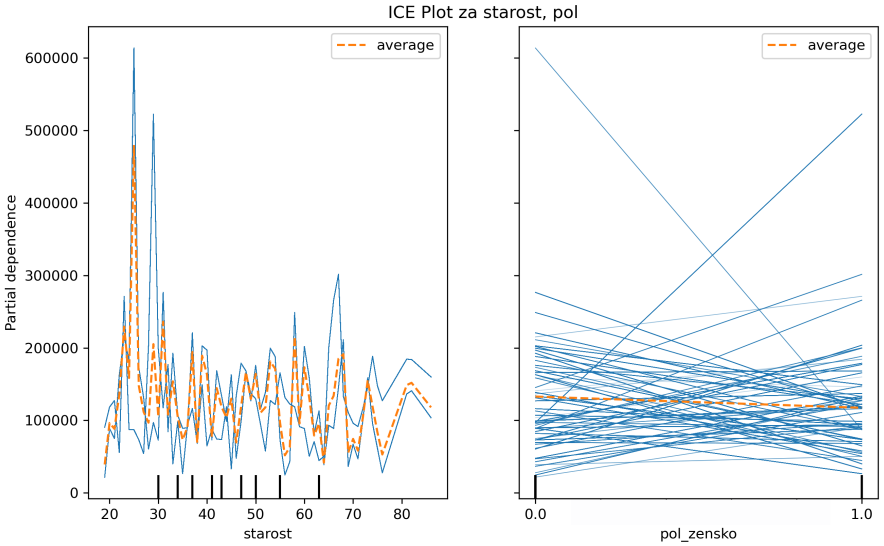
- **Otkrivanje heterogenosti**: Za razliku od PDP-a koji prikazuje prosečan efekat, ICE omogućava identifikaciju varijacija u efektima promenljivih na različite instance, što može ukazivati na interakcije između promenljivih ili nelinearne odnose.
- **Detaljna interpretacija**: ICE pruža uvid u ponašanje modela na nivou pojedinačnih uzoraka, što je posebno korisno u praksi u industrijama gde je razumevanje individualnih predikcija ključno, poput medicine ili finansija.
- **Vizuelna složenost**: Kod velikih skupova podataka, ICE plot može postati pretrpan i teško interpretabilan zbog velikog broja krivih. U takvim slučajevima, preporučuje se uzorkovanje ili agregacija krivih radi bolje preglednosti.
- **Jednodimenzionalnost**: ICE plotovi su najefikasniji za analizu efekata pojedinačnih promenljivih. Analiza višedimenzionalnih efekata zahteva složenije vizualizacije koje mogu biti teže za interpretaciju.

Primer primene ICE metode

U praksi, ICE plotovi se često koriste zajedno sa PDP-ovima kako bi se dobila kompletna slika o efektima ulaznih promenljivih na predikcije modela. Dok PDP pruža globalni pregled prosečnih efekata, ICE omogućava detaljan uvid u individualne varijacije, što je ključno za identifikaciju potencijalnih interakcija i nelinearnosti u modelu.

Napravljen je model slučajnih šuma (eng. Random Forest model) za procenu visine štete kasko osiguranja u zavisnosti od starosti i pola vozača. Grafik na slici 2.1 pokazuje kako se predikcije modela (visine štete) menjaju u zavisnosti od starosti (leva strana) i pola (desna strana) kada se ostale promenljive drže konstantnim. Promenljiva pol_zensko predstavlja binarnu promenljivu koja ima vrednost 1 za ženski pol, a 0 za muški pol.

Na levom grafiku na slici 2.1 prikazano je kako se predviđena visina štete menja u zavisnosti od starosti vozača. Svaka plava linija predstavlja pojedinačnog osiguranika, tj. kako se predikcija menja kada se starost menja za tog osiguranika. Na-



Slika 2.1: ICE plot za starost i pol vozača i iznos štete

randžasta isprekidana linija predstavlja prosečnu zavisnost (PDP). Velike oscilacije mogu ukazivati na to da starost značajno utiče na štetu, ali da postoji i značajan nivo varijacije među osiguranicima.

Na desnom grafiku na slici 2.1 prikazano je kako visina štete zavisi od pola vozača. Plave linije prikazuju individualne osiguranike, dok isprekidana linija daje prosečnu zavisnost. Linije koje se preklapaju i ne pokazuju jaku tendenciju sugerišu da pol ima manji uticaj na predviđenu visinu štete u poređenju sa starošću.

Možemo zaključiti da starost vozača ima izražen uticaj na visinu štete, pri čemu su predikcije najnestabilnije i potencijalno najviše za mlađe osiguranike. Takođe vidimo da pol ima relativno slabiji efekat na visinu štete u poređenju sa starošću.

2.3 Metoda akumuliranih lokalnih efekata

Metoda akumuliranih lokalnih efekata (eng. Accumulated Local Effects, ALE), je metoda za interpretaciju modela statističkog učenja koja poboljšava PDP metodu rešavanjem problema koje PDP ima u prisustvu zavisnosti između promenljivih. ALE meri lokalni uticaj promena određenih promenljivih $\mathbf{X}_S$ na predikciju modela i zatim akumulira te efekte duž skale vrednosti $\mathbf{X}_S$. Preostale promenljive su označene sa $\mathbf{X}_C$.

Matematička formulacija

Zbog jednostavnosti zapisa, ovde ćemo pokazati slučaj u kojem analiziramo jednu nezavisnu promenljivu. Neka je x_s konkretna vrednost posmatrane promenljive. ALE računa očekivani lokalni gradijent modela $f(x_S, \mathbf{X}_C)$ u malim intervalima i zatim ih akumulira:

$$ALE_{x_S}(x_S) = \int_{x_{S,\min}}^{x_S} \mathbb{E}_{\mathbf{X}_C|x_S=z}\left[\frac{\partial f(x_S, \mathbf{X}_C)}{\partial x_S}\right] dz,$$

gde:

- $x_{S,\min}$ predstavlja minimalnu vrednost promenljive X_S,
- $\mathbf{X}_C$ su preostale promenljive osim X_S,
- $\mathbb{E}_{\mathbf{X}_C|x_S=z}$ označava uslovno očekivanje lokalnog gradijenta.

U praksi, pod pretpostavkom da imamo uzorak od n posmatranja, tj. da za svako $i = 1, 2, \ldots, n$ znamo realizaciju $\mathbf{x}_C^{(i)}$, ALE se izračunava kroz sledeće korake:

1. **Diskretizacija promenljive** x_S na K intervala (z_k, z_{k+1}).
2. **Računanje lokalnog efekta** u svakom intervalu kao prosečna promena predikcije:

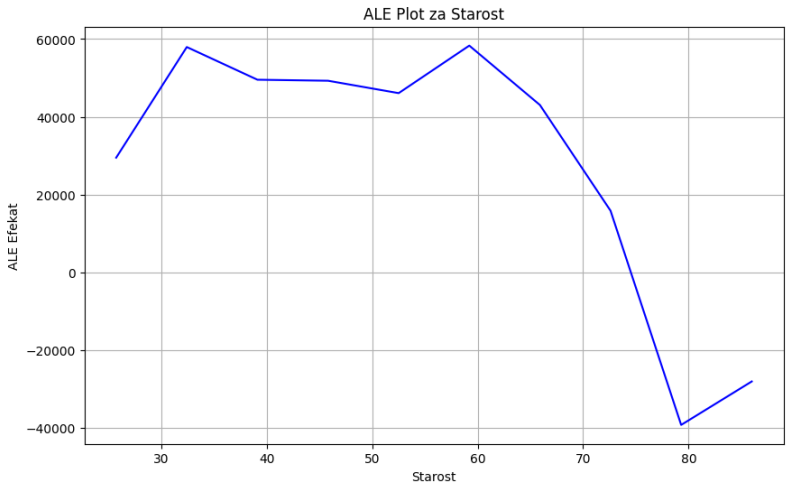
   $$\widehat{ALE}_{x_S}(z_k, z_{k+1}) = \frac{1}{\Delta z_k}\frac{1}{n_k}\sum_{i \in S_k}\Big[f(z_{k+1}, \mathbf{x}_C^{(i)}) - f(z_k, \mathbf{x}_C^{(i)})\Big],$$

   gde je S_k skup instanci za koje vrednost x_S pripada intervalu (z_k, z_{k+1}), n_k njihov broj, i $\Delta z_k = z_{k+1} - z_k$. U praksi su svi intervali iste dužine, pa je član $\frac{1}{\Delta z_k}$ konstanta i može se izostaviti, jer ne menja oblik ALE krive, već samo njenu vertikalnu skalu.
3. **Akumulacija lokalnih promena** duž cele skale vrednosti x_S, čime se dobija globalni efekat.

Dakle, ALE grafici su zasnovani na lokalnim gradijentima funkcije modela po izabranoj promenljivoj, koji se zatim akumuliraju. To znači:

- nagib ALE linije (pozitivan ili negativan) pokazuje pravac promene predikcije kada se posmatrana promenljiva poveća,

- vrednost ALE funkcije u određenoj tački nije gradijent, već akumulirani efekat prethodnih lokalnih gradijenata,
- kada ALE kriva raste, to znači da je u tom intervalu prosečan gradijent pozitivan – predikcija u proseku raste kad se ta promenljiva poveća,
- kada ALE kriva opada, prosečan gradijent je negativan – predikcija u proseku opada s povećanjem te promenljive,
- ako je kriva ravna, gradijenti su u proseku nula – odnosno model u tom intervalu ne reaguje na promenu te promenljive.



Slika 2.2: ALE plot za starost vozača i iznos štete

Za isti primer koji je prikazan za ICE metodu, na slici 2.2 prikazan je ALE grafik uticaja starosti na visinu štete.

- **Rast u ranijim godinama:** Između 30. i 40. godine, ALE efekat raste, što ukazuje na povećanje predikcija za osobe u toj starosnoj grupi.
- **Najveći ALE efekat:** Najveći ALE efekat se javlja između 50. i 60. godine, što znači da je u tom periodu starosti predikcija za štete ili zahteve najviša.
- **Pad u starijim godinama:** Nakon 60. godine, ALE efekat počinje da opada, što može ukazivati na smanjenje predikcija za osobe u toj starosnoj grupi.

Poređenje ALE i PDP

- **Globalni i lokalni efekti**: PDP prikazuje globalni efekat promene promenljive x_S na predikciju modela, dok ALE analizira lokalne efekte, odnosno kako promena neke promenljive utiče na predikciju modela za svaku instancu podataka. PDP daje uvid u prosečan efekat promenljive, dok ALE omogućava detaljan uvid u individualne reakcije modela, prikazujući lokalne efekte pre nego što ih akumulira.

- **Računanje i aproksimacija**: PDP se računa kao uslovno očekivanje predikcija modela, dok ALE koristi kumulativnu procenu lokalnih efekata kroz integralnu funkciju. PDP je jednostavniji za izračunavanje, dok ALE zahteva dodatne korake za diskretizaciju promenljivih i računanje lokalnih promena. Takođe, ALE može biti numerički nestabilan u oblastima sa malo podataka.

- **Interakcije između promenljivih**: PDP ne uzima u obzir interakcije između promenljivih i može biti manje tačan u prisustvu jakih interakcija. ALE, sa druge strane, neutralizuje efekte korelisanih promenljivih. Pomoću ove metode, može se bolje identifikovati kako promena jedne promenljive utiče na predikciju modela u odnosu na druge promenljive, naročito kada se koristi za analizu više promenljivih u isto vreme.

- **Preciznost u interpretaciji**: PDP je efikasan za analizu prosečnih efekata ulaznih promenljivih na model, dok je ALE korisniji za analizu individualnih efekata i može pružiti precizniji uvid u to kako model reaguje na promene u različitim promenljivama za različite instance podataka. Međutim, zbog svoje složenosti, ALE može biti teži za interpretaciju u odnosu na PDP.

Poređenje ALE i ICE

- **Lokalni i kumulativni efekti**: Kao i PDP, ICE prikazuje lokalne efekte promena neke promenljive na model, dok ALE računa kumulativne efekte lokalnih promena za svaku promenljivu. ICE daje uvid u to kako promena u promenljivim utiče na predikciju za svaku pojedinačnu instancu podataka, dok ALE prikazuje kumulativnu sumu tih efekata.

- **Računanje i aproksimacija**: ICE se obično računa za svaku instancu u skupu podataka, tako što se beleži kako se predikcija modela menja za različite

vrednosti promenljive x_S. Za svaku instancu računa se predikcija pri različitim vrednostima x_S, dok ALE koristi integralnu funkciju i aproksimira promenu za sve vrednosti promenljive.

- **Prosečni i lokalni efekat**: ALE računa prosečan lokalni efekat, odnosno kako promena neke promenljive utiče na model u proseku, dok ICE daje detaljan uvid u to kako ona utiče na predikciju za svaku pojedinačnu instancu.
- **Korišćenje interakcija**: ALE može bolje da identifikuje interakcije između promenljivih, dok ICE ne uzima u obzir međusobne interakcije promenljivih i fokusira se samo na efekat posmatrane promenljive na model.
- **Vizuelizacija efekata**: ICE prikazuje individualne krive za svaku instancu, što omogućava detaljnu analizu i otkrivanje varijabilnosti među instancama, dok ALE izračunava i akumulira lokalne efekte, pružajući jedinstvenu, glatku krivu koja odražava prosečan uticaj posmatrane promenljive na izlaz modela.
- **Osetljivost na korelacije**: ICE može biti osetljiva na jake korelacije između promenljivih koje analiziramo i komplementarnog skupa promenljivih, što može dovesti do nepredvidivih i ponekad nerealnih efekata, dok ALE neutralizuje ove efekte koristeći lokalne promene unutar realnih intervala podataka, čime se postiže stabilniji prikaz.

2.4 Šeplijeva metoda

Šeplijeva metoda (eng. SHapley Additive exPlanations metoda, SHAP) je pristup interpretaciji modela statističkog učenja zasnovan na teoriji kooperativnih igara. Cilj je dodeliti doprinos svakoj nezavisnoj promenljivoj tako da se ukupna predikcija modela $f(\mathbf{x})$ dekomponuje u baznu vrednost i individualne doprinose:

$$f(\mathbf{x}) = \phi_0 + \sum_{i=1}^{M} \phi_i,$$

gde:

- ϕ_0 predstavlja baznu (osnovnu) vrednost modela, obično matematičko očekivanje modela,
- ϕ_i označava doprinos promenljive x_i (od ukupno M promenljivih) za dati ulaz $\mathbf{x}$.

Šeplijeve vrednosti

U teoriji kooperativnih igara, Šeplijeva vrednost ϕ_i dodeljuje fer doprinos svakom igraču (u našem slučaju, nezavisnoj promenljivoj) na osnovu njihovog marginalnog doprinosa u svim mogućim kombinacijama igrača [19]. Za promenljivu x_i definisana je sledeća formula:

$$\phi_i(f, \mathbf{x}) = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!\,(M - |S| - 1)!}{M!} \left[f_{S \cup \{i\}}(\mathbf{x}_{S \cup \{i\}}) - f_S(\mathbf{x}_S) \right],$$

gde:

- $N = \{1, 2, \ldots, M\}$ je skup indeksa svih nezavisnih promenljivih,
- S je proizvoljan podskup skupa N koji ne sadri indeks i,
- $f_S(\mathbf{x}_S)$ predstavlja očekivanu vrednost funkcije f kada su poznate vrednosti promenljivih sa indeksima iz skupa S, odnosno:

  $$f_S(\mathbf{x}_S) = \mathbb{E}_{\mathbf{X}_C}\big[f(\mathbf{x}_S, \mathbf{X}_C)\big],$$

  pri čemu je C komplement skupa S u N.

Osnovna svojstva Šeplijevih vrednosti

Težinski koeficijenti:

$$\frac{|S|!\,(M - |S| - 1)!}{M!}$$

osiguravaju da Šeplijeve vrednosti imaju sledeće osobine, što garantuje jedinstvenu i doslednu dekompoziciju predikcije:

1. **Efikasnost:**
   $$\phi_0 + \sum_{i=1}^{M} \phi_i = f(\mathbf{x}).$$
2. **Simetričnost:** Ako dve promenljive x_i i x_j daju isti doprinos u svim podskupovima, onda je $\phi_i = \phi_j$.
3. **Neutralnost:** Ako dodavanje promenljive x_i ne utiče na izlaz modela, tj. $f_{S \cup \{i\}} = f_S$ za svako $S \subseteq N \setminus \{i\}$, onda je $\phi_i = 0$.
4. **Aditivnost:** Za dve funkcije f i g, važi:
   $$\phi_i(f + g, \mathbf{x}) = \phi_i(f, \mathbf{x}) + \phi_i(g, \mathbf{x}).$$

Ove aksiome čine Šeplijevu metodologiju konzistentnom i fer, omogućavajući interpretaciju doprinosa svakog ulaza na predikciju.

Izazovi u praksi

Izračunavanje tačnih Šeplijevih vrednosti zahteva sumiranje preko svih 2^{M-1} podskupova, što postaje računski neizvodljivo za veći broj promenljivih. Zbog toga se u praksi često koriste različite aproksimativne metode. Neke od njih koje ćemo ovde objasniti su Monte Karlo pristup, Šeplijeve vrednosti zasnovane na jezgrima (eng. Kernel SHAP) i Šeplijeve vrednosti zasnovane na stablima (eng. Tree SHAP) [11].

Monte Karlo pristup

Umesto da razmatramo sve moguće podskupove, možemo nasumično birati *permutacije* svih promenljivih. U svakoj permutaciji posmatramo *marginalni doprinos* promenljive i u trenutku kada se prvi put pojavi u permutaciji.

Neka je $\mathcal{M} = \{1, 2, \ldots, p\}$ skup indeksa svih ulaznih promenljivih u modelu, i neka je π permutacija skupa $\mathcal{M}$. Neka $pos(i, \pi)$ označava poziciju promenljive x_i u permutaciji π, a $Pred(i, \pi)$ skup indeksa promenljivih koji se nalaze *ispred* i u permutaciji π (pojavljuju se pre i u redosledu π). Tada je marginalni doprinos promenljive x_i u toj permutaciji:

$$\Delta_i(\pi) = v\Big(Pred(i, \pi) \cup \{i\}\Big) - v\Big(Pred(i, \pi)\Big),$$

gde je

$$v(S) = \mathbb{E}\left[f\mathbf{x}_S, \mathbf{x}_C)\right].$$

Šeplijeva vrednost se može aproksimirati uzimanjem proseka marginalnih doprinosa preko R nasumično odabranih permutacija:

$$\widehat{\phi_i} = \frac{1}{R} \sum_{r=1}^{R} \Delta_i(\pi^{(r)}).$$

Broj permutacija R bira se u skladu sa željenom tačnošću aproksimacije i dostupnim resursima. U praksi se često koristi od nekoliko stotina do nekoliko hiljada permutacija.

Međutim, ako posmatramo R permutacija, računanje zahteva da za svaku permutaciju izračunamo vrednost funkcije $v(\cdot)$ *dva puta* (sa i bez promenljive x_i) za

svaku relevantnu poziciju. U zavisnosti od definicije $v(\cdot)$ (npr. model crne kutije koji treba ponovo trenirati), to može biti i dalje zahtevno, ali je značajno efikasnije od algoritama sa vremenskom složenošću $\mathcal{O}(2^p)$.

Šeplijeve vrednosti zasnovane na jezgrima

Pristup Šeplijevih vrednosti zasnovanih na jezgrima je pristup nezavisan od tipa modela, koji uvodi ideju da se Šeplijeve vrednosti mogu posmatrati kao rešenje problema težinske linearne regresije.

Posmatramo izraz:

$$\min_{\phi_0,\phi_1,\ldots,\phi_p} \sum_{S\subseteq\mathcal{M}} \pi(S)\Big(f(\mathbf{x}_S) - \big[\phi_0 + \sum_{j\in S}\phi_j\big]\Big)^2, \tag{2.3}$$

gde je $f(\mathbf{x}_S)$ vrednost modela kada su poznate samo promenljive sa indeksima iz S, dok su ostale „maskirane"[1]. Funkcija $\pi(S)$ je *Šeplijevo jezgro*, i prestavlja težinski koeficijent podskupa S:

$$\pi(S) = \frac{p-1}{\binom{p}{|S|}\,|S|(p-|S|)},$$

koji je konstruisan tako da zadovoljava osnovne Šeplijeve aksiome, što se može pokazati.

Neka je X matrica svih mogućih binarnih vektora dužine p, dimenzija $2^p \times p$. Koristimo Šeplijevo jezgro za računanje Šeplijevih vrednosti primenom težinske linearne regresije:

$$\phi = (X^T W X)^{-1} X^T W y,$$

gde je W dijagonalna matrica sa težinama Šeplijevog jezgra za svaki red matrice X, a $y_i = f_x(S_i)$ su vrednosti izlazne funkcije za skup promenljivih sa indeksima iz S_i.

Primetimo da je $\pi(p) = \pi(0) = \infty$, pa je W beskonačna za sve redove od X koji se sastoje samo od nula ili samo od jedinica. Međutim, ako postavimo sve ove beskonačne težine na veliku konstantu, onda je $(X^T W X)^{-1} = \frac{1}{M-1} I + cJ$, za neku pozitivnu konstantu c (I je jedinična matrica, a J matrica jedinica). Kad $c \to \infty$, prethodni izraz postaje:

$$(X^T W X)^{-1} = I + \frac{1}{M-1}(I - J).$$

[1]Maskiranje promenljivih se odnosi na postupak u kome se određenim promenljivama (koji nemaju indeks u podskupu S´) S) dodeljuju neutralne vrednosti – kao da nisu poznate modelu. Najčešće to podrazumeva i da ih zamenjujemo prosečnim vrednostima iz podataka, ili ih nasumično uzorkujemo iz marginalne raspodele (nezavisno od ostatka podataka).

Izraz X^TW predstavlja matricu u kojoj su svi elementi jedinice u X^T zamenjeni vrednostima $\pi(S)$, gde je $|S|$ broj jedinica u toj koloni matrice X^T. Množenjem X^TW sa $(X^TWX)^{-1}$ dobija se matrica težina koje se primenjuju na izlazne vrednosti funkcije y. Ako posmatramo samo Šeplijevu vrednost za jednu promenljivu ϕ_j, tada je dovoljno razmatrati samo jedan red ove $2^p \times p$ matrice, što je ekvivalentno korišćenju samo j-tog reda iz $(X^TWX)^{-1}$. Kada to uradimo, dobijamo da je vrednost težine za red i:

$$\pi(s_i)\Big[\mathbb{1}_{\{X_{i,j}=1\}} - \frac{s_i - \mathbb{1}_{\{X_{i,j}=1\}}}{p-1}\Big] = \frac{p-1}{\binom{p}{s_i} s_i(p-s_i)}\mathbb{1}_{\{X_{i,j}=1\}} - \frac{(s_i - \mathbb{1}_{\{X_{i,j}=1\}})}{\binom{p}{s_i} s_i(p-s_i)}$$

$$\begin{aligned}
&= \frac{(p-1)(p-s_i)!s_i!}{p!s_i(p-s_i)}\mathbb{1}_{\{X_{i,j}=1\}} - \frac{(s_i - \mathbb{1}_{\{X_{i,j}=1\}})(p-s_i)!s_i!}{p!s_i(p-s_i)} \\
&= \frac{(p-1)(p-s_i-1)!(s_i-1)!}{p!}\mathbb{1}_{\{X_{i,j}=1\}} \\
&- \frac{(s_i - \mathbb{1}_{\{X_{i,j}=1\}})(p-s_i-1)!(s_i-1)!}{p!} \\
&= \frac{(p-s_i-1)!(s_i-1)!}{p!}\Big[(p-1)\mathbb{1}_{\{X_{i,j}=1\}} - (s_i - \mathbb{1}_{\{X_{i,j}=1\}})\Big],
\end{aligned}$$

gde je s_i broj jedinica u i-tom redu matrice X, i $\mathbb{1}_{\{X_{i,j}=1\}}$ je indikator da li je $X_{i,j} = 1$.

Kada je $\mathbb{1}_{\{X_{i,j}=1\}} = 0$, dobijamo:

$$-\frac{(p-s_i-1)!s_i!}{p!}.$$

Kada je $\mathbb{1}_{\{X_{i,j}=1\}} = 1$, dobijamo:

$$\frac{(p-s_i-1)!(s_i-1)!}{p!}[(p-1)-(s_i-1)] = \frac{(p-s_i-2)!(s_i-1)!}{p!}.$$

Uzimanjem skalarnog proizvoda ovih vrednosti sa y dolazimo do sledeće jednačine:

$$\phi_j = \sum_{S \subseteq \mathcal{M}\setminus\{j\}} \frac{(p-s_i-1)! \cdot s_i!}{p!}[f_{S\cup\{j\}}(x_{S\cup\{j\}}) - f_S(x_S)].$$

što predstavlja klasičan oblik za računanje Šeplijevih vrednosti ϕ_j (gde je $\mathcal{M}$ skup indeksa svih promenljivih).

