

# Predviđanje cijena dionica korištenjem ansambla algoritama strojnog učenja

---

Žugec, Antonio

Master's thesis / Diplomski rad

2024

*Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj:* **Josip Juraj Strossmayer University of Osijek, Faculty of Electrical Engineering, Computer Science and Information Technology Osijek / Sveučilište Josipa Jurja Strossmayera u Osijeku, Fakultet elektrotehnike, računarstva i informacijskih tehnologija Osijek**

*Permanent link / Trajna poveznica:* <https://um.nsk.hr/um:nbn:hr:200:861058>

*Rights / Prava:* [In copyright](#)/[Zaštićeno autorskim pravom.](#)

*Download date / Datum preuzimanja:* **2024-11-20**

*Repository / Repozitorij:*

[Faculty of Electrical Engineering, Computer Science and Information Technology Osijek](#)



**SVEUČILIŠTE JOSIPA JURJA STROSSMAYERA U OSIJEKU**  
**FAKULTET ELEKTROTEHNIKE, RAČUNARSTVA I**  
**INFORMACIJSKIH TEHNOLOGIJA OSIJEK**

**Sveučilišni diplomski studij Računarstvo**

**PREDVIĐANJE CIJENA DIONICA KORIŠTENJEM**  
**ANSAMBLA ALGORITAMA STROJNOG UČENJA**

**Diplomski rad**

**Antonio Žugec**

**Osijek, 2024.**

**Obrazac D1: Obrazac za ocjenu diplomskog rada na sveučilišnom diplomskom studiju**

**Ocjena diplomskog rada na sveučilišnom diplomskom studiju**

<b>Ime i prezime pristupnika:</b>	Antonio Žugec
<b>Studij, smjer:</b>	Sveučilišni diplomski studij Računarstvo
<b>Mat. br. pristupnika, god.</b>	D1344R, 07.10.2022.
<b>JMBAG:</b>	0165079667
<b>Mentor:</b>	doc. dr. sc. Dražen Bajer
<b>Sumentor:</b>	dr. sc. Mario Dudjak
<b>Sumentor iz tvrtke:</b>	
<b>Predsjednik Povjerenstva:</b>	doc. dr. sc. Bruno Zorić
<b>Član Povjerenstva 1:</b>	dr. sc. Mario Dudjak
<b>Član Povjerenstva 2:</b>	izv. prof. dr. sc. Zdravko Krpić
<b>Naslov diplomskog rada:</b>	Predviđanje cijena dionica korištenjem ansambla algoritama strojnog učenja
<b>Znanstvena grana diplomskog rada:</b>	<b>Umjetna inteligencija (zn. polje računarstvo)</b>
<b>Zadatak diplomskog rada:</b>	U radu je potrebno opisati problem predviđanja cijena dionica kao regresijski problem u području nadziranog strojnog učenja. Potrebno je istražiti relevantnu literaturu kako bi se utvrdio uobičajeni tijek učenja iz skupova podataka koji opisuju ovaj problem. U praktičnom dijelu rada nužno je razviti programsko rješenje koje prikuplja raznovrsne podatke o tržištu dionica i na temelju njih gradi odgovarajuće regresijske modele za predviđanje njihove cijene. Uz to, potrebno je implementirati odgovarajuće postupke za obiediniavanje ishoda pojedinačnih modela u vidu
<b>Datum ocjene pismenog dijela diplomskog rada od strane mentora:</b>	18.09.2024.
<b>Ocjena pismenog dijela diplomskog rada od strane mentora:</b>	Izvrstan (5)
<b>Datum obrane diplomskog rada:</b>	30.9.2024.
<b>Ocjena usmenog dijela diplomskog rada (obrane):</b>	Izvrstan (5)
<b>Ukupna ocjena diplomskog rada:</b>	Izvrstan (5)
<b>Datum potvrde mentora o predaji konačne verzije diplomskog rada čime je pristupnik završio sveučilišni diplomski studij:</b>	30.09.2024.



**FERIT**

FAKULTET ELEKTROTEHNIKE, RAČUNARSTVA  
I INFORMACIJSKIH TEHNOLOGIJA **OSIJEK**

## IZJAVA O IZVORNOSTI RADA

Osijek, 30.09.2024.

**Ime i prezime Pristupnika:**

Antonio Žugec

**Studij:**

Sveučilišni diplomski studij Računarstvo

**Mat. br. Pristupnika, godina upisa:**

D1344R, 07.10.2022.

**Turnitin podudaranje [%]:**

2

Ovom izjavom izjavljujem da je rad pod nazivom: **Predviđanje cijena dionica korištenjem ansambla algoritama strojnog učenja**

izrađen pod vodstvom mentora doc. dr. sc. Dražen Bajer

i sumentora dr. sc. Mario Dudjak

moj vlastiti rad i prema mom najboljem znanju ne sadrži prethodno objavljene ili neobjavljene pisane materijale drugih osoba, osim onih koji su izričito priznati navođenjem literature i drugih izvora informacija.

Izjavljujem da je intelektualni sadržaj navedenog rada proizvod mog vlastitog rada, osim u onom dijelu za koji mi je bila potrebna pomoć mentora, sumentora i drugih osoba, a što je izričito navedeno u radu.

Potpis pristupnika:



# SADRŽAJ

<b>1. UVOD</b> .....	<b>1</b>
<b>2. PREDVIĐANJE CIJENA DIONICA</b> .....	<b>3</b>
2.1. Tehnička analiza .....	4
2.2. Fundamentalna analiza .....	5
2.3. Analiza vremenskih serija .....	7
2.4. Analiza sentimenta.....	8
2.5. Kritički osvrt .....	9
<b>3. ANSAMBLI U STROJNOM UČENJU</b> .....	<b>11</b>
3.1. Ansambli zasnovani na glasanju.....	11
3.2. Ansambli zasnovani na slaganju.....	12
3.3. Ansambli zasnovani na pakiranju i pojačavanju .....	13
3.4. Kritički osvrt .....	14
<b>4. OSTVARENO PROGRAMSKO RJEŠENJE</b> .....	<b>16</b>
4.1. Odabir pojedinih modela.....	17
4.2. Težinsko glasanje .....	18
4.3. Optimizacija težina u ansamblu .....	19
4.3.1. Algoritam diferencijalne evolucije .....	19
4.3.2. Primjena algoritma DE za optimizaciju težina .....	21
<b>5. EKSPERIMENTALNA ANALIZA I REZULTATI</b> .....	<b>24</b>
5.1. Postavke eksperimenta .....	24
5.2. Usporedba pojedinih modela .....	35
5.2.1. Usporedba modela za analizu vremenskih serija .....	35
5.2.2. Usporedba modela za tehničku analizu .....	40
5.2.3. Usporedba modela za fundamentalnu analizu .....	46
5.2.4. Usporedba modela za analizu sentimenta .....	51
5.3. Usporedba pojedinih modela s predloženim ansamblom .....	57
5.4. Analiza učinka optimizacije težina .....	58
<b>6. ZAKLJUČAK</b> .....	<b>62</b>

<b>LITERATURA .....</b>	<b>64</b>
<b>POPIS SLIKA.....</b>	<b>66</b>
<b>POPIS TABLICA.....</b>	<b>67</b>
<b>SAŽETAK.....</b>	<b>68</b>
<b>ABSTRACT .....</b>	<b>69</b>
<b>ŽIVOTOPIS.....</b>	<b>70</b>
<b>PRILOZI.....</b>	<b>71</b>

## 1. UVOD

Izreka "svijet se vrti oko novca" jedna je od često spominjanih u današnjem društvu, i to s dobrim razlogom. Novac je, bez sumnje, bitan dio naših života. Otkad je novac izmišljen, ljudi su stalno tražili učinkovitije načine za njegovo stjecanje. Dok su u prošlosti prevladavali razmjena dobara i usluga, danas se taj proces odvija na sofisticiraniji, kompleksniji i brži način. Burze su postale središta globalne trgovine, gdje se svakodnevno razmjenjuju milijarde dolara. Za velike tvrtke, banke i globalnu ekonomiju, precizno prognoziranje budućih kretanja cijena i ekonomskih trendova predstavlja ključan faktor. Svaka prednost u ovom području može značiti razliku između bankrota i bogatstva. Stoga je ovo područje ne samo izuzetno zanimljivo, već i intenzivno istraživano.

Procjena cijene dionica predstavlja kompleksan izazov zbog niza nepredvidivih čimbenika koji mogu značajno utjecati na njihovu vrijednost. Na primjer, informacije i komentari na društvenim mrežama, poput pozitivnih pohvala poznatih osoba na platformi X, mogu izazvati nagli porast cijene dionica. Nasuprot tome, negativni događaji, poput izvještaja o ozbiljnim katastrofama ili krizama, mogu uzrokovati značajan pad cijene dionica. Identifikacija trendova putem različitih tehničkih indikatora često se pokazuje ključnom, dok svježi objavljeni financijski izvještaji mogu pružiti dragocjene uvide u stanje dionice, otkrivajući prednosti i nedostatke koji nisu odmah očiti. Također, fluktuacije u cijenama komoditeta poput zlata ili nafte mogu utjecati na cijene dionica. Dodatno, neki trendovi mogu biti skriveni unutar same cijene dionice, te samo sofisticirani algoritmi mogu prepoznati obrasce koje ljudski analitičari možda ne mogu uočiti. Neke od ovih situacija mogu imati samo blage ili nikakve utjecaje na cijenu dionice, što dodatno otežava analizu. Ovi strukturirani i nestrukturirani podaci značajno povećavaju volumen, složenost, raznolikost i točnost podataka.

Zbog ove raznolikosti i kompleksnosti podataka, precizna i efikasna analiza ogromnih količina informacija postaje izuzetno zahtjevna. Stoga je važno odabrati najprikladnije metode analize i razviti odgovarajuće modele. U praksi se ističu fundamentalna analiza, tehnička analiza, analiza sentimenta i analiza vremenskih serija. Svaka od njih koristi specifične skupove podataka, a njihova kombinacija unutar ansambla može pružiti inovativna rješenja jer omogućuje obuhvaćanje i analizu širokog raspona validnih informacija. Na taj način ansambl može ponuditi dublje uvide nego što bi to omogućili pojedinačni modeli.

Također, postoje situacije kada su određene informacije relevantnije od drugih. Na primjer, tehnički indikatori mogu biti važniji u stabilnim razdobljima, dok fundamentalni faktori postaju ključni u kriznim situacijama. Stoga je smisleno prilagoditi značaj različitih modela u skladu s trenutnom relevantnošću. Ansambl se pokazuje kao prikladno rješenje za ovakve situacije jer omogućuje prilagodbu različitih modela prema specifičnim okolnostima i uvjetima.

Cilj ovog rada jest ispitati metode koje se često primjenjuju u praksi i pronaći najbolja rješenja unutar svake analize. Nadalje, rad istražuje mogućnosti kombiniranja ovih metoda u ansamblu kako bi se utvrdila prednost takvih kombinacija i ispitala mogućnost implementacije ovakvih metoda u stvarnom svijetu. U drugom poglavlju pruža se pregled najčešće korištenih metoda u procjeni vrijednosti dionica, uz naglasak na ključne parametre unutar tih analiza. Treće poglavlje detaljno obrađuje različite ansambl metode koje se mogu primijeniti na zadani problem. Četvrto poglavlje usmjereno je na opis ostvarenog programskog rješenja za procjenu dionica. U petom poglavlju dani su implementacijski detalji, zajedno s analizom i usporedbom postignutih rezultata.

## 2. PREDVIĐANJE CIJENA DIONICA

Burza je mjesto gdje se trguje dionicama poduzeća. To je mreža za trgovanje dionicama gdje investitori i trgovci kupuju i prodaju dionice. Ove trgovine određuju cijene dionica, odražavajući percipiranu vrijednost tvrtke i tržišne uvjete [1]. Poznavanjem tržišta i drugih ekonomskih faktora moguće je preciznije procijeniti buduću cijenu tržišta ili specifične dionice. Uspješno i točno predviđanje tržišnih cijena donosi značajnu dobit investorima, čineći ovaj zadatak vrlo privlačnim za primjenu strojnog učenja. Ljudski mozak ne može obraditi toliku količinu informacija, stoga je računalna obrada izuzetno prikladna za ovaj problem, jer računala mogu analizirati velike količine podataka i pružiti preciznije procjene od ljudi.



Slika 2.1: Utjecajni čimbenici na cijenu [2]

Međutim, na tržištima dionica diljem svijeta svakodnevno se generiraju milijarde strukturiranih i nestrukturiranih podataka, povećavajući volumen, brzinu, raznolikost i točnost burzovnih podataka, čineći ih složenima za analizu. Općenito se prihvaćaju četiri metode za analizu velikih količina informacija u svrhu procjene vrijednosti dionica: tehnička analiza, fundamentalna analiza, analiza vremenskih serija i analiza sentimenta, što je prikazano na slici 2.1. Fundamentalna analiza i tehnička analiza dva su najčešće korištena pristupa za analizu i predviđanje cijena dionica [3]. Također je pokazano da analiza sentimenta i analiza vremenskih serija pružaju valjane procjene. Korištenjem ove četiri vrste analiza obuhvaća se širok raspon informacija koje mogu utjecati na cijenu dionica.

## 2.1. Tehnička analiza

Tehnička analiza temelji se na pretpostavci da su sve informacije o tržištu sadržane u cijenama dionica u stvarnom vremenu. Tehnički analitičari koriste grafikone, tehnike modeliranja i tehničke indikatore kako bi identificirali trendove u cijenama dionica. Ova metoda oslanja se na povijesne podatke za predviđanje budućih kretanja cijena dionica [4]. Ustaljeni parametri koji se obično koriste za predviđanje pomoću tehničkih indikatora su:

- **Razlika cijena:** Vrijednost koja se dobiva oduzimanjem cijene prethodnog dana od trenutnog dana kako bi se vidjelo za koliko je cijena pala odnosno narasla.
- **Konvergencija / divergencija pokretnog prosjeka** (engl. *moving average convergence / divergence*, **MACD**): Indikator koji prikazuje odnos između dva pomična prosjeka cijene dionice, pomažući u identificiranju trendova i potencijalnih preokreta.
- **Indeks relativne snage** (engl. *relative strength index*, **RSI**): Mjera brzine i promjene kretanja cijena koja se koristi za procjenu je li dionica prekupljena ili preprodana.
- **Stohastički oscilator** (engl. *stochastic oscillator*): Indikator koji uspoređuje zatvaranje cijene dionice s rasponom cijena tijekom određenog vremenskog razdoblja, pomažući u prepoznavanju prekupljeni i preprodanih uvjeta.
- **Jednostavan pomični prosjek** (engl. *simple moving average*, **SMA**): Prosjek cijene dionice tijekom određenog broja dana, koji izjednačuje kratkoročne oscilacije i pomaže u identificiranju osnovnog trenda.
- **Eksponencijalni pomični prosjek** (engl. *exponential moving average*, **EMA**): Sličan SMA-u, ali pridaje veću težinu novijim cijenama, čime se brže reagira na nove informacije i promjene u cijeni dionice.

Najčešće korišteni algoritmi za tehničku analizu uključuju model duge kratkoročne memorije (engl. *long short-Term memory*, LSTM), stroj potpornih vektora (engl. *support vector machine*, SVM), regresor nasumične šume (engl. *random forest regressor*, RF) i umjetnu neuronsku mrežu (engl. *artificial neural network*, ANN). Primjer tehničkog indikatora RSI prikazan je na slici 2.2

## EURAUD, Daily Chart 5 Classic RSI Signals



Slika 2.2: Tehnički indikatori (RSI) [5]

### 2.2. Fundamentalna analiza

Fundamentalna analiza koristi ekonomski položaj tvrtke, zaposlenike, upravni odbor, financijski status, godišnja izvješća tvrtke, bilance, izvješća o prihodima, te prirodne pojave i klimatske okolnosti, uključujući prirodne katastrofe i političke promjene, kako bi predvidjela buduću cijenu dionica [6]. U ovom radu korišteni su financijski izvještaji dostupni na stranici U.S. Securities and Exchange Commission (SEC) [7]. Primjer financijskog izvještaja sa stranice SEC prikazan je na slici 2.3. Kao i kod tehničkih indikatora, postoji obilje podataka vezanih za fundamentalnu analizu. Najučestaliji ulazni podaci za fundamentalnu analizu su:

- **Razlika cijena:** Vrijednost koja se dobiva oduzimanjem cijene prethodnog dana od trenutnog dana kako bi se vidjelo za koliko je cijena pala odnosno narasla.

- **Prihod:** Ukupni iznos novca koji kompanija ostvaruje prodajom svojih proizvoda ili usluga tijekom određenog razdoblja. Također poznat kao "prodaja" ili "ukupni prihodi".
- **Neto dobit:** Krajnji profit kompanije nakon odbitka svih troškova, uključujući operativne troškove, poreze, kamate i druge obveze. Ovo je ključna mjera profitabilnosti kompanije.
- **Dobit po dionici** (engl. *earnings per share*, **EPS**): Dobit po dionici predstavlja neto dobit podijeljenu s ukupnim brojem izdanih dionica. Ovo mjeri koliko zarade pripada svakoj dionici, što je važno za investitore.
- **Bruto profitna marža:** Odnos bruto dobiti (prihoda minus trošak prodane robe) prema ukupnim prihodima, izražen u postocima. Pokazuje koliko kompanija zarađuje nakon odbitka troškova proizvodnje.
- **Operativna marža:** Odnos operativne dobiti (prihoda minus operativni troškovi kao što su plaće i najamnine) prema ukupnim prihodima, izražen u postocima. Ova marža mjeri efikasnost kompanije u poslovanju prije poreza i kamata.
- **Neto profitna marža:** Odnos neto dobiti prema ukupnim prihodima, izražen u postocima. Pokazuje koliki dio prihoda ostaje kao neto dobit nakon svih troškova.
- **Novčani tok iz poslovanja:** Mjeri novčani tok generiran iz redovnog poslovanja kompanije, isključujući investicijske i financijske aktivnosti. Ovo je važno za procjenu koliko novca kompanija generira za pokrivanje operativnih troškova i obveza.
- **Ukupna aktiva:** Ukupna vrijednost svega što kompanija posjeduje, uključujući gotovinu, potraživanja, inventar, nekretnine i opremu. Predstavlja resurse kojima kompanija raspolaže.
- **Ukupne obveze:** Ukupni iznos dugovanja kompanije prema vjerovnicima, uključujući kratkoročne i dugoročne obveze. To su obaveze koje kompanija mora podmiriti.



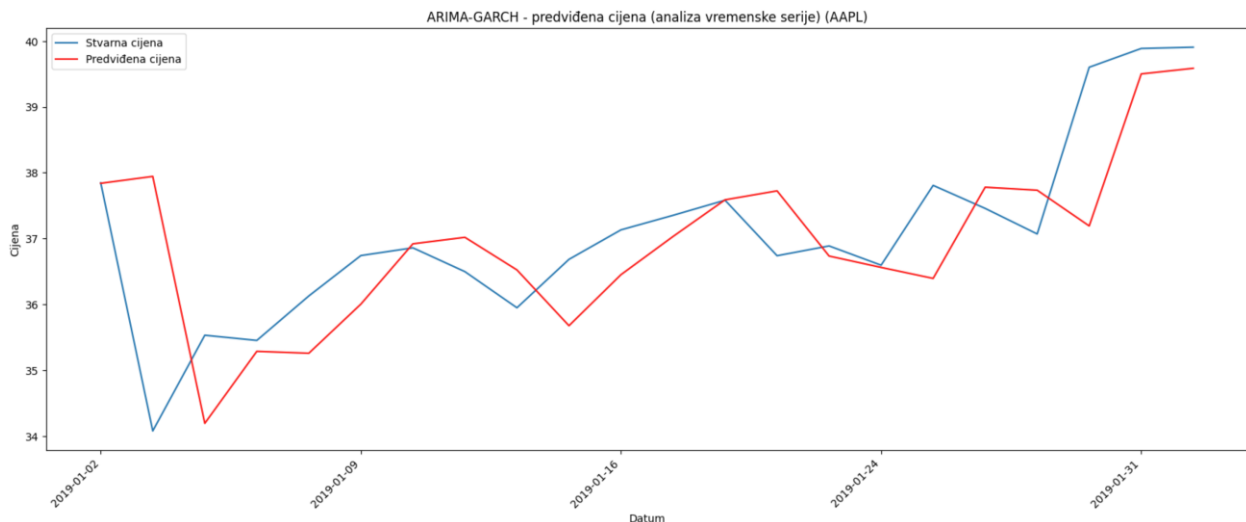
	March 30, 2024
Net sales:	
Products	\$ 66,886
Services	23,867
Total net sales	90,753
Cost of sales:	
Products	42,424
Services	6,058
Total cost of sales	48,482
Gross margin	42,271
Operating expenses:	
Research and development	7,903
Selling, general and administrative	6,468
Total operating expenses	14,371
Operating income	27,900
Other income/(expense), net	158
Income before provision for income taxes	28,058
Provision for income taxes	4,422
Net income	\$ 23,636
Earnings per share:	
Basic	\$ 1.53
Diluted	\$ 1.53

Slika 2.3: Dio financijskog izvještaj tvrtke Apple [8]

Najčešće korišteni algoritmi za fundamentalnu analizu su LSTM, SVM, RF i ANN.

### 2.3. Analiza vremenskih serija

Analiza vremenskih serija temelji se na pretpostavci da su financijske vremenske serije generirane iz linearnog procesa te se mogu modelirati kako bi se prognozirala buduća vrijednost. Drugim riječima, pretpostavlja se da svi podaci potrebni za procjenu buduće vrijednosti već postoje u povijesnim podacima dionice, slično kao kod tehničke analize. Međutim, u ovom pristupu koristi se isključivo cijena dionice. Zbog toga što podaci dionica sadrže mnogo šuma i imaju vrlo nelinearne, složene, dinamične, neparametričke i kaotične karakteristike, procjena budućih vrijednosti ovih vremenskih serija postaje značajno otežana [9]. Najčešće korištene tradicionalne statističke metode jesu eksponencijalno izgladivanje (engl. *exponential smoothing*, ES) [10], autoregresivni integrirani pomični prosjek (engl. *autoregressive integrated moving average*, ARIMA) [9] i generalizirana autoregresivna uvjetna heteroskedastičnost (engl. *generalized autoregressive conditional heteroskedasticity*, GARCH) [9]. Primjer procjene vrijednosti dionice putem ARIMA-GARCH modela prikazano je na slici 2.4.

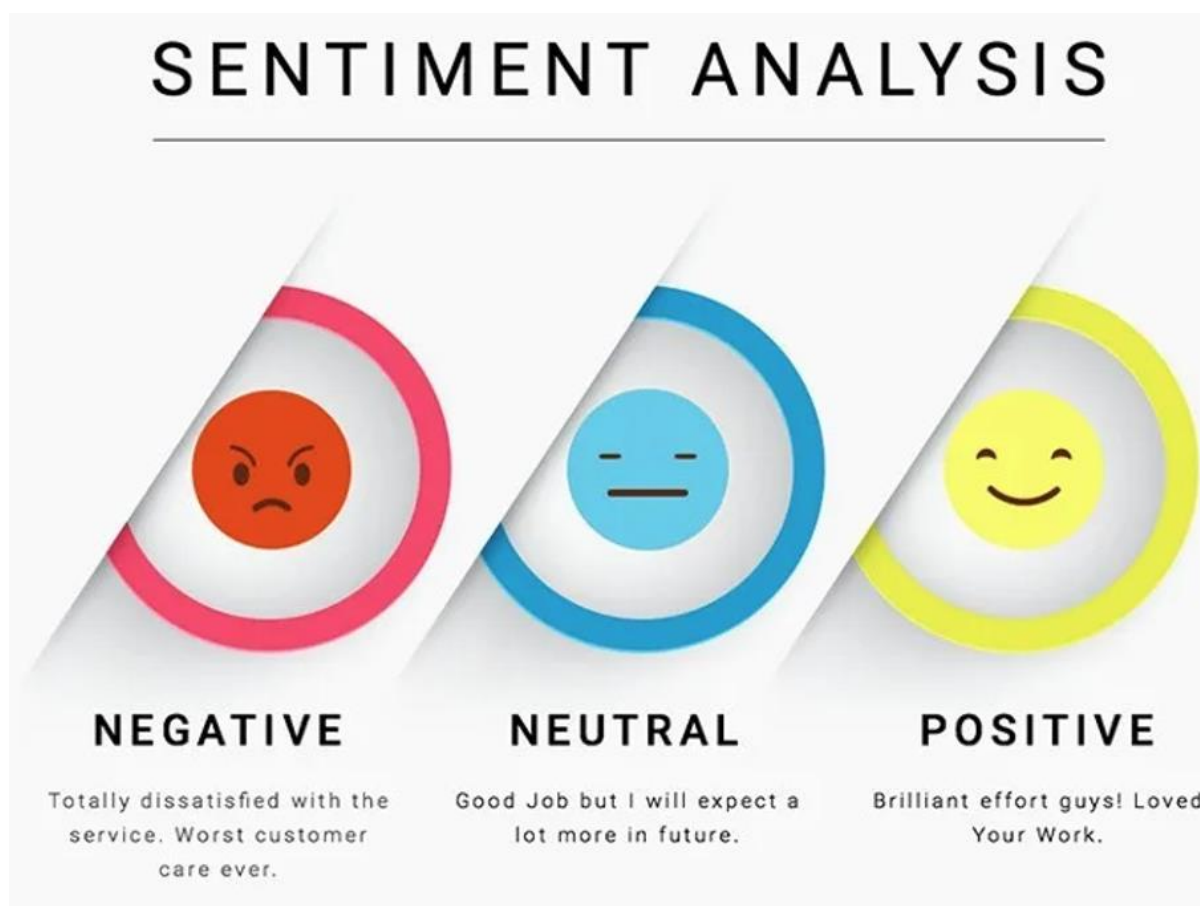


Slika 2.4: Primjer procjene dionice Apple putem ARIMA-GARCH modela

Ulazni parametri za ovu vrstu modela su same dnevne cijene dionica, što čini model vrlo jednostavnim. To može biti prednost u odnosu na druge metode i modele, ali također može predstavljati nedostatak, ovisno o specifičnom zadatku.