Metoda zatim rešava linearnu regresiju da dobije ϕ_j.

Pojednostavljen algoritam implementacije u praksi je:

1. Generišemo *uzorke* podskupova S.

2. Za svaki podskup izračunamo $f(\mathbf{x}_S)$ (npr. tako što „maskiramo" nebitne promenljive) i dobijamo par $\big(\mathbf{x}_S, f(\mathbf{x}_S)\big)$.

3. U linearnu regresiju uključimo težine $\pi(S)$.

4. Minimizacijom težinske regresije, dobijamo koeficijente $\phi_0, \phi_1, \ldots, \phi_p$ koji aproksimiraju Šeplijeve vrednosti.

Međutim, i ovaj pristup ima svoje prednosti i ograničenja:

- **Prednost**: Može se primeniti na bilo koji model crne kutije, odnosno agnostičan je na tip modela.
- **Ograničenje**: Za veće p i broj uzoraka podskupova, i dalje može biti računski zahtevno.

Šeplijeve vrednosti zasnovane na stablima

Ovaj pristup koristi strukturu *stabala odluke* (i ansambala kao što su slučajne šume ili gradijentno pojačavanje) kako bi efikasno izračunao Šeplijeve vrednosti.

Ideja je da se za svaku permutaciju promenljivih i svaki čvor stabla izračuna verovatnoća da instanca dođe do tog čvora, pod pretpostavkom da su poznate samo vrednosti promenljivih sa indeksima iz podskupa S. Na osnovu tih verovatnoća kombinuju se predikcije iz listova stabla, kako bi se izračunao doprinos svake promenljive prema Šeplijevoj metodi.

Ovaj pristup nije aproksimacija, već je *tačan* za modele stabala, što ga čini vrlo efikasnim i čestim izborom za `XGBoost` (ekstremno gradijentno ubrzano stablo, eng. Extreme Gradient Boosting) i `LightGBM` (laka implementacija gradijentnog ubrzavanja, eng. Light Gradient Boosting Machine), popularnim Python bibliotekama za modele zasnovane na stablima, i druge slične modele.

Navedimo pojednostavljen prikaz rekurzivnog algoritma kojim se u praksi računaju Šeplijeve vrednosti:

1. Za svako stablo u ansamblu i svaku permutaciju, prolazimo kroz čvorove na osnovu poznatog podskupa promenljivih sa indeksima iz skupa S.

2. U svakoj grani računamo verovatnoću prelaska u levo ili desno podstablo, $\mathbb{P}(\text{grana} \mid S)$.

3. Kada dođemo do lista, znamo njegovu vrednost.

4. Agregacijom dobijamo doprinos promenljive x_i na nivou stabla, a zatim sabiranjem preko svih stabala dobijamo konačne ϕ_i.

Još neke metode koje se koriste za izračunavanje ili aproksimaciju Šeplijevih vrednosti su aproksimacija zasnovana na neuronskim mrežama (eng. Deep SHAP), Šeplijeve vrednosti zasnovane na linearnim modelima (eng. Linear SHAP), Šeplijeve vrednosti zasnovane na hijerarhijskoj strukturi unutar stabala (Partition SHAP) [11].

Praktična primena Šeplijeve metode biće detaljno prikazana u glavi 3.

2.5 Poređenje Šeplijeve metode sa PDP, ICE i ALE metodama

- **PDP** metoda računa prosečan efekat promenljive x_S na izlaz modela marginalizacijom ostalih promenljivih. Međutim, PDP može proizvesti nerealne kombinacije vrednosti kada su promenljive međusobno korelisani. Šeplijeve vrednosti, s druge strane, dodeljuju doprinos svakoj promenljivoj na osnovu svih mogućih kombinacija, čime se postiže dosledna dekompozicija predikcije.

- **ICE** metoda prikazuje kako se predikcija modela menja za svaku instancu kada se menja x_S, čime se otkriva heterogenost efekata među instancama. Šeplijeve vrednosti takođe omogućavaju interpretaciju na nivou pojedinačnih instanci, ali osiguravaju aditivnost i fer raspodelu doprinosa pomoću aksioma Šeplijevih vrednosti.

- **ALE** metoda računa lokalne promene u predikciji kada se x_S menja i akumulira te promene, što omogućava robusniji prikaz globalnog efekta u prisustvu korelacija među promenljivama. Za razliku od ALE, Šeplijeva metoda daje detaljnu dekompoziciju predikcije na nivou pojedinačnih promenljivih uz strogu matematičku osnovu koja zadovoljava zadate aksiome.

Možemo zaključiti da Šeplijeva metoda predstavlja moćan alat za interpretaciju modela statističkog učenja, jer omogućava preciznu i fer dekompoziciju predikcija u doprinos pojedinačnih promenljivih. Zahvaljujući matematičkoj pozadini u teoriji kooperativnih igara i aksiomima (efikasnost, simetrija, neutralnost, linearnost), Šeplijeva metoda nudi doslednu interpretaciju koja kombinuje prednosti lokalnih (kao

kod ICE) i globalnih (kao kod ALE) metoda, istovremeno rešavajući nedostatke PDP-a.

Glava 3

Primer u praksi

U ovom poglavlju, bavićemo se praktičnom primenom nelinearnih modela statističkog učenja u kontekstu industrije osiguranja. Koristićemo bazu podataka osiguranika dobrovoljnog zdravstvenog osiguranja jednog od osiguravača na tržištu u Srbiji.

Nelinearni modeli, poput modela zasnovanih na stablima odlučivanja i modela neuronskih mreža, omogućavaju bolje razumevanje kompleksnih i nelinearnih odnosa među promenljivama u podacima iz osiguranja, u odnosu na tradicionalne linearne pristupe.

U ovom delu rada, prvo su na različite tipove modela primenjene prethodno opisane metode interpretacije, poput grafika parcijalne zavisnosti (PDP), metode individualnog uslovnog očekivanja (ICE), metode akumulranih lokalnih efekata (ALE) i Šeplijeve metode (SHAP). Ovo nam je omogućilo da uočimo njihove specifičnosti i ograničenja. Kako su svi modeli pokazali određene slabosti, bilo u performansama, bilo u interpretabilnosti, pristup je naknadno usmeren ka konstruisanju što boljeg prediktivnog modela. U njegovom formiranju korišćeni su uvidi iz prethodnih analiza, uključujući informacije o važnosti promenljivih, interakcijama i obrascima ponašanja modela. U završnom delu rada predstavljene su interpretacije upravo konačnog, unapređenog modela.

Metode interpretacije omogućavaju da se model razvije i da se učini razumljivim za krajnje korisnike u industriji osiguranja. Ovakav pristup je od ključne važnosti za osiguravajuće kompanije jer omogućava ne samo bolje modeliranje rizika, već i donošenje informisanih i transparentnih odluka u vezi sa strategijama upravljanja rizicima i optimizacijom premija.

3.1 Baza podataka

Radićemo sa bazom osiguranika dobrovoljnog zdravstvenog osiguranja jednog od osiguravača na tržištu u Srbiji. Pogledajmo detaljnije koji su nam podaci dostupni.

```
Broj redova: 30069
Broj nedostajućih vrednosti: 201
Kolone: Index(['BROJ_POJEDINACNE_POLISE', 'Velike polise', 'Drugo mišljenje',
       'Fizikalna terapija', 'Komplementarna medicina',
       'Oftalmološki pregled i usluge', 'Posebno pokriće u slučaju tumora',
       'Prepisivanje lekova od strane ovlašćenog lekara', 'Sistematski',
       'Stomatološke usluge', 'Vanbolničko i bolničko lečenje',
       'Vanbolničko lečenje',
       'Zdravstvena zaštita trudnica + Troškovi porođaja i zdravstvena zaštita novorođenčeta',
       'Opšta participacija', 'Opšta participacija opis',
       'Bel Medic participacija', 'Bel Medic participacija opis',
       'Medigroup participacija', 'Medigroup participacija opis',
       'Van mreže participacija', 'Van mreže participacija opis',
       'Dodatna participacija opis', 'Ugovarač', 'Ugovarač mesto', 'Region',
       'Veliki grad', 'Organizaciona jedinica', 'Interna/Eksterna prodaja',
       'Datum knjiženja', 'IDLICE', 'Stranac', 'Pol osiguranika',
       'Starost osiguranika', 'JMBG', 'Broj šteta', 'IME', 'PREZIME',
       'BROJ_KARTICE'],
      dtype='object')
```

Slika 3.1: Pregled osnovnih informacija o bazi podataka

Dakle, baza ima 30.069 redova, i 19 kolona.

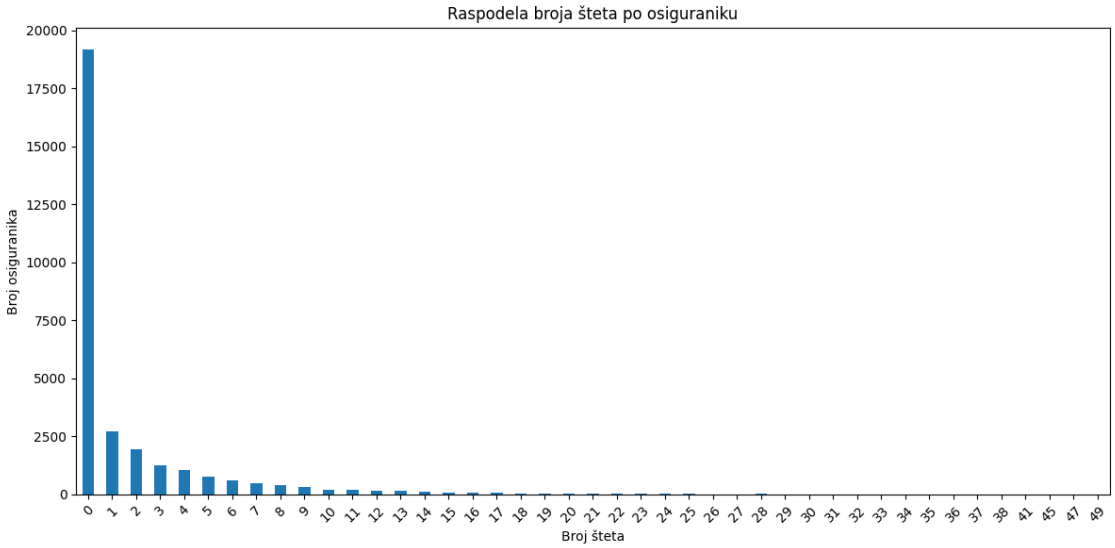
Postoji 201 red koji sadrži nedostajuće vrednosti, koje ćemo isključiti jer čine manje od 1% svih podataka.

Sada ćemo definisati kolone koje ćemo koristiti, formirati zavisne i nezavisne promenljive, i uraditi osnovne transformacije. Neka je:

- Y = 'Broj šteta' zavisna, odnosno ciljna promenljiva, koja predstavlja broj šteta koje je osiguranik imao,
- $\mathbf{X} = (X_1, X_2, \ldots, X_{20})$ predstavlja vektor nezavisnih promenljivih, gde je:
  - X_1 = 'Drugo mišljenje' indikator (0 ili 1) koji označava da li je ugovoreno pokriće usluge drugog mišljenja
  - X_2 = 'Fizikalna terapija' indikator (0 ili 1) koji označava da li je ugovoreno pokriće usluge fizikalne terapije

- X_3 = 'Komplementarna medicina' indikator (0 ili 1) koji označava da li je ugovoreno pokriće komplementarne medicine
- X_4 = 'Oftalmološki pregled i usluge' indikator (0 ili 1) koji označava da li je ugovoreno pokriće oftalmoloških usluga
- X_5 = 'Posebno pokriće u slučaju tumora' indikator (0 ili 1) koji označava da li je ugovoreno pokriće u slučaju tumora
- X_6 = 'Prepisivanje lekova od strane ovlašćenog lekara' indikator (0 ili 1) koji označava da li je ugovoreno pokriće ove usluge
- X_7 = 'Sistematski' indikator (0 ili 1) koji označava da li osiguranik ima ugovoren sistematski pregled
- X_8 = 'Stomatološke usluge' indikator (0 ili 1) koji označava da li je ugovoreno pokriće stomatoloških usluga
- X_9 = 'Vanbolničko i bolničko lečenje' indikator (0 ili 1) koji označava da li je ugovoreno pokriće Vanbolničko i bolničko lečenje
- X_{10} = 'Vanbolničko lečenje' indikator (0 ili 1) koji označava da li je ugovoreno pokriće Vanbolničko lečenje
- X_{11} = 'Zdravstvena zaštita trudnica + Troškovi porođaja' indikator (0 ili 1) koji označava da li je ugovoreno pokriće Zdravstvena zaštita trudnica + Troškovi porođaja
- X_{12} = 'Opšta participacija' procenat obaveznog učešća osiguranika u šteti u svim medicinskim ustanovana
- X_{13} = 'Bel Medic participacija' procenat obaveznog učešća osiguranika u šteti u medicinskim ustanovana Bel Medic
- X_{14} = 'Medigroup participacija' procenat obaveznog učešća osiguranika u šteti u medicinskim ustanovana Medigroup
- X_{15} = 'Van mreže participacija' procenat obaveznog učešća osiguranika u šteti u medicinskim ustanovana van mreže ugovarača
- X_{16} = 'Stranac' oznaka da li je osiguranik strani državljanin (NE/DA)
- X_{17} = 'Pol osiguranika' pol osiguranika (M ili F)
- X_{18} = 'Starost osiguranika' - Godine osiguranika
- X_{19} = 'Veliki grad' oznaka da li je osiguranik iz Beograda, Novog Sada, Niša, ili nekog drugog mesta

– X_{20} = 'Interna/Eksterna prodaja' prodajna mreža koja je ugovorila polisu osiguranika



Slika 3.2: Raspodela broja šteta po osiguraniku

Na slici 3.2 prikazana je raspodela broja šteta po osiguraniku. Uočava se izražena asimetričnost, sa dominacijom osiguranika koji nisu imali nijednu prijavljenu štetu. Ovakva raspodela je karakteristična za podatke iz oblasti osiguranja.

Nakon vrednosti od 5–6 šteta, broj osiguranika postaje veoma mali, a ekstremi sa više od 20 šteta su retki i predstavljaju izuzetke. Ovo će biti značajno u nastavku kada budemo gradili model.

Uklonićemo redove sa nedostajućim podacima i uraditi referentno (eng. dummy) kodiranje za kategoričke promenljive.

Kodiranje sa referentnom kategorijom (eng. *dummy coding*) predstavlja tehniku pretvaranja kategoričkih promenljivih u numerički format, pogodan za korišćenje u regresionim modelima.

Za kategoričku promenljivu koja ima k različitih vrednosti, kreira se $k-1$ binarnih (eng. dummy) promenljivih. Jedna od kategorija se izostavlja i koristi se kao referentna kategorija. Koeficijenti modela za ostale kategorije tada izražavaju efekat tih kategorija u odnosu na referentnu. Na primer, ako posmatramo promenljivu *Veliki grad* sa četiri kategorije (Beograd, Novi Sad, Niš, Ostalo), i izaberemo Beograd kao referentnu kategoriju, dobijamo sledeće kodiranje:

Veliki grad	Novi Sad	Niš	Ostalo
Beograd	0	0	0
Novi Sad	1	0	0
Niš	0	1	0
Ostalo	0	0	1

Tabela 3.1: Referentno kodiranje kategoričke promenljive Veliki grad

Ocenjena regresiona funkcija u ovom slučaju je:

$$\hat{y} = \beta_0 + \beta_1 \cdot \text{Novi Sad} + \beta_2 \cdot \text{Niš} + \beta_3 \cdot \text{Ostalo},$$

gde je:

- β_0 – očekivana vrednost za referentnu kategoriju (Beograd),
- β_1 – razlika u odnosu na Beograd ako je mesto Novi Sad,
- β_2 – razlika u odnosu na Beograd ako je mesto Niš,
- β_3 – razlika u odnosu na Beograd ako je vrednost Ostalo.

Ovo kodiranje omogućava interpretabilnost modela i sprečava multikolinearnost, koja bi nastala kada bi se uključilo svih k kategorija (što se dešava u klasičnom one-hot kodiranju, o kojem će biti reči u nastavku rada).

Nakon toga je slučajnim izborom izvršena podela podataka na test i trening skup, gde se u test skupu nalazi 20% svih instanci.

Standardizacija podataka nije primenjena jer su ulazne promenljive većinski kategoričke, dok su preostale numeričke promenljive izražene u procentima, pa njihovo skaliranje nije bilo neophodno.

3.2 Linearan model

Prvo ćemo na trening skupu napraviti i trenirati jednostavan linearan model. Formula linearnog modela data je u nastavku:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_{20} + \varepsilon,$$

gde je:

- Y broj šteta,

- $X_1, X_2, \ldots, X_{20}$ nezavisne promenljive,
- β_0 slobodan član,
- $\beta_1, \ldots, \beta_p$ koeficijenti regresije,
- ε slučajna greška (razlika između stvarne vrednosti i vrednosti predviđene modelom).

Neka je y_i stvarna vrednost zavisne promenljive, $\hat{y}_i$ predikcija modela, $\bar{y}$ srednja vrednost stvarnih vrednosti zavisne promenljive, i n ukupan broj posmatranja, $i \in \{1, 2, ..., n\}$. Model smo na test skupu evaluirali na osnovu sledećih metrika:

- Srednja apsolutna greška (MAE) – prosečna apsolutna razlika između stvarnih vrednosti i predikcija (reziduala). Manje je osetljiva na ekstremne greške, i računa se sledećom formulom:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i| \,. \tag{3.1}$$

- Srednja kvadratna greška (MSE) – prosečna kvadratna razlika između stvarnih vrednosti i predikcija. Zbog kvadriranja, veće greške imaju značajno veći uticaj. Računa se sledećom formulom:

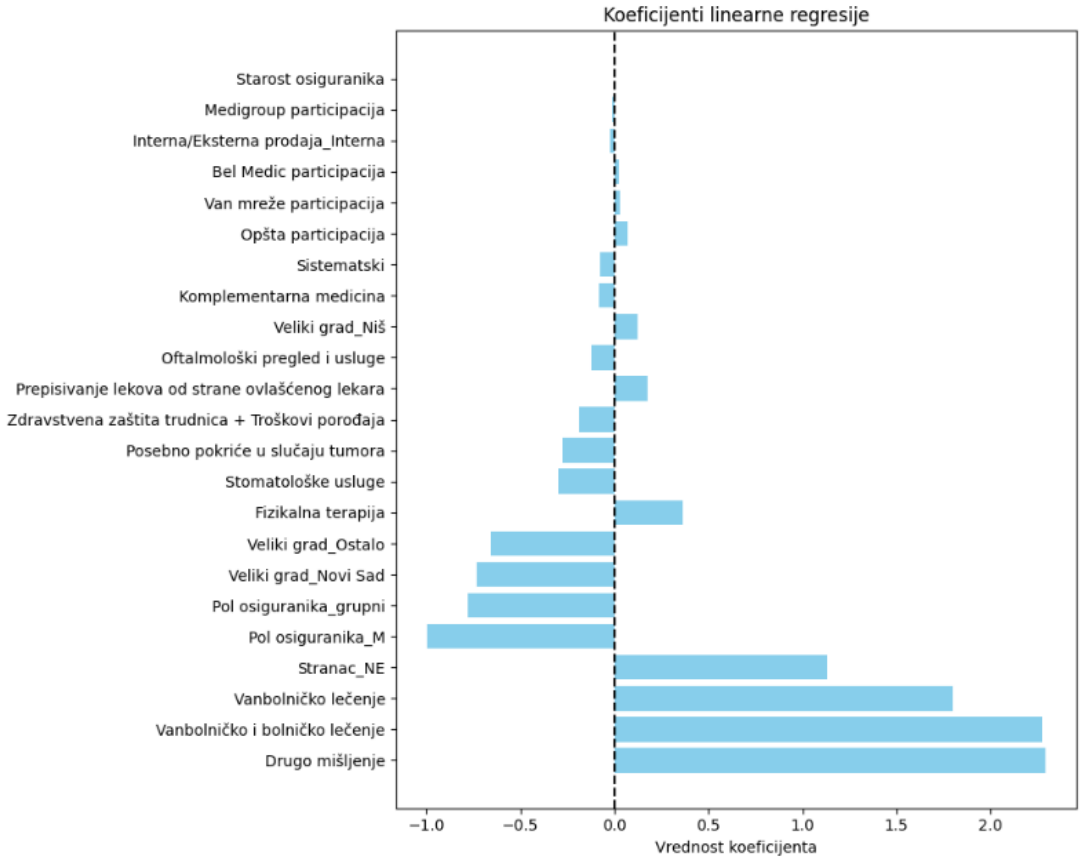
$$\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \,. \tag{3.2}$$

- Koeficijent determinacije (R^2) – meri udeo varijabilnosti ciljne promenljive koji je objašnjen modelom. Vrednosti bliže 1 ukazuju na bolju prilagođenost modela podacima. Računa se preko sledeće formule:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}.$$

Metrika	**Vrednost**
MSE	12.213752
MAE	2.162788
R^2	0.051493

Tabela 3.2: Evaluacione metrike linearnog modela



Slika 3.3: Koeficijenti linearne regresije

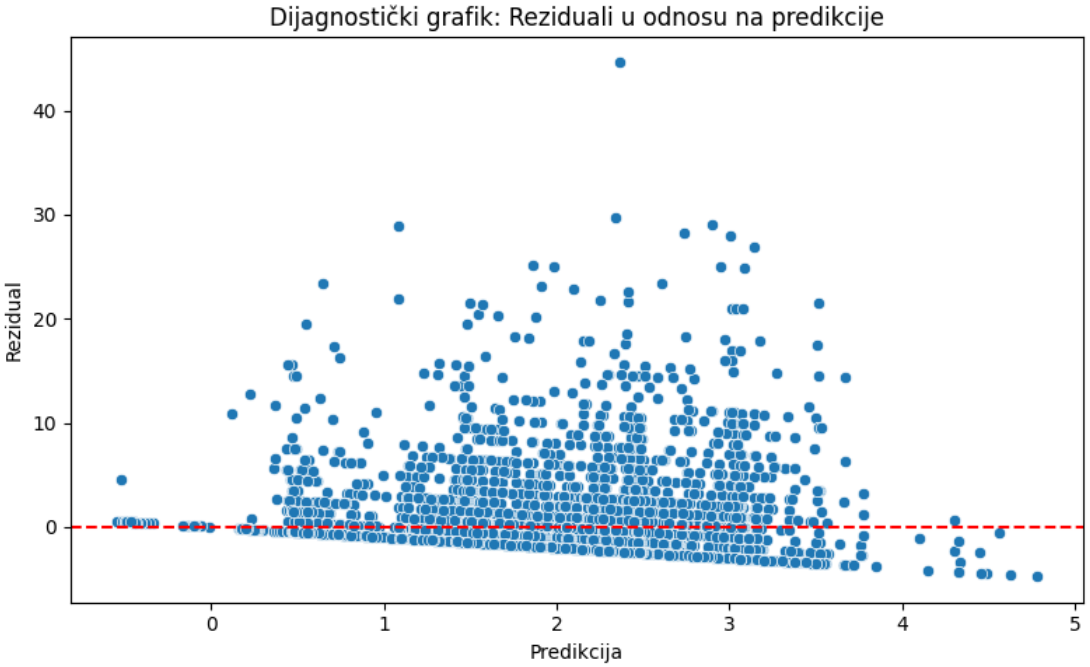
Koeficijenti linearne regresije (slika 3.3) prikazani su u obliku stubičastog grafika, gde svaka traka odgovara jednoj promenljivoj iz modela. Dužina i pravac trake ukazuju na jačinu i smer uticaja na zavisnu promenljivu, odnosno broj šteta.

Pozitivni koeficijenti ukazuju na to da povećanje vrednosti promenljive dovodi do povećanja predikcije (povećanje broja šteta), dok negativni koeficijenti znače da povećanje vrednosti promenljive smanjuje predikciju (smanjuje broja šteta).

Drugo mišljenje i vanbolničko i bolničko lečenje imaju najveći pozitivan uticaj, što znači da su u ovom modelu osiguranici koji imaju ugovorenu uslugu drugog mišljenja ili vanbolničko lečenje skloniji većem broju šteta.

Sa druge strane, promenljive kao što su Veliki_grad_Novi_Sad i Pol osiguranika_M imaju negativne koeficijente, što sugeriše da su muški osiguranici i osiguranici iz Novog Sada manji nosioci rizika u ovom modelu. Promenljivu Pol osiguranika_grupni ne posmatramo zbog veoma malog broja osiguranika u ovoj kategoriji.

U tabeli 3.5 su prikazane evaluacione metrike modela. $R^2 = 0.051$ znači da model objašnjava samo 5.1% varijacije ciljne promenljive. Ovo je nizak R^2, što može ukazivati da je model možda suviše jednostavan za podatke, pa ga dalje nećemo razmatrati. Na slici 3.4 prikazan je dijagnostički grafik (reziduali u odnosu na predikcije), koji dodatno potvrđuje lošu prilagođenost modela. Reziduali nisu raspoređeni nasumično u odnosu na predikciju i prisutna su sistematska odstupanja.



Slika 3.4: Dijagnostički grafik linearnog modela

3.3 Uopšten linearan model

Sledeći model koji ćemo napraviti jeste uopšten linearan model, konkretno *Poasonov regresioni model*, koji se koristi u slučajevima kada je zavisna promenljiva celobrojna i diskretna, kao što je broj prijavljenih šteta.

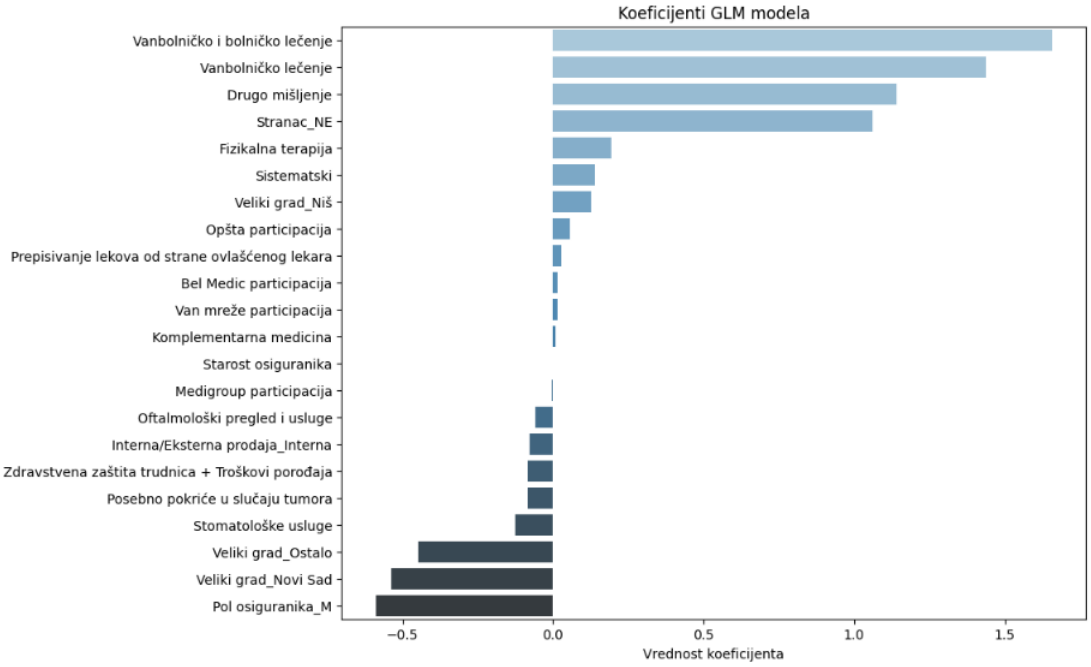
Poasonov model pripada klasi uopštenih linearnih modela, i pretpostavlja da zavisna promenljiva ima Poasonovu raspodelu, odnosno da logaritam očekivane vrednosti zavisi linearno od ulaznih promenljivih:

$$\mathbb{E}[Y \mid \mathbf{X}] = \mu = e^{\beta_0 + \sum_{i=1}^{p} \beta_i X_i}, \text{ odnosno}$$

$$\log(\mathbb{E}[Y \mid \mathbf{X}]) = \log(\mu) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_{20} X_{20},$$

gde je:

- Y broj šteta,
- $X_1, X_2, \ldots, X_{20}$ nezavisne promenljive,
- β_0 slobodan član,
- $\beta_1, \ldots, \beta_{20}$ koeficijenti modela.



Slika 3.5: Koeficijenti uopštene linearne regresije

Model je treniran na trening skupu, a na slici 3.5 je prikazan bar-plot njegovih koeficijenata. Svaki stubić na grafiku predstavlja jednu promenljivu iz modela, dok njena dužina pokazuje vrednost procenjenog koeficijenta regresije za tu promenljivu. Vrednosti koeficijenata predstavljaju uticaj pojedinačnih promenljivih na logaritmovanu vrednost očekivanog broja šteta, tj. na $\log(\mu)$, gde je μ očekivani broj šteta za jednog osiguranika.

Nezavisne promenljive uz koje se nalaze najveći pozitivni koeficijenti su:

- Vanbolničko i bolničko lečenje – pokazuje snažnu pozitivnu povezanost sa brojem šteta, što je očekivano imajući u vidu da osiguranici sa ovakvim pokrićima češće koriste osiguranje.
- Drugo mišljenje, Stranac_NE i Fizikalna terapija – takođe imaju pozitivan uticaj, sugerišući da osiguranici sa ovim karakteristikama češće podnose zahteve za štetom.

Uz neke od promenljivih nalaze se negativni koeficijenti:

- Pol osiguranika_M – ukazuje da muškarci, u poređenju sa ženama, imaju manji očekivani broj šteta.
- Veliki grad_Novi Sad i Veliki grad_Ostalo – imaju negativan uticaj u odnosu na referentni grad (najverovatnije Beograd).
- Stomatološke usluge i posebno pokriće u slučaju tumora – takođe pokazuju slabiji negativan uticaj.

Određene promenljive, poput starosti osiguranika, komplementarne medicine i Bel Medic participacije, imaju koeficijente bliske nuli, što ukazuje na to da nemaju izražen uticaj na očekivani broj šteta u ovom modelu.

Ukoliko se model koristi za aktuarske svrhe ili odlučivanje o strukturi osiguravajućih paketa, rezultati poput značajnog uticaja pojedinih pokrića mogu ukazivati na mogućnosti za segmentaciju klijenata, optimizaciju premija ili prilagođavanje proizvoda.

Takođe, važno je imati u vidu da ovi rezultati nisu uzročni – ne može se tvrditi da npr. „pol uzrokuje" manji broj šteta, već samo da postoji uočena razlika između kategorija u okviru posmatranih podataka.

Performanse ovog modela procenjene su na test skupu korišćenjem standardnih regresionih metrika - srednje kvadratne, srednje apsolutne greške i srednje devijacije. Srednja devijacija pokazuje prosečno odstupanje modela u odnosu na zasićeni model[1] po instanci. Manja vrednost srednje devijacije ukazuje na bolje slaganje modela sa podacima, a računa se prema sledećoj formuli:

$$D = \frac{2}{n} \sum_{i=1}^{n} \left[y_i \log \left(\frac{y_i}{\hat{\mu}_i} \right) - (y_i - \hat{\mu}_i) \right],$$

gde je:

[1] model koji savršeno predviđa posmatrane vrednosti

- y_i broj šteta za i-tu instancu,
- $\hat{\mu}_i$ predikcija modela za i-tu instancu,
- broj instanci u skupu na kojem računamo meru (u našem slučaju test skup).

Dobijeni rezultati su prikazani u tabeli 3.5:

Metrika	**Vrednost**
MSE	12.271251
MAE	2.151479
Srednja devijacija	4.290343

Tabela 3.3: Evaluacione metrike uopštenog linearnog modela

Metrika MAE od 2.15 ukazuje da model u proseku greši za nešto više od dve štete po osiguraniku. MSE iznosi 12.27, što je viša vrednost jer kvadrira greške i time dodatno penalizuje veća odstupanja između predikcija i stvarnih vrednosti.

Srednja devijacija iznosi 4.29 na test skupu, dok je prosečan broj šteta po osiguraniku 1.68. Budući da je devijacija znatno veća od prosečne vrednosti ciljne promenljive, možemo zaključiti da Poasonov model u ovom slučaju pokazuje slabu prilagođenost podacima. Ovo ukazuje da predikcije modela značajno odstupaju od stvarnih vrednosti broja šteta, što može biti posledica visoke varijabilnosti i kompleksnosti podataka.

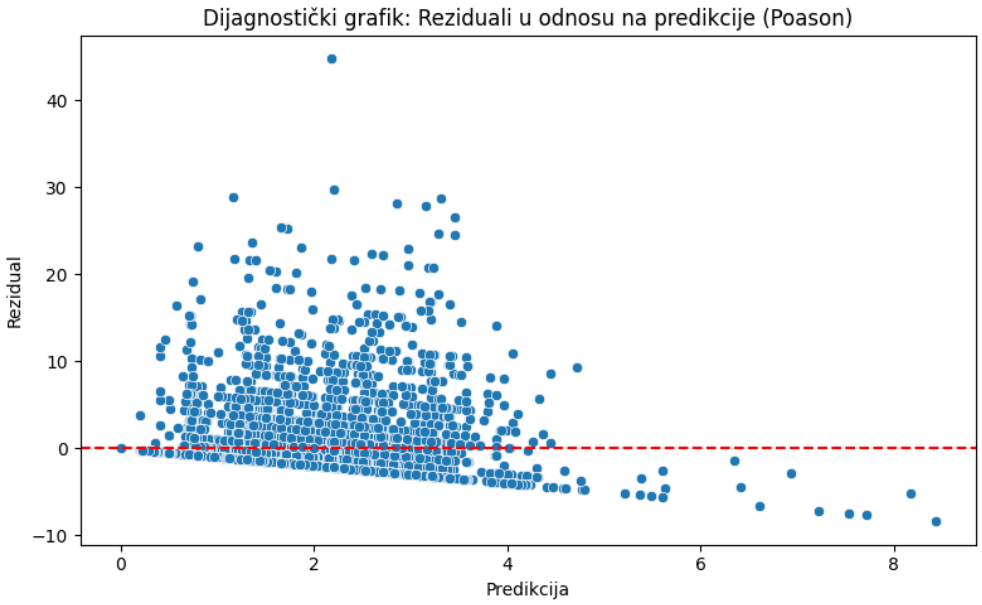
Na slici 3.6 prikazan je dijagnostički grafik, koji potvrđuje lošu prilagođenost modela.

Ipak, model pruža korisne uvide u uticaj pojedinačnih promenljivih.

3.4 Model slučajnih šuma

Kao sledeći korak, primenićemo nelinearne modele, koji su u nekim slučajevima sposobniji da uoče složenije odnose među promenljivama. Počećemo sa *modelom slučajnih šuma*(eng. Random Forest model).

Model slučajnih šuma je metoda koja kombinuje veći broj stabala odlučivanja kako bi se dobio stabilniji i tačniji model. Svako stablo se trenira na različitom podskupu podataka i koristi nasumično izabrane promenljive prilikom grananja. U regresionim zadacima, krajnja predikcija se dobija kao prosečna vrednost predikcija svih pojedinačnih stabala.



Slika 3.6: Dijagnostički grafik Puasonovog modela

Za nelinearne modele, poput slučajnih šuma i neuronskih mreža, korišćeno je one-hot kodiranje kategoričkih promenljivih. Za razliku od referentnog (eng. dummy) kodiranja, koje izostavlja jednu kategoriju kao referentnu, one-hot kodiranje eksplicitno prikazuje sve kategorije, omogućavajući modelima da ravnopravno uče odnose između svih vrednosti bez gubitka informacije. Budući da ovi modeli nisu osetljivi na multikolinearnost, nema potrebe za izbacivanjem kategorija radi stabilnosti modela. Kao primer, one-hot kodiranje promenljive Veliki grad je prikazano u tabeli 3.4

Beograd	1	0	0	0
Novi Sad	0	1	0	0
Niš	0	0	1	0
Ostalo	0	0	0	1

Tabela 3.4: One-hot kodiranje kategoričke promenljive Veliki grad

Na trening skup je primenjen model slučajnih šuma, sa ograničenjem maksimalne dubine stabala na 10 i brojem stabala postavljenim na 100, što su podrazumevane vrednosti u klasi `sklearn.ensemble.RandomForestRegressor`. Ovaj model omogućava uočavanje nelinearnih veza između nezavisnih i ciljne promenljive, bez potrebe za pretpostavkama o raspodeli podataka.

Performanse modela slučajnih šuma procenjene su na test skupu pomoću metrika prikazanih u tabeli 3.5.

Metrika	**Vrednost**
MSE	11.715791
MAE	2.062314
R^2	0.090164

Tabela 3.5: Evaluacione metrike modela slučajnih šuma

Koeficijent determinacije modela slučajnih šuma je nizak, objašnjavajući oko 9% varijanse ciljne promenljive. Na slici 3.7 je prikazana važnost zasnovana na smanjenju greške, atribut prethodno pomenute klase `RandomForestRegressor`, `feature_importances_`. Ove vrednosti ukazuju na to koliko je svaka promenljiva doprinela smanjenju greške predikcije kroz celokupnu strukturu stabala u šumi.

Važnost neke promenljive se računa kao prosečno poboljšanje koje ona donosi modelu prilikom njegovog treniranja, odnosno koliko se smanji nepreciznost (npr. varijansa kod regresije) svaki put kada se koristi za podelu u stablima. Ovo se sabira kroz sva stabla u šumi, a zatim se sve vrednosti prilagode tako da njihov zbir bude jednak 1.

Slika 3.7: Važnost nezavisnih promenljivih modela slučajnih šuma

Dakle, možemo izvesti sledeće zaključke:

- Starost osiguranika – ubedljivo najvažnija promenljiva, što ukazuje da je broj šteta snažno povezan sa starosnom strukturom korisnika. Ovde je zanimljivo primetiti da kod prethodnih modela ova promenljiva gotovo da nije imala nikakav značaj, što može ukazivati na složene nelinearne odnose koje oni nisu mogli da uoče.
- Pol osiguranika_F – značajan uticaj pola na broj šteta.
- Veliki grad_Beograd – geografska lokacija se pokazala kao relevantna, sa uticajem grada Beograd.
- Fizikalna terapija – ukazuju na uticaj ovog pokrića na učestalost prijava šteta.
- Opšta participacija, Bel Medic participacija i Van mreže participacija – obavezno učešće u šteti utiče na učestalost prijava šteta.

Model slučajnih šuma je pokazao bolju sposobnost predikcije u odnosu na uopšten linearan model, omogućivši uvid u nelinearne odnose. Najvažnije promenljive, kao što su starost i pol osiguranika, mogu biti korisni za dalju segmentaciju osiguranika i optimizaciju proizvoda u oblasti dobrovoljnog zdravstvenog osiguranja.

Međutim, iako pruža brzu i globalnu sliku o važnosti promenljivih, ova metoda ne uzima u obzir interakcije među promenljivama i može favorizovati promenljive sa većim brojem jedinstvenih vrednosti.

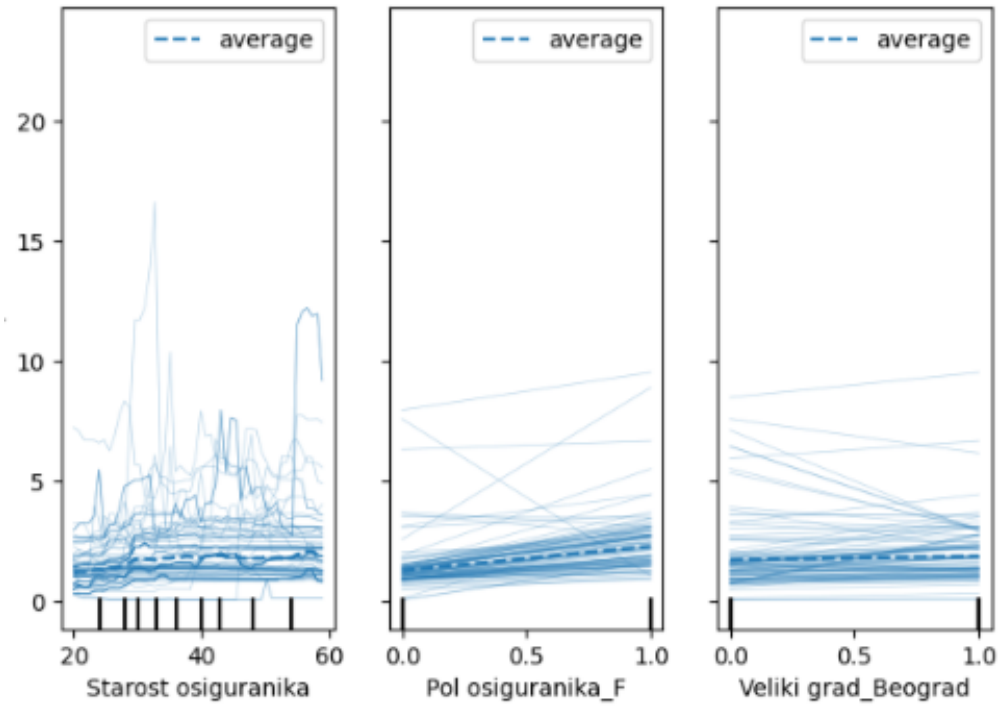
Zbog toga ćemo u nastavku koristiti i naprednije metode interpretacije modela, kao što su grafici parcijalne zavisnosti (PDP), metoda individualnog uslovnog očekivanja (ICE), metoda akumuliranih lokalnih efekata (ALE) i Šeplijeve vrednosti, koje omogućavaju detaljniju i robusniju interpretaciju uticaja pojedinačnih nezavisnih promenljivih na predikciju modela, kako na globalnom, tako i na lokalnom nivou.

Grafici parcijalne zavisnosti i metoda individualnog uslovnog očekivanja

Pogledajmo prvo kako izgledaju PDP grafici i ALE metoda za tri najznačajnije nezavisne promenljive prema važnosti promenljivih zasnovanoj na smanjenju greške (slika 3.7).

Koristićemo metodu klase `PartialDependenceDisplay` iz biblioteke scikit-learn, i podesiti parametar kind='both', kako bismo dobili i PDP i ICE grafik. Na slučajan način ćemo za ICE ograničiti broj osiguranika na 100, kako bi grafik bio pregledniji.

Dok PDP prikazuje prosečan uticaj promenljive na model, ICE krive omogućavaju uvid u varijacije među pojedinačnim osiguranicima.



Slika 3.8: Grafici parcijalne zavisnosti i metoda individualnog uslovnog očekivanja za odabrane promenljive

Na slici 3.8 primećujemo:

- Starost osiguranika - na grafiku se uočava da je uticaj starosti nelinearan, s izraženim varijacijama među osiguranicima (što je prikazano kroz raspršene ICE krive). Iako prosečan PDP sugeriše blagi rast očekivanog broja šteta sa starošću, ICE krive pokazuju da kod određenih pojedinaca može doći do naglih skokova, što ukazuje na potencijalne interakcije sa drugim promenljivama.
- Pol osiguranika_F - i PDP i ICE krive ukazuju na konzistentan pozitivan efekat ženskog pola na broj šteta. ICE krive su prilično paralelne i stabilne, što sugeriše da ova promenljiva ima sličan uticaj kod većine osiguranika, bez značajnijih interakcija.

- Veliki grad_Beograd - ovde je efekat takođe pozitivan u odnosu na očekivanu vrednost broja šteta. ICE krive pokazuju nešto veću disperziju nego kod pola, ali su i dalje prilično dosledne, što ukazuje na postojanje globalnog trenda nešto većeg broja šteta kod osiguranika iz Beograda.

Dakle, kombinovani prikaz PDP i ICE krivih omogućava istovremenu analizu prosečnih efekata i individualnih odstupanja. Posebno kod promenljive `Starost osiguranika` vidi se da prosečna vrednost može da maskira značajnu varijabilnost, što dodatno potvrđuje korisnost nelinearnih modela kao što je model slučajnih šuma u aktuarskim analizama.

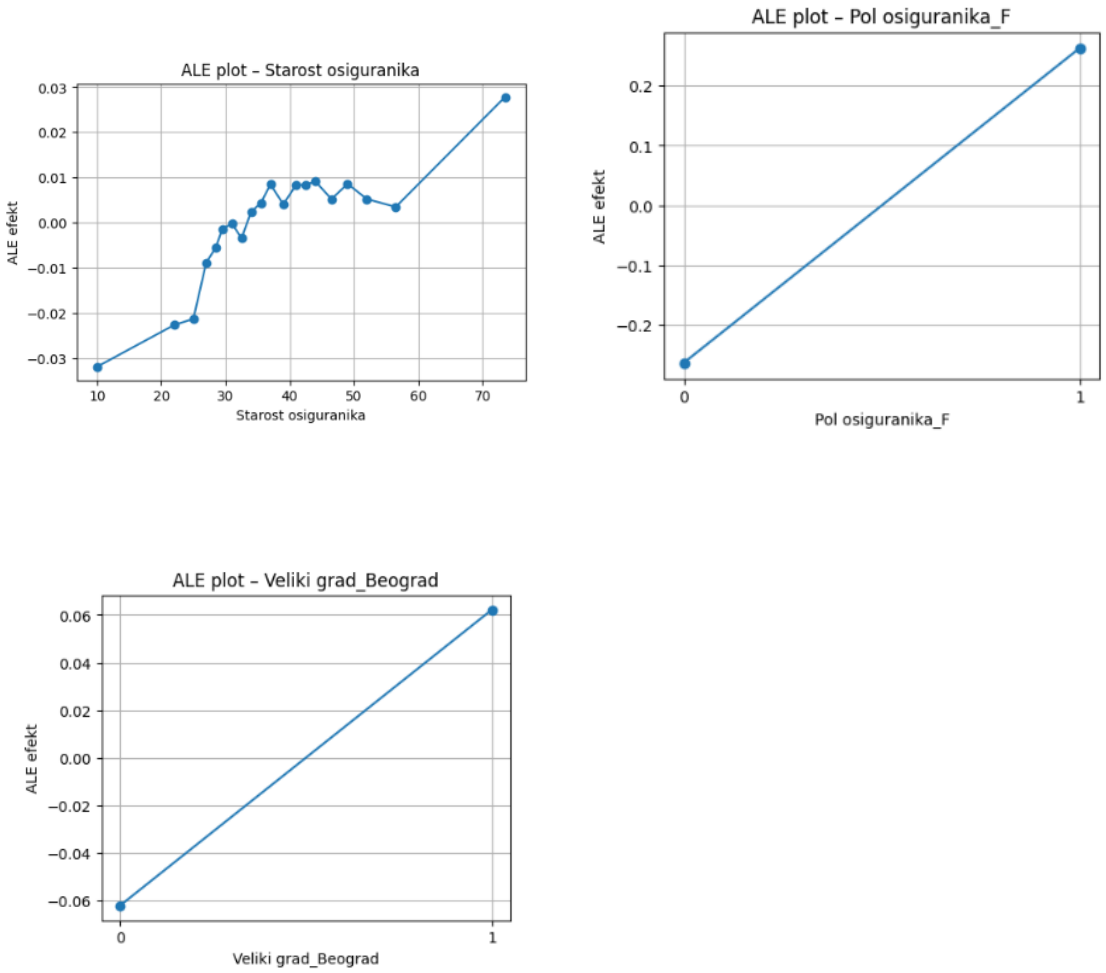
Metoda akumuliranih lokalnih efekata

Na slici 3.9 prikazani su ALE grafici za promenljive Starost osiguranika i Pol osiguranika_F. Posebne funkcije za neprekidne i binarne promenljive, plot_ale_1d i plot_ale_binary, mogu se pronaći u GitHub repozitorijumu [9].

Rezultati ukazuju da prelazak sa muškog (0) na ženski pol (1) ima pozitivan ALE efekat, što znači da ženski osiguranici u proseku imaju veći očekivani broj šteta. Ovaj rezultat je u skladu sa rezultatima uopštenog linearnog modela, kao i sa prethodnim interpretacijama PDP i ICE prikaza. To ukazuje na to da različiti metodološki pristupi (linearni i nelinearni) upućuju na isti zaključak o uticaju pola, čime se povećava poverenje u stabilnost tog zapažanja.

Moguće objašnjenje zbog čega žene češće koriste zdravstvene usluge, može biti rezultat višeg nivoa zdravstvene odgovornosti ili specifične potrebe, kao što su preventivni pregledi i usluge vezane za reproduktivno zdravlje.

Slično, ALE za Veliki grad_Beograd pokazuje pozitivan efekat prelaska kategorije Beograd, što ukazuje na to da osiguranici iz Beograda imaju viši očekivani broj šteta. Ova lokalna interpretacija dodatno potvrđuje zapažanja iz drugih analiza.