## 2.4. Analiza sentimenta

Analiza sentimenta je proces analize digitalnog teksta kako bi se utvrdilo je li emocionalni ton poruke pozitivan, negativan ili neutralan. Danas tvrtke imaju velike količine tekstualnih podataka kao što su e-poruke, transkripti razgovora korisničke podrške, komentari na društvenim mrežama i recenzije [11]. Ova informacija može se obraditi pomoću modela za obradu prirodnog jezika (engl. *natural language processing*, NLP) kako bi se ocijenio emocionalni ton poruke. Nadalje, rezultati NLP-a mogu se koristiti u modelima gdje su najčešće korišteni LSTM, SVM, RF i ANN za procjenu buduće cijene dionica. Primjer negativnog, neutralnog i pozitivnog komentara prikazano je na slici 2.5.



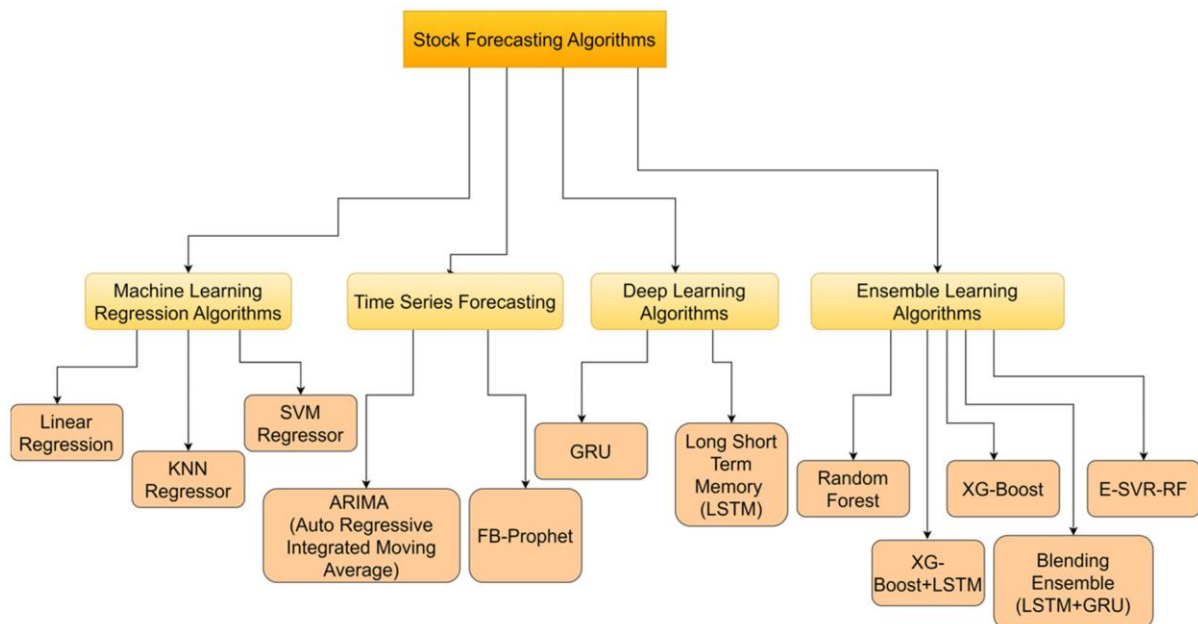
Slika 2.5: Analiza sentimenta [12]

Postoji mnogo izvora informacija koji se mogu koristiti za analizu sentimenta, kao što su razni časopisi, financijski blogovi, društvene mreže i drugi. X (bivši Twitter) jedna je od relevantnijih društvenih mreža u kojoj se svakodnevno komentira tržište dionica.

## 2.5. Kritički osvrt

Svaka vrsta analize pruža svoju logičnu i validnu procjenu, zbog čega je izazovno odrediti optimalni model za ovaj specifičan problem. S obzirom na raznolikost modela, gdje svaki može imati svoju specifičnu konfiguraciju, i obilje informacija koje se mogu koristiti kao ulazni parametri, teško je utvrditi koji parametri i modeli najpreciznije procjenjuju cijenu dionica. Na slici 2.6 je prikazana raznolikost ovih modela.

Prema istraživanju [13], 66% pregledanih radova temeljilo se na tehničkoj analizi. U tom istraživanju, tehnička analiza i analiza vremenskih serija svrstane su u istu kategoriju. Stoga, tehnička analiza i analiza vremenskih serija često se primjenjuju, dok se analiza sentimenta također koristi relativno često. S druge strane, fundamentalna analiza najmanje je zastupljena u dosadašnjim radovima.



Slika 2.6: Različite tehnike za procjenu cijene dionica [14]

Prema spomenutom istraživanju [13], samo 10 % radova koristilo je kombinaciju različitih analiza, što ukazuje na potencijalno nedovoljno istraženo područje u analizi i procjeni cijena dionica. Jedno od rješenja za ovakav problem može biti uporaba ansambla algoritma, koji objedinjuje rezultate različitih analitičkih pristupa zbog potencijalno boljih rezultata. Pretpostavka je da svaka vrsta analize ima svoje prednosti i nedostatke, no integracijom tih metoda moguće je eliminirati nedostatke i pojačati prednosti svakog modela. Ovakav pristup može poboljšati rezultate jer ansambl pristupa većoj količini informacija, no isto tako može povećati složenost modela, što može negativno utjecati na preciznost rezultata.

### 3. ANSAMBLI U STROJNOM UČENJU

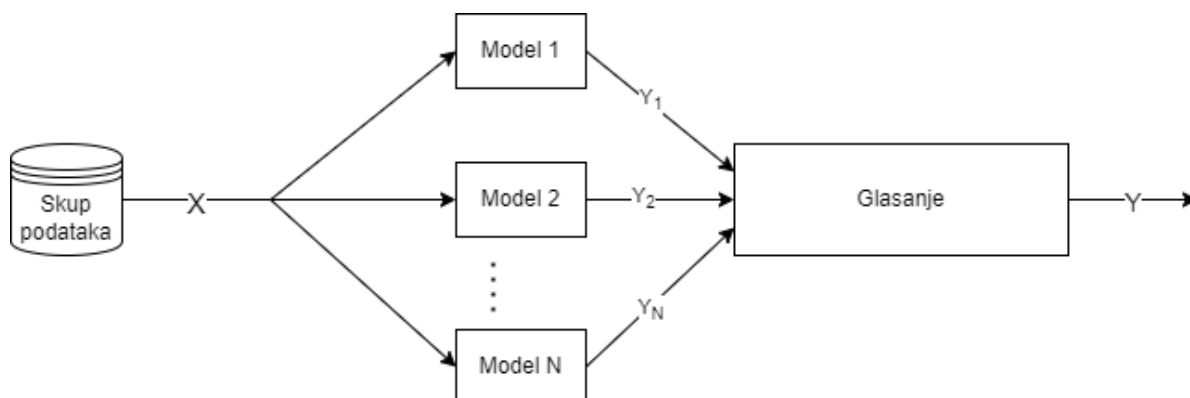
Ansambli u strojnom učenju predstavljaju tehniku koja uključuje kombiniranje predviđanja više modela radi poboljšanja točnosti i robusnosti. Kombiniranjem izlaza više modela ova metoda može smanjiti rizik od prenaučivosti (engl. *overfitting*) i poboljšati generalizaciju. Metode ansambla pokazale su se učinkovitima u širokom rasponu zadataka strojnog učenja uključujući klasifikaciju, regresiju i grupiranje podataka [15].

Modeli unutar ansambla ne moraju nužno biti isti i mogu koristiti različite algoritme te različite skupove podataka. Upotreba raznovrsnih modela osobito je korisna kada nije jasno koji tip modela je najprikladniji za rješavanje određenog problema. Problem procjene cijene dionica je upravo takav slučaj, jer ne postoji jedna univerzalna metoda koja je najbolja za ovaj zadatak. Zbog toga ansambl, kombiniranjem različitih pristupa, može bolje obuhvatiti složenost podataka i poboljšati preciznost predviđanja.

Ansambli se dijele na osnovne i napredne. Osnovni ansambli temelje se na glasanju (engl. *voting*), dok se napredni ansambli oslanjaju na složene metode poput slaganja (engl. *stacking*), pakiranja (engl. *blending*) i pojačavanja (engl. *bagging*) [16].

#### 3.1. Ansambli zasnovani na glasanju

Ansambli zasnovani na glasanju, kao što je prikazano na slici 3.1, predstavljaju vrlo popularnu tehniku koja kombinira rezultate različitih modela koristeći specifičan sustav glasanja. Iako se ova metoda najčešće primjenjuje za klasifikacijske probleme, ona ima značajnu primjenu u regresijskim problemima. Glavne varijante ansambla koji koriste sustav glasanja uključuju većinsko glasanje (engl. *majority voting*), prosječno glasanje (engl. *average voting*) i težinsko glasanje (engl. *weighted voting*).

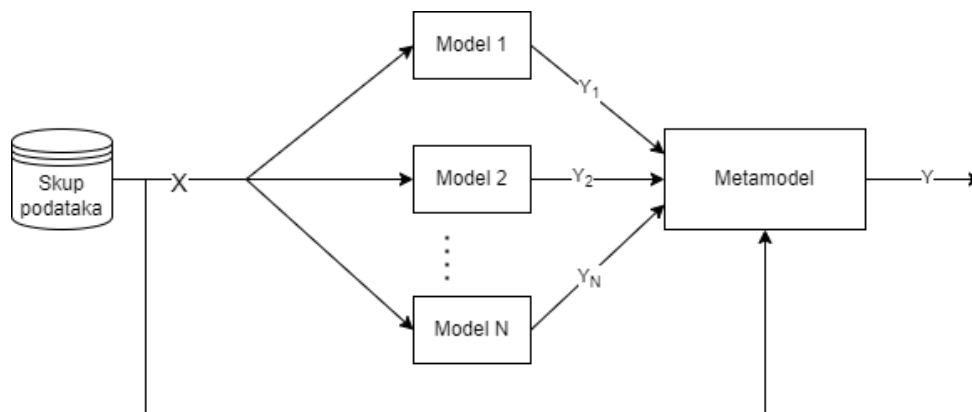


Slika 3.1: Način rada ansambla zasnovanog na glasanju

Većinsko glasanje, koje se koristi isključivo za klasifikacijske probleme, funkcionira na jednostavnom principu gdje se izlaz svakog modela tretira kao glas, a konačna odluka donosi se na temelju odgovora većine. S druge strane, prosječno glasanje izračunava prosjek svih izlaza modela kako bi se dobio konačan rezultat. Težinsko glasanje određuje važnost svakog modela putem dodijeljenih težina. Izlaz ovog ansambla zapravo predstavlja linearnu kombinaciju izlaza pojedinih modela, pri čemu se određuju odgovarajuće težine. Težine mogu biti jednake, što rezultira prosječnim glasanjem, ili različite, ovisno o specifičnom doprinosu svakog modela. Cilj je pronaći optimalne težine, za što se mogu primijeniti razni algoritmi optimizacije.

### 3.2. Ansambli zasnovani na slaganju

Slaganje je napredna tehnika u strojnom učenju koja kombinira predikcije nekoliko različitih modela kako bi se izgradio novi, snažniji i točniji model. Ovaj novi model, koji se često naziva meta model, koristi se za donošenje konačnih predikcija na testnom skupu podataka. Meta model obično ima sposobnost prepoznavanja obrazaca u predikcijama osnovnih modela i na temelju tih informacija generirati konačan rezultat.



Slika 3.2: Način rada ansambla zasnovanog na slaganju

Proces slaganja se odvija u nekoliko ključnih koraka, koji su prikazani na dijagramu na slici 3.2. Prvi korak uključuje razdvajanje skupa podataka na dva neovisna dijela. Ovo razdvajanje je ključno kako bi se osiguralo da modeli trenirani na prvom dijelu skupa podataka ne budu testirani na istim tim podacima, čime se izbjegava prenaučenosť. Nakon toga, osnovni modeli, koji mogu uključivati različite algoritme poput regresije, stabala odlučivanja, ili neuronskih mreža, treniraju se na prvom dijelu skupa podataka.

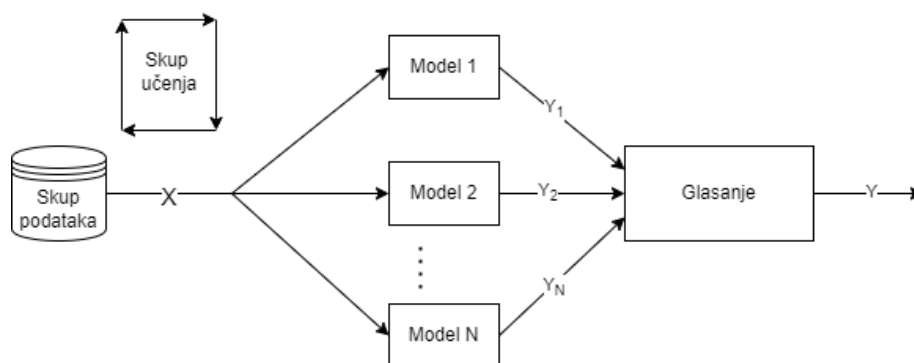
Nakon što su osnovni modeli istrenirani, provodi se vrednovanje njihovih performansi na drugom dijelu skupa podataka. U ovom koraku, predikcije koje generiraju osnovni modeli se koriste kao ulazi za trening meta modela. Meta model koristi ove predikcije kao ulazne podatke, dok se stvarni

rezultati iz drugog dijela skupa koriste kao ishodi. Tako, meta model uči kako najbolje kombinirati predikcije osnovnih modela kako bi se dobila najtočnija moguća predikcija.

Ova tehnika slaganja omogućava iskorištavanje prednosti različitih modela i smanjenje njihovih individualnih slabosti, što često rezultira boljim performansama u usporedbi s korištenjem samo jednog modela.

### 3.3. Ansambli zasnovani na pakiranju i pojačavanju

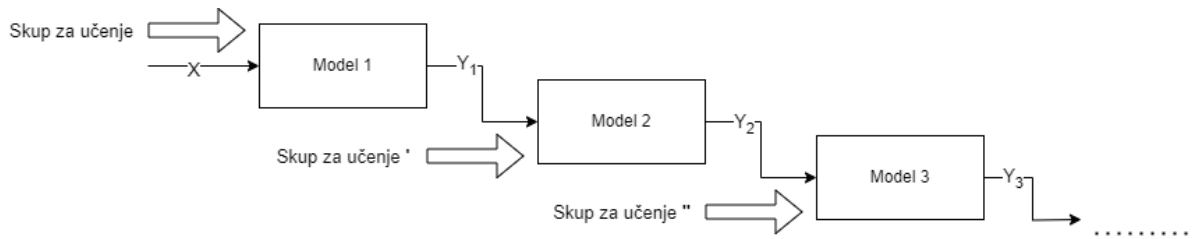
Pakiranje je tehnika u strojnom učenju koja koristi isti modele za treniranje, ali s različitim podskupovima podataka. Ovi podskupovi, poznati kao „paketi“, stvaraju se nasumičnim odabirom iz cijelog skupa podataka. Svaki model treniran je na vlastitom paketu podataka, a rezultati svih modela se potom kombiniraju kako bi se dobio konačan ishod. U praksi, najčešće se koristi metoda glasanja za određivanje konačne odluke, gdje svaki model ima pravo glasa, a najčešće odabrana predikcija se uzima kao konačan rezultat. Ovaj proces je ilustriran na slici 3.3.



Slika 3.3: Način rada ansambla zasnovanog na pakiranju

Cilj ansambla zasnovanog na pakiranju je poboljšanje robusnosti modela, posebno u slučajevima kada pojedini modeli imaju sklonost ka prenaučivosti. Korištenjem različitih podskupova podataka za trening, pakiranje pomaže u ublažavanju problema prenaučivosti jer svaki model uči na različitim dijelovima skupa podataka i tako daje različite predikcije.

Pojačavanje je tehnika u strojnom učenju koja se razlikuje od pakiranja po tome što trenira niz modela u kojem svaki novi model ispravlja greške prethodnog. U pakiranju, osnovni modeli se treniraju paralelno, dok se u pojačavanju oni treniraju slijedno. Umjesto da koristi različite podskupove podataka, pojačavanje koristi cijeli skup podataka, ali pridaje veću težinu onim podacima koje su prethodni modeli predvidjeli pogrešno. Na taj način, svaki model u nizu postaje sve precizniji, usmjeravajući se na podatke gdje su prethodni modeli imali slabije performanse. Ovaj proces je ilustriran na slici 3.4.



Slika 3.4: Način rada ansambla zasnovanog na pojačavanju

Proces pojačavanja odvija se u nekoliko ključnih koraka. Prvi korak uključuje treniranje inicijalnog modela na cijelom skupu podataka. Nakon toga, procjenjuje se točnost ovog modela, a greške koje je napravio se identificiraju. Sljedeći model u nizu se trenira s većim naglaskom na primjerima koje je prethodni model pogrešno predvidio. Ovaj postupak se ponavlja za svaki novi model u nizu, pri čemu svaki model nastoji ispraviti greške prethodnih modela. Na kraju, predikcije svih modela se kombiniraju kako bi se dobio konačan rezultat, pri čemu svaki model doprinosi ovisno o svojoj točnosti. Cilj pojačavanja je smanjenje pristranosti modela i povećanje njihove točnosti, posebno kada osnovni modeli imaju slabije performanse. Ova metoda, koja se često koristi u praksi, poznata je kao adaptivno pojačanje (engl. *AdaBoost*).

### 3.4. Kritički osvrt

Ansamblu zasnovani na glasanju predstavljaju jednostavnu i intuitivnu tehniku koja kombinira predviđanja više modela kako bi se postigla konačna odluka. Ova metoda se često koristi zbog svoje lakoće implementacije, manje računalne složenosti i robusnosti. Većinsko glasanje posebno je korisno u klasifikacijskim problemima, dok prosječno i težinsko glasanje mogu biti primjenjivi u regresijskim zadacima. Prednost težinskog glasanja je u tome što omogućuje prilagodbu doprinosa svakog modela, čime se povećava točnost predikcija, što ga može učiniti pogodnim za procjenu cijena dionica. Međutim, nedostatak ove metode je njena osjetljivost na loše modele unutar ansambla, koji mogu negativno utjecati na konačni rezultat.

Slaganje, s druge strane, predstavlja sofisticiraniju tehniku gdje se predikcije nekoliko modela koriste za treniranje meta modela. Ova metoda omogućava integraciju različitih modela, čime se povećava sposobnost generalizacije. Slaganje je posebno korisno kada su dostupni modeli koji se različito ponašaju na različitim dijelovima skupa podataka. Iako slaganje može postići vrlo visoku točnost, njegova složenost i potreba za većom količinom podataka mogu predstavljati izazov u primjeni. Uz to, slaganje može zahtijevati značajnu računalnu snagu i vrijeme za treniranje, što može biti ograničavajuće u nekim okruženjima.

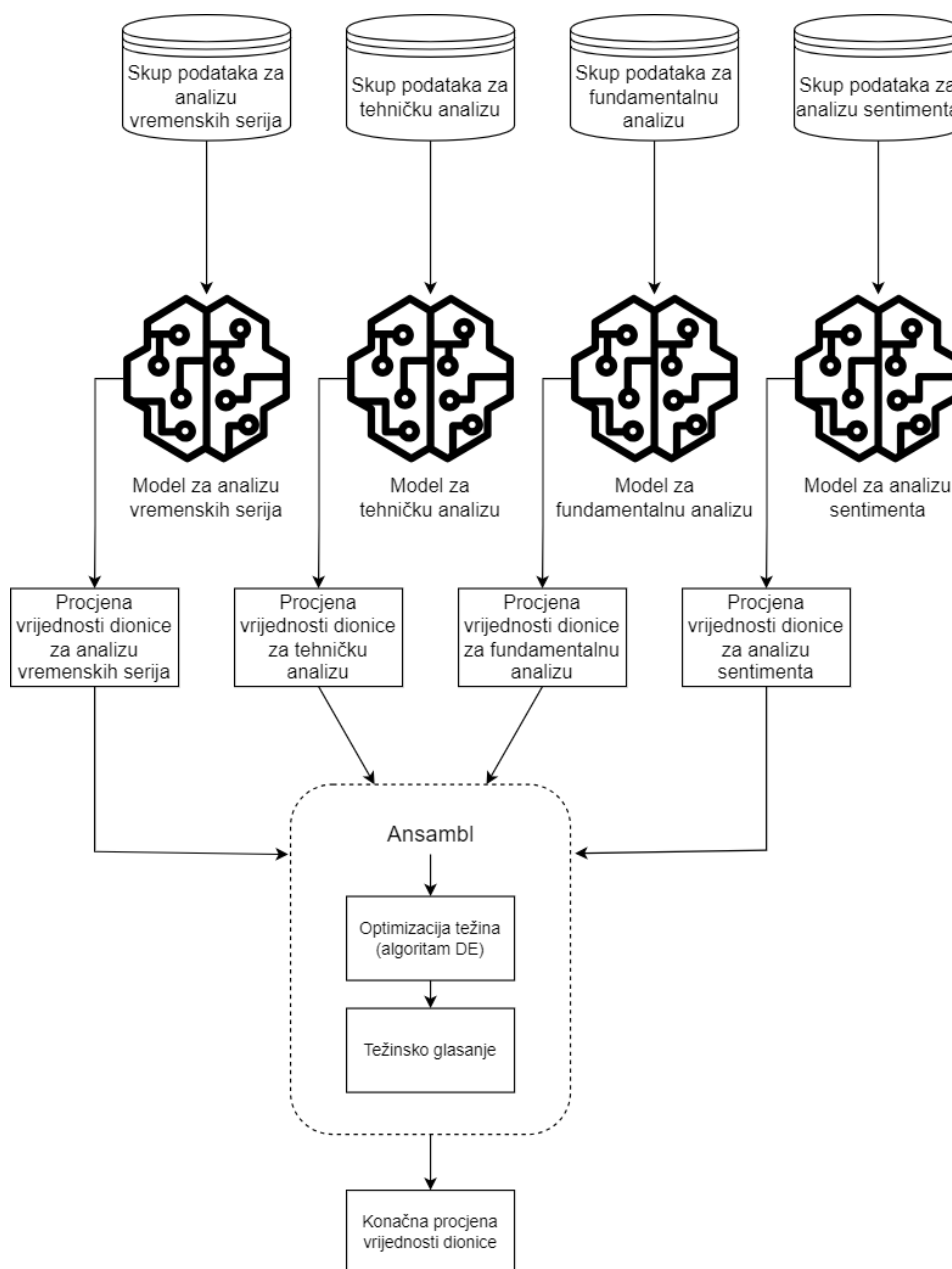


Pakiranje i pojačavanje su također robusne tehnike ansambla. Pakiranje koristi *bootstrap* agregiranje kako bi smanjilo varijabilnost modela, što može dovesti do značajnog poboljšanja performansi. Njegova glavna prednost je u smanjenju prenaučivosti, ali je mana što može biti računski intenzivno. Pojačavanje je izuzetno učinkovito u poboljšavanju performansi modela na specifičnim zadacima, no također je sklono prenaučivosti ako se ne primijeni pažljivo. Također, obje metode, zbog svoje složenosti, zahtijevaju više vremena i resursa u usporedbi s glasanjem.

Iako su tehnike poput slaganja, pakiranja i pojačavanja vrlo učinkovite, njihova složenost i potreba za računalnim resursima mogu predstavljati izazov u praksi. U usporedbi s njima, glasanje nudi jednostavniju implementaciju i manju računalnu zahtjevnost, što ga čini privlačnom opcijom u mnogim slučajevima. Jedan od takvih slučajeva je predviđanje cijena dionica, koje je samo po sebi izuzetno kompleksno, pa se odabir jednostavnijih algoritama za predviđanje nameće kao prikladniji izbor.

## 4. OSTVARENO PROGRAMSKO RJEŠENJE

Procjena cijena dionica je kompleksan izazov zbog mnogih čimbenika poput društvenih medija, financijskih izvještaja i raznih tehničkih indikatora. Analizirati ove podatke precizno je teško, stoga se koriste metode poput fundamentalne analize, tehničke analize, analize sentimenta i vremenskih serija. Kombiniranje tih metoda u ansambl može omogućiti dublje uvide i prilagodbu različitim uvjetima. Na slici 4.1 se nalazi dijagram ansambla korišten u ovom radu. Tu je prikazano kako se ujedinjuju prediktivne sposobnosti različitih modela koji koriste različite izvore podataka, s ciljem postizanja veće točnosti u predviđanju.



Slika 4.1: Ansambl za procjenu cijene dionice koristeći četiri vrste analiza

## 4.1. Odabir pojedinih modela

U literaturi je razvijeno mnoštvo modela u svrhu procjene vrijednosti dionice, međutim, njihovi rezultati često variraju zbog složenosti problema. Neki modeli su precizniji u određenim vremenskim razdobljima, dok su drugi bolji za specifične vrste analiza, a treći donose dobre rezultate za određene skupove podataka. Stoga je teško sa sigurnošću reći koji je model najbolji za određenu situaciju i upravo zbog toga je važno razviti više modela za svaku vrstu analize.

Za analizu povijesnih cijena dionica koriste se različiti modeli za analizu vremenskih serija. Prikladni modeli vremenskih serija za ovu vrstu problema su spomenuti modeli ES, ARIMA i ARIMA-GARCH. Ovi modeli su pogodni jer se oslanjaju na pretpostavku da će se prošli trendovi ponavljati u budućnosti, te se fokusiraju isključivo na povijesne cijene dionica, bez potrebe za dodatnim informacijama. Na primjer, model ES se koristi za modeliranje podataka koji pokazuju sezonalnost i trendove kroz vrijeme, pružajući jednostavne i efikasne projekcije budućih kretanja cijene. Model ARIMA se primjenjuje jer učinkovito obrađuje vremenske serije koje mogu imati trendove i sezonalnosti, dok diferenciranje pomaže u postizanju stacionarnosti, što je ključno za precizno predviđanje budućih cijena na temelju povijesnih podataka. Proširenje ARIMA modela je ARIMA-GARCH model, koji uključuje modeliranje volatilnosti, što ga čini prikladnim za analizu financijskih podataka koji često pokazuju promjenjivu volatilnost.

S druge strane, složeniji regresijski modeli su prikladni za analizu sentimenta, tehničkih i fundamentalnih faktora, gdje je potrebno obraditi veći broj značajki. Ovi modeli omogućuju prepoznavanje odnosa između zavisne varijable i jedne ili više nezavisnih varijabli, što je ključno kada se analizira više faktora. Česti modeli korišteni u ovoj vrsti analizi su modeli SVM, ANN, LSTM i RF. Konkretno, SVM i ANN modeli su posebno učinkoviti u analizi složenih i nelinearnih odnosa među faktorima koji utječu na cijene. Nadalje, LSTM modeli su korisni zbog svoje sposobnosti da dugoročno zadrže i analiziraju važne povijesne informacije, prepoznajući obrasce u vremenskim serijama. Modeli RF su dobri za procjenu cijene dionica zbog svoje jednostavnosti i sposobnosti obrade različitih vrsta podataka te fleksibilnosti u modeliranju složenih odnosa između značajki.

Nakon što su odabrani modeli za upotrebu, potrebno ih je podesiti i trenirati. Ovi koraci uključuju podešavanje parametara (engl. *parameter tuning*), pripremu i selekciju podataka, te samo treniranje i vrednovanje modela. Za potrebe pojednostavljenja modela odabrani su najčešći i najutjecajniji podaci za fundamentalnu i tehničku analizu, koji su navedeni u drugom poglavlju. Važno je napomenuti da se podaci za fundamentalnu analizu ažuriraju samo kvartalno, što može

predstavljati nedostatak jer su novi podaci dostupni tek svaka tri mjeseca. Nasuprot tome, podaci iz drugih vrsta analiza ažuriraju se svakodnevno. Za ocjenjivanje sentimenta koriste se prosječna ocjena sentimenta i standardna devijacija ocjena. U ansamblu se koristi samo jedan model u okviru svake analize, te je nužno usporediti različite modele kako bi se izabrao najbolji. Najprikladniji modeli se zatim integriraju u ansambl gdje se provodi optimizacija težina i težinsko glasanje.

## 4.2. Težinsko glasanje

Za kombiniranje izlaza pojedinih modela koristi se metoda težinskog glasanja, koja je prethodno opisana kao prikladna metoda za ovu vrstu problema. Kao što je navedeno, izlaz ansambla predstavlja linearnu kombinaciju izlaza pojedinih modela. Izlaze pojedinih modela možemo predstaviti kao vektorom

$$\mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \in \mathbb{R}^n,$$

gdje su  $y_i$  procjene pojedinih modela, a  $n$  ukupan broj modela. Na drugoj strani, radi se optimizacija težina za što treba optimizacijski algoritam, a njegov izlaz je stoga predstavljen vektorom

$$\mathbf{W} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} \in [0,1]^n,$$

gdje su  $w_i$  optimizirane težine za pojedine modele. Oznaka  $n$  predstavlja broj modela odnosno broj težina koji je u ovom radu jednak četiri. Izlaz ansambla predstavlja linearnu kombinaciju elemenata vektora  $\mathbf{Y}$  gdje su njihove odgovarajuće težine sadržane u vektoru  $\mathbf{W}$  prema

$$\hat{y} = \sum_{i=1}^n (y_i \cdot w_i). \quad (4.1)$$

Za ilustraciju, pretpostavimo da su odabrane težine 0.1, 0.2, 0.3 i 0.4, pri čemu je svaka težina dodijeljena izlazu pojedinog modela. Izlazi modela bit će pomnoženi s pripadajućom težinom, a potom zbrojeni kao

$$\text{Rezultat} = ((0.1y_1) + (0.2y_2) + (0.3y_3) + (0.4y_4)), \quad (4.2)$$

gdje su  $y_1, y_2, y_3, y_4$  izlazi pojedinih modela.

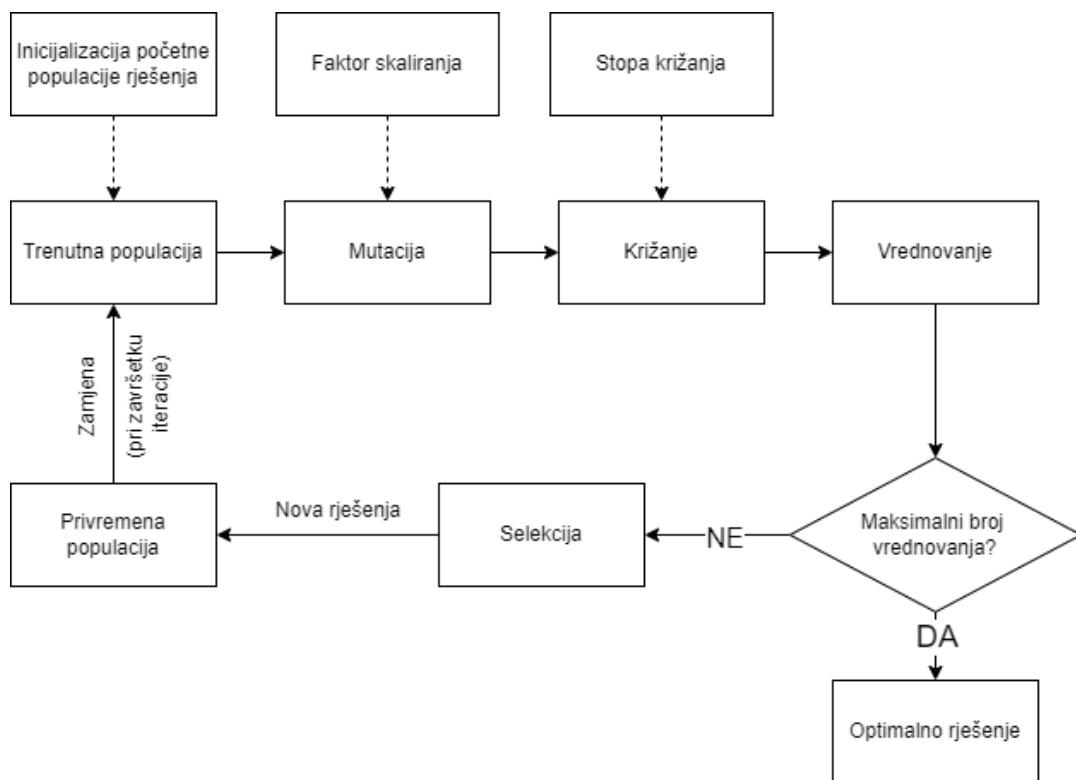
### 4.3. Optimizacija težina u ansamblu

Glavni izazov ovog ansambla leži u koraku optimizacije težina prije postupka težinskog glasanja, gdje se pravilno trebaju odrediti težine za svaki model. Za to postoji niz različitih algoritama optimizacije, a u ovom radu se koristi algoritam diferencijalne evolucije (engl. *differential evolution*, DE) za optimizaciju težina.

Diferencijalna evolucija je optimizacijski algoritam temeljen na populaciji rješenja koji istražuje prostor pretrage kroz postupno poboljšavanje rješenja primjenom evolucijskih procesa kao što su mutacija, križanje i selekcija. Ovi algoritmi zahtijevaju minimalne ili nikakve pretpostavke o osnovnom problemu optimizacije i sposobni su za brzo istraživanje vrlo opsežnih prostora pretrage. Algoritam DE se često smatra jednim od svestranijih i stabilnijih algoritama pretraživanja temeljenih na populaciji [17]. U ovom radu, standardni algoritam DE (DE/rand/1/bin) [18] (uz manju modifikaciju operatora mutacije predloženu u [19]) se primjenjuje u fazi ansambla, gdje se koristi za optimizaciju težina.

#### 4.3.1. Algoritam diferencijalne evolucije

Na slici 4.2 prikazan je dijagram rada algoritma DE. Kao što je prethodno spomenuto, algoritam koristi procese koji simuliraju evolucijske procese u nastojanju da pronađe optimalno rješenje.



Slika 4.2: Dijagram algoritma DE

Na početku algoritma DE, potrebno je inicijalizirati početnu populaciju, što se često radi nasumično. Svaki pojedinac u populaciji generira se tako da njegove vrijednosti za svaku varijablu budu unutar unaprijed definiranih granica prostora pretrage.

Mutacija je ključna operacija u DE algoritmu. U ovom programskom rješenju, mutacija se izvodi kao i u [20], gdje se stvara novi vektor (mutant) kao

$$\mathbf{u}^{j,g} = \mathbf{v}^{r1,g} + F \cdot (\mathbf{v}^{r2,g} - \mathbf{v}^{r3,g}), \quad (4-2)$$

gdje su  $\mathbf{v}^{r1,g}$ ,  $\mathbf{v}^{r2,g}$  i  $\mathbf{v}^{r3,g}$  nasumično odabrani vektori iz populacije (uz uvjet  $j \neq r1 \neq r2 \neq r3$ ), pri čemu je  $\mathbf{v}^{r1,g}$  najkvalitetniji među njima. Mutirani vektor  $\mathbf{u}^{j,g}$  se stvara za svaki član trenutne populacije  $\mathbf{v}^{j,g}$ ,  $j = 1, \dots, N_p$  (veličina populacije), gdje  $g$  predstavlja broj generacije. Razlika dvaju vektora određuju intenzitet mutacije i dodatno je ponderirana faktorom skaliranja (engl. *scale factor*)  $F \in < 0, 1]$ .

Križanje se koristi za kombiniranje mutiranog vektora s trenutnim kandidatom rješenja tako što svaki element iz mutiranog vektora može zamijeniti odgovarajući element u trenutnom rješenju, prema definiranoj stopi križanja (engl. *crossover rate*). Da bi se osigurala zamjena barem jednog elementa, jedan nasumični element se uvijek mijenja.

Nakon mutacije i križanja, dodaje se mali Gaussov šum faktoru skaliranja  $F$  i stopi križanja  $CR$ . Ova modifikacija primjenjuje se na razini svakog člana populacije, prema sljedećim jednadžbama

$$F_i^g = F + N_i^g(0, 0.1)$$

$$CR_i^g = CR + N_i^g(0, 0.1), \quad (4-3)$$

gdje  $g$  označava broj generacije, dok  $i = 1, \dots, N_p$  predstavlja indeks člana trenutne populacije. Gaussova varijabla ima distribucijske parametre  $\mu = 0$  i  $\sigma = 0.1$ . Korištenje Gaussovog šuma uvodi određenu varijabilnost u vrijednosti faktora skaliranja i stope križanja, što znači da ovi parametri neće uvijek biti fiksni, već će se blago mijenjati tijekom vremena. Takav pristup doprinosi fleksibilnosti algoritma, čime se poboljšava istraživački potencijal u različitim fazama evolucijskog procesa.

Selekcija odlučuje hoće li novo rješenje, nastalo mutacijom i križanjem, zamijeniti postojeće rješenje u populaciji, pri čemu će novo rješenje zamijeniti staro samo ako pokazuje bolju kvalitetu (engl. *fitness*), dok će u suprotnom staro rješenje ostati. Rješenja koja prođu selekciju se pohranjuju

u privremenu populaciju, a na kraju iteracije zamjenjuju rješenja trenutne populacije. Optimalno rješenje na kraju algoritma naziva se rješenje s najboljom kvalitetom (engl. *best fitness*). Drugim riječima, optimalno rješenje zapravo predstavlja najbolje rješenje u konačnoj populaciji.

Cijeli proces algoritma DE prikazan je pseudokodom na slici 4.3, koji ilustrira način izvršavanja ovog postupka unutar računalnog okvira.

### **Linija    Kod**

```

1:        /*Algoritam DE*/
2:        Postavi parametre algoritma;
3:        Inicijaliziraj početnu populaciju rješenja;
4:        g = 0
5:        ponavljaj
6:        #Prođi kroz sva rješenja populacije
7:        za j = 1, ..., Np čini
8:            Mutacijom stvori novi vektor (mutant)  $\mathbf{u}^{j,g}$  ;
9:            Križanjem mutanta  $\mathbf{u}^{j,g}$  i trenutnog rješenja  $\mathbf{v}^{j,g}$  stvori vektor  $\mathbf{t}^{j,g}$  ;
10:          #Selekcija
11:          ako  $\mathbf{t}^{j,g}$  kvalitetniji ili jednak  $\mathbf{v}^{j,g}$  onda
12:             $\mathbf{v}^{j,g+1} = \mathbf{t}^{j,g}$ ;
13:          inače
14:             $\mathbf{v}^{j,g+1} = \mathbf{v}^{j,g}$ ;
15:          kraj
16:        kraj
17:        g = g + 1;
18:        dok se ne dosegne maksimalan broj vrednovanja;

```

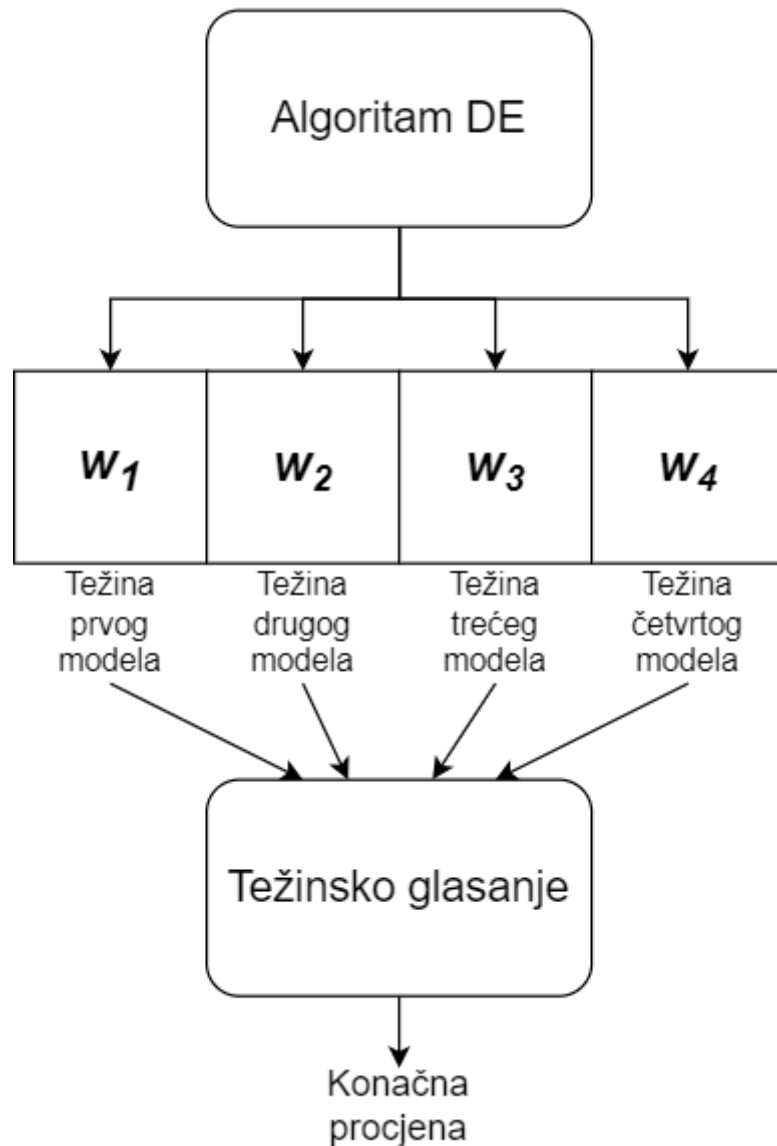
Slika 4.3: Pseudokod algoritma DE

Glavni parametri algoritma DE uključuju veličinu populacije ( $N_p$ ), koja određuje broj kandidata rješenja, te maksimalan broj vrednovanja, nakon čega se algoritam zaustavlja. Parametri poput stope križanja i faktora skaliranja kontroliraju kako se mutirani vektor i trenutni kandidat kombiniraju i skaliraju. Metoda za upravljanje granicama osigurava da rješenja ostanu unutar zadanih granica.

### **4.3.2. Primjena algoritma DE za optimizaciju težina**

Kako se u ovom radu koriste četiri različita modela, od kojih svaki daje svoju individualnu procjenu, tako i algoritam DE generira svoje rješenje. Svako rješenje unutar algoritma DE predstavlja četverodimenzionalni (4D) vektor, koji sadrži odgovarajuće težine za svaki od četiri modela unutar ansambla. Te se težine zatim primjenjuju u postupku težinskog glasanja, kao što je

opisano pomoću (4.1). Ovaj korak predstavlja završnu fazu ansambla, pri čemu se dobiva konačna procjena cijene dionice.



Slika 4.4: Dijagram primjene algoritma DE za optimizaciju težina

Slika 4.4 prikazuje dijagram primjene algoritma DE za optimizaciju težina. Optimalno rješenje nastoji se pronaći pomoću ranije opisanog algoritma DE, gdje se tijekom selekcije vrednuju moguća rješenja. Na primjer, algoritam DE nasumično generira četverodimenzionalni vektor. Istodobno, postoje procjene od različitih modela koje su isto predstavljene kao vektori, pri čemu svaki vektor odgovara jednom danu za koji se vrši procjena. Kada se ti vektori uključe u procesu težinskog glasanja, dobiju se rezultati za svaki dan koji se procjenjuje. Zatim se ta predviđanja uspoređuju sa stvarnim vrijednostima kako bi se ocijenila učinkovitost trenutnog



četverodimenzionalnog vektora težina. Takva procjena kvalitete se može izraziti nekim uobičajenim mjerama za vrednovanje učinkovitosti modela kao što su srednja kvadratna pogreška ili korijen srednje kvadratne pogreške. Na temelju te procjene, odlučuje se hoće li se rješenje odbaciti ili zadržati. Ovaj se proces ponavlja sve dok se ne postigne maksimalni broj vrednovanja, s ciljem pronalaska optimalnog rješenja na kraju postupka.

## 5. EKSPERIMENTALNA ANALIZA I REZULTATI

U ovom eksperimentu razvijeni su modeli za svaku vrstu analize. Korištenjem skupa prikupljenih podataka, modeli su podešavani s ciljem postizanja što boljih rezultata. Nakon treniranja modela, rezultati su uspoređeni unutar svake pojedine analize, te je najbolji model odabran za daljnju integraciju u ansambl. Dva različita ansambla zatim su uspoređena kako bi se temeljitije analizirali rezultati i procijenila njihova učinkovitost.

Modeli i ansambl procjenjuju cijenu dionice za sljedeći dan, što je prikladan vremenski interval s obzirom na dostupne podatke. Za procjenu cijena na tjednoj, mjesečnoj ili godišnjoj razini, bio bi potreban znatno duži vremenski raspon podataka. Također, pri analizi rezultata treba uzeti u obzir kratki vremenski interval procjene, budući da se cijene mogu lako mijenjati iz dana u dan bez jasnih pravila.

Pretpostavka je da će ansambl, kombinirajući najbolje modele iz svake analize, dati preciznije procjene od bilo kojeg pojedinog modela. Tijekom analize korišteni su različiti grafovi kako bi se vizualizirale razne perspektive rezultata, čime se omogućilo detaljnije razumijevanje izvedbe modela. To je nužno jer sami rezultati ne pružaju uvijek potpun uvid u sve aspekte analize. Vizualni prikazi pomogli su u interpretaciji rezultata, omogućujući jasniji uvid u učinkovitost i uspješnost pojedinih modela i ansambla.

### 5.1. Postavke eksperimenta

Prikupljeni su podaci za tri dionice, Google (GOOG), Amazon (AMZN) i Apple (AAPL) u razdoblju od 2015. do 2020. godine. Podaci su obuhvatili svaki radni dan burze tijekom tih pet godina te su korišteni za različite vrste analize. Povijesni podaci odabranih dionica preuzeti su s platforme Yahoo Finance [21], uključujući podatke o cijeni otvaranja (engl. *open*), najvišoj (engl. *high*) i najnižoj (engl. *low*) cijeni, cijeni zatvaranja (engl. *close*) te volumenu trgovanja za svaki radni dan burze. Na temelju tih podataka izračunati su tehnički indikatori pomoću *TA-Lib* biblioteke [22]. Parametri korišteni za izračun tehničkih indikatora detaljno su opisani u drugom poglavlju. Za fundamentalnu analizu iskorišteni su financijski izvještaji sa stranice *SEC* [7], a ključni parametri navedeni su također u drugom poglavlju. Budući da financijski izvještaji izlaze kvartalno, podaci su za svaki radni dan popunjeni s istim vrijednostima sve dok ne dođe novi financijski izvještaj. Za analizu sentimenta korišten je skup podataka objava s društvene mreže X (bivši Twitter). Objave za svaku dionicu su najprije filtrirane, a potom analizirane pomoću unaprijed istreniranog modela za obradu prirodnog jezika roBERTa [23], koji je ocijenio sentiment

svake objave. Zatim su izračunati prosječni sentiment i standardna devijacija za svaki radni dan burze.

Cilj modela bio je predvidjeti kretanje cijene dionice sljedećeg dana u odnosu na današnju cijenu. Primjerice, ako je trenutna cijena 50 dolara, a model predviđa -2, to znači da se očekuje pad cijene za 2 dolara, odnosno da će sutra cijena iznositi 48 dolara. Stoga je važno u skupu podataka uključiti varijablu koja prati razliku cijene između dva uzastopna dana. Ova razlika dodana je kao parametar koji prikazuje koliko se cijena promijenila na određen dan u odnosu na prethodni. Time svaki dan ima svoj kompletan skup podataka za tehničku, fundamentalnu i sentiment analizu, uz stvarnu cijenu, razliku cijene i razliku cijene za idući dan. Razlika cijene za idući dan se koristila kao izlazna varijabla (Y) za treniranje modela, dok su ostali podaci ulazni (X).

Kao što je navedeno, korišteni su ARIMA, ES i ARIMA-GARCH modeli za analizu vremenskih serija. Svaki model prilagođen je specifičnostima dionice koju analizira, jer ovi modeli mogu prepoznati trendove i sezonalnosti karakteristične za pojedine dionice. Parametri za ARIMA i ARIMA-GARCH modele su redosljed autoregresije (engl. *autoregressive order*, p), stupanj diferenciranja (engl. *degree of differencing*, d) i redosljed pomičnog prosjeka (engl. *moving average order*, q), dok GARCH ima dodatne parametre redosljed autoregresivnih uvjetnih varijanci (engl. *order of the autoregressive conditional variances*, p) i redosljed pomičnih prosjeka uvjetnih varijanci (engl. *order of the moving averages of conditional variances*, q). Za određivanje p, d, q ARIMA parametre poslužio je prošireni Dickey-Fullerov test (engl. *augmented Dickey-Fuller test*, ADF test) [24] koji se koristi za provjeru stacionarnosti vremenskih serija. Ukoliko serija nije stacionarna, potrebno je primijeniti diferenciranje kako bi se pripremila za ARIMA modeliranje. Akaikeov informacijski kriterij (engl. *Akaike information criterion*, AIC) i Bayesov informacijski kriterij (engl. *Bayesov informacijski kriterij*, BIC) su korišteni za određivanje p i q parametre u GARCH dijelu [25]. AIC traži ravnotežu između točnosti i složenosti modela, dok BIC dodatno penalizira složenost ovisno o broju uzoraka. Oba kriterija pomažu u odabiru optimalnog modela, osobito u ARIMA-GARCH modeliranju, analizirajući kako različiti parametri utječu na prilagodbu podacima i složenost modela. Kod ES modela, ključan je parametar zaglađivanja (engl. *smoothing parametar*,  $\alpha$ ), koji je optimiziran pomoću metode pretrage mreže (engl. *grid search*). Podešeni parametri za modele vremenskih serija u okviru svake dionice prikazani su u tablici 5.1.

Tablica 5.1: Modeli za analizu vremenskih serija

Analiza vremenskih serija za Apple		
Model	Parametri	Skup podataka
ARIMA	$p = 3, d = 1, q = 3$	Stvarna cijena
ES	$\alpha = 0.6$	Stvarna cijena
ARIMA-GARCH	ARIMA: $p = 3, d = 1, q = 3$ GARCH: $p = 4, q = 1$	Stvarna cijena

Analiza vremenskih serija za Amazon		
Model	Parametri	Skup podataka
ARIMA	$p = 8, d = 1, q = 1$	Stvarna cijena
ES	$\alpha = 0.6$	Stvarna cijena
ARIMA-GARCH	ARIMA: $p = 8, d = 1, q = 1$ GARCH: $p = 1, q = 4$	Stvarna cijena

Analiza vremenskih serija za Google		
Model	Parametri	Skup podataka
ARIMA	$p = 3, d = 1, q = 3$	Stvarna cijena
ES	$\alpha = 0.6$	Stvarna cijena
ARIMA-GARCH	ARIMA: $p = 3, d = 1, q = 3$ GARCH: $p = 4, q = 1$	Stvarna cijena

Ostale tri analize koristile su modele SVM, LSTM, ANN i RF, pri čemu je svaki model imao specifične parametre unutar svake vrste analize. Parametri nisu prilagođeni svakoj pojedinoj dionici, kao što je kod analize vremenskih serija.

Za implementaciju SVM modela korišten je SVR algoritam iz biblioteke *sklearn* [26]. Prije treniranja modela, podaci su normalizirani upotrebom funkcije *StandardScaler* iz iste biblioteke, kako bi se olakšalo treniranje. Ključni parametri regularizacije (engl. *regularization parameter*,  $C$ ), gama (engl. *kernel coefficient*,  $\gamma$ ) i jezgre (engl. *kernel*) modela SVM određeni su primjenom metode pretrage mreže. Nakon što su dobiveni rezultati, nužno je provesti inverznu transformaciju kako bi se dobio stvarni rezultat. Parametri SVM modela za svaku pojedinu vrstu analize detaljno su prikazani u tablici 5.2.

Tablica 5.2: Upotrebljeni parametri i skup podataka za SVM modele u sklopu svake analize

SVM model		
Analiza	Parametri	Skup podataka
Tehnička analiza	Jezgra= rbf, $\gamma = 0.1$ , C = 100	Stvarna cijena, razlika cijene, MACD, RSI, SMA, EMA, Stohastički oscilator
Fundamentalna analiza	Jezgra = rbf, $\gamma = 0.2$ , C = 100	Stvarna cijena, razlika cijene, prihod, neto dobit, EPS, bruto profitna marža, operativna marža, neto profitna marža, novčani tok iz poslovanja, ukupna aktiva, ukupne obveze
Analiza sentimenta	Jezgra = rbf, $\gamma = 0.2$ , C = 100	Stvarna cijena, razlika cijene, prosječna ocjena, standardna devijacija ocjena

Za model RF korišten je *RandomForestRegressor* iz biblioteke *sklearn*. Parametri kao što su maksimalna dubina (engl. *max depth*), maksimalan broj značajki (engl. *max features*), minimalan broj uzoraka po listi (engl. *min samples leaf*), minimalan broj uzoraka za podjelu (engl. *min samples split*) i broj stabala (engl. *n estimators*) određeni su primjenom metode pretraga mreže. Ovi ključni parametri optimizirani su kako bi se postigla najbolja točnost modela. Parametri za RF model detaljno su prikazani u tablici 5.3.