Slika 3.9: ALE grafici za model slučajnih šuma - odabrane promenljive

ALE analiza omogućava stabilnu i robusnu interpretaciju i za binarne promenljive, bez preterane zavisnosti od raspodele podataka. Dobijeni rezultati su konzistentni sa prethodnim metodama interpretacije i dodatno potvrđuju pouzdanost modela.

Prikazan je i ALE grafik za neprekidnu promenljivu Starost osiguranika. Uočen je nelinearan odnos između starosti i broja šteta:

- U mlađim godinama (do 25 godina), ALE efekat je negativan, što ukazuje na manji očekivani broj šteta kod mlađih osiguranika.
- U srednjim godinama (25–60), efekat postaje pozitivan i stabilan, sa blagim porastom očekivanih šteta.

- Nakon 60. godine primećuje se izraženiji porast efekta, što može biti posledica povećane zdravstvene potrebe starijih osiguranika.

Ovaj rezultat potvrđuje korisnost ALE analiza u uočavanju nelinearnih obrazaca koje linearni modeli ne mogu da uoče. Takođe, rezultat je u skladu sa prethodnim rezultatima dobijenim primenom PDP metode, uz dodatnu robusnost na korelacije među promenljivama.

Šeplijeva metoda

Sada ćemo upotrebiti Šeplijevu metodu. Implementacija je urađena korišćenjem biblioteke `shap`.

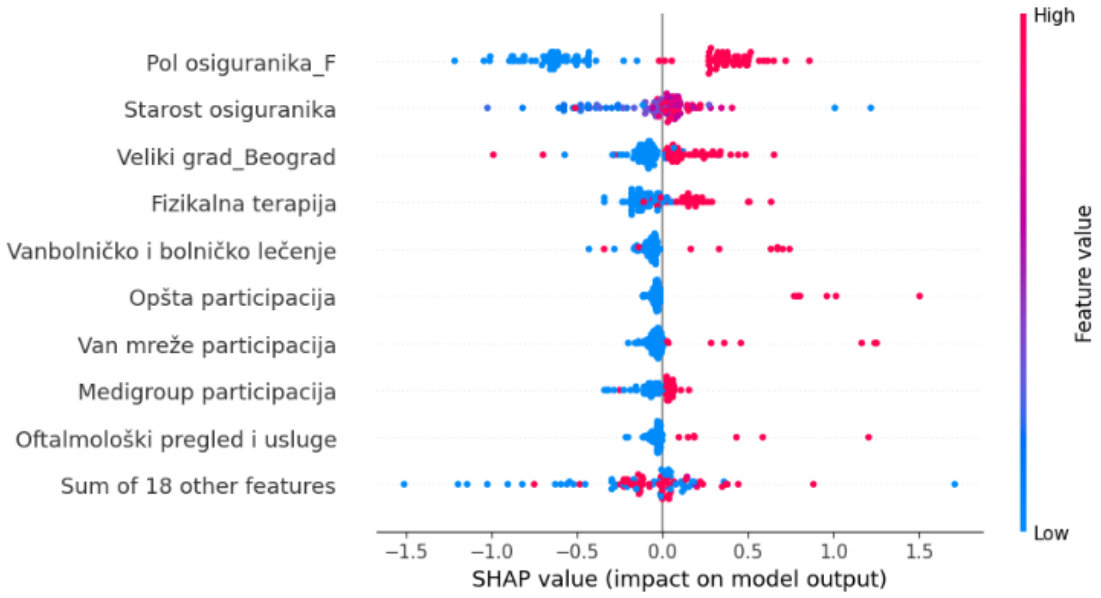
Nakon inicijalizacije `Explainer` objekta nad treniranim modelom i skupom podataka, izračunate su Šeplijeve vrednosti za prvih 100 primera iz trening skupa. Prikazaćemo tri ključna grafika:

- **Grafik raspodele uticaja nezavisnih promenljivih (eng. Beeswarm plot)** — prikazuje globalni značaj promenljivih i njihov doprinos predikciji, zajedno sa raspodelom vrednosti.
- **Bar plot** — prikazuje prosečne apsolutne Šeplijeve vrednosti po promenljivoj, čime omogućava rangiranje značaja.
- **Grafik toka doprinosa (eng. Waterfall plot)** — omogućava lokalnu interpretaciju doprinosa za konkretnu instancu, prikazujući kako se od bazne vrednosti (npr. prosečne vrednosti ciljne promenljive) dolazi do konačne predikcije, pri čemu se vidi koliko je svaka promenljiva pozitivno ili negativno uticala na rezultat modela.

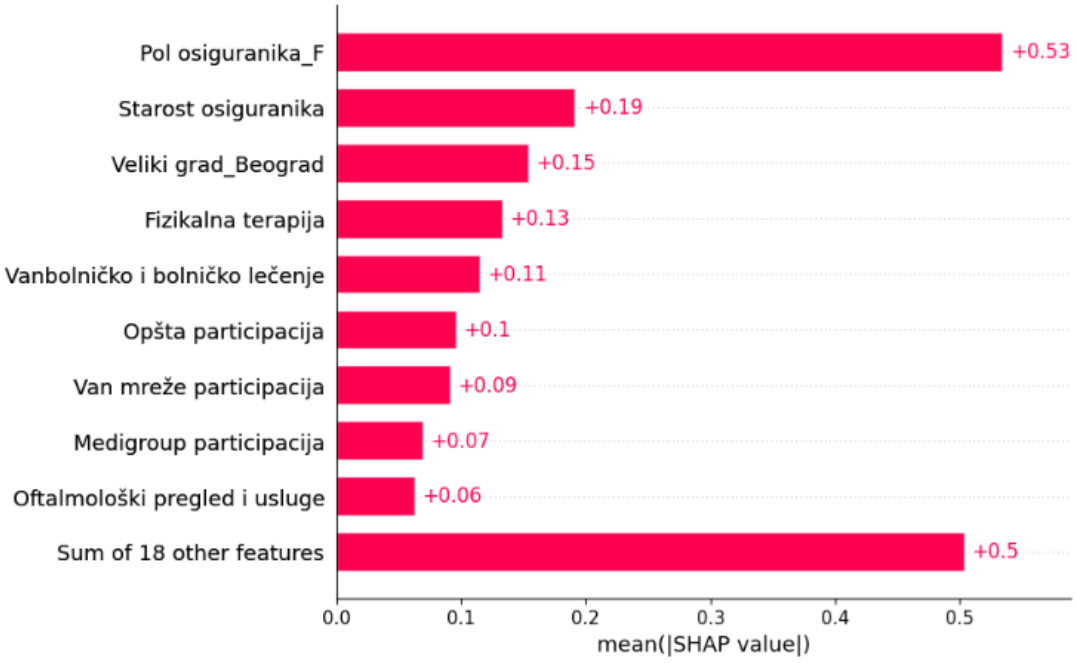
Na slici 3.10 je prikazana raspodela Šeplijevih vrednosti za sve instance i sve najvažnije promenljive. Boja označava vrednost promenljive (od niske - plava, do visoke - crvena), a horizontalna pozicija predstavlja uticaj na izlaz modela.

Prikazan je i bar-plot sa prosečnim apsolutnim Šeplijevim vrednostima ulaznih promenljivih, slika 3.11. Na osnovu ova dva grafika zaključujemo:

- Pol osiguranika_F – najuticajnija promenljiva, gde vidimo da ako je osoba ženskog pola (crveno) imamo veće Šeplijeve vrednosti i veći broj šteta, a kada je osoba muškog pola (plavo), imamo manje Šeplijeve vrednosti i manji broj šteta.

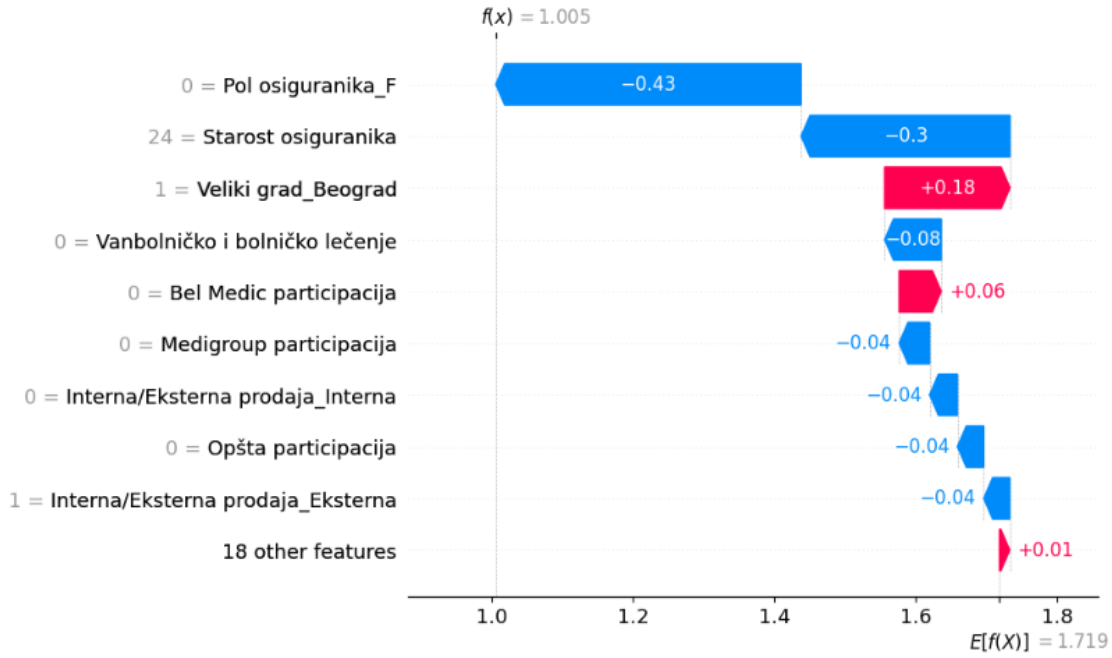


Slika 3.10: Grafik raspodele uticaja nezavisnih promenljivih - model slučajnih šuma



Slika 3.11: Bar plot - model slučajnih šuma

- Starost osiguranika – druga najuticajnija promenljiva, gde vidimo da manja starost (plavo) ima tendenciju ka nižim vrednostima predikcije, a veća starost (crveno) doprinosi većem broju šteta, ali efekat je umeren.
- Veliki grad_Beograd, Fizikalna terapija, Vanbolničko i bolničko lečenje – takođe visoko rangirani po važnosti.



Slika 3.12: Grafik toka doprinosa nezavisnih promenljivih - model slučajnih šuma

Na slici 3.12 prikazan je grafik toka doprinosa za jednu konkretnu instancu (osiguranika, muškarac, starosti 24 godine iz Beograda), koji prikazuje kako pojedinačna promenljive doprinose ukupnoj predikciji. Vidi se da Pol osiguranika_F i Starost osiguranika imaju najveći negativni doprinos, dok određeni faktori kao što su Veliki_grad_Beograd doprinosi povećanju vrednosti predikcije.

Šeplijeve vrednosti omogućavaju jasnu interpretaciju kako pojedinačne promenljive utiču na model, kako na globalnom nivou (važno za opštu analizu osiguranika), tako i lokalno (važno za donošenje odluka na nivou pojedinačnog klijenta). Ovi rezultati su konzistentni sa prethodnim zapažanjima iz PDP i ALE analiza, što potvrđuje robusnost modela i interpretacija.

3.5 Neuronska mreža

Sada ćemo ponoviti postupak, samo sa modelom neuronske mreže za regresiju broja šteta.

Neuronske mreže su modeli inspirisani načinom na koji funkcionišu ljudski neuroni. Sastoje se od više slojeva neurona (čvorova), pri čemu svaki neuron vrši jednostavnu matematičku operaciju nad ulaznim podacima. Podaci se prosleđuju kroz ulazni sloj, jedan ili više skrivenih slojeva, i na kraju izlazni sloj. Svaka veza između neurona ima svoju težinu, koja se tokom učenja prilagođava kako bi se smanjila greška modela. Zahvaljujući složenoj strukturi, neuronske mreže su sposobne da modeluju veoma kompleksne obrasce i nelinearne odnose među podacima.

Napravićemo i trenirati model, uraditi evaluaciju i prikazati značaj permutacije nezavisnih promenljivih.

Značaj permutacije (eng. permutation importance) predstavlja metodu za procenu značaja nezavisnih promenljivih tako što meri pad performansi modela kada se vrednosti određene promenljive nasumično permutuju. Ako permutacija neke promenljive značajno pogorša performanse modela, to znači da je ta promenljiva važna za model.

U poređenju sa atributom `feature_importances_` kod modela kao što je model slučajnih šuma, značaj permutacije je metoda agnostična na tip modela i može se primeniti i na neuronske mreže, s obzirom da ne daju direktne koeficijente.
u Pythonu se koristi funkcija `permutation_importance` iz biblioteke `sklearn.inspection`.

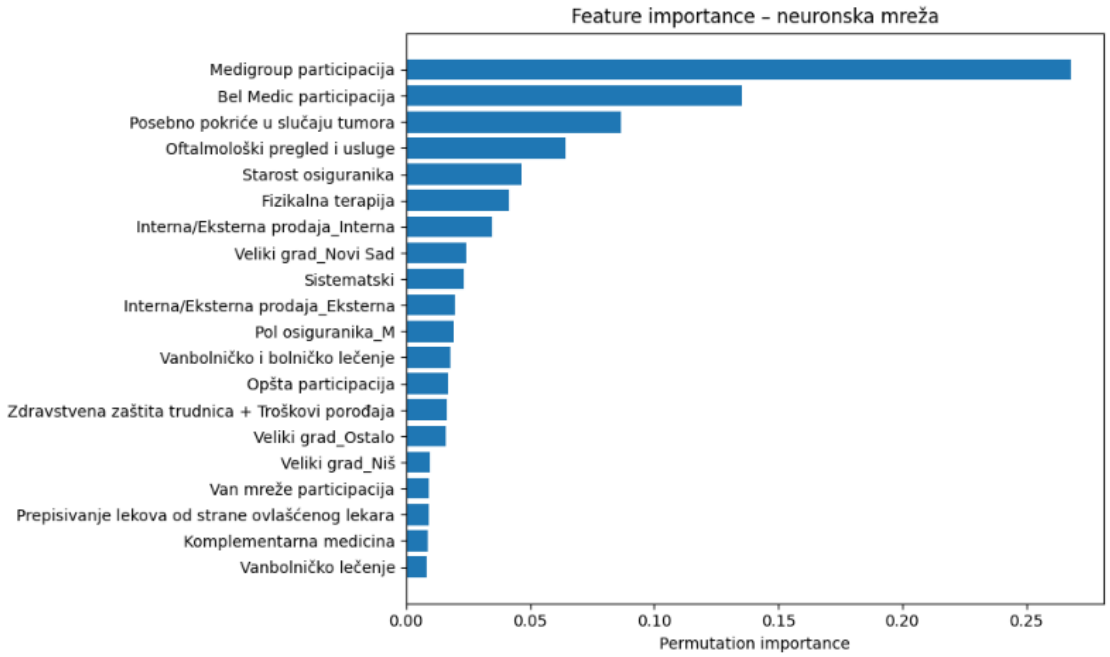
Metrika	**Vrednost**
MSE	12.336679
MAE	2.233036
R^2	0.041946

Tabela 3.6: Evaluacione metrike neuronske mreže

Trenirana je višeslojna perceptronska neuronska mreža (*Multi-layer Perceptron – MLP*)[6] za regresiju broja šteta. Model je treniran sa dva skrivena sloja (64 i 32 neurona), a funkcija aktivacije je ReLU.

Rezultati evaluacije na test skupu su sledeći prikazani u tabeli 3.6.

Model objašnjava oko 4.2% disperzije ciljne promenljive, što je slabije od modela slučajnih šuma.



Slika 3.13: Permutation importance neuronske mreže

Na slici 3.13 je prikazana važnost 20 najuticajnijih nezavisnih promenljivih na osnovu metode značaja permutacije.

Najveći uticaj na predikciju modela imaju sledeće nezavisne promenljive:

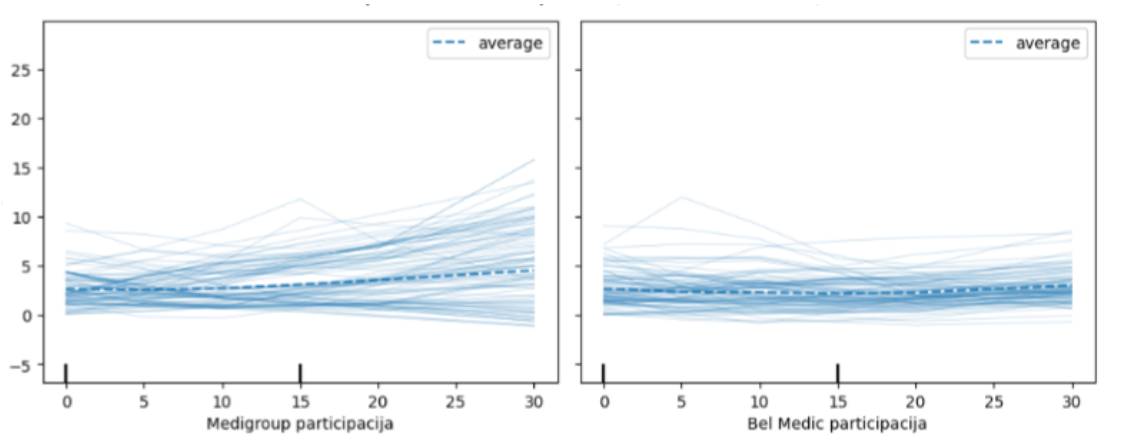
- Medigroup participacija i Bel Medic participacija – visoko rangirani, što može ukazivati na to da pristup ovim ustanovama utiče na učestalost korišćenja osiguranja. Ovo je model koji za sada pokazuje najznačajniji uticaj ovih promenljivih, što ukazuje na nelinearnu vezu sa ostalim nezavisnim promenljivim.
- Posebno pokriće u slučaju tumora i Oftalmološki pregled i usluge – takođe sa izraženim uticajem.
- Starost osiguranika, Fizikalna terapija i Interna/Eksterna prodaja_Interna – predstavljaju promenljive sa umerenim doprinosom.

Iako neuronska mreža nije postigla bolji rezultat u pogledu R^2, ona pokazuje sposobnost da prepozna kompleksne obrasce među promenljivama. Značaj permutacije pojedinačnih promenljivih omogućava uvid u relativni doprinos promenljivih, čak i kod modela koji ne nude transparentan uvid u svoju unutrašnju strukturu.

Grafici parcijalne zavisnosti i metoda individualnog uslovnog očekivanja

Sada ćemo nacrtati kombinovane grafike parcijalne zavisnosti (PDP) i krive individualnog uslovnog očekivanja (ICE) za dve najvažnije nezavisne promenljive u okviru neuronske mreže. PDP prikazuje prosečan uticaj vrednosti pojedinačne promenljive na izlaz modela, dok ICE linije prikazuju varijabilnost po instanci (osiguraniku).

Na isti način kao i za model slučajnih šuma, koristićemo metodu klase PartialDependenceDisplay iz biblioteke scikit-learn.



Slika 3.14: PDP i ICE grafici za odabrane promenljive

Za promenljive Medigroup paricipacija i Bel Medic participacija, koje predstavljaju obaveznu paricipaciju u privatnim zdravstvenim ustanovama, sa slike 3.14 primećujemo da model pokazuje suptilan, ali pozitivan trend. Veće vrednosti promenljivih blago povećavaju očekivani broj šteta. Ova veza može ukazivati na to da osiguranici sa pristupom ovim ustanovama češće koriste osiguranje, uprkos pretpostavci da bi participacija smanjila broj prijava šteta.

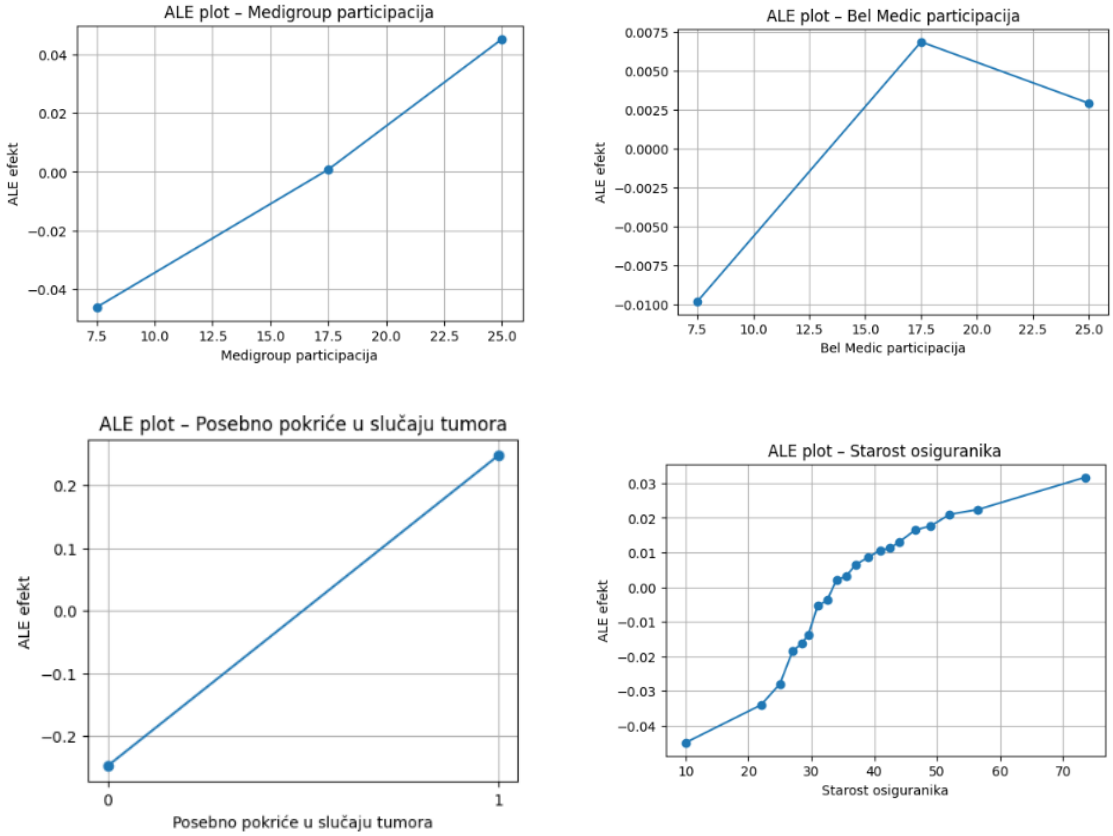
Za ostale promenljive, PDP i ICE grafici pokazali su nisku varijabilnost među osiguranicima, što ukazuje da neuronska mreža nije uspela da identifikuje njihov nelinearni uticaj na broj šteta.

U poređenju sa uopštenim linearnim modelima i modelima slučajnih šuma, neuronska mreža pokazuje manju osetljivost na pojedinačne nezavisne promenljive, bar kada se posmatra kroz PDP i ICE pristup.

Sada ćemo prikazati metode interpretacije kao što su ALE i Šeplijeva metoda, i videti da li mogu pomoći u dubljoj analizi interakcija i lokalnih doprinosa.

Metoda akumuliranih lokalnih efekata

Za ALE koristićemo funkcije koje smo definisali prilikom rada sa modelom slučajnih šuma.



Slika 3.15: ALE grafici za neuronsku mrežu – odabrane promenljive

ALE omogućava da se vizuelizuju prosečni lokalni uticaji promenljivih na predikciju modela, bez pretpostavke o nezavisnosti promenljivih (za razliku od PDP-a). Na slici 3.15 su prikazani ALE grafici za četiri odabrane promenljive. Zaključujemo:

- Medigroup participacija - efekat je pozitivan, odnosno što je veći iznos Medigroup participacije, to je veći očekivani broj šteta.
- Bel Medic participacija - za niže vrednosti participacije u Bel Medic-u, broj šteta se smanjuje, ali kod većih vrednosti blago raste – moguće je da postoji nelinearna veza.

- Starost osiguranika - uočen je izraženi rastući efekat, posebno izražen kod starijih od 40 godina.
- Posebno pokriće u slučaju tumora - osobe koje imaju uključeno ovo pokriće imaju veću očekivanu vrednost broja šteta (možda zato što se češće koristi ili zato što su rizičniji osiguranici).

Iako su ALE grafici za pojedine promenljive, poput Medigroup i Bel Medic participacije, prikazali određene obrasce, interpretacija ovih rezultata je otežana. Vrednosti ovih promenljivih su diskretne i ograničenog broja, što umanjuje pouzdanost lokalnih efekata prikazanih na grafiku. Takođe, ove promenljive verovatno deluju posredno, kroz povezanost sa drugim socioekonomskim ili karakteristikama samih paketa, a ne kao direktan uzrok nastanka štete. Stoga se ovi rezultati ne mogu smatrati pouzdanim pokazateljima u interpretaciji neuronske mreže. Dakle, neuronska mreža, iako potencijalno snažnija u predikciji, pokazuje ograničenu interpretabilnost za ovaj tip promenljivih.

Šeplijeva metoda

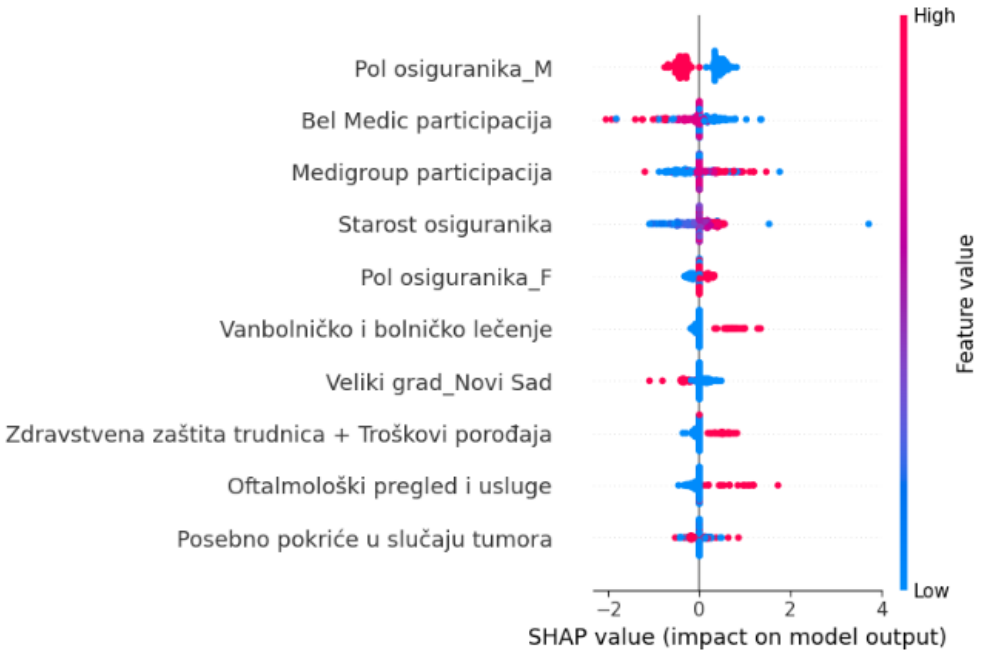
Sada ćemo prikazati Šeplijevu metodu. Koristićemo `KernelExplainer`[2], koji je nezavisan od tipa modela (model-agnostic), ali znatno sporiji, pa je analiza izvedena nad manjim uzorkom podataka.

Nacrtaćemo ista tri grafika kao i za model slučajnih šuma.

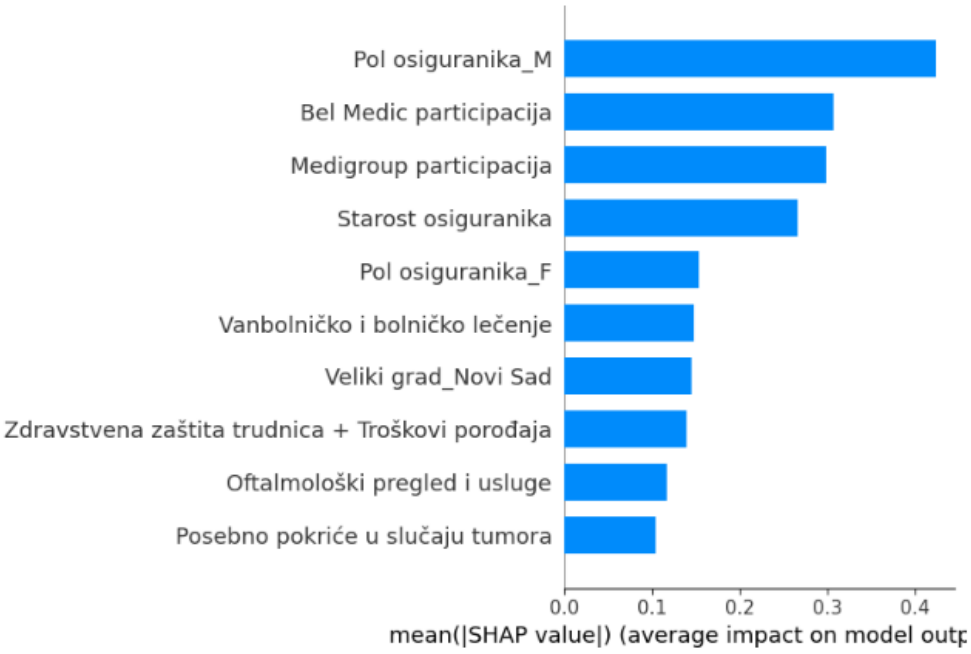
Najveći globalni uticaj (Grafik raspodele uticaja nezavisnih promenljivih i bar plot Šeplijevih vrednosti nezavisnih promenljivih, prikazani na slikama 3.16 i 3.17) na predikcije modela imaju sledeće nezavisne promenljive:

- Pol osiguranika_M — ima najnegativniji uticaj, što ukazuje da muški osiguranici imaju u proseku manji broj šteta.
- Bel Medic participacija — doprinosi modelu u negativnom smeru pri višim vrednostima.
- Medigroup participacija — doprinosi modelu u pozitivnom smeru pri višim vrednostima.

[2]KernelExplainer je implementacija u okviru Python biblioteke SHAP, koja aproksimira Šeplijeve vrednosti korišćenjem linearnog regresionog modela sa odgovarajućom funkcijom jezgra.



Slika 3.16: Grafik raspodele uticaja nezavisnih promenljivih - Neuronska mreža

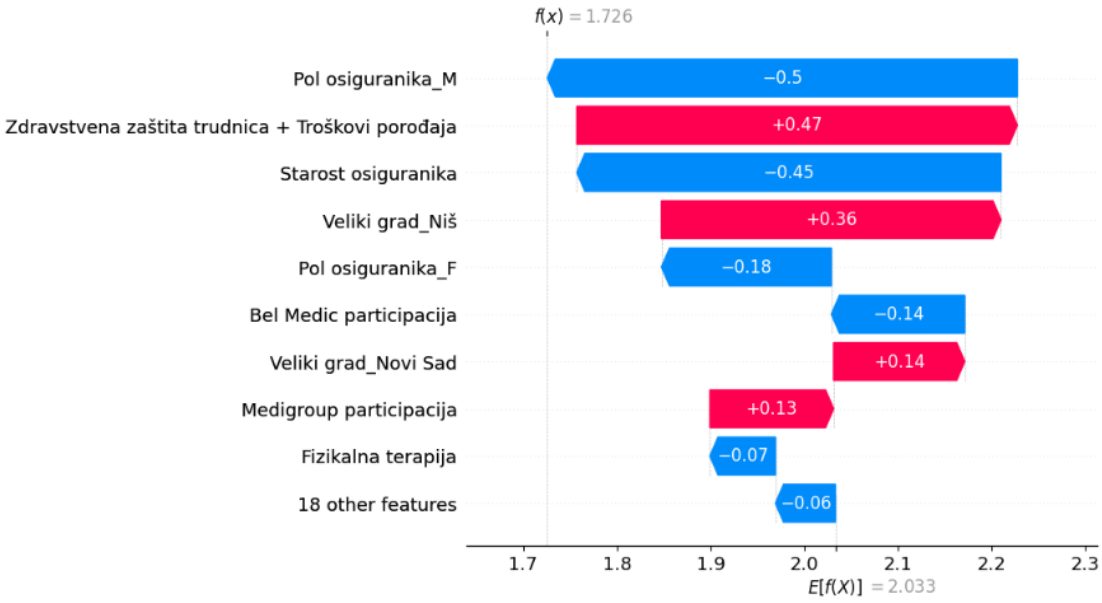


Slika 3.17: Bar plot - Neuronska mreža

- Starost osiguranika — pokazuje manji, ali prisutan negativan uticaj pri nižim vrednostima.

Za većinu ostalih promenljivih efekat je neutralan ili slab.

Na primeru jedne instance, na slici 3.18, model je predvideo vrednost $f(x) = 1.726$, dok je osnovna vrednost modela $E[f(x)] = 2.033$. Glavni negativni doprinosi potiču od nezavisnih promenljivih Pol osiguranika_M i Starost osiguranika. S druge strane, pozitivni doprinos daju Zdravstvena zaštita trudnica i Veliki_grad_Niš.



Slika 3.18: Grafik toka doprinosa nezavisnih promenljivih - Neuronska mreža

Šeplijeva analiza pokazuje da neuronska mreža najviše koristi pol, participaciju i starost u predikcijama, što je u skladu sa rezultatima prethodnih interpretacionih tehnika kao što je ALE. Iako su pol i starost i kod neuronske mreže i modela slučajnih šuma dosledno identifikovani kao važni faktori, doprinosi pojedinih promenljivih, poput participacija, značajno variraju između modela. Kod neuronske mreže oni su manje izraženi i ne pokazuju konzistentan pravac uticaja kao kod modela slučajnih šuma, što može ukazivati na manju interpretabilnost u ovom delu modela.

3.6 Poređenje interpretacije modela: Uopšteni linearan model, model slučajnih šuma i neuronska mreža

Interpretacija prediktivnih modela predstavlja ključni aspekt ovog rada, zbog praktične potrebe, da se osim tačnih predikcija obezbedi i transparentnost u vezi sa faktorima koji utiču na broj šteta po osiguraniku. Analizirana su tri konceptualno različita modela, uopšten linearan model, model slučajnih šuma i višeslojna perceptronska neuronska mreža (MLP). Na njih je primenjeno više interpretacionih metoda, poput metode koja meri pad performansi modela prilikom permutacije pojedinačnih nezavisnih promenljivih (eng. permutation importance), PDP/ICE, ALE i Šepllijeve analize.

Uopšten linearan model - kao linearan i transparentan model, uopšten linearan model omogućava direktno tumačenje koeficijenata u smislu logaritamskog odnosa između nezavisnih promenljivih i ciljne promenljive. Ova interpretacija je jednostavna i intuitivna, ali podrazumeva linearnost i aditivnost odnosa, što može ograničiti detekciju složenijih nelinearnih obrazaca. Na primer, uopšten linearan model je pokazao da promenljive poput Pol osiguranika_M imaju značajan negativan uticaj na broj šteta, ali nije uspeo da detektuje složene interakcije između promenljivih kao što su starost i tip pokrića.

Model slučajnih šuma - model slučajnih šuma kao ansambl model pokazao je veliku sposobnost u detekciji nelinearnih i interaktivnih odnosa među nezavisnim promenljivama. Metoda koja meri pad performansi modela prilikom permutacije pojedinačnih nezavisnih promenljivih i Šeplijeva analiza pokazale su da ovaj model efektivno identifikuje najvažnije faktore (npr. starost, pol, tip participacije), dok su ALE i PDP prikazi omogućili detaljan uvid u njihov efekta. Zahvaljujući svojoj robusnosti i prirodi zasnovanoj na mnoštvu stabala, model slučajnih šuma uspeva da zadrži stabilnu interpretabilnost i u prisustvu korelisanih promenljivih.

Neuronska mreža (MLP) - MLP model je pokazao određeni nivo prediktivne moći, ali sa ograničenom interpretabilnošću. Zbog složenosti arhitekture i „black-

box“ karaktera[3], interpretacija se oslanjala isključivo na model-agnostičke metode poput metode koja meri pad performansi modela prilikom permutacije pojedinačnih nezavisnih promenljivih i Šeplijeve metode zasnovane na jezgrima. Iako su rezultati Šeplijeve analize ukazali na slične dominantne nezavisne promenljive kao i kod prethodna dva modela, efekti su bili slabije diferencirani. Takođe, PDP i ALE grafici za neuronsku mrežu pokazali su da model u manjoj meri uspeva da uoči izražene nelinearnosti u podacima, verovatno usled ograničenih kapaciteta učenja na datom uzorku i bez podešavanja hiperparametara.

Dakle, u pogledu interpretabilnosti, uopšten linearan model nudi najdirektnije i teorijski najpotpunije objašnjenje, ali uz ograničenja u pogledu fleksibilnosti. Model slučajnih šuma predstavlja kompromis između interpretabilnosti i modelskog kapaciteta, dok MLP model, iako u teoriji sposoban za detekciju kompleksnih obrazaca, pokazuje smanjenu transparentnost i osetljivost na ulazne varijacije u okviru korišćenih interpretacionih tehnika. U nastavku rada biće razmotreno kako objediniti ove rezultate u cilju izgradnje optimalnog prediktivnog modela.

3.7 Izrada najoptimalnijeg modela

Sada ćemo pokušati, na osnovu informacija dobijenih primenom metoda interpretacije, da napravimo najbolji mogući model za predviđanje broja šteta.

Prvo ćemo napraviti `LightGBM` model, kako bismo videli da li ima bolje performanse od modela slučajnih šuma.

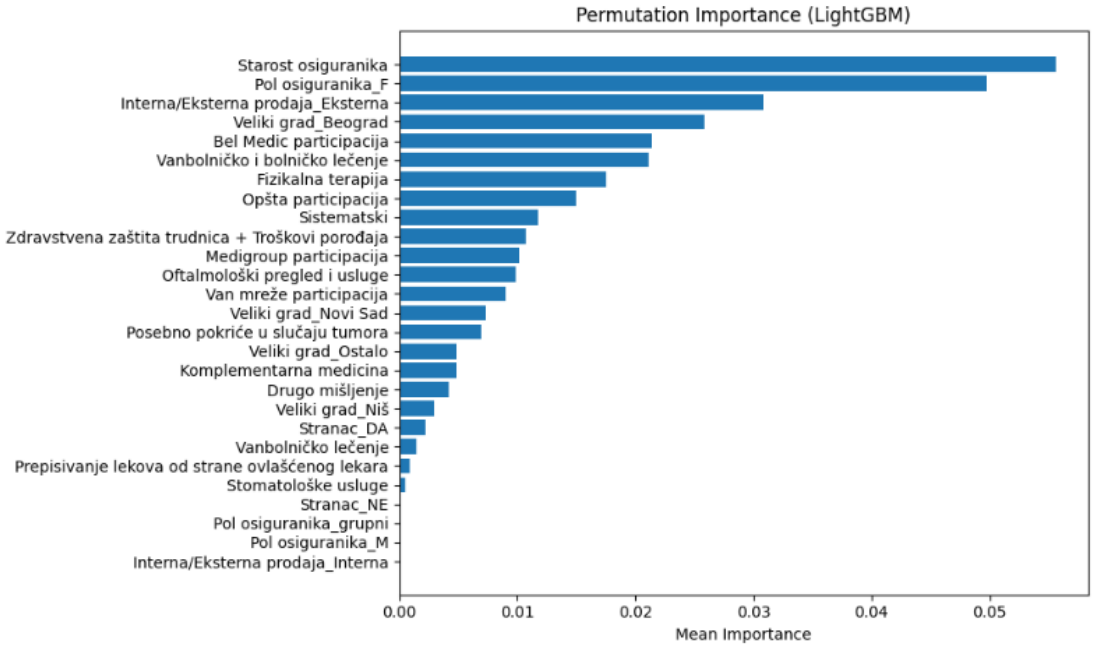
`LightGBM` (model za ubrzano gradijentno poboljšanje, eng. Light Gradient Boosting Machine) je model baziran na principu gradijentnog unapređivanja. Zasniva se na ansamblu stabala odlučivanja, pri čemu se nova stabla uzastopno dodaju kako bi se ispravile greške prethodnih. Za razliku od tradicionalnih algoritama, koristi strategiju grananja po listovima (tzv. *leaf-wise* rast stabala), što znači da se uvek širi onaj list koji trenutno ima najveći potencijal da smanji grešku modela. To dovodi do dubljih i često asimetričnih stabala, ali sa boljom preciznošću. `LightGBM` je poznat po velikoj brzini i efikasnosti, posebno kod velikih skupova podataka.

Rezultati evaluacije su prikazani u tabeli 3.7. MSE i MAE su slični kao u prethodnim modelima, dok je R^2 nešto bolji.

[3]Model čiju je unutrašnju logiku teško ili nemoguće direktno tumačiti, pa nije jasno na koji način pojedinačne ulazne promenljive utiču na predikciju.

Metrika	Vrednost
MSE	11.609925
MAE	2.055771
R^2	0.098385

Tabela 3.7: Evaluacione metrike `LightGBM` modela



Slika 3.19: Feature importance `LightGBM` modela

Najvažnije promenljive po `LightGBM` modelu (slika 3.19) su Starost osiguranika, Pol osiguranika_F, Interna/Eksterna prodaja_Eksterna, Veliki_grad_Beograd i Bel Medic participacija i Vanbolničko i bolničko lečenje, što se najvećim delom poklapa sa modelom slučajnih šuma, i daje stabilnost interpretacije.

Ovi rezultati se mogu dodatno poboljšati podešavanjem hiperparametara prikazanih u tabeli 3.8.

Parametar	Opis	Skup vrednosti
`n_estimators`	Broj stabala koje model gradi; direktno utiče na preciznost i brzinu treniranja	{100, 200, 300, 500}
`learning_rate`	Korak kojim model uči – manja vrednost znači sporije učenje, ali potencijalno bolji rezultat (uz više stabala)	{0.01, 0.05, 0.1, 0.2}
`num_leaves`	Maksimalan broj listova po stablu – veći broj omogućava veću složenost modela, ali i veći rizik od preprilagođavanja	{15, 31, 50, 70}
`max_depth`	Maksimalna dubina stabla – pomaže u kontroli složenosti modela; ako je 1, dubina nije ograničena	{-1, 3, 5, 7}
`min_child_samples`	Minimalan broj podataka koji mora postojati u listu da bi on ostao u modelu – veće vrednosti vode ka jednostavnijem modelu	{5, 10, 20}
`subsample`	Procenat slučajnog uzorka instanci koje se koriste za treniranje svakog stabla; pomaže u sprečavanju preprilagođavanja	{0.6, 0.8, 1.0}
`colsample_bytree`	Procenat kolona (promenljivih) koje se nasumično uzimaju za svako stablo; pomaže u sprečavanju preprilagođavanja	{0.6, 0.8, 1.0}

Tabela 3.8: Opis hiperparametara koje optimizujemo

Uradićemo:

- nasumičnu pretragu sa unakrsnom validacijom (eng. RandomizedSearchCV) - tehnika koja nasumično pretražuje kombinacije hiperparametara iz zadatih opsega i za svaku kombinaciju trenira model koristeći ukrštenu validaciju, a zatim bira onu koja daje najbolje rezultate prema izabranoj metričkoj funkciji (u ovom slučaju koristimo koeficijent determinacije, R^2). Za razliku od klasične pretrage sa ukrštenom validacijom (eng. GridSearch), koji iscrpno pretražuje sve moguće kombinacije, pa tehnika nasumično ispituje određeni broj kombinacija, čime značajno ubrzava proces.

- petostruku ukrštenu validaciju - metoda validacije modela pri kojoj se podaci dele na pet delova, od kojih se četiri koriste za treniranje, a peti za testiranje.

Proces se ponavlja pet puta, svaki put sa drugim delom za testiranje, a rezultat je prosečna vrednost metrika.

- R^2 - za potrebe evaluacije koristi se koeficijent determinacije, koji pokazuje koliko dobro model objašnjava varijaciju u podacima.

Metrika	**Vrednost**
MSE	11.638208
MAE	2.060676
R^2	0.096189

Tabela 3.9: Evaluacione metrike `LightGBM` modela nakon RandomizedSearchCV

Rezultati evaluacije su prikazani u tabeli 3.9. Kao što možemo videti, model nije poboljšan. Ovo ukazuje na to da su podaci veoma zahtevni za predikciju, što može biti posledica izražene nepredvidivosti u broju šteta.

Pokušaćemo da dodamo interakcije između promenljivih. Targetiraćemo one promenljive koje su se pokazale kao značajne tokom interpretacije, pri čemu simbol + označava sabiranje promenljivih, dok * označava njihov proizvod, odnosno interakcioni član:

- Starost * Pol
- Bel Medic participacija + Medigroup participacija
- Bel Medic participacija * Vanbolničko i bolničko lečenje
- Interna prodaja * grad Beograd

Rezultati evaluacije su prikazani na slici 3.10.

Metrika	**Vrednost**
MSE	11.596013
MAE	2.055942
R^2	0.099466

Tabela 3.10: Evaluacione metrike `LightGBM` modela sa interakcijama

Dakle, metrike kažu da su se performanse poboljšale, ali ne značajno, odnosno dodate interakcije i transformacije nisu uvele novu informaciju koju `LightGBM` već nije uspevao da uhvati sam.

`LightGBM` je već sposoban da uoči nelinearnosti i interakcije, pa ručne kombinacije (osim vrlo specifičnih) često ne prave veliku razliku.

Dvostepeni pristup

Kako nijedan od prethodnih modela nije davao dovoljno dobre rezultate, probaćemo sada drugačiji pristup.

U pitanju je *dvostepeni pristup* (eng. Hurdle ili Two-part model), koji se u literaturi i industriji često koristi.

Dvostepeni modeli koriste se za modelovanje zavisnih promenljivih koje predstavljaju brojčane podatke sa velikim brojem nula (tzv. *zero-inflated* podaci). Ovi modeli pretpostavljaju da proces generisanja nula i proces generisanja pozitivnih vrednosti potiču iz dva različita mehanizma, pa se model koncipira u dva dela:

- **Prvi deo:** Binarni model (npr. logistička regresija) koji modeluje verovatnoću da je posmatrana vrednost veća od nule, tj. $\mathbb{P}(Y > 0)$.
- **Drugi deo:** tzv. model sa odsecanjem (npr. Poisson ili Negativna binomna regresija) koji modeluje vrednosti Y uslovljene na $Y > 0$.

Ukupna verovatnoća se modeluje kao kombinacija ova dva procesa. Za neki broj $y \in \{1, 2, 3, \dots\}$, verovatnoća se računa kao:

$$\mathbb{P}(Y = y) = \mathbb{P}(Y > 0) \cdot \mathbb{P}(Y = y \mid Y > 0).$$

Za nulu važi:

$$\mathbb{P}(Y = 0) = 1 - \mathbb{P}(Y > 0).$$

Glavna prednost dvostepenog modela u odnosu na klasične modele (npr. Puasonova ili Negativna binomna regresija) jeste sposobnost da se odvojeno modeluje prisustvo nula i raspodela pozitivnih vrednosti, što ga čini naročito pogodnim za situacije gde su nule posledica posebnog procesa (npr. neaktivnosti osiguranika) koji je kvalitativno različit od procesa koji generiše pozitivne vrednosti (npr. broj šteta kod aktivnih osiguranika).

U prvom koraku dvostepenog modela vršimo binarnu klasifikaciju da li osiguranik ima bar jednu prijavljenu štetu ($Y > 0$). Za ovu svrhu koristi se niz različitih modela i

strategija balansiranja klase, s obzirom da podaci imaju izrazitu neravnotežu između klasa (dominira klasa sa $Y = 0$).