Tablica 5.3: Upotrebjeni parametri i skup podataka za RF modele u sklopu svake analize

RF model		
Analiza	Parametri	Skup podataka
Tehnička analiza	Maksimalna dubina = 50, Maksimalni broj značajki = None, Minimalni broj uzoraka po listi = 2, Minimalni broj uzoraka za podjelu 50, Broj stabala = 100,	Stvarna cijena, razlika cijene, MACD, RSI, SMA, EMA, Stohastički oscilator
Fundamentalna analiza	Maksimalna dubina = 50, Maksimalni broj značajki = None, Minimalni broj uzoraka po listi = 2, Minimalni broj uzoraka za podjelu 50, Broj stabala = 100,	Stvarna cijena, razlika cijene, prihod, neto dobit, EPS, bruto profitna marža, operativna marža, neto profitna marža, novčani tok iz poslovanja, ukupna aktiva, ukupne obveze
Analiza sentimenta	Maksimalna dubina = 50, Maksimalni broj značajki = None, Minimalni broj uzoraka po listi = 2, Minimalni broj uzoraka za podjelu 50, Broj stabala = 100,	Stvarna cijena, razlika cijene, prosječna ocjena, standardna devijacija ocjena

Kompleksniji model ANN izgrađen je koristeći biblioteku *tensorflow.keras* [27] koji se koristi za kreiranje slojeva modela. Svaki model sadrži ulazni sloj, jedan ili više skrivenih slojeva te izlazni sloj koji daje konačnu predikciju. Svaki sloj ima određeni broj neurona i aktivacijsku funkciju koja aktivira neurone. Treniranje ovog modela odvijalo se pomoću optimizacije *adam*, a ključni parametri koji su specificirani uključuju broj epoha, veličina skupine i validacijski skup te mehanizam ranog zaustavljanja (engl. *early stopping*), koji prekida treniranje ako se rezultati modela ne poboljšavaju nakon određenog broja iteracija. Prije treniranja, podaci su normalizirani upotrebom funkcije *MinMaxScaler* iz biblioteke *sklearn*, a nakon treniranja, rezultat modela potrebno je inverzno transformirati kako bi se dobio stvarni rezultat. Parametri ANN modela detaljno su prikazani u tablici 5.4.

Tablica 5.4: Upotrebljeni parametri i skup podataka za ANN modele u sklopu svake analize

ANN model		
Analiza	Parametri	Skup podataka
Tehnička analiza	Prvi sloj: Dense(256, 'relu'), Drugi sloj: Dense(16, 'relu'), Treći sloj: Dense(1), Optimizacija = Adam, Broj epoha = 50, Veličina skupine = 32, Validacijski skup = 0.1, Rano zaustavljanje: strpljenje=5	Stvarna cijena, razlika cijene, MACD, RSI, SMA, EMA, Stohastički oscilator
Fundamentalna analiza	Prvi sloj: Dense(256, 'relu'), Drugi sloj: Dense(16, 'relu'), Treći sloj: Dense(1), Optimizacija = Adam, Broj epoha = 50, Veličina skupine = 32, Validacijski skup = 0.1, Rano zaustavljanje: strpljenje=5	Stvarna cijena, razlika cijene, prihod, neto dobit, EPS, bruto profitna marža, operativna marža, neto profitna marža, novčani tok iz poslovanja, ukupna aktiva, ukupne obveze
Analiza sentimenta	Prvi sloj: Dense(256, 'relu'), Drugi sloj: Dense(16, 'relu'), Treći sloj: Dense(1), Optimizacija = Adam, Broj epoha = 50, Veličina skupine = 32, Validacijski skup = 0.1, Rano zaustavljanje: strpljenje=5	Stvarna cijena, razlika cijene, prosječna ocjena, standardna devijacija ocjena

Za izradu složenijeg modela LSTM korištena je biblioteka *tensorflow.keras* koja je omogućila izradu potrebnih slojeva. Uz standardne slojeve, implementiran je i sloj ispuštanja (engl. *dropout layer*), koji privremeno deaktivira određeni broj neurona kako bi se spriječilo prenaučnost modela. Kao i kod ANN modela, definirani su ključni parametri, uključujući broj epoha, veličinu serije, validacijski skup te mehanizam ranog zaustavljanja, uz korištenje *adam* optimizacije. Podaci su normalizirani prije samog treniranja modela upotrebom funkcije *MinMaxScaler*, a nakon treniranja inverzno transformirati rezultate kako bi se dobile konačne predikcije. Parametri LSTM modela detaljno su prikazani u tablici 5.5.

Tablica 5.5: Upotrebjeni parametri i skup podataka za LSTM modele u sklopu svake analize

<b>LSTM model</b>		
<b>Analiza</b>	<b>Parametri</b>	<b>Skup podataka</b>
Tehnička analiza	Prvi sloj: LSTM(units = 50), Drugi sloj: Dropout(0.2), Treći sloj: LSTM(units = 50), Četvrti sloj: Dropout(0.2), Peti sloj: Dense(1), Optimizacija = Adam, Broj epoha = 50, Veličina skupine = 32, Validacijski skup = 0.1, Rano zaustavljanje: strpljenje=5	Stvarna cijena, razlika cijene, MACD, RSI, SMA, EMA, Stohastički oscilator
Fundamentalna analiza	Prvi sloj: LSTM(units = 100), Drugi sloj: Dropout(0.2), Treći sloj: LSTM(units = 100), Četvrti sloj: Dropout(0.2), Peti sloj: Dense(1), Optimizacija = Adam, Broj epoha = 50, Veličina skupine = 32, Validacijski skup = 0.1, Rano zaustavljanje: strpljenje=5	Stvarna cijena, razlika cijene, prihod, neto dobit, EPS, bruto profitna marža, operativna marža, neto profitna marža, novčani tok iz poslovanja, ukupna aktiva, ukupne obveze
Analiza sentimenta	Prvi sloj: LSTM(units = 100), Drugi sloj: Dropout(0.2), Treći sloj: LSTM(units = 100), Četvrti sloj: Dropout(0.2), Peti sloj: Dense(1), Optimizacija = Adam, Broj epoha = 50, Veličina skupine = 32, Validacijski skup = 0.1, Rano zaustavljanje: strpljenje=5	Stvarna cijena, razlika cijene, prosječna ocjena, standardna devijacija ocjena

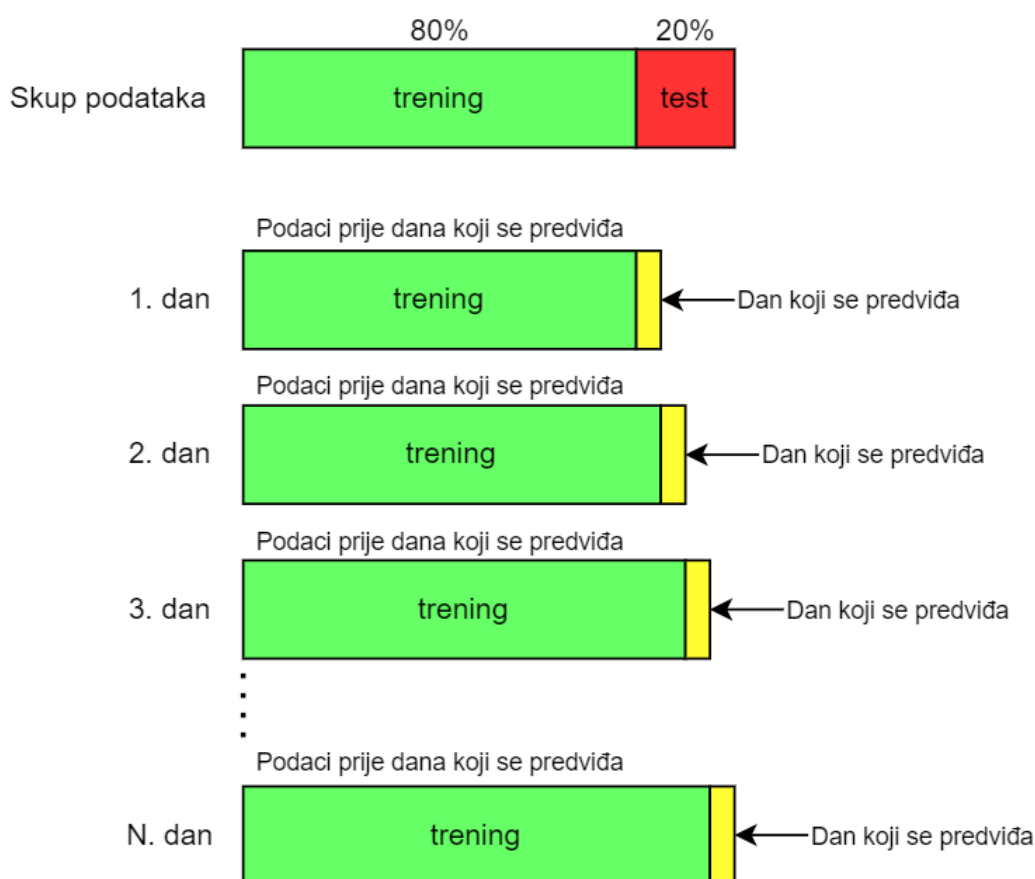
Sada se razmatraju postavljene parametri za algoritam DE, koji su prikazani u tablici 5.6. Ovaj algoritam koristi rezultate modela iz svake pojedinačne analize kao ulazni skup podataka. Parametri algoritma DE detaljno su opisani u četvrtom poglavlju.



Tablica 5.6: Upotrebjeni parametri i skup podataka za algoritam DE

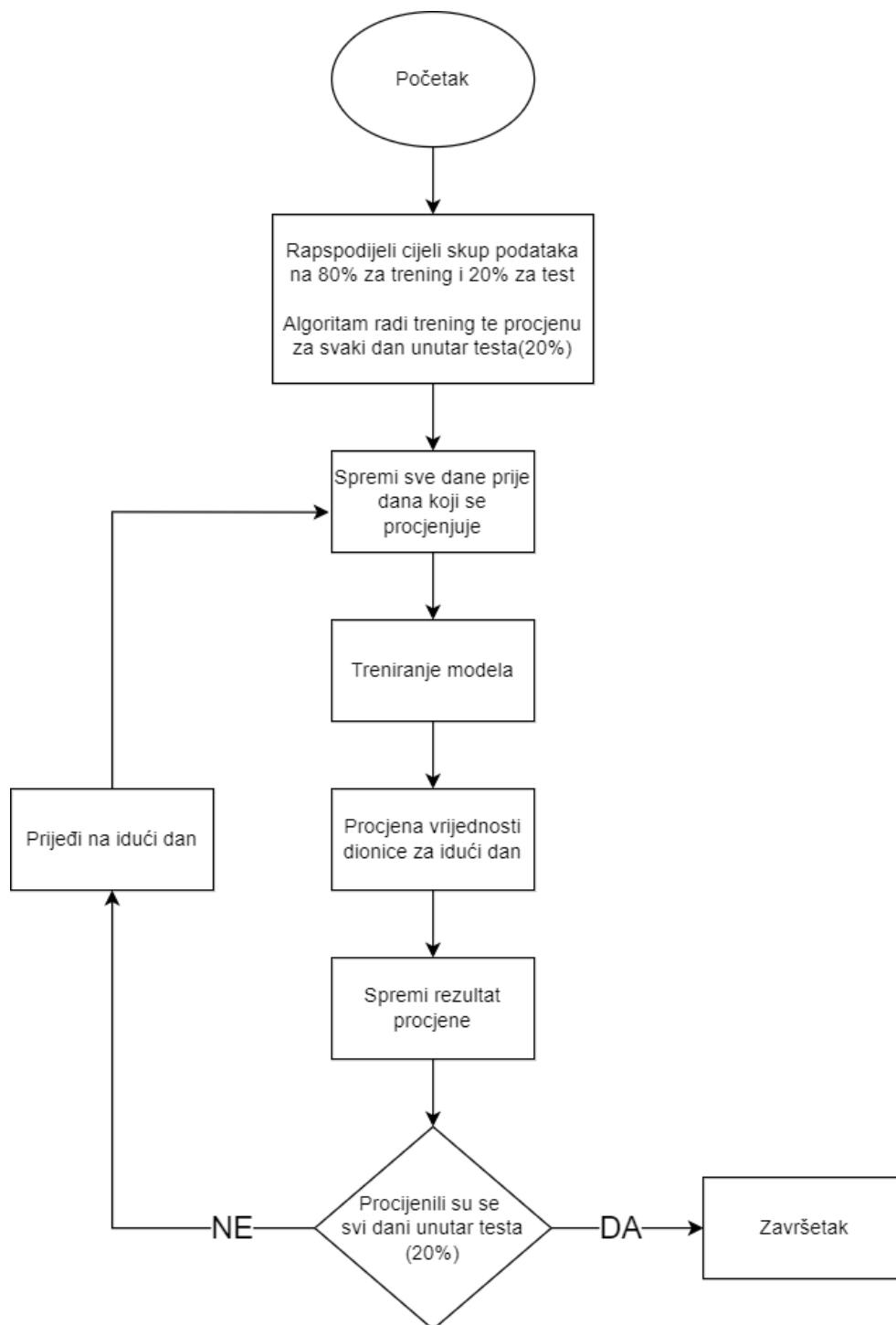
Algoritam DE	
Parametri	Skup podataka
Veličina populacije = 50, Broj vrednovanja = 15000, Stopa križanja = 0.9, Faktor skaliranja = 0.5,	Procjene fundamentalne analize, procjene tehničke analize, procjene analize sentimenta, procjene analize vremenskih serija

Svi modeli koristili su metodu validacije s vremenskim pomicanjem unaprijed (engl. *forward-walk validation*) za predviđanje vrijednosti dionica za sljedeći dan. Ova metoda omogućava modelima treniranje na svim dostupnim podacima do dana za koji se vrši predikcija. Validacija s vremenskim pomicanjem unaprijed provodi se tako da modeli predviđaju dane iz testnog skupa jedan po jedan dan, pri čemu se svakom iteracijom proširuje skup podataka za treniranje. Proces je ilustriran na slici 5.1. U ovom eksperimentu, podaci su podijeljeni tako da 80% čini skup za treniranje, a 20% skup za testiranje.



Slika 5.1: Način rada validacije uz vremensko pomicanje unaprijed

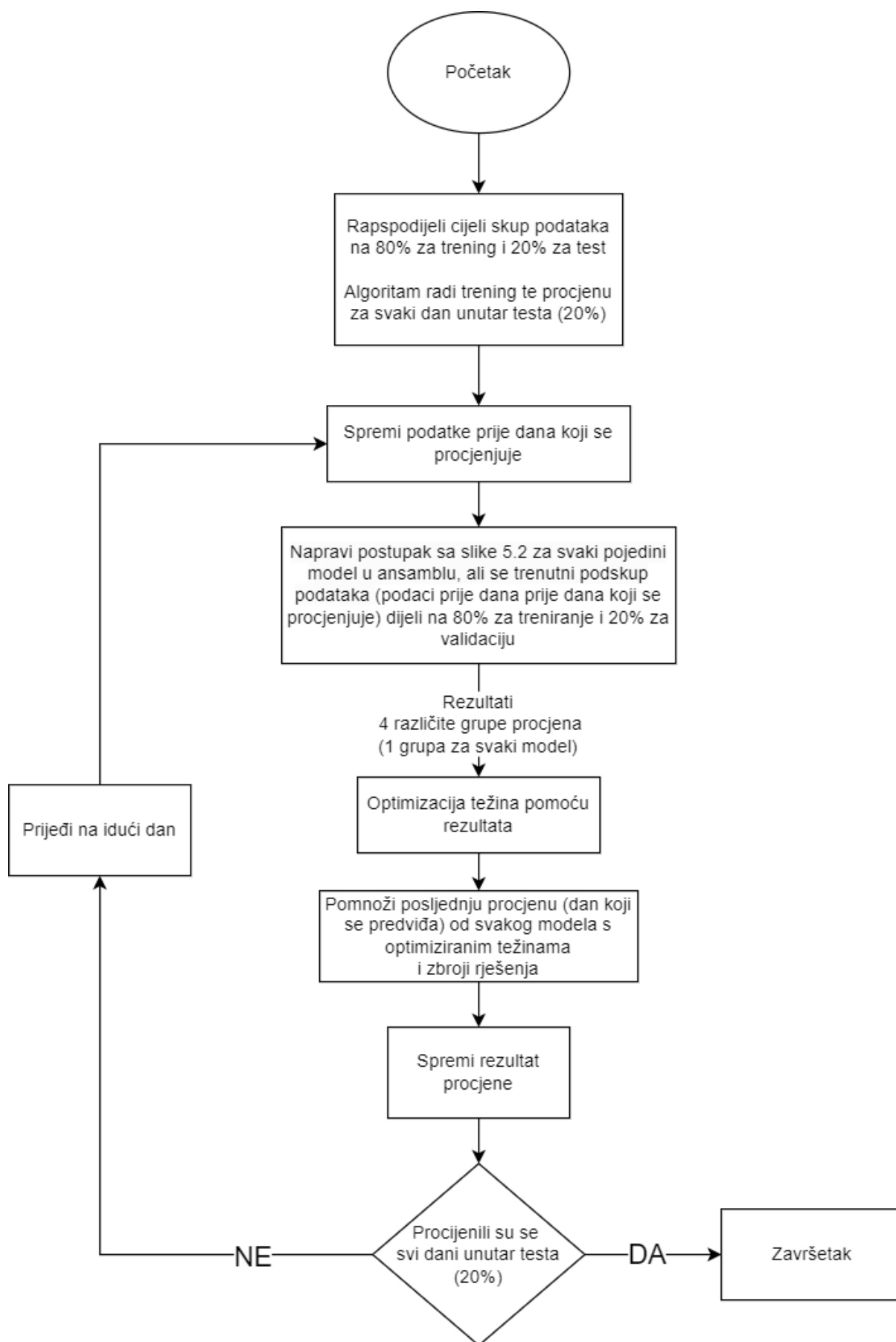
Podjela skupova i implementacija rješenja za svaki model prikazana je na slici 5.2. Prije svakog koraka, modeli se treniraju na svim dostupnim podacima prije dana za koji se vrši predikcija. Ovim pristupom, dolaskom svakog novog dana, podaci se kontinuirano ažuriraju, što omogućava modelu preciznije predviđanje budućih kretanja cijena. Nakon svakog koraka, novotrenirani model vrši procjenu, rezultat se sprema, a proces se ponavlja sve dok se ne izvrše procjene za sve dane iz testnog skupa podataka.



Slika 5.2: Način treniranja i vrednovanja modela

S druge strane, isti postupak validacije s vremenskim pomicanjem unaprijed primijenjen je na ansamblu. Ključno je naglasiti da se validacija s vremenskim pomicanjem unaprijed i dalje provodi za svaki pojedini model unutar ansambla. Na početku, podaci se dijele na skup za treniranje (80%) i skup za testiranje (20%). Zatim se primjenjuje postupak dan po dan, pri čemu se svi podaci koji prethode danu koji se procjenjuje prosljeđuju modelima. Ovaj podskup podataka dodatno se dijeli na skupove za treniranje (80%) i validaciju (20%). Ovaj pristup je presudan zbog optimizacijskog algoritma, koji zahtijeva zaseban podskup na kojemu će provjeriti kvalitetu rješenja, a taj podskup je onih 20% validacije što se podijelio unutar svakog zasebnog modela. Na taj način, svaki model generira skup procjena koje služe kao ulazni podaci za optimizacijski algoritam. Uz te procjene, dostupne su i stvarne razlike cijena, što omogućava vrednovanje rješenja tijekom treniranja. Na kraju postupka primjenjuje se metoda težinskog prosjeka, prikazana u formuli 4.1, kojom se kombiniraju predikcije svih modela u ansamblu. Cjelokupni proces detaljno je prikazan na slici 5.3.

Zbog ograničenih resursa bilo je nužno skratiti korak prikazan na slici 5.2 i istrenirati modele samo jednom na početku, koristeći povijesne podatke. Nakon toga, prethodno istrenirani modeli primali su podatke za svaki novi dan i na temelju njih generirali svoje procjene. Iako ovaj pristup uzrokuje određeni gubitak u preciznosti u usporedbi s prvotnim rješenjem, dvostruka validacija s vremenskim pomicanjem unaprijed, uz ponovno treniranje modela u svakom koraku, zahtijevala bi previše vremena i resursa za provedbu, što bi bilo neizvedivo u zadanom okruženju.



Slika 5.3: Način treniranja i vrednovanja ansambla

## 5.2. Usporedba pojedinih modela

U ovom poglavlju uspoređuju se različiti modeli unutar svake analize s ciljem identifikacije najboljih rješenja. Kao mjerilo za vrednovanje korišten je RMSE (engl. *root mean squared error*), koji predstavlja korijen srednje kvadratne pogreške između stvarnih i predviđenih vrijednosti. Uz prosječni RMSE, rezultati su prikazani na tri različita grafička prikaza.

Prvi graf prikazuje dvije krivulje: jednu koja prikazuje stvarnu razliku cijene, a drugu koja prikazuje predviđenu razliku cijene. Ostala dva grafa su prikazani u prilogu. Drugi graf prikazuje vrijednost RMSE za svaki dan koji se predviđa, što omogućuje uvid u odstupanje predviđenih vrijednosti od stvarnih. Veće odstupanje od nule ukazuje na veću pogrešku modela. Treći graf uspoređuje stvarnu cijenu s predviđenom cijenom.

S obzirom na to da se koristi vrlo kratki interval procjene, samo jedan dan, očekuje se da modeli neće pokazivati značajna odstupanja u procjenama u odnosu na trenutni dan. Stoga su grafovi koji prikazuju RMSE i razliku cijena značajniji i prikladniji za analizu, budući da će većina modela pretpostaviti da cijena neće značajno varirati u tako kratkom vremenskom okviru. Recimo da je trenutna cijena dionice 100 dolara, a sutradan cijena poraste na 102 dolara. Zatim, model procijeni da će cijena porasti za samo 0.5 dolara, graf s prikazom stvarnih cijena može sugerirati da je model pružio dobru procjenu, budući da se razlika između 102 i 100.5 dolara neće značajno primijetiti golim okom. Međutim, u stvarnosti, model je promašio za 1.5 dolara, što predstavlja značajnu pogrešku. Ova razlika bit će jasnije prikazana na grafu koji prikazuje samo razlike cijena, dok će graf RMSE pružiti precizniju sliku o ukupnoj pogrešci modela.

### 5.2.1. Usporedba modela za analizu vremenskih serija

Rezultati modela ARIMA, ARIMA-GARCH i ES uspoređeni su za sve tri promatrane dionice. Model ARIMA-GARCH pokazao se najpreciznijim, što je vidljivo iz tablice 5.7, u kojoj su prikazani prosječni RMSE vrijednosti za svaki model. Iako ARIMA i ARIMA-GARCH modeli ostvaruju vrlo slične rezultate, ARIMA-GARCH model ima blagu prednost zahvaljujući GARCH komponenti koja uključuje volatilitnost kretanja cijena. Ova usporedba jasno ukazuje da je ARIMA-GARCH model najprikladniji za analizu vremenskih serija te će biti korišten u daljnjoj izradi ansambla modela.