Modeli koje ćemo prvo razmotriti su:

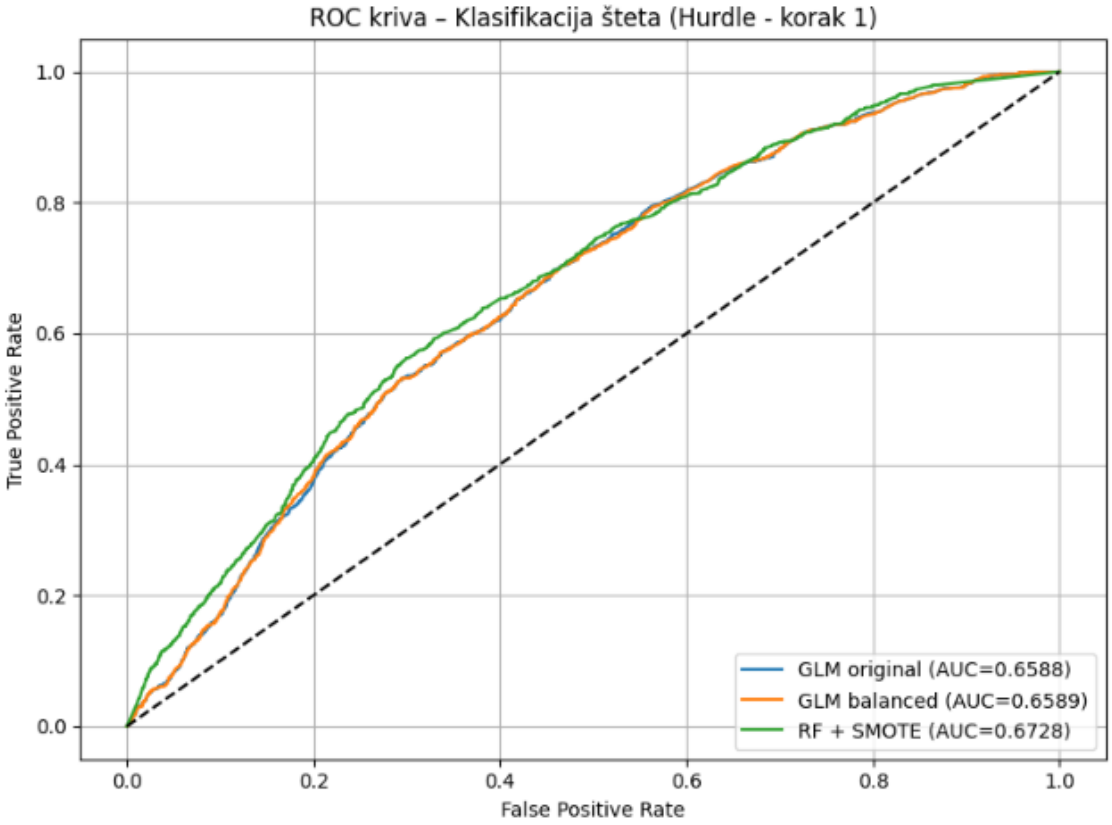
- uopšteni linearni model - klasična logistička regresija bez balansiranja klasa.
- balansiran uopšteni linearan model - logistička regresija sa parametrom $class\_weight$ = 'balanced', čime se automatski kompenzuje disbalans klasa.
- klasifikacioni model slučajnih šuma uz korišćenje SMOTE tehnike (eng. Synthetic Minority Over-sampling Technique), predstavlja tehniku generisanja novih sintetičkih primera iz manjinske klase u cilju rešavanja problema neuravnoteženosti klasa. Koristi se za proširenje skupa podataka pre treniranja modela. Više detalja o ovoj tehnici može se pronaći u [16].

Za svaki model izračunaćemo sledeće metrike:

- Tačnost (eng. Accuracy) – udeo tačno klasifikovanih primera u odnosu na ukupan broj primera.
- Preciznost (eng. Precision) – udeo tačno predviđenih pozitivnih primera u odnosu na sve predviđene pozitivne primere.
- Odziv (eng. Recall) – udeo tačno predviđenih pozitivnih primera u odnosu na sve stvarne pozitivne primere.
- F1 skor – harmonijska sredina preciznosti i odziva; koristan kada postoji neravnoteža u broju primera po klasama.
- ROC kriva – grafički prikaz odnosa između stope lažno pozitivnih i tačno pozitivnih predikcija pri različitim pragovima klasifikacije.
- AUC (eng. Area Under the Curve) – površina ispod ROC krive; pokazuje koliko je model uspešan u razlikovanju klasa.
- Matrica konfuzije – tabela koja prikazuje broj tačnih i pogrešnih klasifikacija po klasama.

Rezultati prikazani u tabelama 3.11 i 3.12 ilustruju uporedne performanse tri klasifikaciona modela.

Iz prikazane ROC krive (slika 3.20) i tabela metrika možemo izvesti sledeće zaključke:



Slika 3.20: ROC kriva za tri kandidata

Model	**Tačnost**	**Preciznost**	**Odziv**	**F1 skor**	**AUC**
uopšteni linearni model	0.6362	0.5006	0.1886	0.2740	0.6588
balansiran uopšteni linearni model	0.6160	0.4781	0.6003	0.5323	0.6589
model slučajnih šuma+SMOTE	0.6502	0.5180	0.5586	0.5376	0.6728

Tabela 3.11: Performanse tri kandidata

Model	**Matrica konfuzije**
uopšteni linearni model	[3405, 411], [1772, 412]
balansiran uopšteni linearni model	[2385, 1431], [873, 1311]
model slučajnih šuma+SMOTE	[2681, 1135], [964, 1220]

Tabela 3.12: Matrice konfuzije kandidata

- uopšten linearan model ostvaruje AUC od 0.6588, sa vrlo niskim odzivom (Recall = 0.1886), što ukazuje da značajan broj osiguranika sa štetom nije pravilno klasifikovan.

- balansiran uopšten linearan model blago popravlja odziv (Recall = 0.6003), ali po cenu smanjenja preciznosti (Precision = 0.4781) i tačnosti. Ipak, AUC ostaje praktično identičan (0.6589), što implicira da je poboljšanje u recall-u kompenzovano većim brojem lažno pozitivnih predikcija.

- klasifikacioni model slučajnih šuma + SMOTE tehnika pokazuje najstabilnije performanse sa najboljim balansom između svih metrika:

  – Najveći AUC: 0.6728

  – Najveći F1 skor: 0.5376

  – Najveću preciznost: 0.5180

  – Najveća tačnost: 0.6502

  Ovo ga čini najefikasnijim modelom u ovom koraku klasifikacije.

Matrice konfuzije takođe pokazuju da model slučajnih šuma u kombinaciji sa SMOTE tehnikom ima bolji odnos tačnih i netačnih predikcija u klasi $y = 1$.

Sada ćemo za prvi korak dvostepenog pristupa pokušati sa modelom `LightGBM`, s obzirom na to da je u prethodnim analizama pokazao najbolje performanse prilikom predviđanja broja šteta.

Model je treniran na originalnoj neuravnoteženoj bazi podataka, bez dodatnog uzorkovanja. Na slici je prikazana je ROC kriva sa odgovarajućom AUC vrednošću.

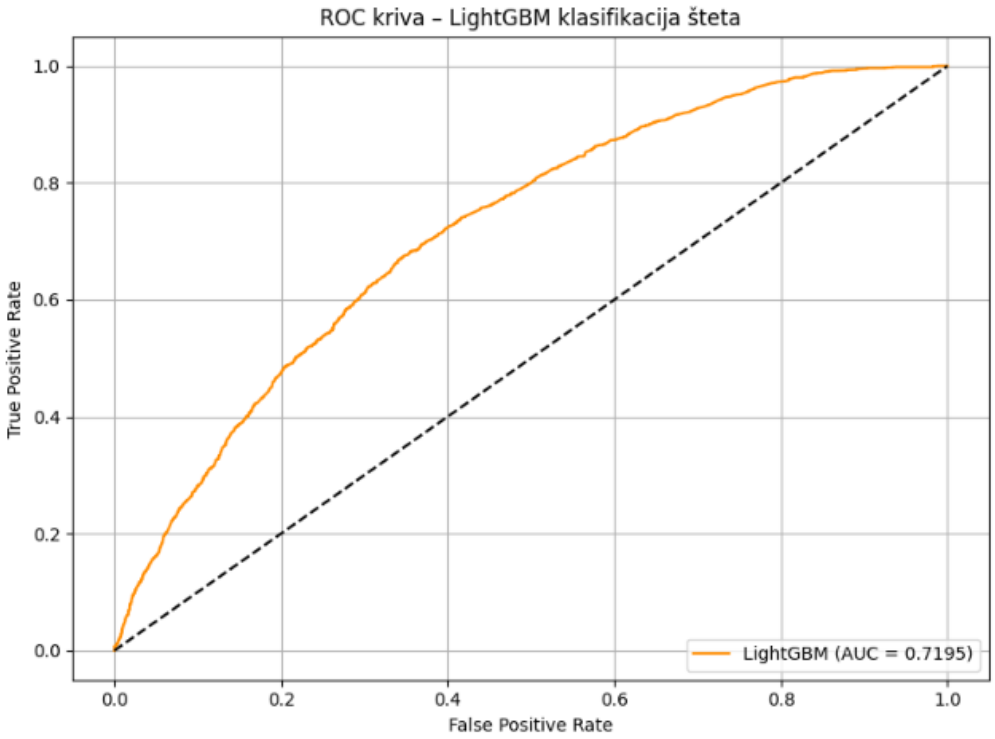
Model je evaluiran na test skupu pomoću standardnih klasifikacionih metrika, sa rezultatima prikazanim u tabelama 3.13 i 3.14.

Metrika	**Vrednost**
Tačnost	0.6818
Preciznost	0.5999
Odziv	0.3782
F1 skor	0.4639
AUC	0.7195

Tabela 3.13: Rezultati `LightGBM` klasifikatora po osnovnim metrikama

	Predikcija 0	**Predikcija 1**
Klasa 0	3265	551
Klasa 1	1358	826

Tabela 3.14: Matrica konfuzije `LightGBM` klasifikatora



Slika 3.21: ROC kriva `LightGBM` modela

Dobijeni rezultati ukazuju na to da `LightGBM` model postiže bolju razliku između pozitivne i negativne klase u poređenju sa prethodno analiziranim modelima (npr. uopšteni linearan model i model slučajnih šuma), što potvrđuje i najveća AUC vrednost od 0.7195.

Međutim, iako je preciznost solidna (0.5999), odziv (Recall = 0.3782) je i dalje nizak, što znači da značajan broj osiguranika koji imaju prijavljenu štetu nije prepoznat od strane modela (TP = 826, FN = 1374).

F1 skor od 0.4639 ukazuje na umeren balans između preciznosti i odziva. Ipak, u kontekstu zadatka gde je cilj identifikovati osiguranike sa potencijalnim štetama, neophodna je dalja optimizacija modela, kao i balansiranje podataka ili prilagođavanje klasifikacionog praga.

Sada ćemo izvršiti optimizaciju hiperparametara koristeći `RandomizedSearchCV` sa petostrukom ukrštenom validacijom (CV=5) unutar trening skupa, na isti način kao i kod prvog `LightGBM` modela koji smo razmatrali pre dvostepenog pristupa. Ciljna funkcija optimizacije bila je `ROC AUC`, a pretraga je izvršena na osnovu sledeće mreže parametara:

- `n_estimators` $\in \{100, 200, 300\}$
- `learning_rate` $\in \{0.01, 0.05, 0.1\}$
- `num_leaves` $\in \{15, 31, 50\}$
- `max_depth` $\in \{-1, 5, 10\}$
- `min_child_samples` $\in \{10, 20, 30\}$
- `subsample` $\in \{0.6, 0.8, 1.0\}$
- `colsample_bytree` $\in \{0.6, 0.8, 1.0\}$

Najbolji pronađeni skup hiperparametara, prema kriterijumu maksimalne prosečne AUC vrednosti na validacionim podacima, prikazan je u Tabeli 3.15.

Parametar	**Vrednost**
subsample	0.6
num_leaves	50
n_estimators	200
min_child_samples	20
max_depth	10
learning_rate	0.05
colsample_bytree	0.6

Tabela 3.15: Najbolji hiperparametri `LightGBM` modela

Najbolji ostvareni rezultat tokom ukrštene validacije iznosi:

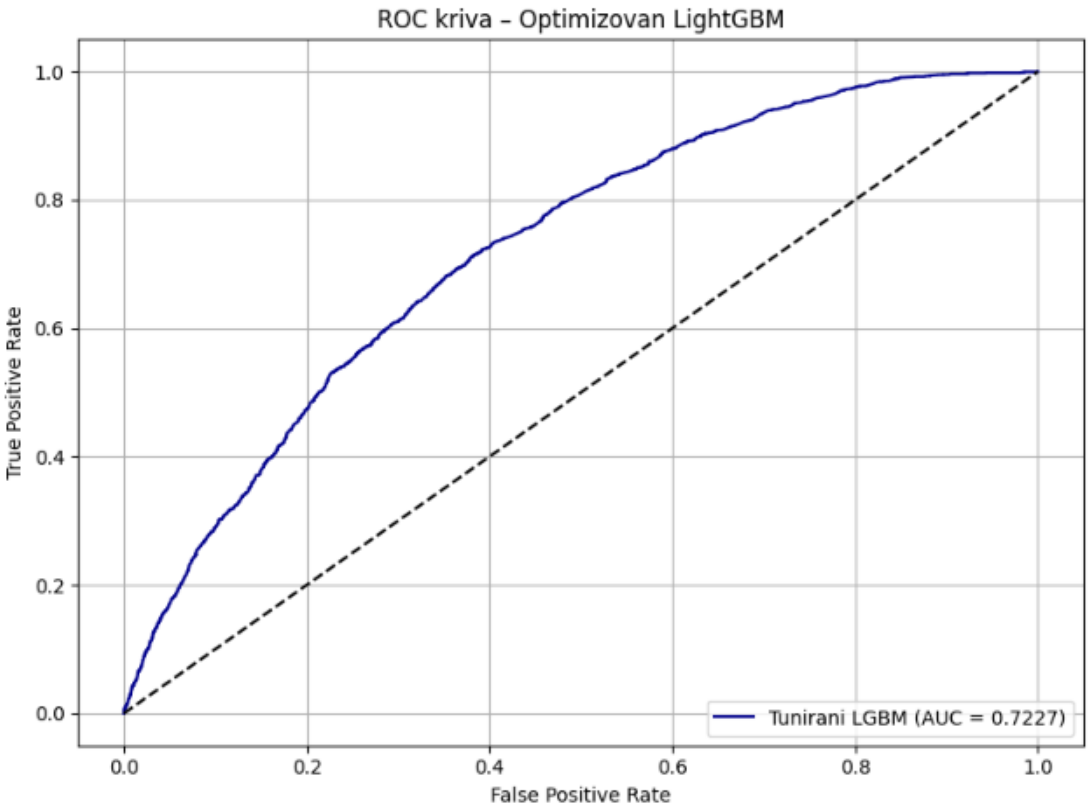
$$\text{AUC}_{\text{CV}} = 0.7214.$$

Model je treniran sa pronađenim optimalnim parametrima, a rezultati evaluacije na test skupu prikazani su u tabelama 3.16 i 3.17. ROC kriva na test skupu prikazana je na slici 3.22.

Optimizovan `LightGBM` model ostvaruje najbolju AUC vrednost do sada (0.7228), što ukazuje da bolje razlikuje osiguranike sa i bez prijavljene štete.

Uprkos i dalje niskom odzivu (što je delimično očekivano kod neravnotežnih klasa), optimizovan `LightGBM` pokazuje najbolje rezultate.

Pošto je cilj klasifikacije u ovom koraku dvostepenog modela da što bolje identifikuje osiguranike sa bar jednom prijavljenom štetom, posebno je važan balans između preciznosti i odziva, odnosno F1 skor. S obzirom da standardna vrednost



Slika 3.22: ROC kriva – Optimizovan `LightGBM` model

Metrika	**Vrednost**
Tačnost	0.6790
Preciznost	0.5921
Odziv	0.3796
F1 skor	0.4626
AUC	0.7228

Tabela 3.16: Rezultati `LightGBM` klasifikatora po osnovnim metrikama

	Predikcija 0	**Predikcija 1**
Klasa 0	3245	571
Klasa 1	1355	829

Tabela 3.17: Matrica konfuzije `LightGBM` klasifikatora

klasifikacionog praga $t = 0.5$ ne mora biti optimalna u uslovima neuravnoteženih klasa, izvršena je dodatna analiza optimizacije praga.

Testirano je više vrednosti praga u intervalu $[0.10, 0.90]$ sa korakom od 0.01, a

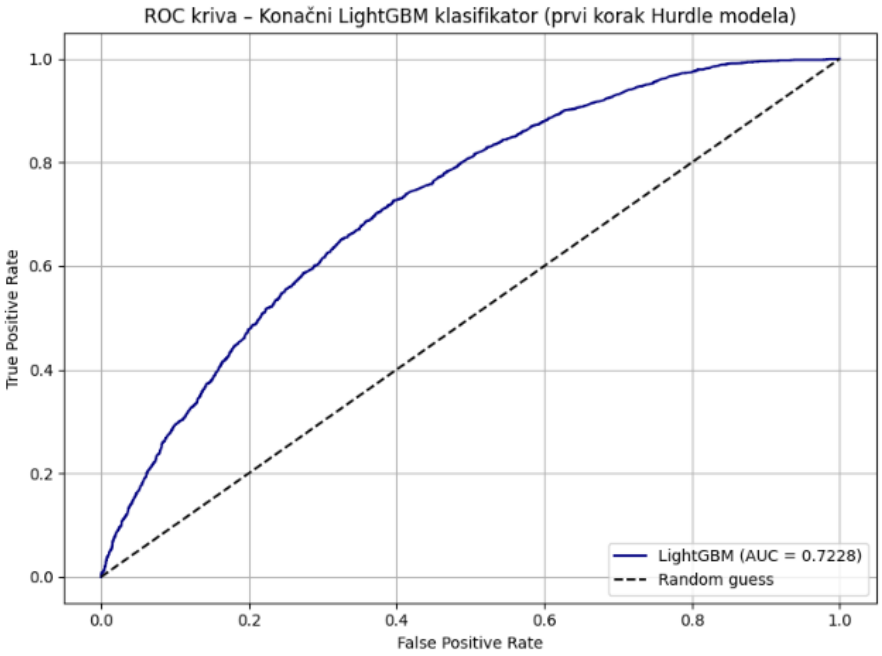
za svaku vrednost izračunat je F1 skor na test skupu. Najviša vrednost F1 skora postignuta je za prag:

$$t^* = 0.31, \quad \text{pri čemu je F1} = 0.6040.$$

Ova vrednost predstavlja značajno poboljšanje u odnosu na prethodni rezultat sa podrazumevanim pragom ($F1 = 0.4626$), čime se potvrđuje da odgovarajuća selekcija praga može znatno poboljšati performanse klasifikacionog modela u zadacima sa neravnotežnim klasama.

Sada ćemo trenirati naš konačan model sa optimalnim pragom i parametrima. Konačan model koristi:

- `LightGBM` klasifikator sa prethodno podešenim hiperparametrima (Tabela 3.15)
- Klasifikacioni prag: $t = 0.31$



Slika 3.23: ROC kriva konačnog `LightGBM` modela (korak 1)

Evaluacija na test skupu dala je rezultate prikazane u tabeli 3.18, konačna matrica konfuzije u tabeli 3.19.

Metrika	Vrednost
Tačnost	0.6052
Preciznost	0.4757
Odziv	0.8274
F1 skor	0.6040
AUC	0.7228

Tabela 3.18: Rezultati konačog `LightGBM` klasifikatora po osnovnim metrikama

	Predikcija 0	**Predikcija 1**
Klasa 0	1824	1992
Klasa 1	377	1807

Tabela 3.19: Matrica konfuzije konačnog `LightGBM` klasifikatora

Uvođenjem optimalnog praga klasifikacije $t = 0.31$, značajno je poboljšan odziv modela (0.8274), uz razumnu preciznost (0.4757), što rezultira F1 skorom od 0.6040. U ovom tipu problema (predviđanje da li će osiguranik imati štetu), važnije je da pravilno odredimo što više onih koji će imati štetu — čak i po cenu da ponekad pogrešno označimo nekoga ko zapravo neće imati štetu. Zato je odziv je prioritet, odnosno važno je da uočimo sve slučajeve osiguranika koji će imati štetu. Naš odabrani model uspešno prepoznaje $82,7\%$ ovih osiguranika. Preciznost može biti nešto manja, jer je manja šteta ako nekog pogrešno označiš kao rizičnog. Tačnost ovde takođe nije najadekvatnija mera, jer klasa onih osiguranika koji nisu imali štetu dominira, pa bi model mogao da bude visoko tačan čak i da ignoriše štete.

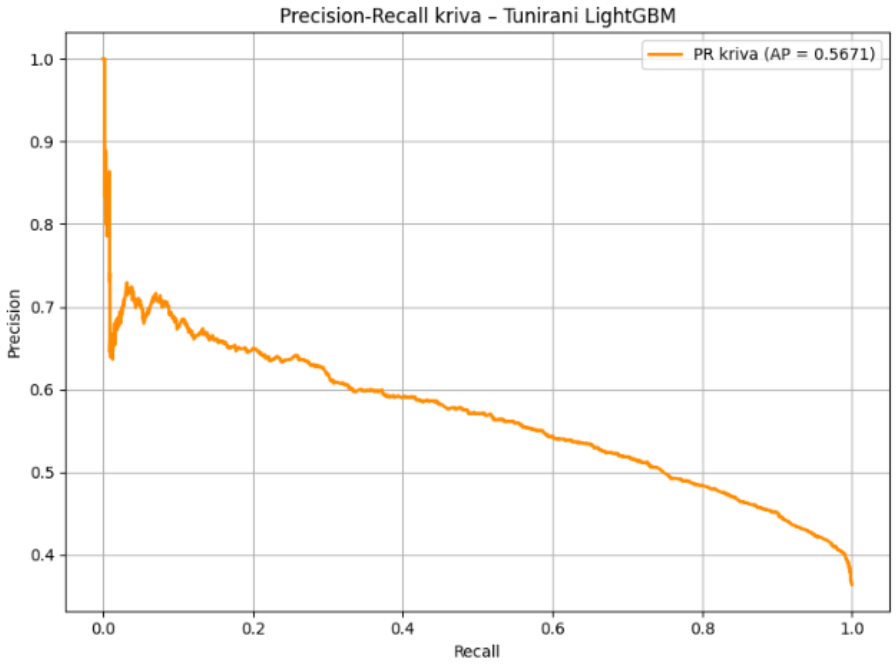
Pored ROC krive, dodatno je analizirana **Precision–Recall (PR) kriva**, koja se posebno preporučuje u kontekstu klasifikacionih problema sa neuravnoteženim klasama. PR kriva prikazuje odnos između preciznosti i odziva modela pri različitim klasifikacionim pragovima, za naš optimizovani model.

Na slici 3.24, možemo videti da prosečna preciznost (eng. Average Precision, AP) iznosi

$$\text{AP} = 0.5671$$

Ova vrednost prikazuje da model uspešno održava visoku preciznost pri različitim nivoima odziva. S obzirom na to da je klasifikacioni zadatak nebalansiran (klasa sa štetama je ređa), PR kriva nudi realističniju procenu performansi od ROC krive.

Kriva pokazuje relativno visok nivo preciznosti pri niskim i srednjim vrednostima odziva, ali i pad preciznosti kako se odziv povećava. Ovo je očekivano, jer povećanjem



Slika 3.24: Precision-Recall kriva konačnog `LightGBM` modela

odziva model sve više instanci klasifikuje kao pozitivne, što povećava broj lažno pozitivnih predikcija.

Zaključno, PR kriva i odgovarajuća vrednost AP potvrđuju da optimizovani `LightGBM` model postiže zadovoljavajuću ravnotežu između tačnosti i obuhvatnosti klasifikacije osiguranika sa štetama, što je od ključne važnosti za prvi korak dvostepenog pristupa.

Kako je važno detektovati što veći broj osiguranika sa štetama, ovaj rezultat predstavlja najbolji kompromis između tačnih pozitivnih i lažnih pozitivnih klasifikacija, pa ćemo ovako podešen `LightGBM` model koristi kao prvi korak dvostepenog pristupa.

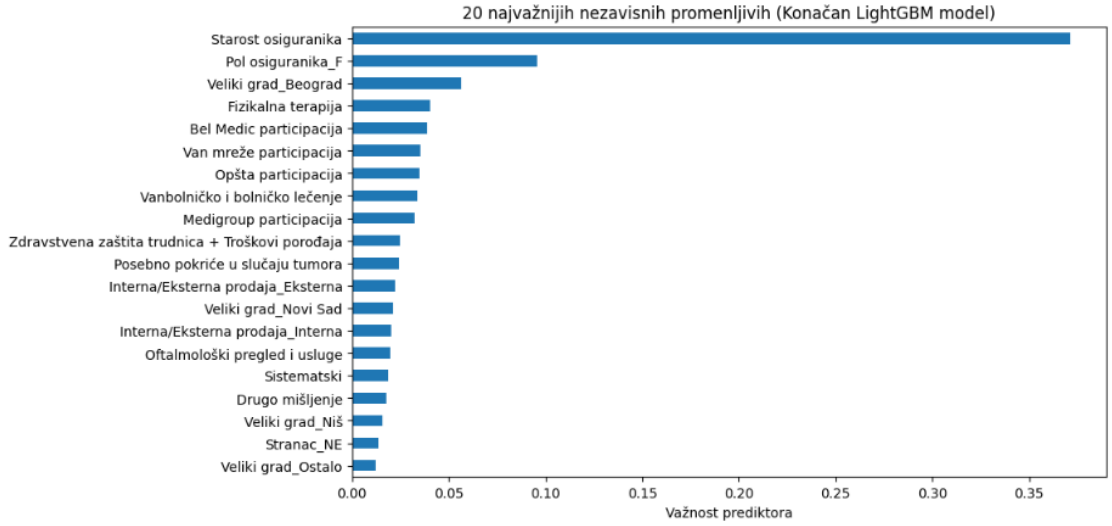
Interpretacija konačnog `LightGBM` modela za prvi korak dvostepenog pristupa

U cilju bolje interpretacije izabranog modela koristićemo sledeće metode:

- važnost nezavisnih promenljivih zasnovana na njihovom doprinosu smanjenju greške,

- metodu akumuliranih lokalnih efekata (ALE): ALE grafici će biti prikazani za ključne numeričke i binarne promenljive.
- Šeplijeve vrednosti, prikazaćemo:
  - grafik raspodele uticaja nezavisnih promenljivih,
  - Bar plot Šeplijevih vrednosti nezavisnih promenljivih,
  - grafik toka doprinosa.