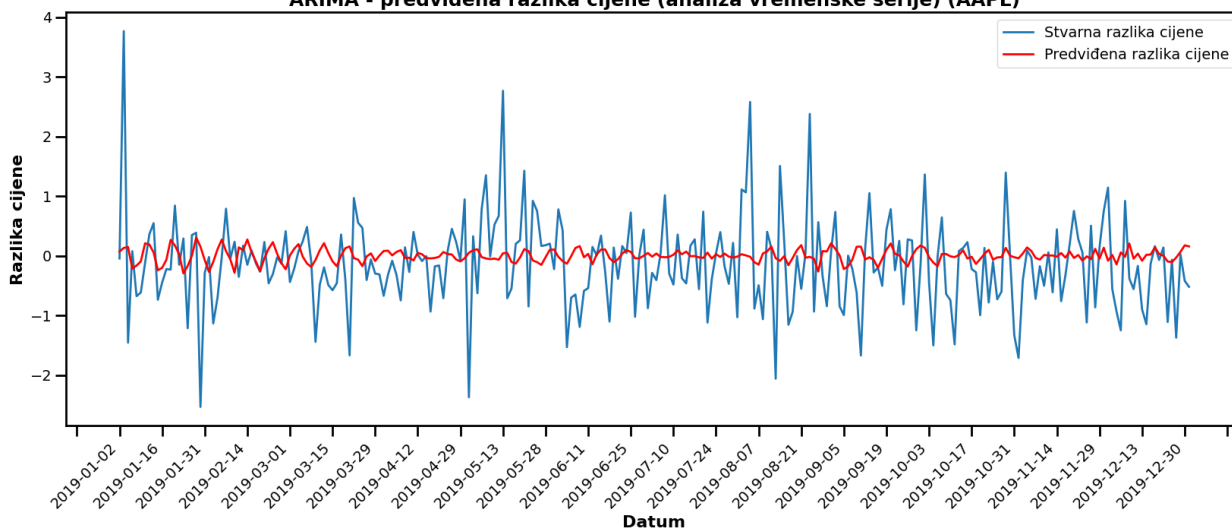
Tablica 5.7: Prosječni RMSE za analizu vremenskih serija

Model	Prosječni RMSE
<b>Apple</b>	
ARIMA	0.78890
<b>ARIMA-GARCH</b>	<b>0.78248</b>
ES	0.83767
<b>Amazon</b>	
ARIMA	1.29427
<b>ARIMA-GARCH</b>	<b>1.29216</b>
ES	1.35879
<b>Google</b>	
ARIMA	0.88512
<b>ARIMA-GARCH</b>	<b>0.88201</b>
ES	0.93126

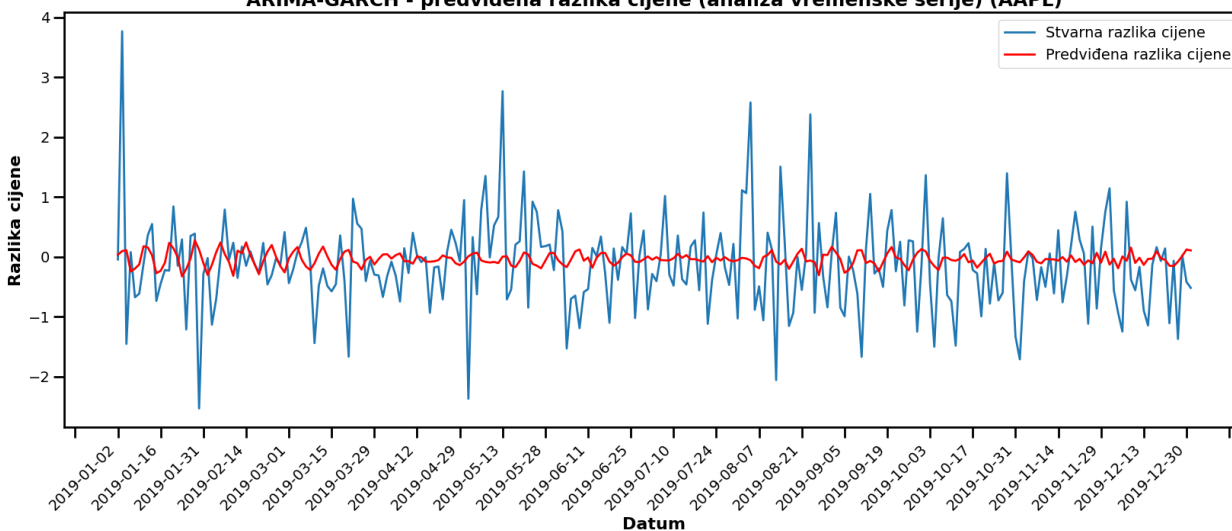
Rezultati za tvrtku Apple prikazani su u slici 5.4, za Amazon u slici 5.5, dok su rezultati za Google prikazani u slici 5.6. Obrasci ponašanja modela za svaku dionicu su vrlo slični. Model ES pokazuje veću varijabilnost u svojim procjenama, dok su ARIMA i ARIMA-GARCH modeli stabilniji, s predviđenim razlikama cijena koje ostaju unutar intervala  $[-1, 1]$ . Ipak, ARIMA-GARCH model pokazuje blagu prednost u točnosti predviđanja u usporedbi s ARIMA modelom. Detaljniji prikazi rezultata mogu se pronaći u prilogima P.1, P.2 i P.3.

# Apple

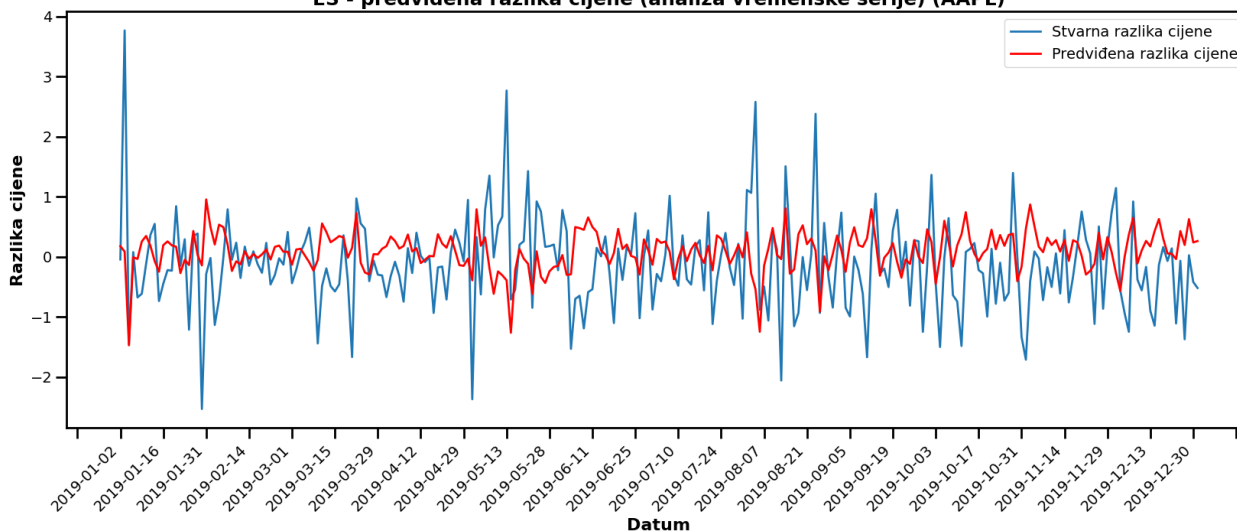
ARIMA - predviđena razlika cijene (analiza vremenske serije) (AAPL)



ARIMA-GARCH - predviđena razlika cijene (analiza vremenske serije) (AAPL)



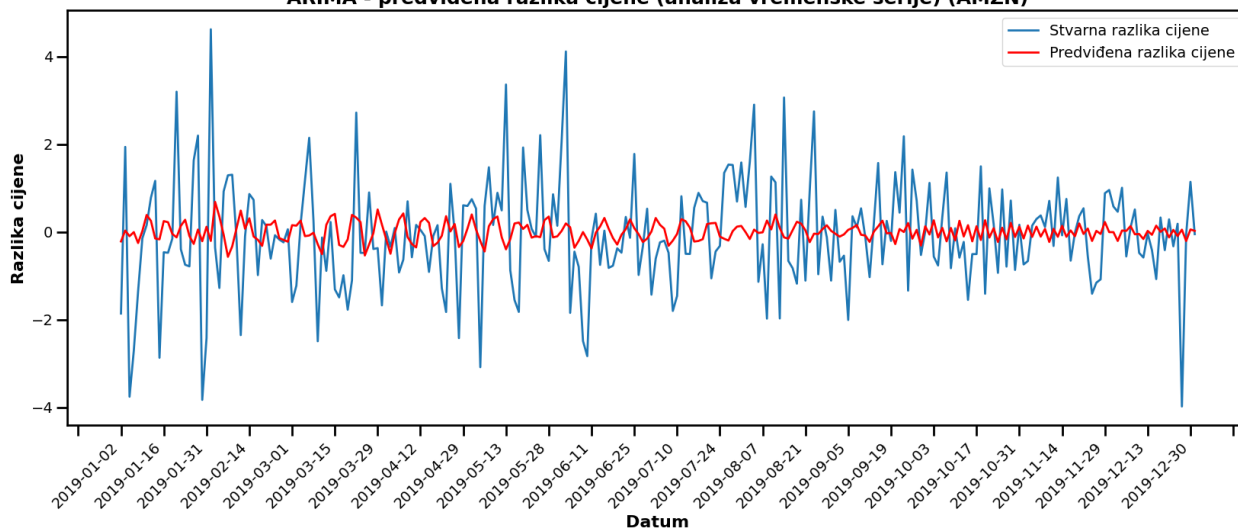
ES - predviđena razlika cijene (analiza vremenske serije) (AAPL)



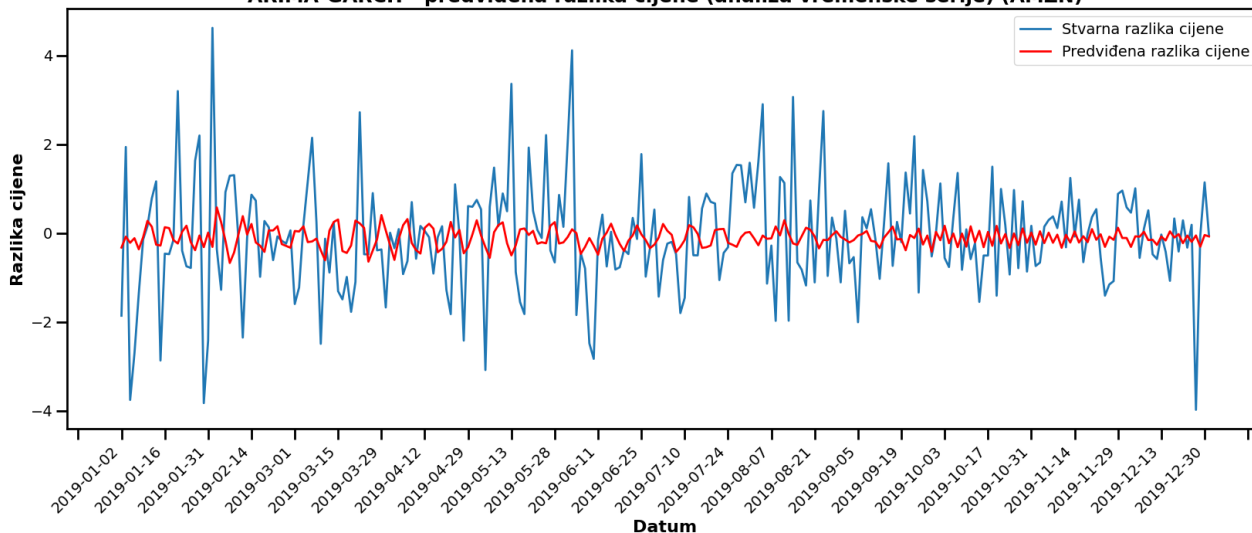
Slika 5.4: Predviđena razlika cijene modela vremenskih serija za tvrtku Apple

## Amazon

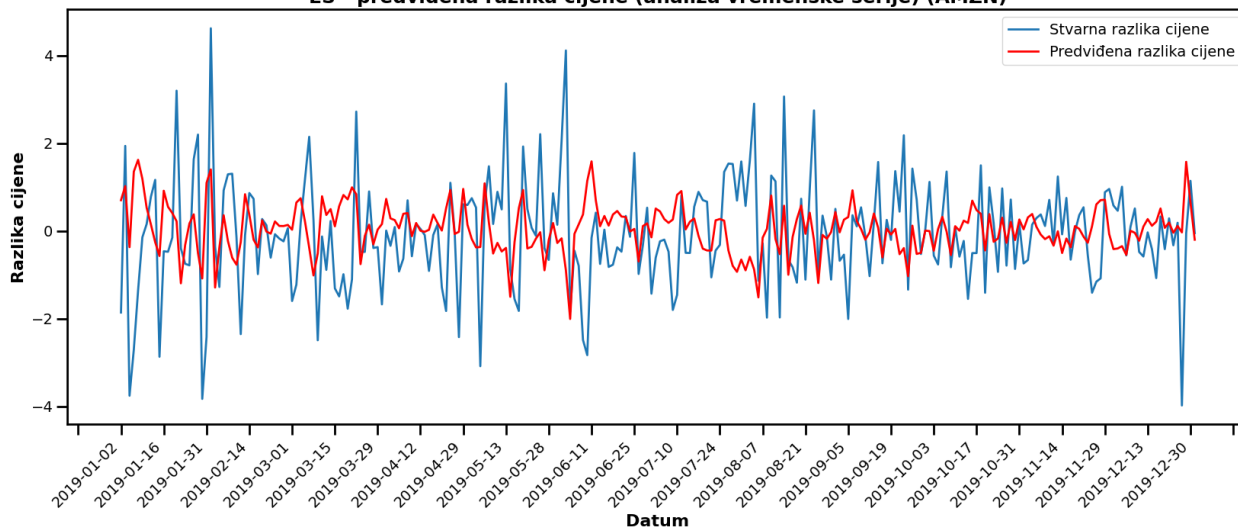
ARIMA - predviđena razlika cijene (analiza vremenske serije) (AMZN)



ARIMA-GARCH - predviđena razlika cijene (analiza vremenske serije) (AMZN)



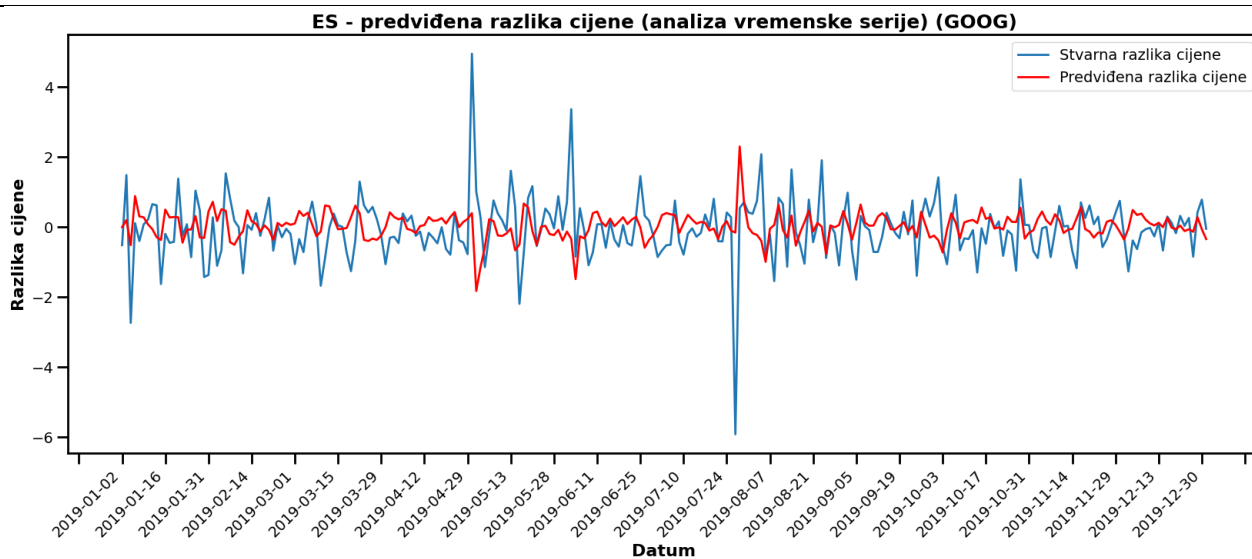
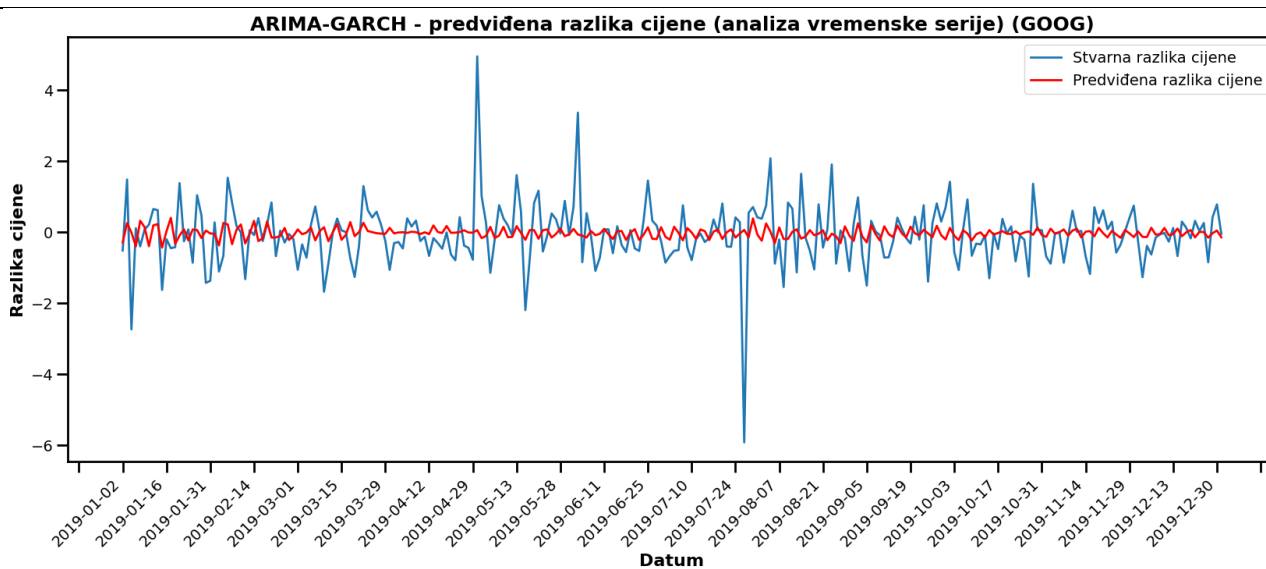
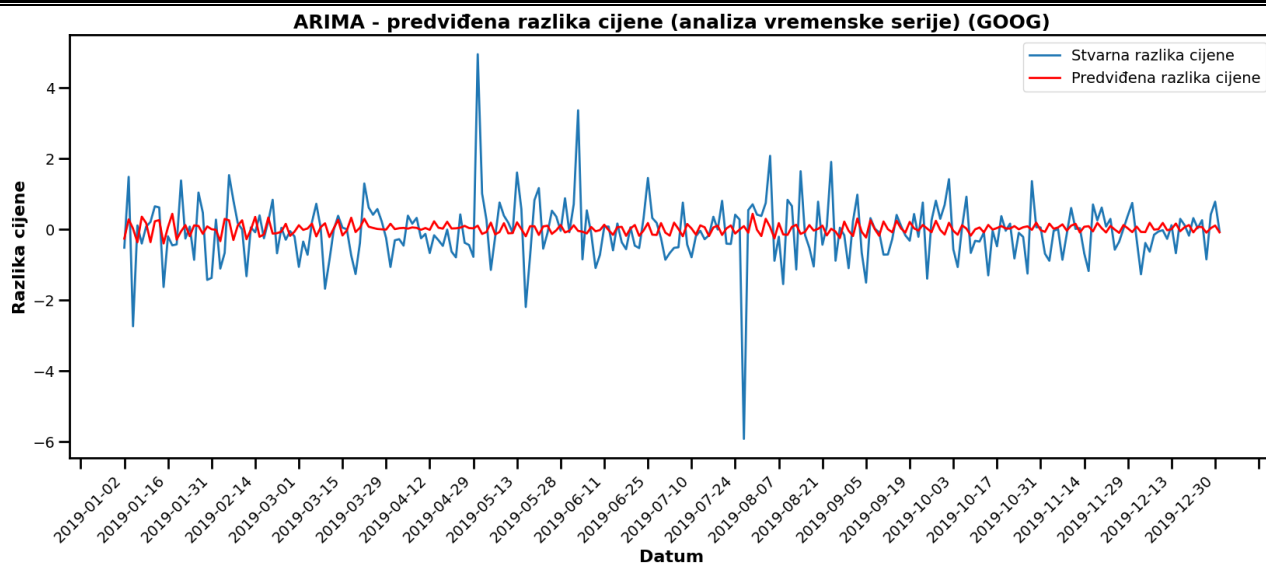
ES - predviđena razlika cijene (analiza vremenske serije) (AMZN)



Slika 5.5: Predviđena razlika cijene modela vremenskih serija za tvrtku Amazon



## Google



Slika 5.6: Predviđena razlika cijene modela vremenskih serija za tvrtku Google

### 5.2.2. Usporedba modela za tehničku analizu

Nakon detaljne usporedbe rezultata SVM, RF, ANN i LSTM modela, odlučeno je da je RF model najpouzdaniji, što je jasno vidljivo iz prikaza u tablici 5.8. Ovaj model je postigao konzistentno bolje rezultate u usporedbi s ostalim modelima, posebno u smislu preciznosti i robusnosti predviđanja. Dok su SVM i ANN modeli pokazali solidne performanse, a LSTM model zabilježio određenu prednost u prepoznavanju složenih obrazaca, RF model se istaknuo zbog svoje jednostavnosti i stabilnosti, te manjeg osciliranja u rezultatima.

S obzirom na pouzdanost i jednostavnost interpretacije, RF model će biti ključni element u tehničkoj analizi te će se koristiti u izradi ansambla modela, gdje se očekuje da će dodatno ojačati ukupne predikcijske sposobnosti sustava.

Tablica 5.8: Prosječni RMSE modela za tehničku analizu

Model	Prosječni RMSE
<b>Apple</b>	
SVM	1.11872
<b>RF</b>	<b>0.80558</b>
ANN	1.02406
LSTM	2.27696
<b>Amazon</b>	
SVM	1.76985
<b>RF</b>	<b>1.31087</b>
ANN	1.58061
LSTM	4.37899
<b>Google</b>	
SVM	1.14879
<b>RF</b>	<b>0.97835</b>
ANN	1.23394
LSTM	2.04196

U slikama 5.7, 5.8 i 5.9 prikazani su rezultati modela za tehničku analizu putem grafova. Na temelju tih grafova može se zaključiti da većina modela, osim RF modela, pokazuje značajnu varijabilnost u svojim procjenama ili određeni pomak koji rezultira većom pogreškom. Modeli poput SVM, ANN i LSTM imaju tendenciju većim oscilacijama u predviđanjima, što može otežati precizno predviđanje kretanja cijena.

S druge strane, RF model se ističe svojim stabilnijim procjenama, gdje se predviđene vrijednosti zadržavaju unutar užeg intervala. Ova karakteristika omogućuje manju pogrešku i veću pouzdanost u predviđanjima. Upravo zbog toga, RF model nudi konzistentnija i sigurnija predviđanja, što ga čini primjerenijim za tehničku analizu u kontekstu ovih podataka. Detaljniji prikazi rezultata mogu se pronaći u prilogima P.4, P.5 i P.6.

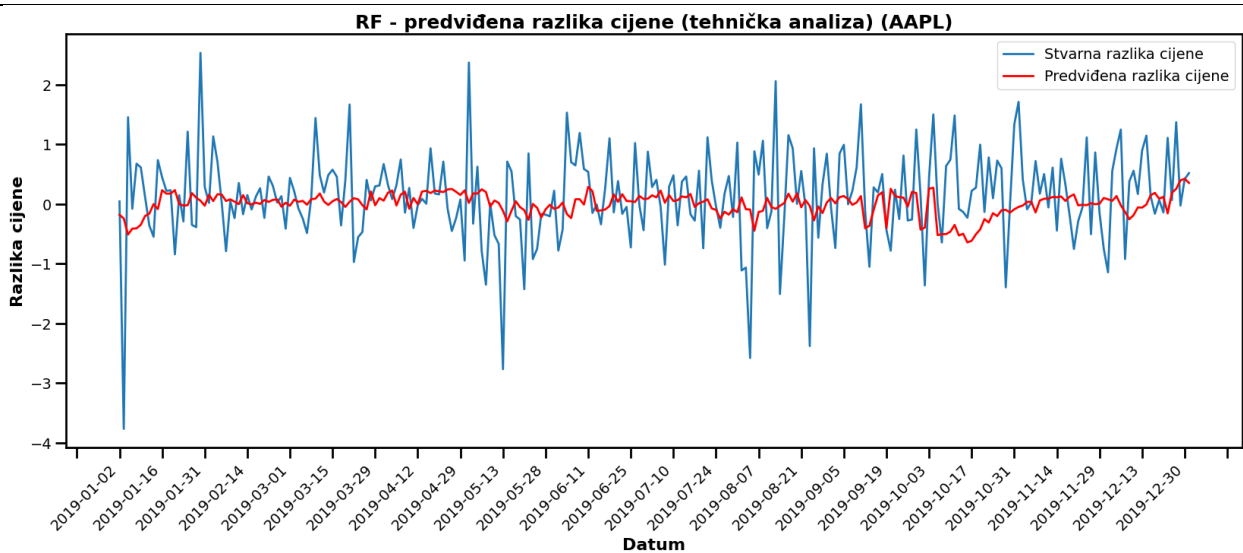
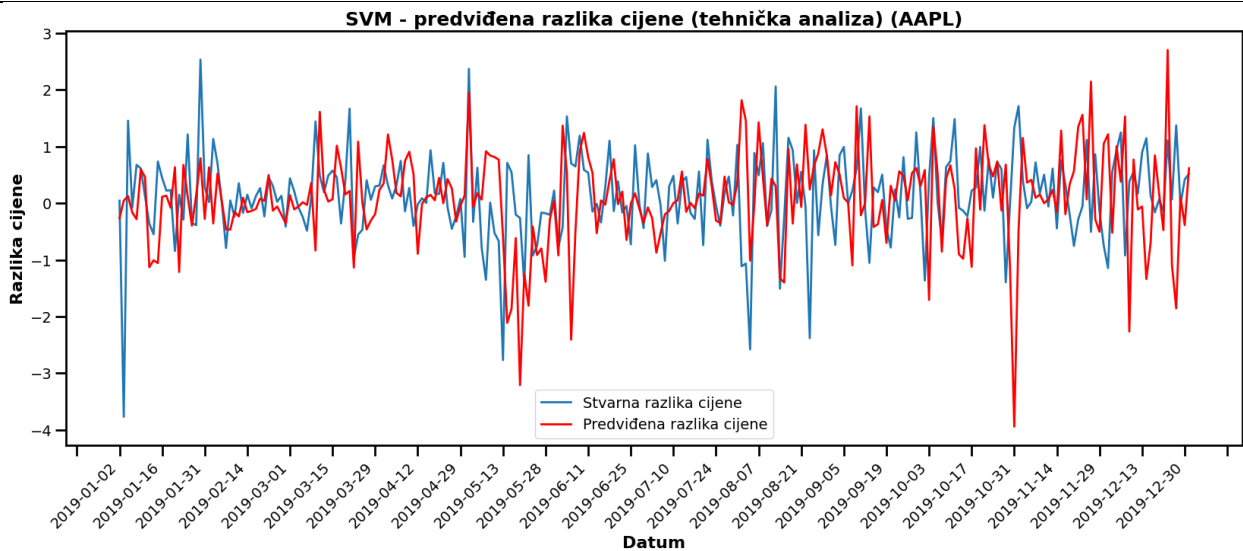
---

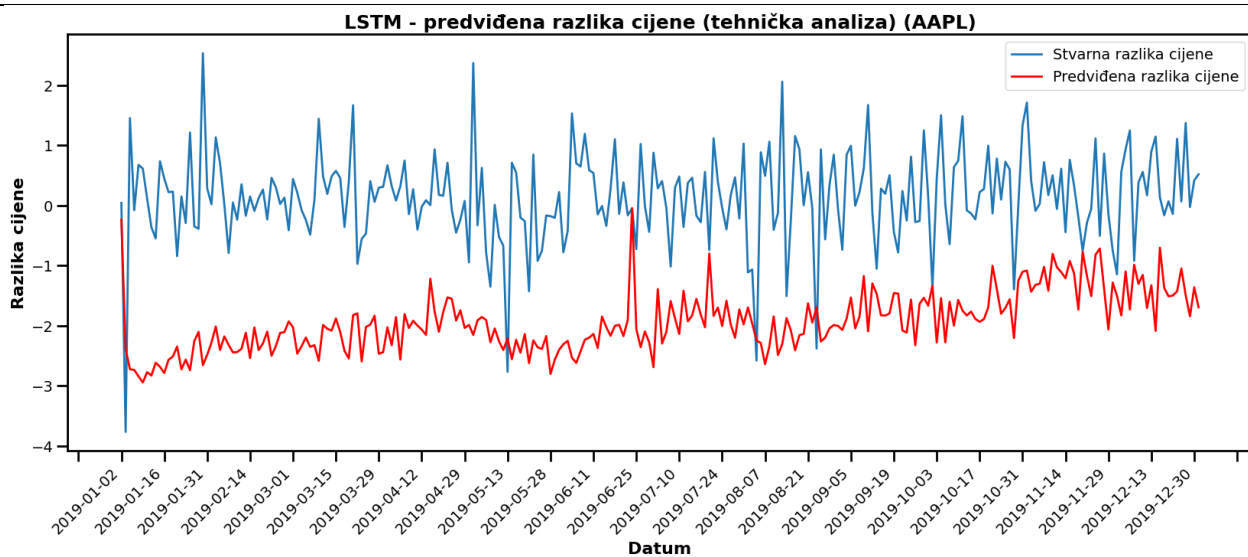
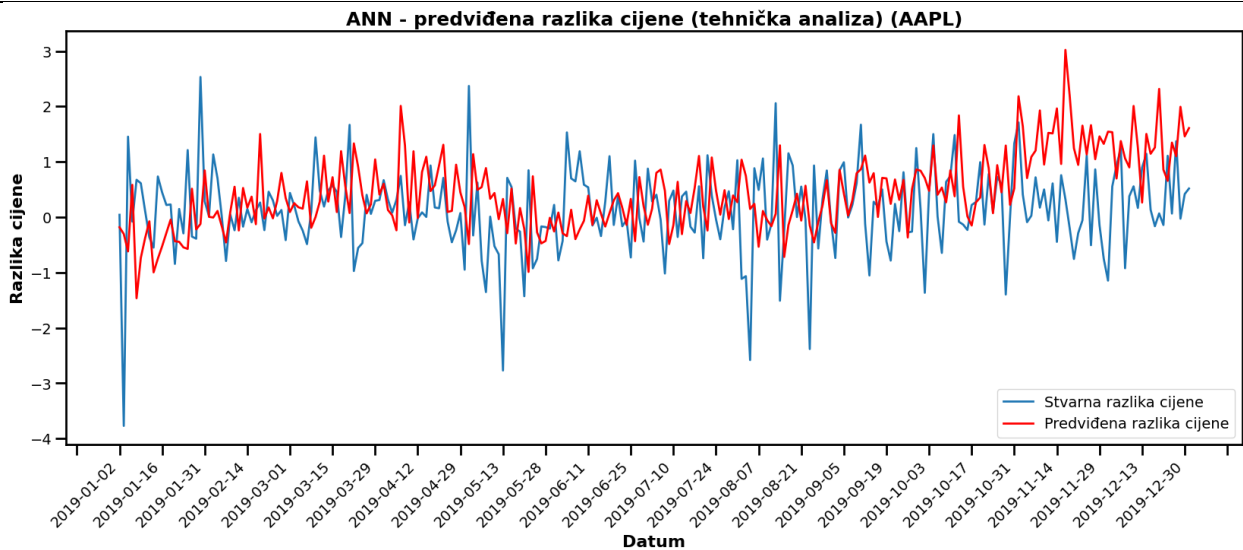
---

## Apple

---

---

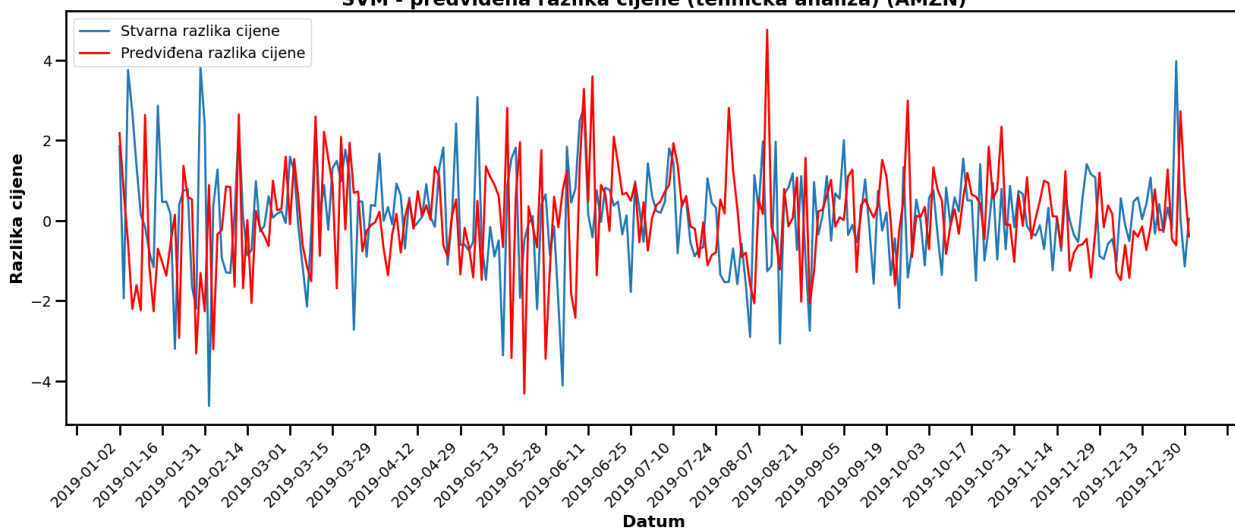




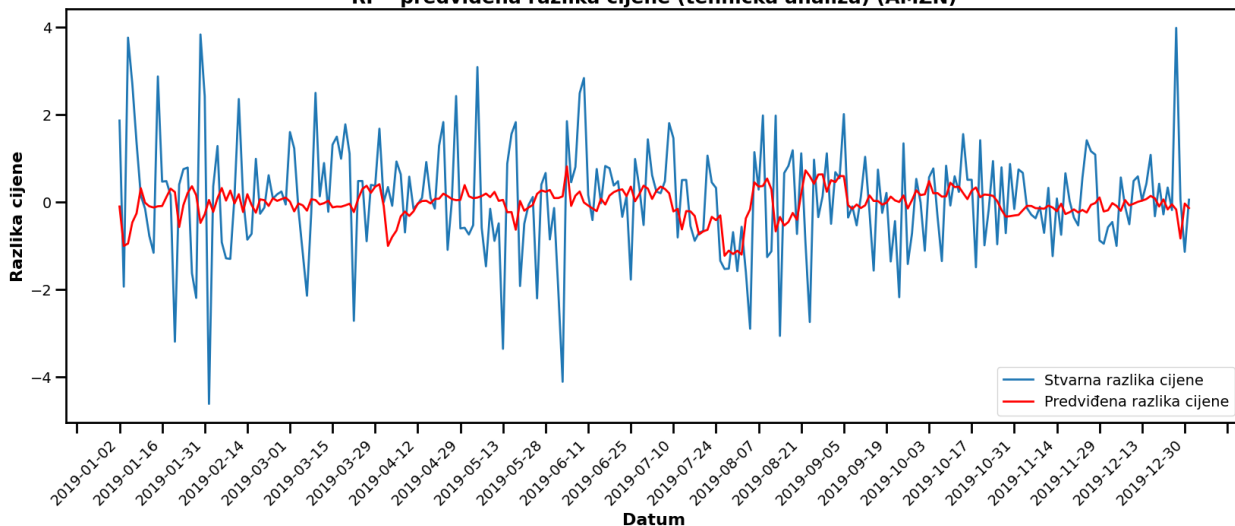
Slika 5.7: Predviđena razlika cijene modela za tehničku analizu za tvrtku Apple

# Amazon

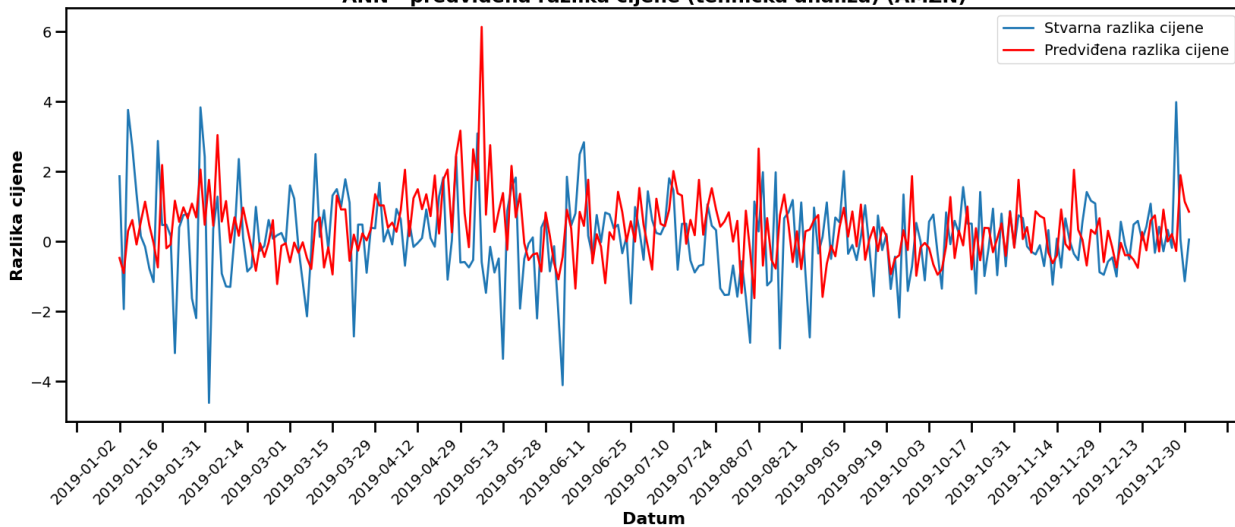
## SVM - predviđena razlika cijene (tehnička analiza) (AMZN)

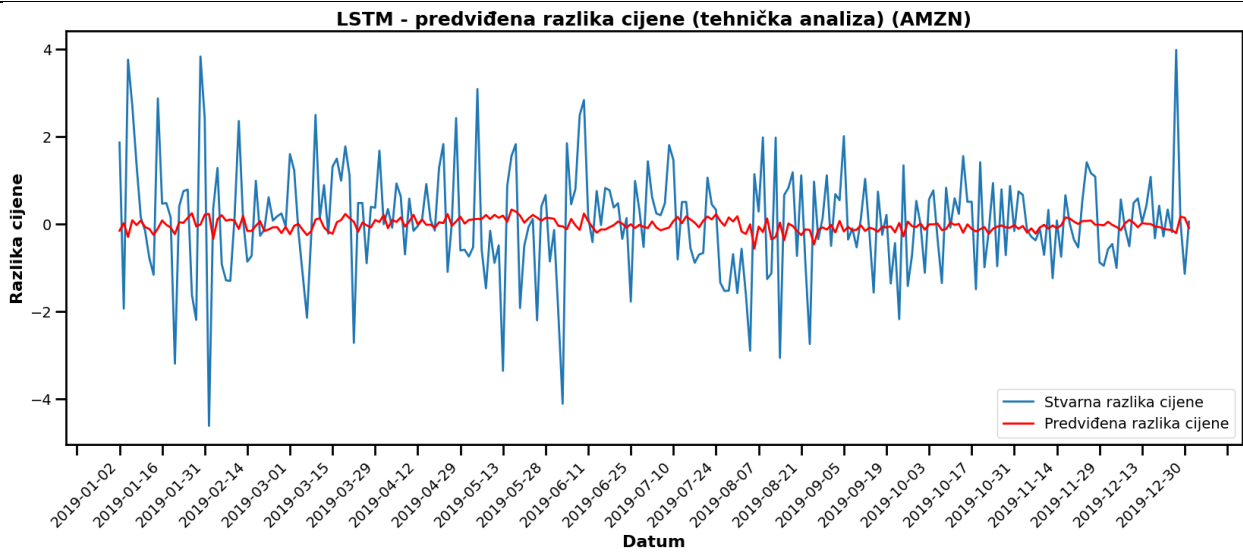


## RF - predviđena razlika cijene (tehnička analiza) (AMZN)



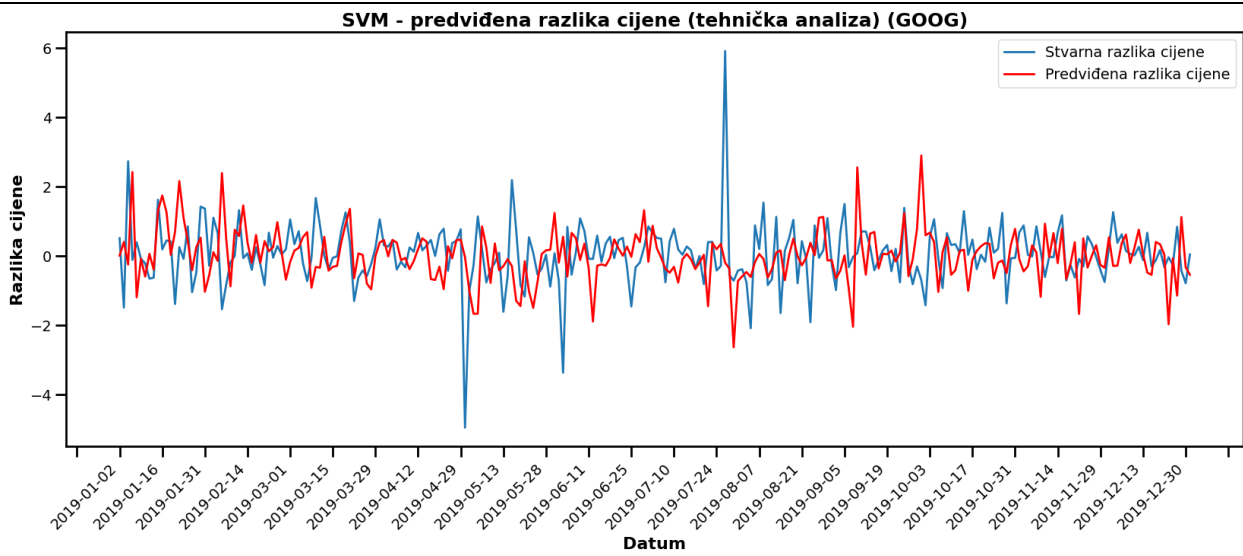
## ANN - predviđena razlika cijene (tehnička analiza) (AMZN)

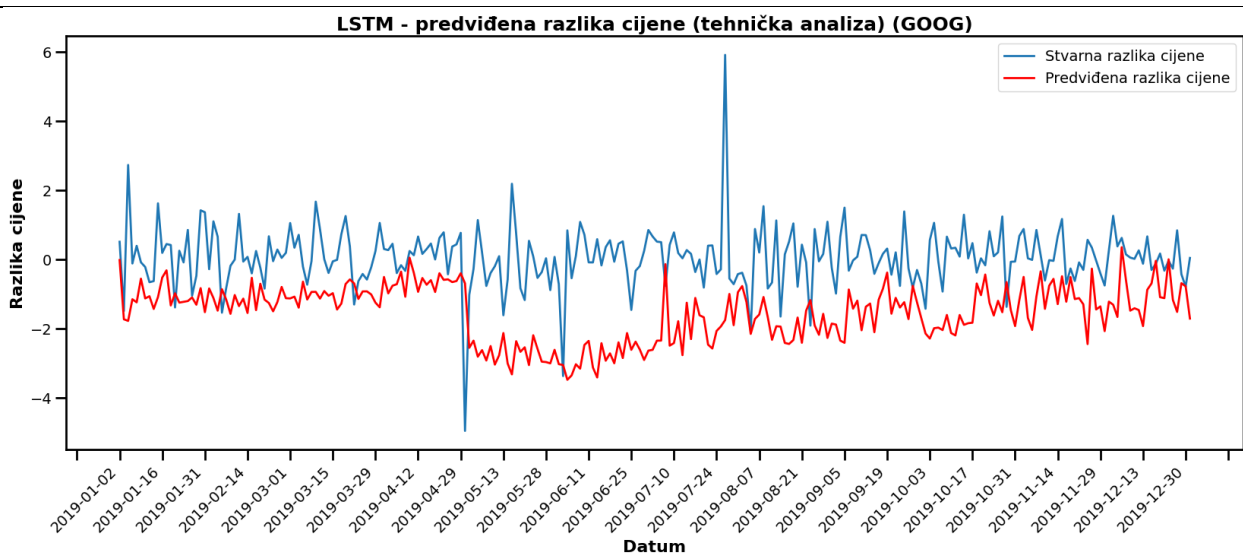
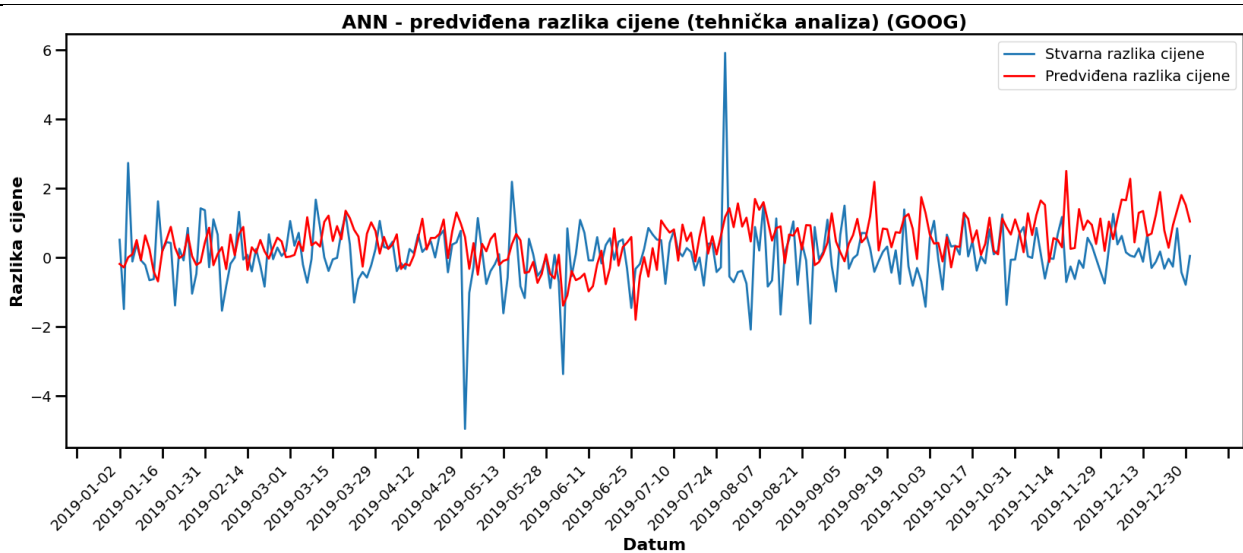
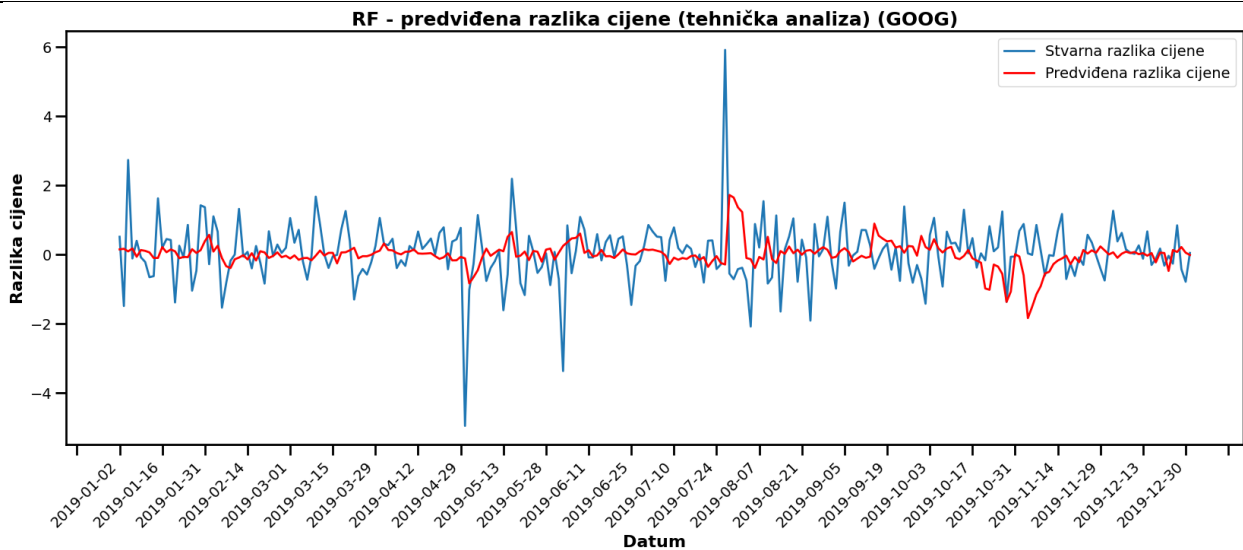




Slika 5.8: Predviđena razlika cijene modela za tehničku analizu za tvrtku Amazon

## Google





Slika 5.9: Predviđena razlika cijene modela za tehničku analizu za tvrtku Google

### 5.2.3. Usporedba modela za fundamentalnu analizu

U fundamentalnoj analizi, koja je koristila modele SVM, RF, ANN i LSTM, ponovno je RF model istaknut kao najpouzdaniji, što je jasno prikazano u tablici 5.9. Ovaj rezultat je u skladu s prethodnim zaključcima iz tehničke analize, gdje je RF model također postigao najbolje rezultate zahvaljujući svojoj stabilnosti i smanjenoj varijanci u procjenama. Dok su modeli SVM, ANN i LSTM pokazali određene prednosti u prepoznavanju kompleksnih obrazaca, RF model se iznova pokazao najučinkovitijim zbog svoje jednostavnosti i konzistentnosti u predviđanjima. Zbog pouzdanih rezultata koje RF model nudi, bit će uključen u ansambl i za fundamentalnu analizu.

Tablica 5.9: Prosječni RMSE modela za fundamentalnu analizu

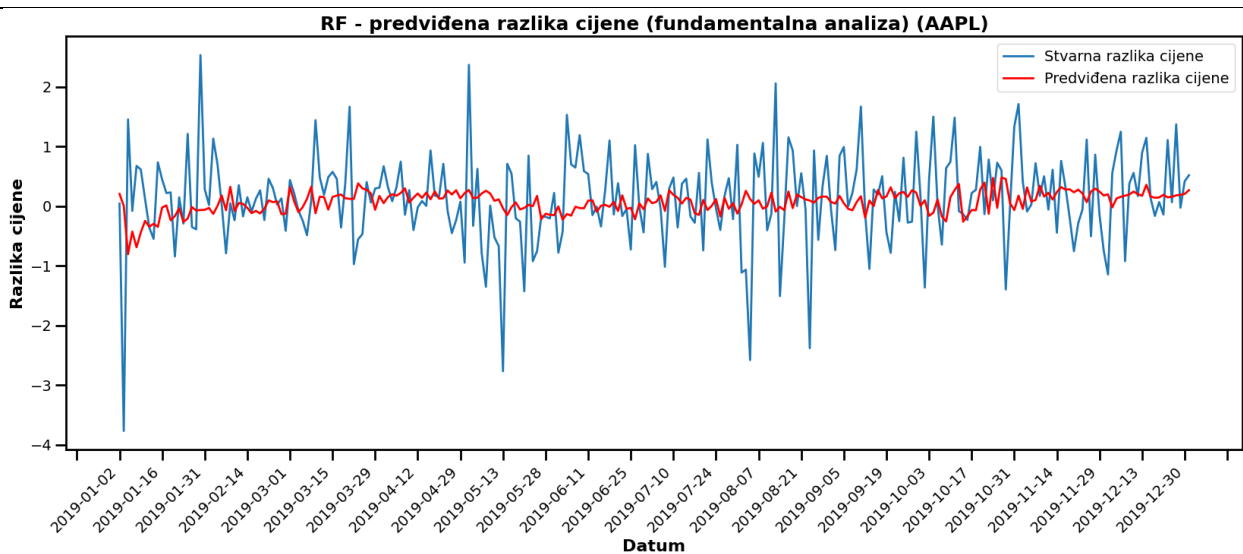
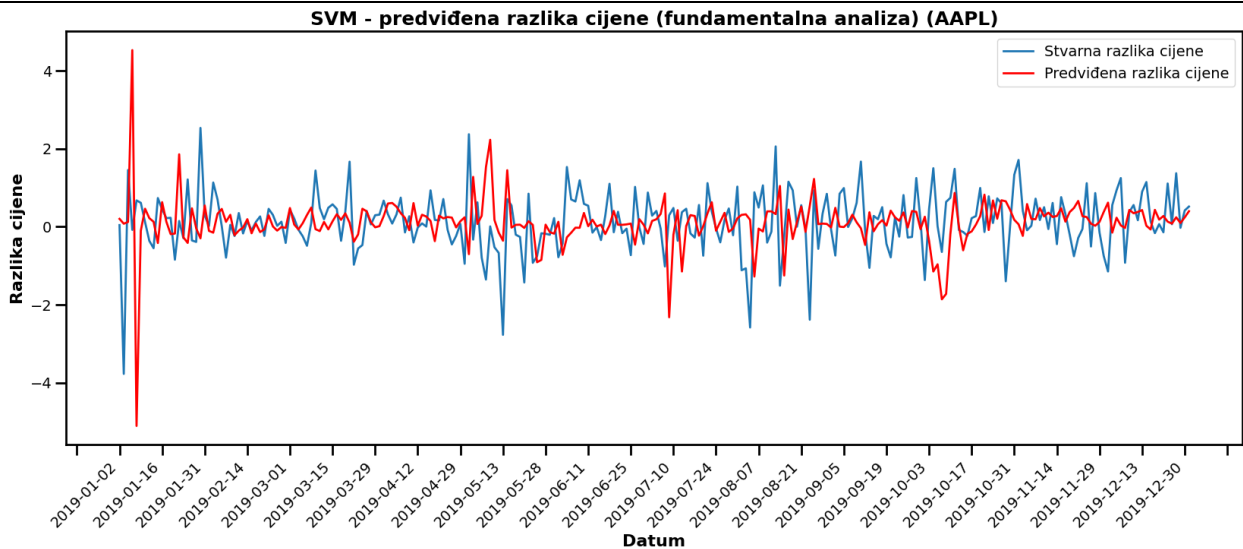
Model	Prosječni RMSE
<b>Apple</b>	
SVM	1.05316
<b>RF</b>	<b>0.80857</b>
ANN	1.36720
LSTM	2.01467
<b>Amazon</b>	
SVM	1.53394
<b>RF</b>	<b>1.27951</b>
ANN	3.06579
LSTM	1.55578
<b>Google</b>	
SVM	0.97554
<b>RF</b>	<b>0.91185</b>
ANN	1.13762
LSTM	1.12816