Kombinovanjem navedenih metoda obezbeđujemo dublje razumevanje načina na koji model donosi odluke, kao i identifikaciju ključnih faktora rizika za nastanak šteta.



Slika 3.25: Važnost nezavisnih promenljivih konačnog `LightGBM` modela

Na slici 3.25 su prikazani najvažnije nezavisne promenljive za model klasifikacije da li će osiguranik imati bar jednu prijavljenu štetu. Vizualizacija je zasnovana na vrednostima važnosti nezavisnih promenljivih iz `LightGBM` modela.

- Najznačajnija nezavisna promenljiva je Starost osiguranika, sa ubedljivo najvećim doprinosom u odnosu na ostale promenljive. To sugeriše da verovatnoća nastanka štete značajno zavisi od starosne dobi osiguranika.
- Sledeći po značaju su Pol osiguranika (kategorija "ženski") i pripadnost velikom gradu Beograd, što ukazuje na moguće razlike u obrascima korišćenja osiguranja između demografskih grupa i regiona.

- U gornjem delu liste nalaze se i promenljive koje opisuju pokriće, kao što su fizikalna terapija, vanbolničko i bolničko lečenje, opšta participacija, kao i pokrivenosti specifičnih mreža zdravstvenih ustanova (Bel Medic, Medigroup, Van mreže).

Dodatne interpretacione tehnike koje ćemo prikazati pokazaće na koji način ove nezavisne promenljive utiču na verovatnoću nastanka štete.

Zaključujemo da model najviše oslanja svoje predikcije na opšte demografske karakteristike (starost, pol, mesto stanovanja), ali i na elemente vezane za strukturu ugovora i pokrivenost zdravstvenih usluga. Ove informacije mogu biti korisne za dalje targetiranje klijenata i analizu rizika.

Metoda akumuliranih lokalnih efekata

Prikazaćemo ALE grafike za neke od najznačajnih nezavisnih promenljivih prema celokupnoj prethodnoj analizi, koristeći funkcije koje se mogu se pronaći u GitHub repozitorijumu [9].

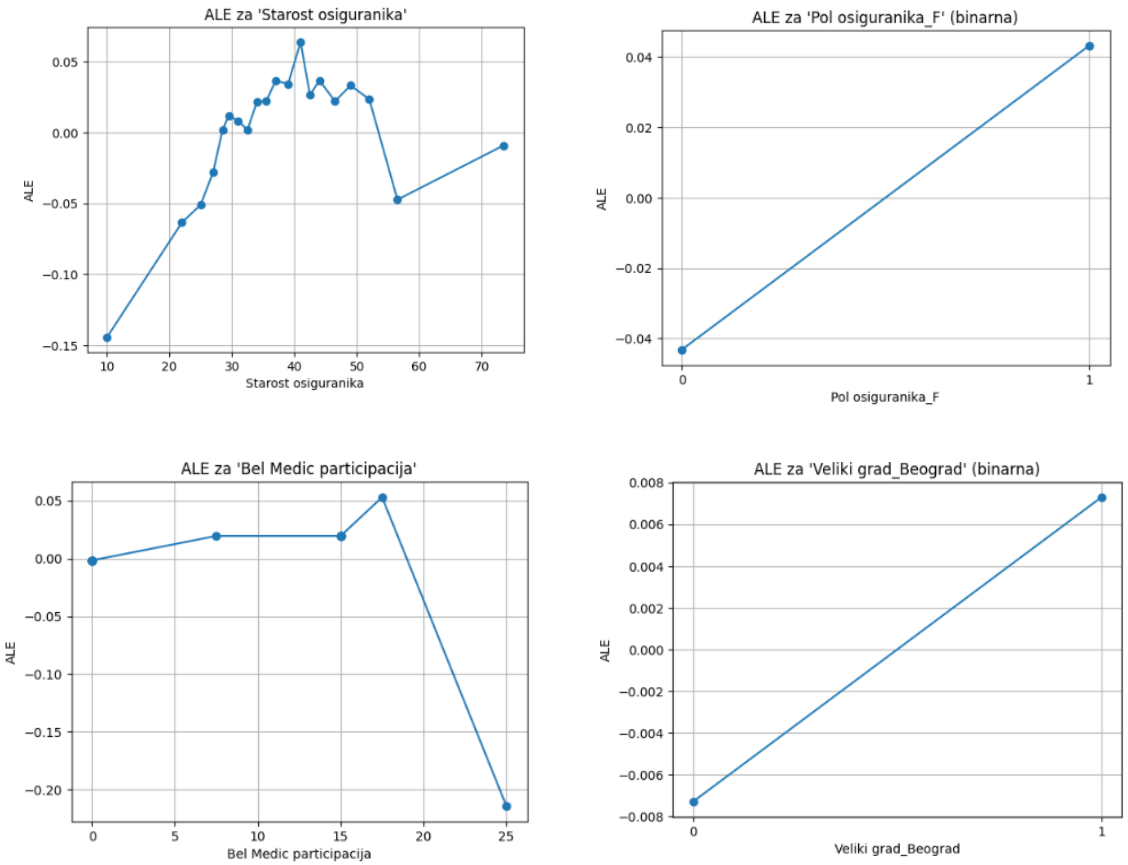
Sa slike 3.26, zaključujemo:

- Starost osiguranika - može se primetiti da se uticaj starosti na verovatnoću nastanka štete menja nelinearno. Verovatnoća raste od najmlađih uzrasta do 40. godine, nakon čega dolazi do postepenog opadanja. Najveći pozitivni efekat zapaža se u uzrastu između 35 i 41 godine, što može ukazivati na fazu života u kojoj se osiguranici najviše oslanjaju na usluge zdravstvenog osiguranja.

- Pol osiguranika - prikazan je uticaj binarne promenljive Pol osiguranika_F. Vidimo da žene (vrednost 1) imaju u proseku veću verovatnoću nastanka štete u odnosu na muškarce (vrednost 0), što može ukazivati na češće korišćenje zdravstvenih usluga među ženskom populacijom.

- Pripadnost velikom gradu Beograd - osiguranici koji žive u Beogradu (vrednost 1) imaju višu verovatnoću nastanka štete u poređenju sa onima koji ne žive u ovom gradu (vrednost 0). Ovakav rezultat može biti posledica dostupnosti određenih mreža zdravstvenih ustanova ili specifične strukture korisnika u ovom regionu.

- Bel Medic participacija - prikazana je analiza za promenljivu Bel Medic participacija, koja označava iznos obavezne participacije (franšize) koju osiguranik



Slika 3.26: ALE grafici za konačan `LightGBM` model – odabrane nezavisne promenljive

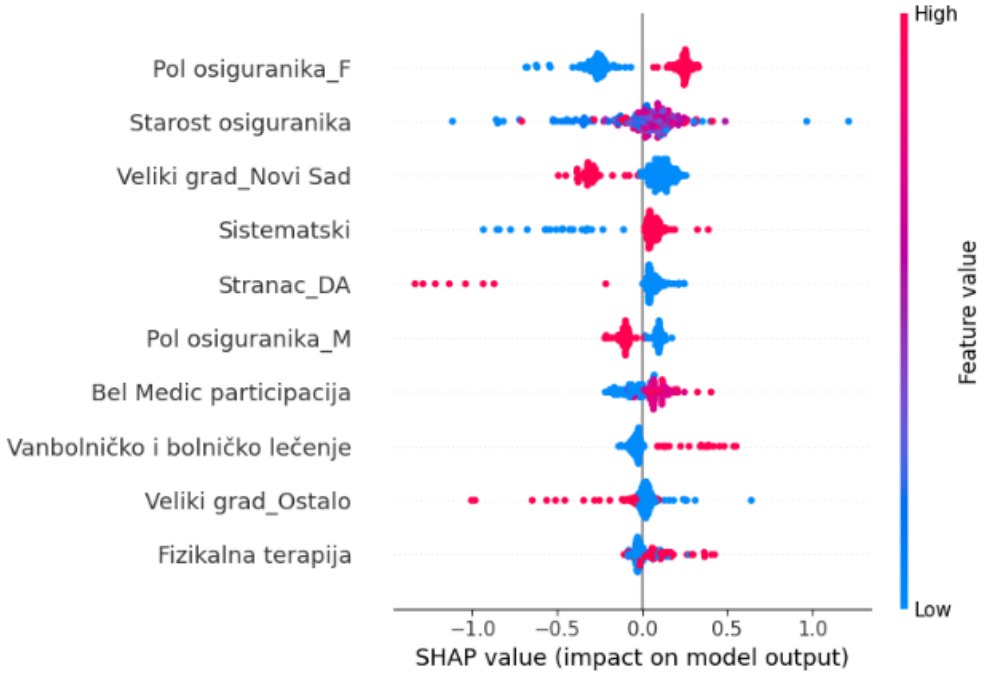
snosi ukoliko koristi usluge klinike Bel Medic. Vidimo da niska ili umerena participacija ima neutralan do blago pozitivan efekat na verovatnoću štete, dok veće vrednosti participacije značajno smanjuju verovatnoću. Ovo može ukazivati na to da veći lični troškovi odvraćaju osiguranike od korišćenja usluga, što u krajnjoj liniji smanjuje prijavu šteta.

Šeplijeva metoda

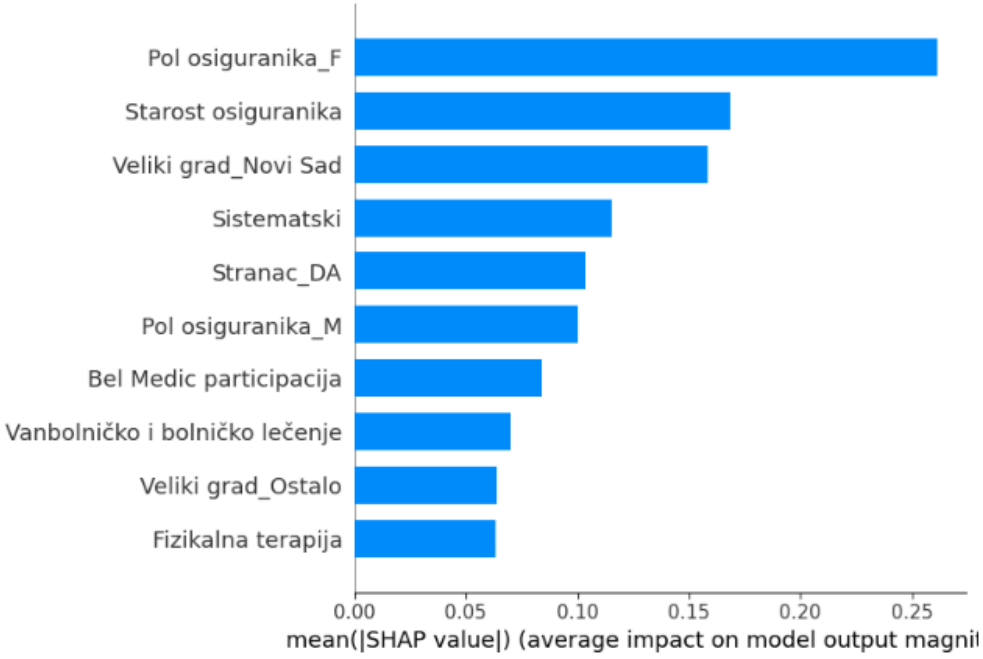
Prikazaćemo grafik raspodele uticaja nezavisnih promenljivih, bar plot uticaja nezavisnih promenljivih i grafik toka doprinosa nezavisnih promenljivih za konkretnu instancu.

Primećujemo:

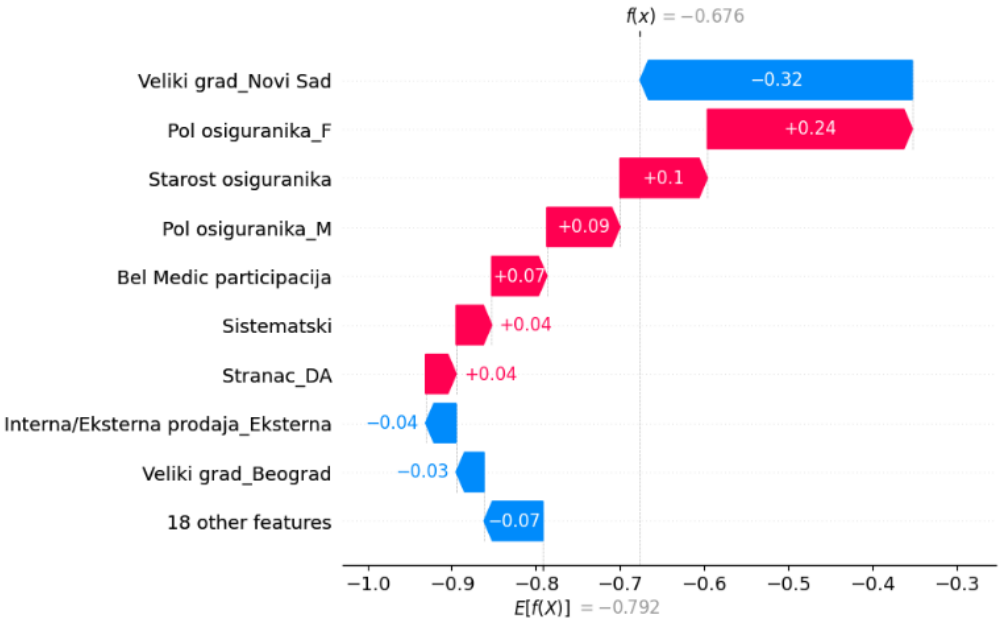
- **Šeplijev bar plot**, slika 3.28 - prikazan je prosečan doprinos nezavisnih promenljivih u apsolutnim Šeplijevim vrednostima. Najveći prosečan doprinos



Slika 3.27: Grafik raspodele uticaja nezavisnih promenljivih - konačan `LightGBM`



Slika 3.28: Bar Plot - Konačan `LightGBM`



Slika 3.29: Grafik toka doprinosa nezavisnih promenljivih - Konačan `LightGBM`

ima promenljiva Pol osiguranika_F, što ukazuje da pol ima najjači globalni uticaj na odluke modela. Slede Starost osiguranika i Veliki grad_Novi Sad.

- **Šeplijev grafik raspodele uticaja nezavisnih promenljivih**, slika 3.27 - prikazan je detaljan uvid u raspodelu Šeplijevih vrednosti za najvažnije nezavisne promenljive. Boja označava vrednost nezavisne promenljive (crveno za visoke vrednosti, plavo za niske).
  - Na primer, za promenljivu Pol osiguranika_F, manje vrednosti (muški pol) uglavnom doprinose smanjenju verovatnoće štete (negativne Šeplijeve vrednosti), dok ženski pol ima pozitivan uticaj.
  - Za Starost osiguranika, niža starost doprinosi smanjenju verovatnoće, dok starija populacija srednjih godina ima tendenciju pozitivnog uticaja.
  - Veliki_grad_Novi_Sad – pozitivna Šeplijeva vrednost kod osiguranika iz Novog Sada ukazuje da pripadnici tog regiona imaju manju verovatnoću prijave štete.
- **Šeplijev grafik toka doprinosa nezavisnih promenljivih za konkretnu instancu**, slika 3.29 - prikazana je lokalna interpretacija našeg modela za jednog osiguranika. Pol i starost ženskog osiguranika najviše doprinose većoj

verovatnoći nastanka štete, dok Veliki grad_Novi Sad doprinosi smanjenju predikcije. Suma svih doprinosa vodi do konačne predikcije za konkretnog osiguranika.

Interpretacija klasifikacionog `LightGBM` modela pokazala je da ključni faktori koji utiču na verovatnoću nastanka bar jedne štete uključuju starost, pol i mesto stanovanja osiguranika, kao i elemente strukture osiguravajuće ponude, poput ugovorenih pokrića i nivoa participacije.

Analiza važnosti nezavisnih promenljivih je identifikovala najuticajnije promenljive, dok su ALE grafici pružili uvid u pravac i intenzitet uticaja pojedinačnih nezavisnih promenljivih. Šeplijeva analiza je dodatno omogućila detaljno tumačenje, kako na globalnom tako i na lokalnom nivou, potvrđujući rezultate prethodnih metoda i nudeći transparentno objašnjenje pojedinačnih predikcija.

Kombinovanjem ovih pristupa dobijen je pouzdan i interpretabilan pregled ponašanja modela, što predstavlja važan korak ka njegovoj upotrebi u realnim poslovnim odlukama u domenu osiguranja.

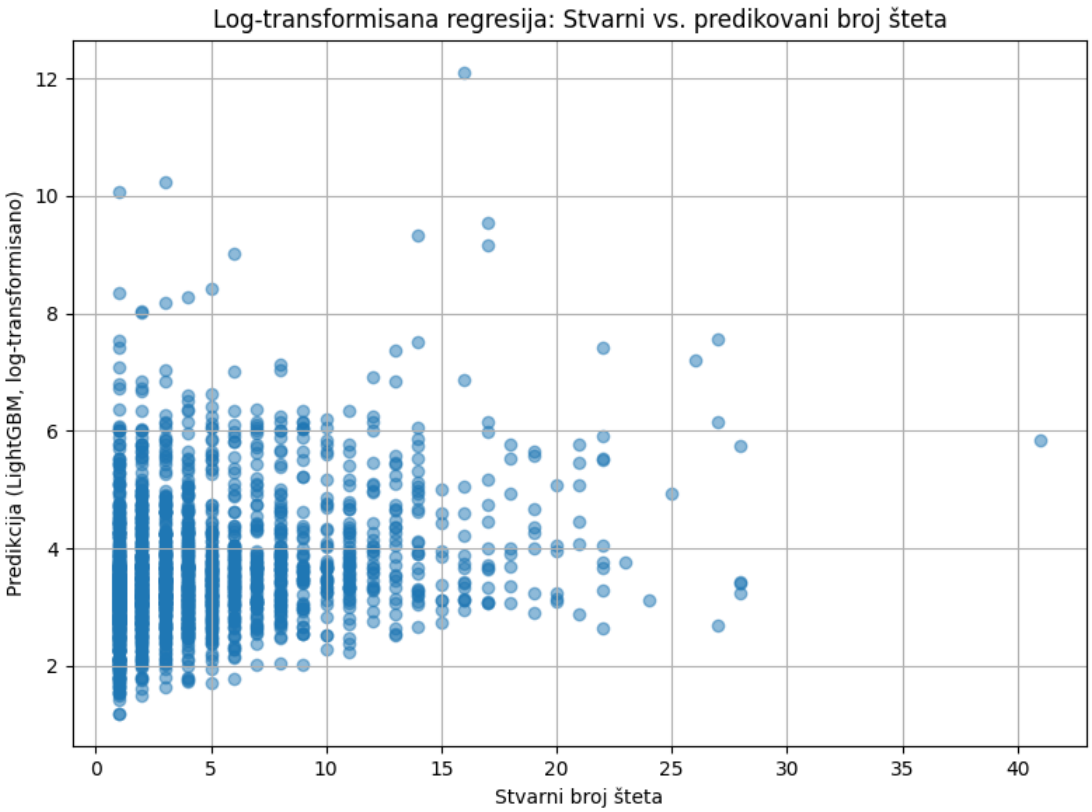
Drugi deo dvostepenog pristupa

Nakon što je u prvom koraku dvostepenog modela procenjena verovatnoća da osiguranik ima barem jednu prijavljenu štetu (tj. $P(Y > 0)$), u drugom koraku fokusiramo se isključivo na one osiguranike kod kojih je utvrđeno da je do štete došlo, i modelujemo *intenzitet šteta*, odnosno očekivanog broja prijava kod osiguranika za koje je procenjeno da će imati štetu. Ovaj deo modela nam omogućava da kvantifikujemo koliko su osiguranici skloni višestrukim prijavama.

U početnim fazama analize razmatrani su modeli poput Puasonove i negativne binomne regresije [8, 3], koja predstavlja proširenje Puasonovog modela i omogućava veću fleksibilnost kroz dodatni parametar disperzije.

Kako broj prijavljenih šteta kod osiguranika predstavlja celobrojnu, asimetričnu na desno promenljivu, sa dugim desnim repom i velikim brojem ekstremnih vrednosti, razmotren je i `LightGBM` regresor, sa logaritamskom transformacijom ciljne promenjive, kako bi bolje pratio nebalansiranost podataka.

Ovaj model je bolje odgovarao podacima, ali i pored toga, nije davao zadovoljavajuće rezultate, što možemo najbolje videti na dijagramu raspršenosti na slici 3.30.



Slika 3.30: Scatter dijagram `LightGBM` regresora (korak 2)

Uočen je obrazac zbijenosti predikcija u relativno uskom opsegu vrednosti, posebno kod osiguranika sa većim brojem prijavljenih šteta, gde model pokazuje smanjenu sposobnost razlikovanja između različitih nivoa ciljne promenljive. Iako log-transformacija ublažava uticaj ekstremnih vrednosti i doprinosi stabilizaciji modela, njena primena nije dovela do značajnog poboljšanja u objašnjavanju varijabilnosti broja šteta. Ovakvi rezultati sugerišu da bi unapređenje performansi zahtevalo drugačiji pristup modelovanju.

Zato smo u ovom radu odlučili za klasifikacioni pristup, gde su osiguranici razvrstani u nekoliko grupa na osnovu broja prijavljenih šteta.

Ovakav pristup se pokazao robusnijim i interpretativno pogodnijim za potrebe modela intenziteta šteta, pa je na kraju odabran kao finalno rešenje.

Odluka da se koristi klasifikacija umesto regresije motivisana je sledećim razlozima:

- Raspodela broja šteta je izrazito asimetrična i ne prati klasične pretpostavke

modela za brojčane podatke.

- Klasifikacioni modeli su pokazali bolje performanse u evaluaciji, naročito u pogledu F1-score metrike.
- Grupisanje osiguranika u kategorije omogućava jasniju poslovnu interpretaciju i donošenje odluka (npr. dodatna provera kod visoko rizičnih korisnika).

Dakle, odlučeno je da primenimo *multiklasifikacioni pristup*[17], u kojem se broj šteta grupiše u diskretne kategorije rizika. Ovaj pristup omogućava bolje razlikovanje između osiguranika na osnovu poslovno relevantnih kriterijuma (npr. nizak, umeren i visok rizik), uz povećanu robusnost modela i interpretabilnost rezultata.

U narednom delu rada testiraće se `LightGBM` klasifikator sa različitim brojem klasa (3 i 2 klase, jer se veći broj klasa pokazao loše, slično prethodnim regresionim modelima), a kao kriterijum izbora koristiće se performanse modela, stabilnost i poslovna primenljivost.

U prvoj klasifikacionoj postavci, osiguranici su grupisani u tri klase na osnovu broja prijavljenih šteta:

- **Klasa 0**: 1 do 3 štete – *umeren rizik*,
- **Klasa 1**: 4 do 6 šteta – *povišen rizik*,
- **Klasa 2**: 7 i više šteta – *visok rizik*.

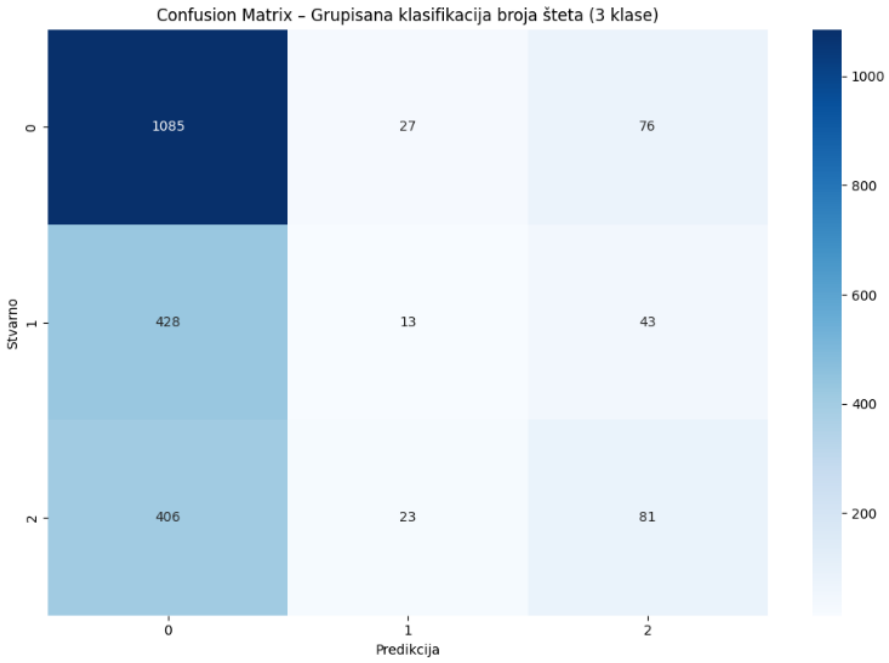
Ova podela omogućava balans između broja klasa i stabilnosti modela, uz dodatnu pogodnost interpretacije za potrebe segmentacije portfolija i upravljanja troškovima osiguranja.