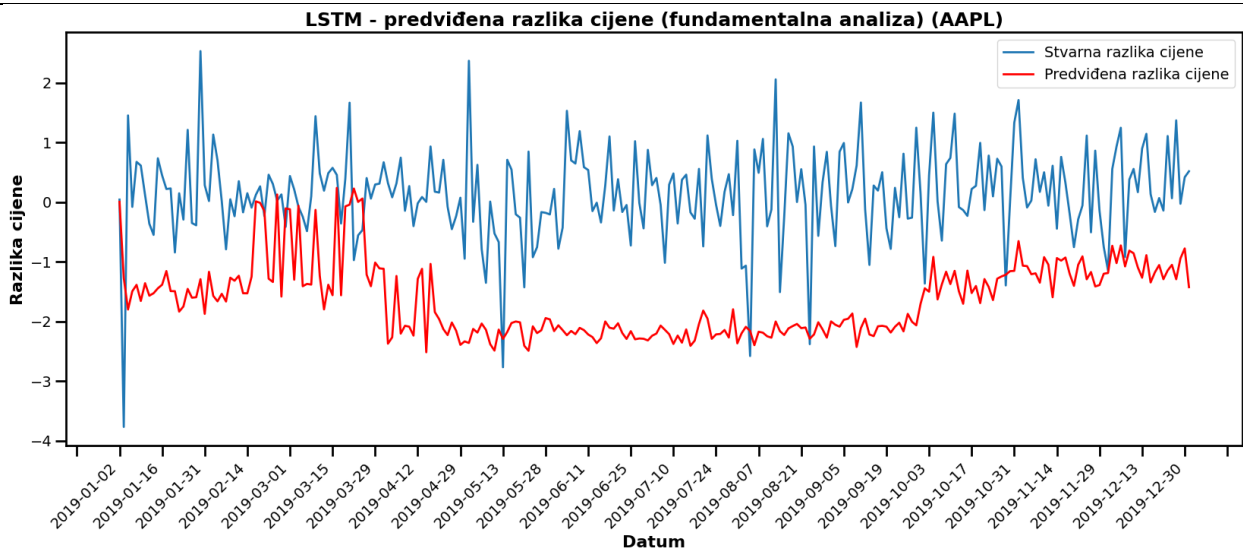
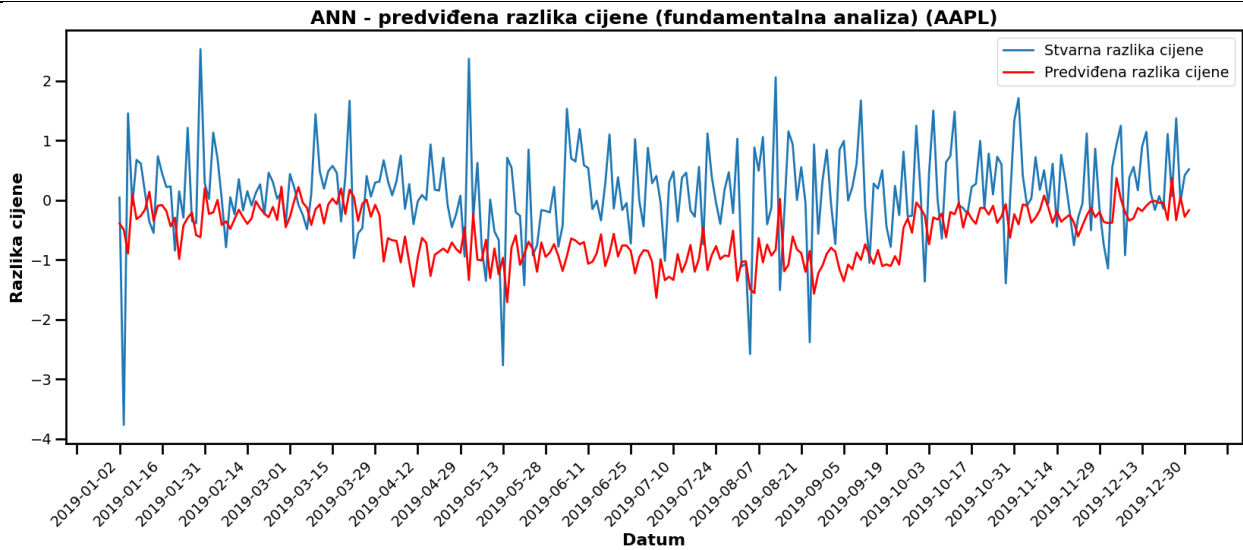
U slikama 5.10, 5.11 i 5.12 prikazani su rezultati za fundamentalnu analizu putem grafova, gdje se mogu uočiti slični obrasci kao u tehničkoj analizi, barem u smislu ponašanja pojedinih modela. RF model se ponovno ističe svojom stabilnošću i manjom varijacijom u procjenama, dok ostali modeli, poput SVM-a, ANN-a i LSTM-a, pokazuju veće oscilacije u predviđanjima. Iako svi modeli prate opće trendove, RF model pruža pouzdanije i preciznije rezultate. U nekim slučajevima, modeli ANN i LSTM uspijevaju relativno dobro pratiti stvarni trend cijena, no često



dolazi do pomaka ili kašnjenja u predikcijama, što umanjuje točnost rezultata. Iako na prvi pogled grafički prikazi sugeriraju da modeli pravilno identificiraju trendove i momentume, ova odstupanja stvaraju veće pogreške u konačnim procjenama, što RF modelu ponovno daje prednost u točnosti i pouzdanosti. Detaljniji prikazi rezultata mogu se pronaći u prilogima P.7, P.8 i P.9.

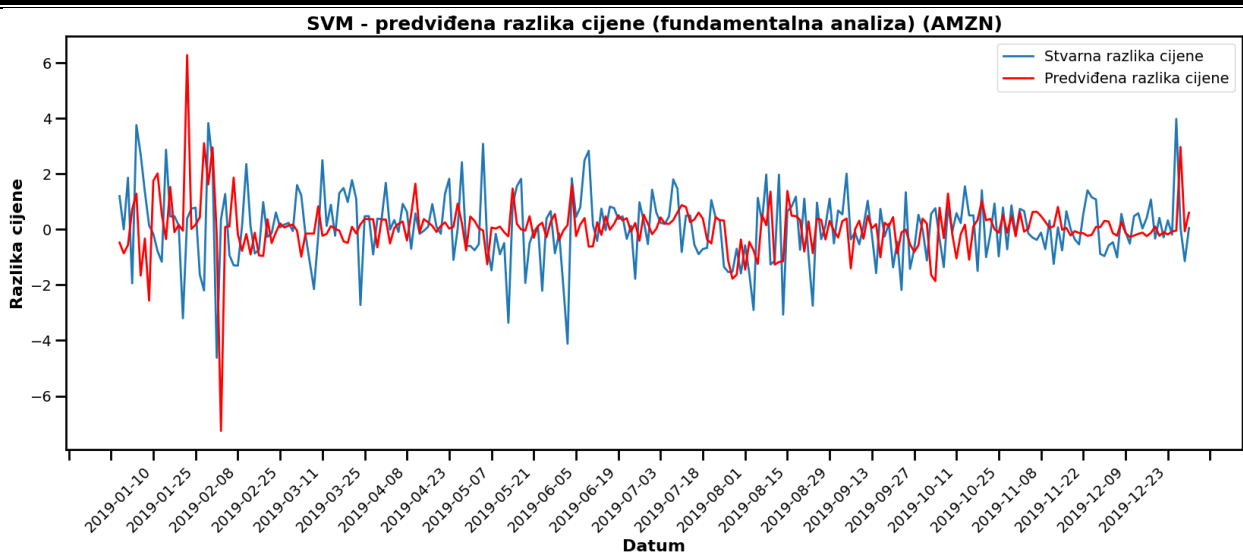
## Apple

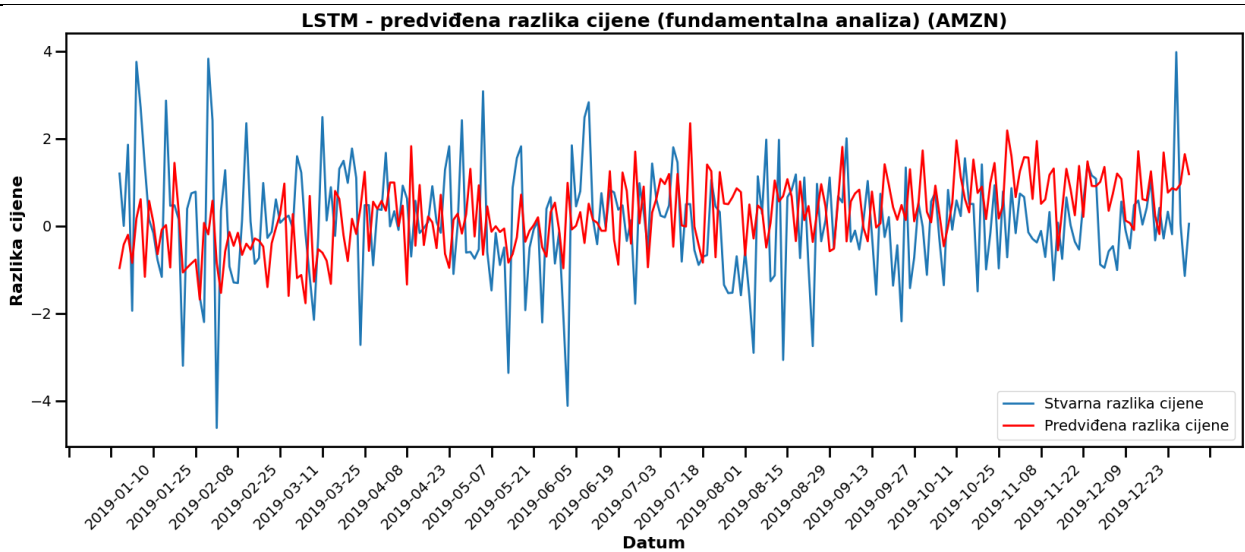
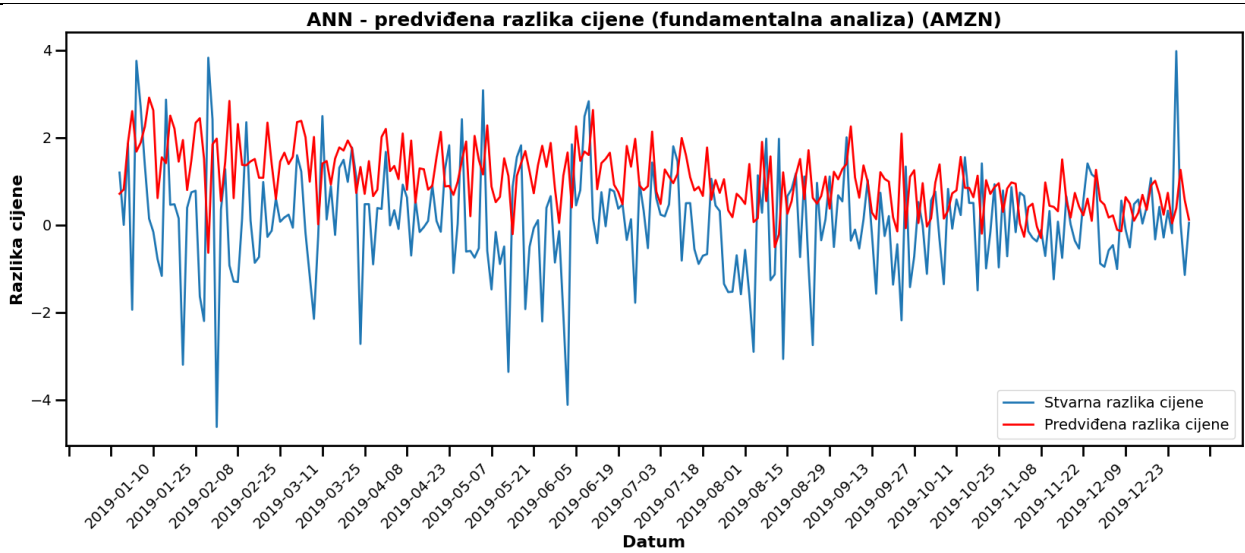
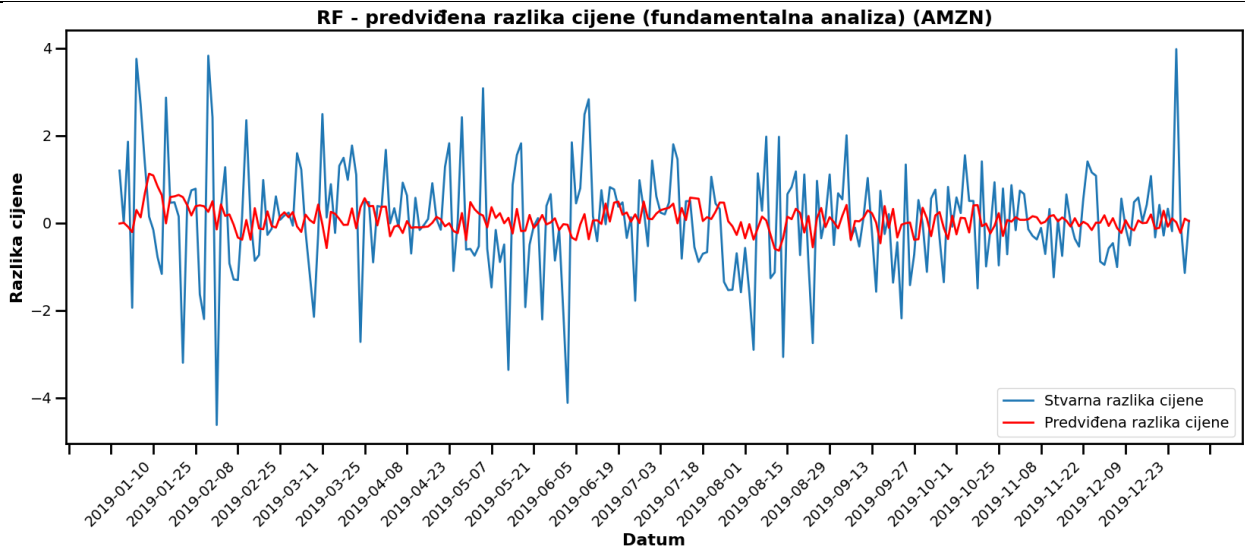




Slika 5.10: Predviđena razlika cijene modela za fundamentalnu analizu za tvrtku Apple

## Amazon

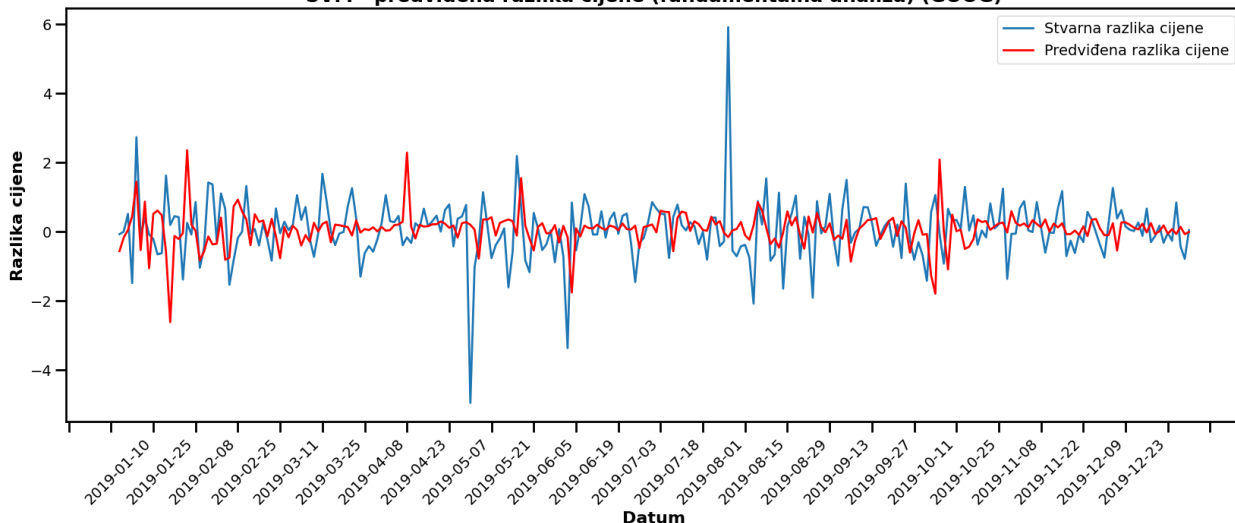




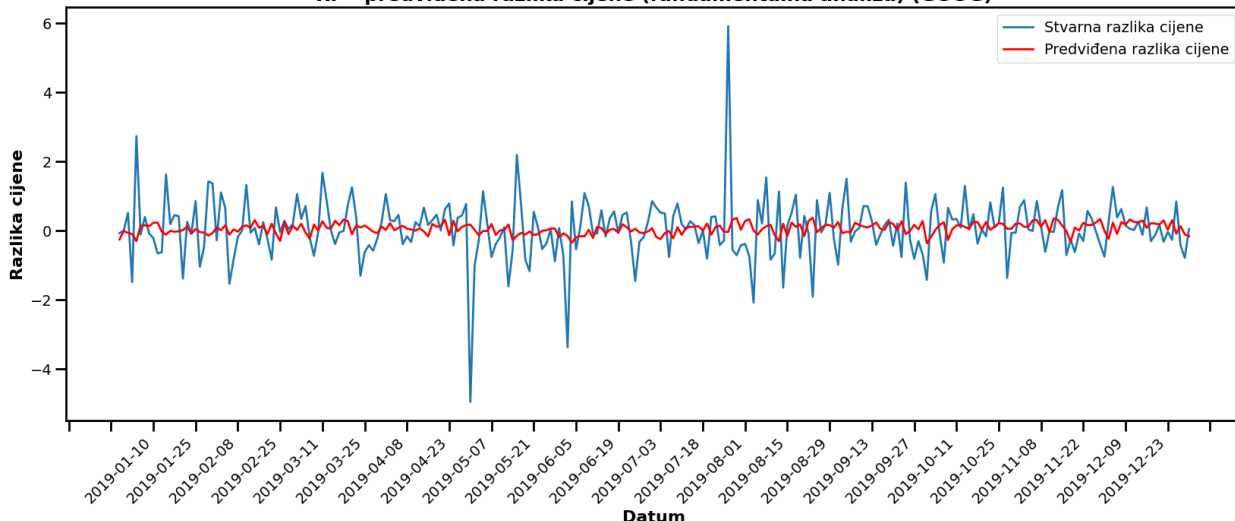
Slika 5.11: Predviđena razlika cijene modela za fundamentalnu analizu za tvrtku Amazon

# Google

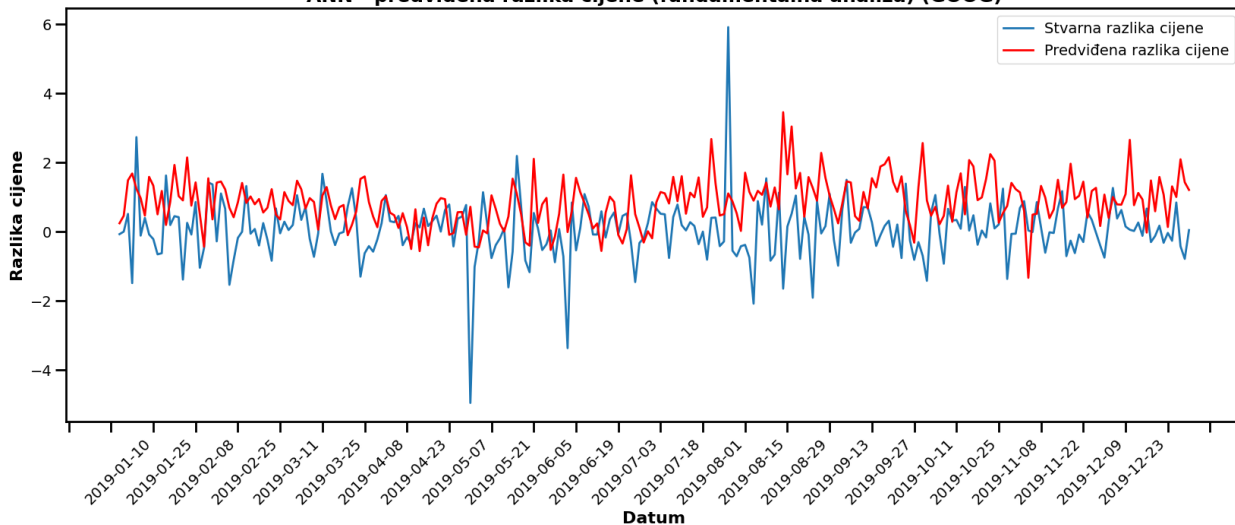
**SVM - predviđena razlika cijene (fundamentalna analiza) (GOOG)**

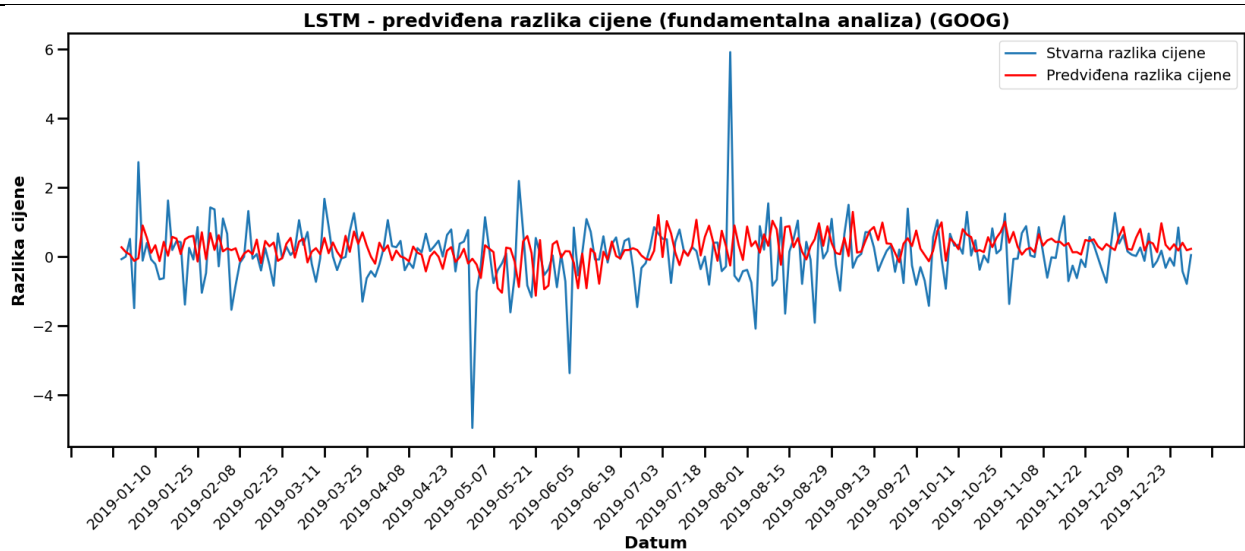


**RF - predviđena razlika cijene (fundamentalna analiza) (GOOG)**



**ANN - predviđena razlika cijene (fundamentalna analiza) (GOOG)**





Slika 5.12: Predviđena razlika cijene modela za fundamentalnu analizu za tvrtku Google

#### 5.2.4. Usporedba modela za analizu sentimenta

Model RF se ponovno istaknuo kao iznimno dobar model za procjenu cijena dionica, što je vidljivo i u analizi sentimenta. Kao i u prethodnim analizama, RF model je pokazao najbolje rezultate, što potvrđuje tablica 5.10, u kojoj ponovno ostvaruje najniži prosječni RMSE. Ova konzistentnost u točnosti predikcija čini RF model najpouzdanijim rješenjem u ovom kontekstu.

Zbog tih izuzetnih rezultata, RF model će biti uključen i u ansambl za analizu sentimenta, čime će dodatno osigurati stabilnost i točnost predikcija u kombinaciji s ostalim modelima. Njegova sposobnost da dosljedno pruža najpreciznije rezultate u različitim vrstama analiza potvrđuje njegovu ključnu ulogu u daljnjoj izradi ansambla modela.