Evaluacija klasifikacionog modela sprovedena je korišćenjem funkcije `classification_report` iz biblioteke `scikit-learn`, koja za svaku klasu prikazuje preciznost, odziv , F1 skor i broj primera. Osim toga, izračunava se i ukupna metrika kao srednja vrednost prethodno pomenutih metrika preko svih klasa, kao i težinska sredina metrika, gde svaka klasa dobija težinu u skladu sa brojem primera.

U tabeli 3.20, možemo videti da najveći broj osiguranika pripada klasi 0 (1–3 štete), koja je modelom najuspešnije identifikovana. U ovoj klasi, preciznost iznosi 0.57, a odziv čak 0.91. Međutim, za klase 1 (4–6 šteta) i 2 (7+ šteta) rezultati su značajno slabiji. Klasa 1 ima vrlo nizak odziv (0.03), dok je kod klase 2 odziv 0.16.

Klasa	Preciznost	Odziv	F1-skor	Broj primera
0	0.57	0.91	0.70	1188
1	0.21	0.03	0.05	484
2	0.41	0.16	0.23	510
Tačnost			0.54	2182
Srednja vrednost metrika	0.39	0.37	0.32	2182
Težinska sredina metrika	0.45	0.54	0.44	2182

Tabela 3.20: Performanse `LightGBM` klasifikatora sa 3 klase



Slika 3.31: Matrica konfuzije `LightGBM` klasifikatora sa 3 klase

Ukupna tačnost modela iznosi 54%, ali niska srednja vrednost metrika (0.32) ukazuje na izrazito neuravnotežene performanse među klasama. Model ima tendenciju da većinu osiguranika klasifikuje u najbrojniju klasu (klasu 0), dok teže razlikuje korisnike sa višim rizikom (klase 1 i 2), što je potvrđeno i uvidom u matricu konfuzije.

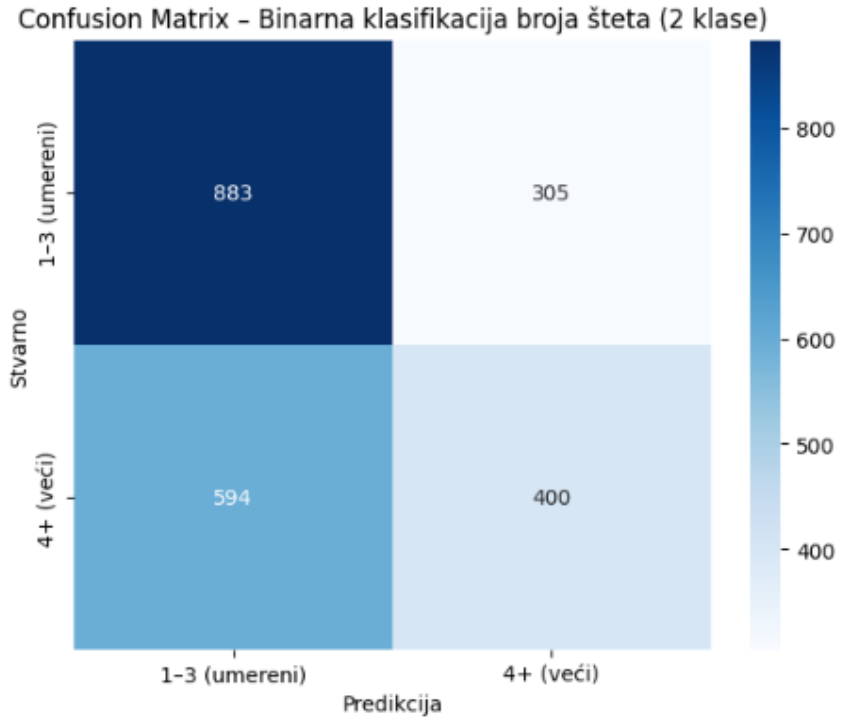
Ovi rezultati ukazuju da je model efikasan u prepoznavanju osiguranika sa umerenim brojem šteta, ali da postoji prostor za unapređenje u tačnoj klasifikaciji korisnika sa povišenim i visokim rizikom. U narednom koraku, razmatra se klasifikacija sa dve klase, u cilju postizanja jednostavnijeg i stabilnijeg modela.

Izvršena je podela osiguranika u dve kategorije:

- **Klasa 0**: 1 do 3 štete - osiguranici sa umerenim rizikom,
- **Klasa 1**: 4 i više šteta - osiguranici sa većim rizikom.

Klasa	Preciznost	Odziv	F1-skor	Broj primera
0	0.60	0.74	0.66	1188
1	0.57	0.40	0.47	994
Tačnost			0.59	2182
Srednja vrednost metrika	0.58	0.57	0.57	2182
Težinska sredina metrika	0.58	0.59	0.58	2182

Tabela 3.21: Performanse `LightGBM` klasifikatora sa 3 klase



Slika 3.32: Matrica konfuzije `LightGBM` klasifikatora sa 2 klase

U tabeli 3.21 vidimo da model ostvaruje ukupnu tačnost od 59%, što je poboljšanje u odnosu na prethodnu podelu u tri klase. Klasa 0 (osiguranici sa 1-3 štete) prepoznata je sa odzivom od 74% i F1-skor vrednošću od 0.66, dok klasa 1 (4+ štete) ima nešto niži odziv (40%) i F1-skor od 0.47.

Iako i dalje postoji određeni broj pogrešnih klasifikacija, posebno kada su osiguranici sa većim brojem šteta klasifikovani kao oni sa umerenim rizikom, rezultati

pokazuju znatno bolji balans između preciznosti i odziva u odnosu na prethodni pristup sa tri klase.

Zbog toga se ovaj model identifikuje kao najpouzdanije rešenje za drugi korak dvostepenog pristupa u okviru ovog rada.

Interpretacija drugog modela u okviru dvostepenog pristupa

U okviru druge faze dvostepenog pristupa, fokus je bio na modelovanju rizika kod onih osiguranika kod kojih je model iz prve faze procenio da će se šteta dogoditi. Iako se i ovde koristi nelinearni model (`LightGBM` klasifikator), njegova poslovna uloga je pomoćna, odnosno on klasifikuje osiguranike na one sa manjim (1–3) i većim (4+) brojem šteta, ali unutar grupe koja je već prepoznata kao rizična.

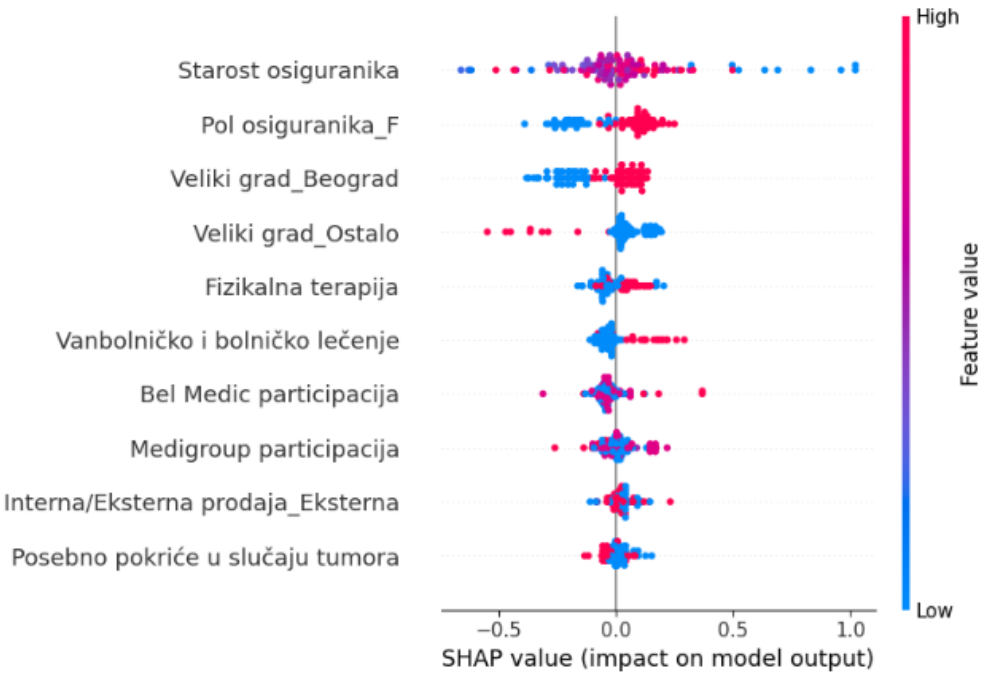
Za razliku od prve faze, koja ima ključnu ulogu u identifikaciji potencijalno štetnih polisa i time direktno utiče na odluke o segmentaciji i prevenciji, drugi model se koristi za dodatnu razradu rizika unutar specifične podgrupe. Zbog toga je poseban akcenat u interpretaciji modela stavljen na prvu fazu.

Ipak, da bi se pružio uvid u faktore koji utiču na veću verovatnoću višestrukih šteta, i ovde se sprovodi interpretacija modela pomoću Šeplijevih vrednosti. Šeplijeva analiza omogućava uvid u značaj pojedinačnih nezavisnih promenljivih i njihovu ulogu u određivanju da li osiguranik pripada grupi sa većim brojem šteta.

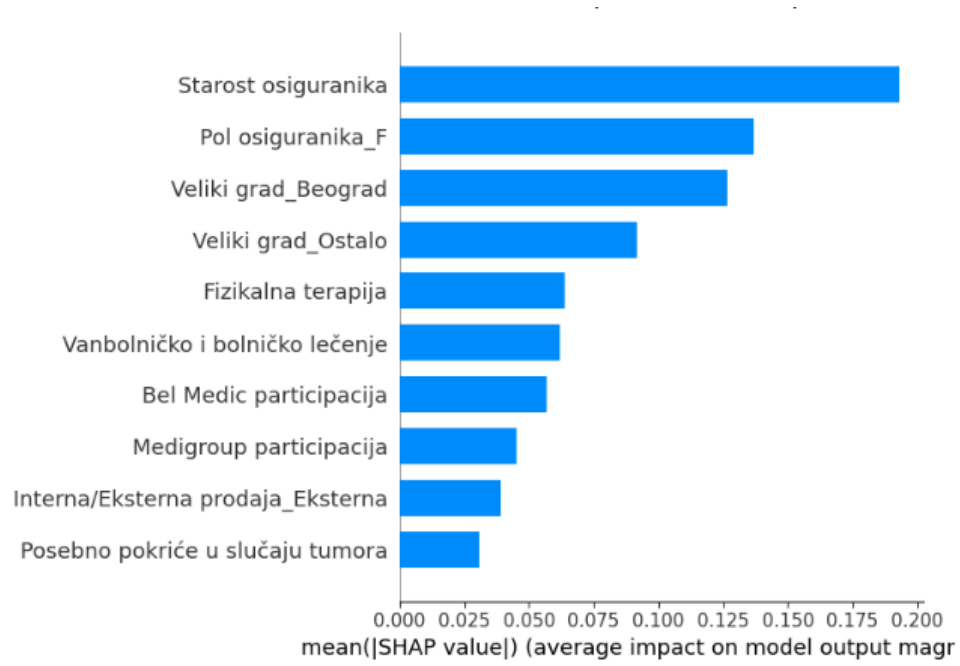
Kao i do sad, prikazujemo grafik raspodele uticaja nezavisnih promenljivih, bar plot uticaja nezavisnih promenljivih i grafik toka doprinosa nezavisnih promenljivih za konkretnu instancu.

Na slici 3.33 prikazan je grafik raspodele uticaja nezavisnih promenljivih koji istovremeno prikazuje uticaj i intenzitet svake promenljive na predikciju. Zaključujemo:

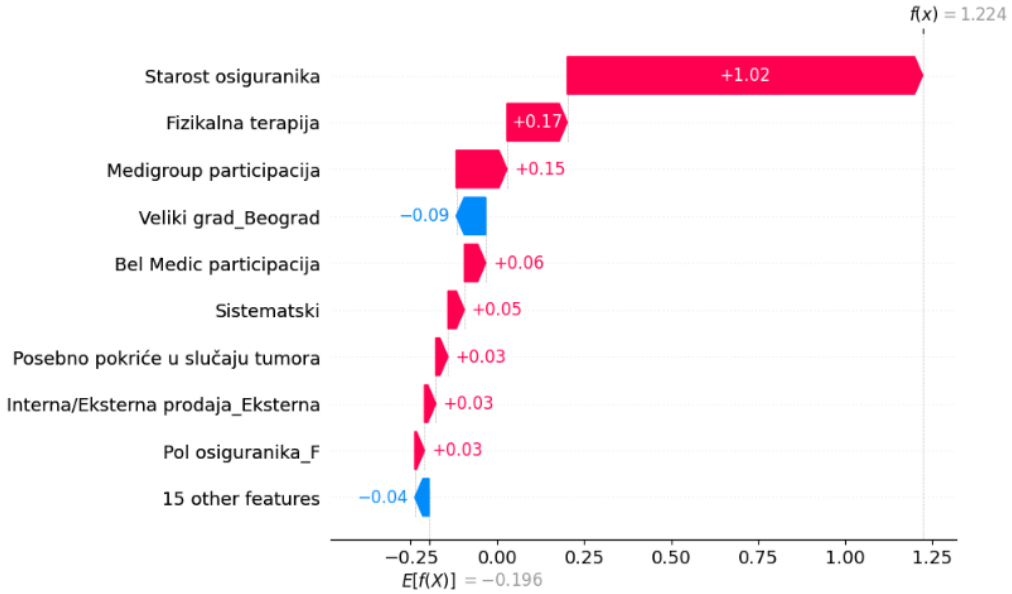
- Starost osiguranika ima najveći doprinos klasifikaciji: viša starost (crvene tačke) povećava verovatnoću za veći broj šteta.
- Veliki_grad_Beograd i Pol_osiguranika_F takođe povećavaju rizik od većeg broja šteta.
- Veliki_grad_Ostalo ima negativne Šeplijeve vrednosti, znači da pripadnost mestima koja nisu Beograd, Novi Sad ili Niš snižava verovatnoću za većim brojem šteta.



Slika 3.33: Grafik raspodele uticaja nezavisnih promenljivih - konačan `LightGBM` (korak2)



Slika 3.34: Bar Plot - Konačan `LightGBM` (korak2)



Slika 3.35: Grafik toka doprinosa nezavisnih promenljivih - Konačan `LightGBM` (korak2)

Na slici 3.34 prikazan je bar plot sa prosečnim apsolutnim Šeplijevim vrednostima, što omogućava globalno rangiranje značaja nezavisnih promenljivih. I ovde je Starost najvažnija, zatim Pol_osiguranika_F, Veliki_grad_Beograd, a zatim slede različite vrste participacija i pokrića.

Konačno, lokalna interpretacija za jednog konkretnog osiguranika prikazana je na slici 3.35. Ovaj grafik toka doprinosa nezavisnih promenljivih za konkretnu instancu jasno prikazuje koje promenljive najviše doprinose predikciji da je konkretni osiguranik u grupi sa manjim rizikom (negativna Šeplijeva vrednost), dok manji broj promenljivih povećava verovatnoću da je u grupi većeg rizika.

Rezultati potvrđuju da je model ne samo koristan prediktivno, već i interpretabilan, što dodatno povećava njegovu vrednost u donošenju poslovnih odluka.

Konačna predikcija, dobija se primenom dva konačna odabrana modela dvostepenog pristupa.

Glava 4

Zaključak

Savremeni modeli statističkog učenja, naročito oni zasnovani na stablima odlučivanja, kao što su model slučajnih šuma i `LightGBM`, obezbeđuju značajna poboljšanja u tačnosti predikcije u poređenju sa klasičnim linearnim modelima. Ipak, njihova složenost predstavlja izazov u pogledu interpretabilnosti i razumevanja doprinosa pojedinačnih ulaznih promenljivih u donošenju odluka modela. Upravo iz tog razloga, cilj ovog rada bio je da se istraže i primene savremene metode interpretacije nelinearnih modela, sa posebnim osvrtom na njihovu primenu u domenu osiguranja.

U radu je predstavljen teorijski okvir najznačajnijih metoda interpretacije modela – grafici parcijalne zavisnosti (Partial Dependence Plots, PDP), metoda individualnog uslovnog očekivanja (Individual Conditional Expectation, ICE), metoda akumuliranih lokalnih efekata (Accumulated Local Effects, ALE) i Šeplijeva metoda (SHAP). Svaka od ovih metoda nudi drugačiji ugao posmatranja i različit nivo interpretacije – od globalnog prosečnog efekta pojedinih nezavisnih promenljivih, do lokalne interpretacije uticaja na nivo pojedinačnih osiguranika.

Korišćen je skup podataka koji sadrži informacije o polisama dobrovoljnog zdravstvenog osiguranja, sa ciljem modelovanja broja šteta po osiguraniku. Prirodna pojava velikog broja nula (osiguranika bez šteta) motivisala je upotrebu dvostepenog modela, koji u prvom koraku klasifikuje da li će šteta nastati, a zatim u drugom koraku modelira broj šteta kod onih osiguranika kod kojih se šteta javlja. Ovakav pristup se pokazao kao efikasan za rešavanje problema asimetrične raspodele ciljne promenljive.

U okviru prvog koraka, upoređeni su logistička regresija, model slučajnih šuma i `LightGBM` model, uz upotrebu metoda za balansiranje klasa (kao što su SMOTE i ponderisanje klasa). `LightGBM` model se izdvojio kao najuspešniji, te je za njega pri-

menjena interpretacija pomoću Šeplijevih vrednosti i ALE prikaza. Na osnovu toga, identifikovani su najznačajnije nezavisne promenljive za predikciju nastanka štete, kao i njihov smer i jačina uticaja na odluku modela. Ove vizualizacije su dodatno omogućile identifikaciju nelinearnih obrazaca i interakcija koje linearni modeli ne mogu da otkriju.

U drugom koraku pristupa, modelovan je broj šteta kod osiguranika kod kojih je šteta nastala. Pokušano je nekoliko pristupa: Poasonova regresija, negativna binomna regresija, log-transformacija ciljne promenljive, kao i `LightGBM` regresioni i klasifikacioni modeli. Dodatno je eksperimentisano sa multi-klasnim pristupima, pri čemu se broj šteta grupisao u kategorije rizika. Na kraju je kao najpraktičniji pristup odabrana binarna klasifikacija osiguranika na one sa manjim (1-3 štete) i većim (4 i više šteta) rizikom.

Kombinovanjem ova dva koraka dobijen je funkcionalan prediktivni model koji ne samo da postiže zadovoljavajuću tačnost, već nudi i interpretaciju odluka modela, što je ključno pri donošenju poslovnih odluka u osiguranju.

Ovakvi modeli mogu se koristiti u osiguravajućim društvima za unapređenje procesa segmentacije klijenata, optimizaciju tarifa i efikasnije upravljanje rizikom. Interpretabilnost omogućava i bolju komunikaciju rezultata prema menadžmentu i regulatorima.

Zaključno, rad pokazuje da je moguće uspešno primeniti i interpretirati nelinearne modele u oblasti osiguranja, čime se dobija balans između tačnosti predikcije i transparentnosti odluka. Metode kao što su ALE i Šeplijeva metoda pokazuju se kao dovoljno robusne i informativne za razumevanje kompleksnih modela, pa bi njihova primena trebalo da postane standardna praksa u aktuarskoj i analitičkoj primeni modela veštačke inteligencije u osiguranju.

Bibliografija

[1] Apley, D. W., & Zhu, J. (2020). *Visualizing the effects of predictor variables in black box supervised learning models.* Journal of the Royal Statistical Society: Series B (Statistical Methodology), 82(4), 1059–1086.

[2] Breiman, L. (2001). *Random forests.* Machine Learning, 45(1), 5–32.

[3] Cameron, A. C., & Trivedi, P. K. (2013). *Regression Analysis of Count Data* (2nd ed.). Cambridge University Press, New York.

[4] Friedman, J. H. (2001). *Greedy function approximation: A gradient boosting machine.* The Annals of Statistics, 29(5), 1189–1232.

[5] Goldstein, A., Kapelner, A., Bleich, J., & Pitkin, E. (2015). *Peeking Inside the Black Box: Visualizing Statistical Learning With Plots of Individual Conditional Expectation.* Journal of Computational and Graphical Statistics, 24(1), 44–65.

[6] Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning.* MIT Press.

[7] Greenwell, B., Boehmke, B., & McCarthy, A. (2018). *Accumulated Local Effects: A New Method for Interpreting Nonlinear Models.* `https://arxiv.org/abs/1806.02983`

[8] Hilbe, J. M. (2011). *Negative Binomial Regression* (2nd ed.). Cambridge University Press, New York.

[9] Jeremić, D. (2025). *Kod za master rad: Interpretacija nelinearnih modela statističkog učenja sa primenom u osiguranju.* GitHub. Dostupno na: `https://github.com/dejanajeremic3/master-rad-kod` [Pristupljeno: 28. avgust 2025.]

[10] Lambert, D. (1992). *Zero-inflated Poisson regression, with an application to defects in manufacturing.* Technometrics, 34(1), 1–14.

[11] Lundberg, S. M., & Lee, S.-I. (2017). *A unified approach to interpreting model predictions.* Advances in Neural Information Processing Systems (NeurIPS), 30.

[12] Lundberg, S. M., Erion, G. G., & Lee, S.-I. (2018). *Consistent Individualized Feature Attribution for Tree Ensembles.* arXiv:1802.03888. `https://arxiv.org/abs/1802.03888`

[13] Molnar, C. (2019). *Interpretable Machine Learning.* Chapter 5.2: Individual Conditional Expectation (ICE), dostupno na `https://christophm.github.io/interpretable-ml-book/`[Pristupljeno: 22. jul 2025.]

[14] Molnar, C. (2022). *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable.* 2nd ed. [Online]. Dostupno na: `https://christophm.github.io/interpretable-ml-book/`[Pristupljeno: 22. jul 2025.]

[15] Neter, J., Kutner, M. H., Nachtsheim, C. J., & Wasserman, W. (1996). *Applied Linear Statistical Models.* 4th ed., McGraw-Hill/Irwin.

[16] Obrenović, N. (2024). *Problem klasifikacije nebalansiranih podataka.* Univerzitet u Beogradu, Matematički fakultet. `http://elibrary.matf.bg.ac.rs/bitstream/handle/123456789/5744/v1_masterNadjaObrenovic.pdf?sequence=1` [Pristupljeno: 5. avgust 2025.]

[17] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, E. (2011). *Scikit-learn: Machine learning in Python.* Journal of Machine Learning Research, 12, 2825–2830.

[18] Ridgeway, G. (2007). *Generalized Boosted Models: A guide to the gbm package.* R vignette.

[19] Shapley, L. S. (1953). *A Value for n-Person Games.* In *Contributions to the Theory of Games* (pp. 307–317). Princeton University Press.

[20] Shrikumar, A., Greenside, P., & Kundaje, A. (2017). *Learning Important Features Through Propagating Activation Differences.* ICML. `http://proceedings.mlr.press/v70/shrikumar17a/shrikumar17a.pdf`[Pristupljeno: 22. jul 2025.]

[21] Winkler, G. (2019). *Understanding and interpreting machine learning models.* Journal of Insurance Analytics, 3(1), 45–68.

[22] Zhu, Z., Xu, Y., & Yu, F. (2020). *Comparison of Zero-Inflated and Hurdle Models for Count Data in Insurance.* Journal of Risk and Insurance Research, 7(2), 91–106.

Biografija autora

Dejana Jeremić rođena je u Beogradu, 5. oktobra 1999. godine, gde je završila osnovno i srednjoškolsko obrazovanje kao nosilac diplome „Vuk Karadžić". 2018. godine upisala je Matematički fakultet u Beogradu, smer Statistika, finansijska i aktuarska matematika. Osnovne akademske studije završila je 2022. godine, i iste godine upisala master akademske studije. 2022. godine zaposlila se kao aktuar u Sava neživotnom osiguranju u Beogradu, gde učestvuje učestvuje u mnogim međunarodnim projektima i implementaciji međunarodnih standarda na domaćem tržištu.