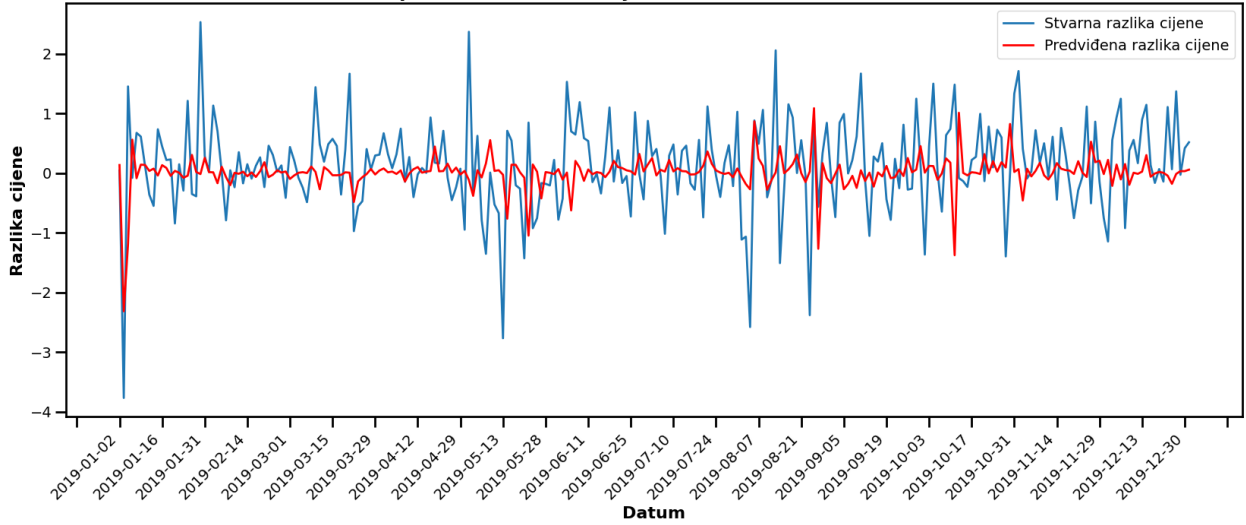
Tablica 5.10: Prosječni RMSE modela za analizu sentimenta

Model	Prosječni RMSE
<b>Apple</b>	
SVM	0.80116
<b>RF</b>	<b>0.78377</b>
ANN	1.01605
LSTM	2.53345
<b>Amazon</b>	
SVM	1.36183
<b>RF</b>	<b>1.2976</b>
ANN	1.80644
LSTM	4.06955
<b>Google</b>	
SVM	0.94361
<b>RF</b>	<b>0.89566</b>
ANN	1.03257
LSTM	2.30668

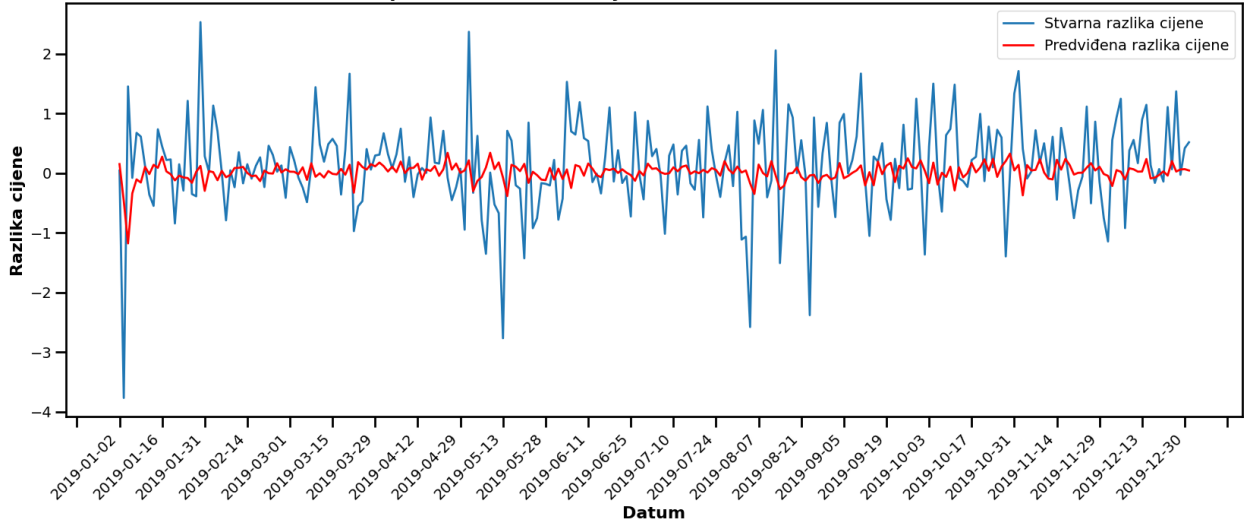
Rezultati prikazani u slikama 5.13, 5.14 i 5.15 još jednom potvrđuju sličan obrazac kao kod tehničke i fundamentalne analize. RF model ponovno se pokazao pouzdanim, s manje riskantnim i blažim procjenama, dok ostali modeli pokazuju oštrije oscilacije ili odstupanja u predviđanjima. Ova stabilnost RF modela omogućuje preciznije i sigurnije procjene, za razliku od drugih modela koji često imaju pomake i kašnjenja u predikcijama. Detaljniji prikazi rezultata mogu se pronaći u priložima P.10, P.11 i P.12.

# Apple

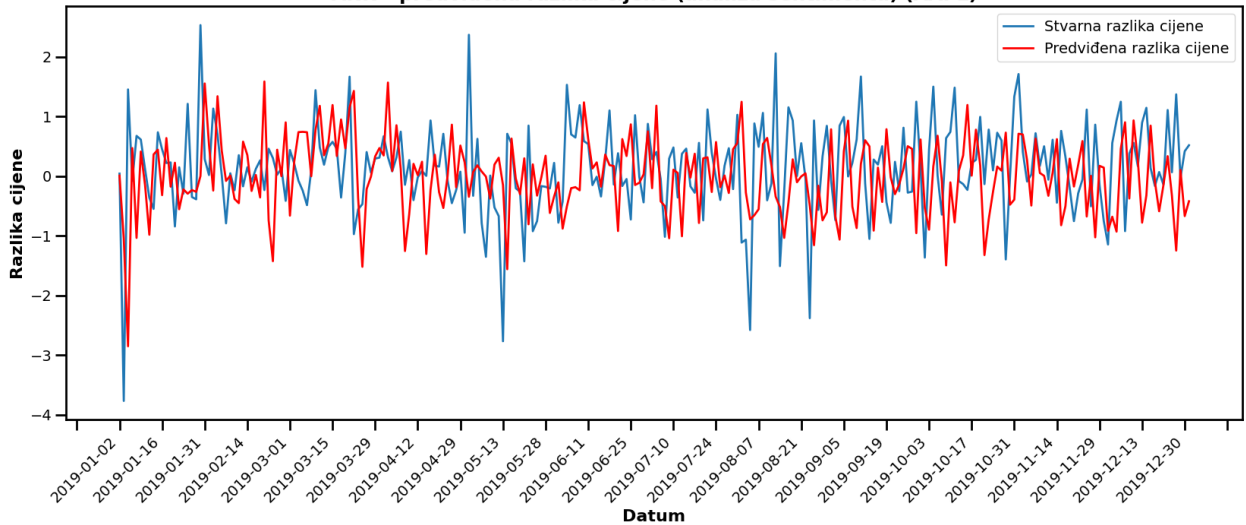
SVM - predviđena razlika cijene (analiza sentimenta) (AAPL)



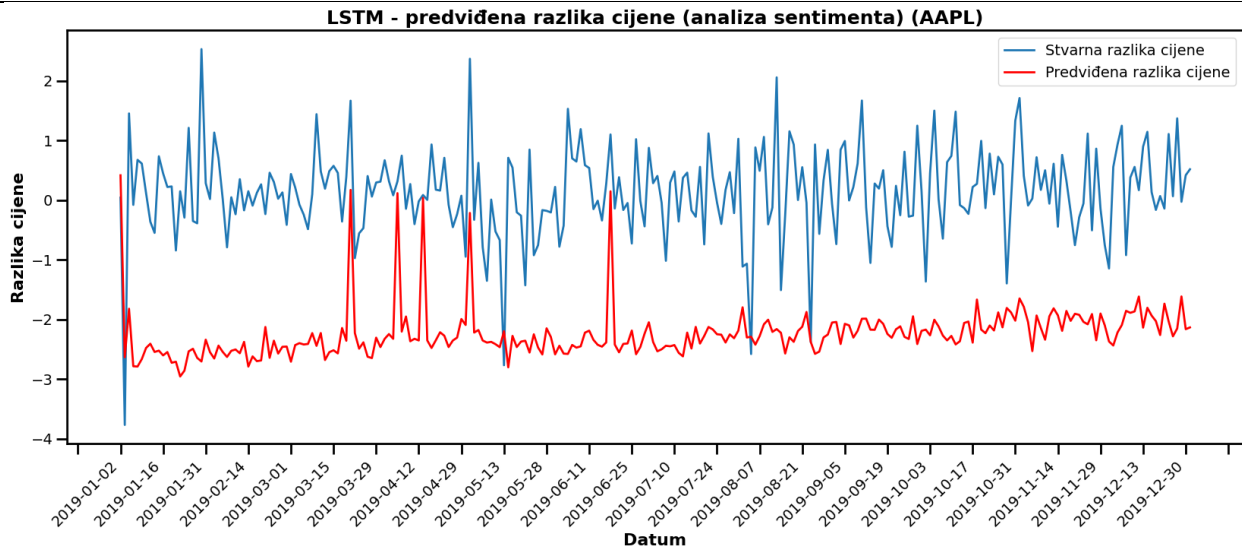
RF - predviđena razlika cijene (analiza sentimenta) (AAPL)



ANN - predviđena razlika cijene (analiza sentimenta) (AAPL)

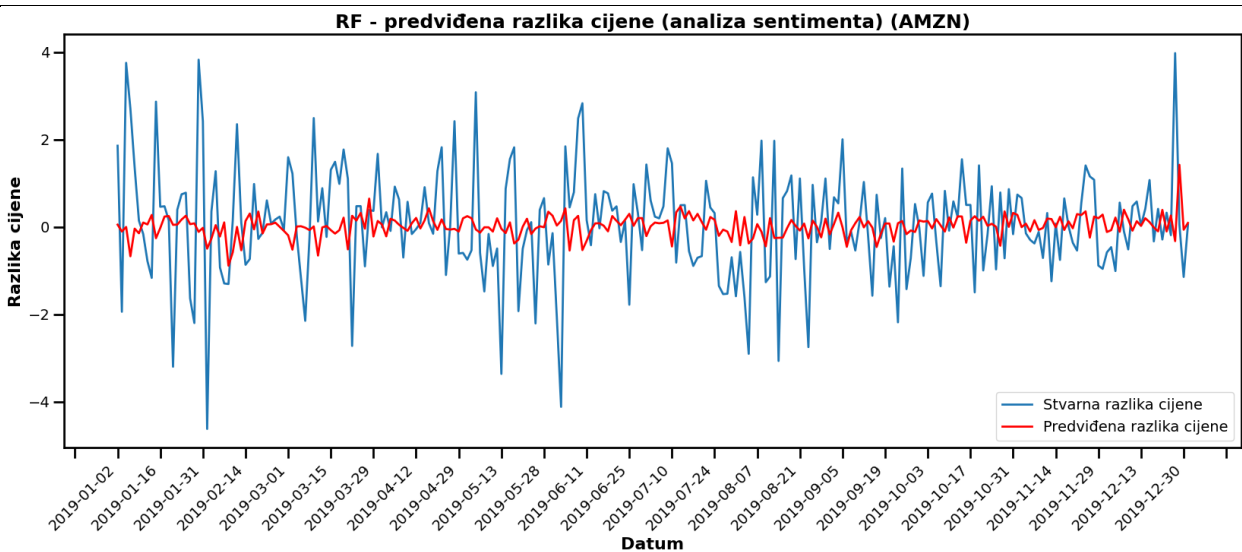
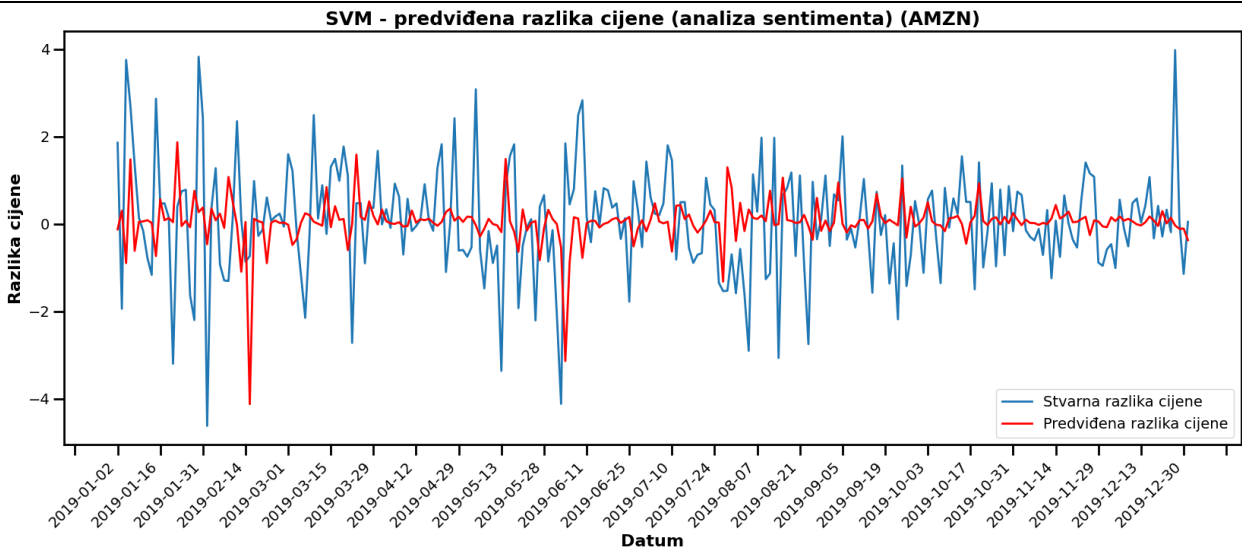




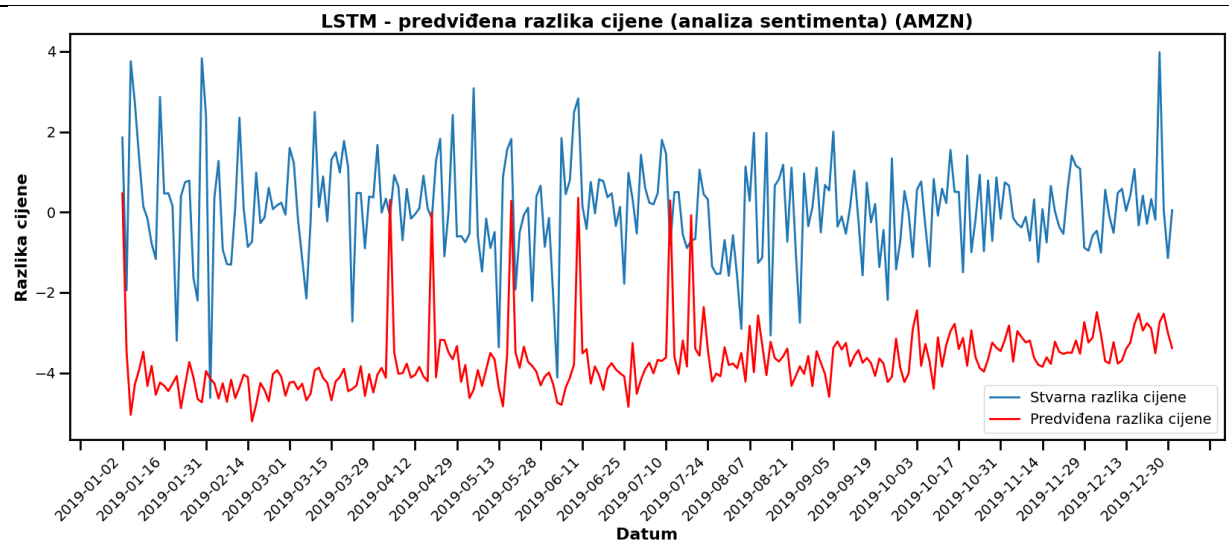
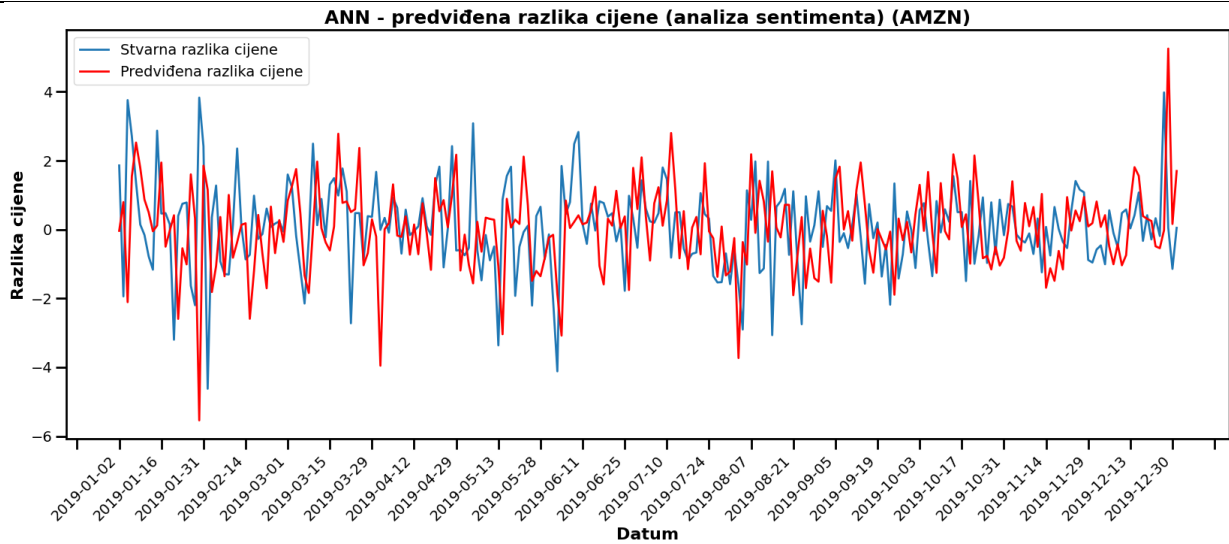


Slika 5.13: Predviđena razlika cijene modela za analizu sentimenta za tvrtku Apple

### Amazon

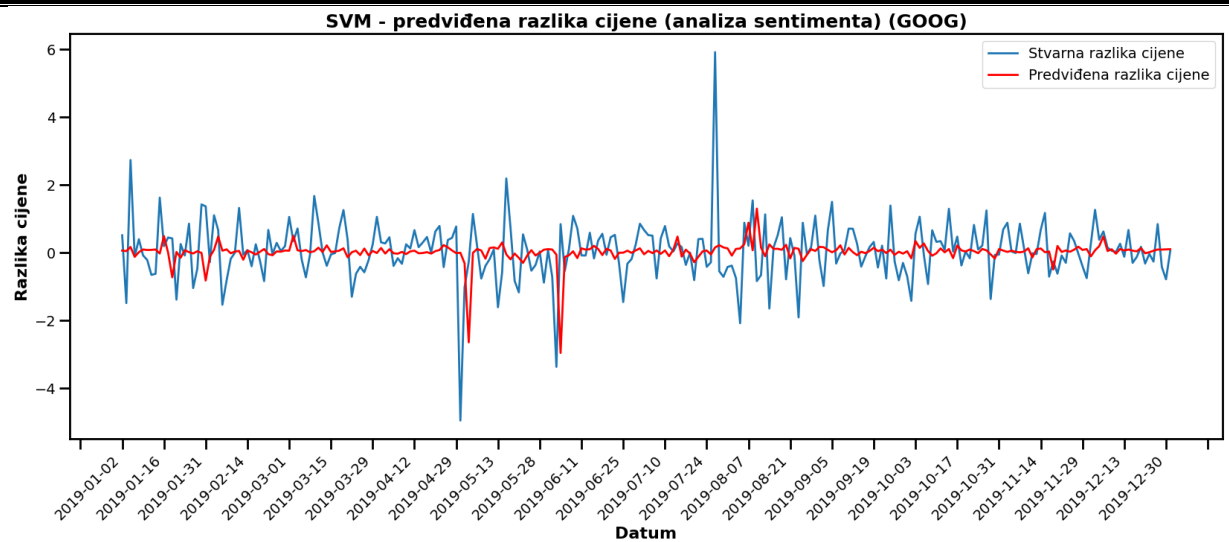


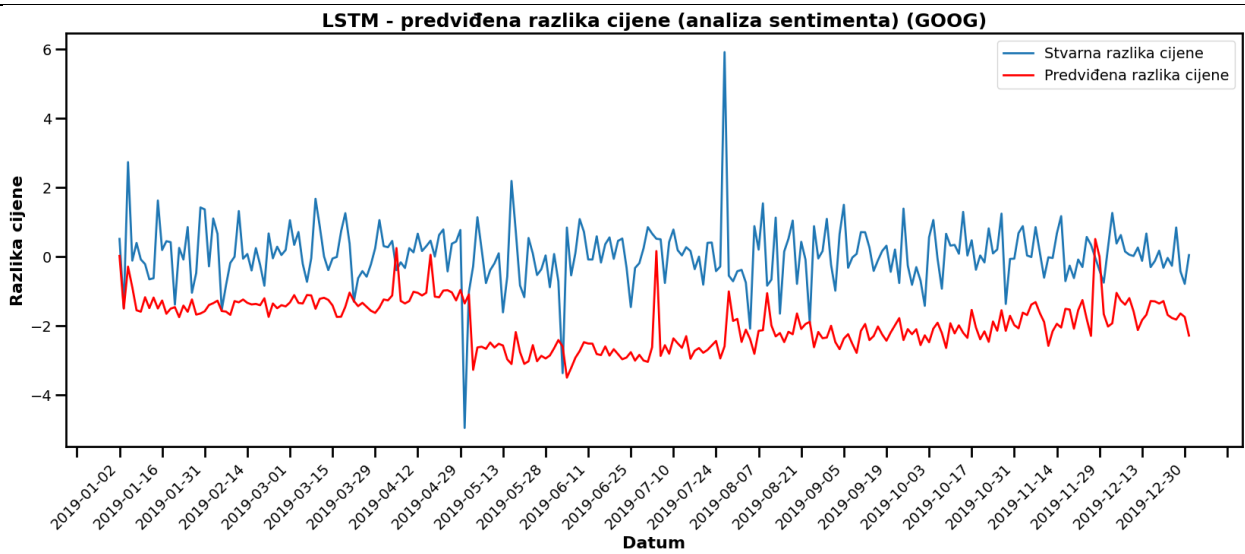
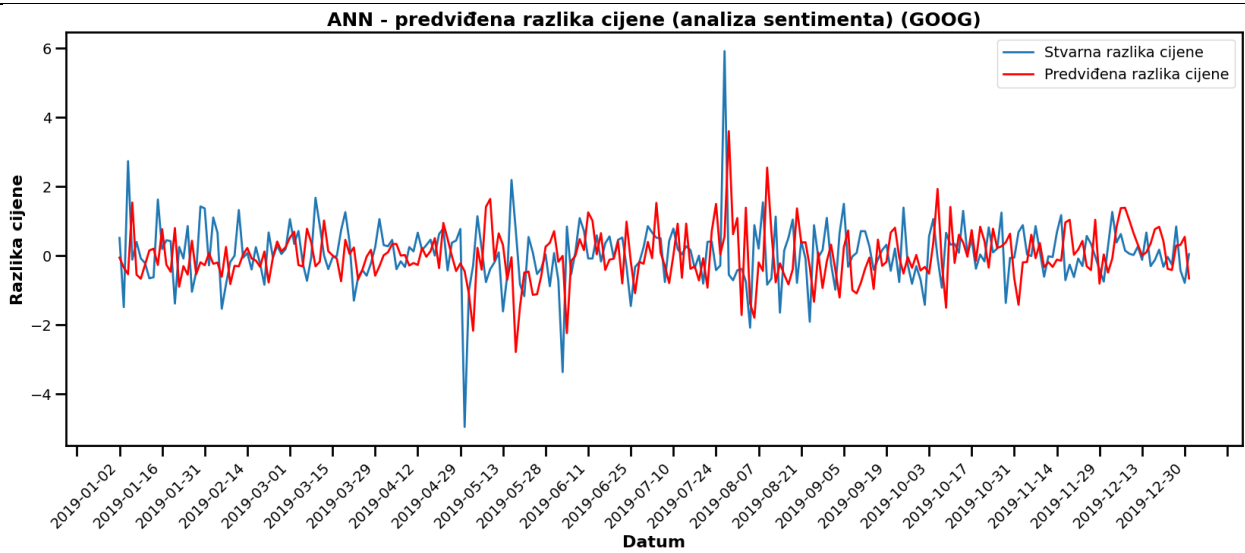
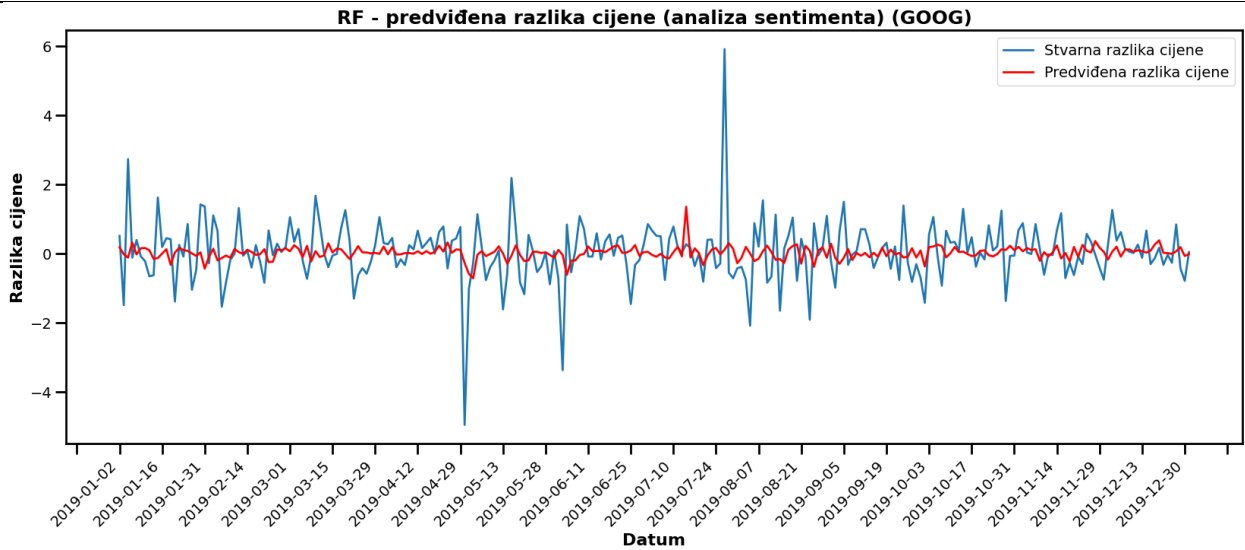




Slika 5.14: Predviđena razlika cijene modela za analizu sentimenta za tvrtku Amazon

## Google





Slika 5.15: Predviđena razlika cijene modela za analizu sentimenta za tvrtku Google

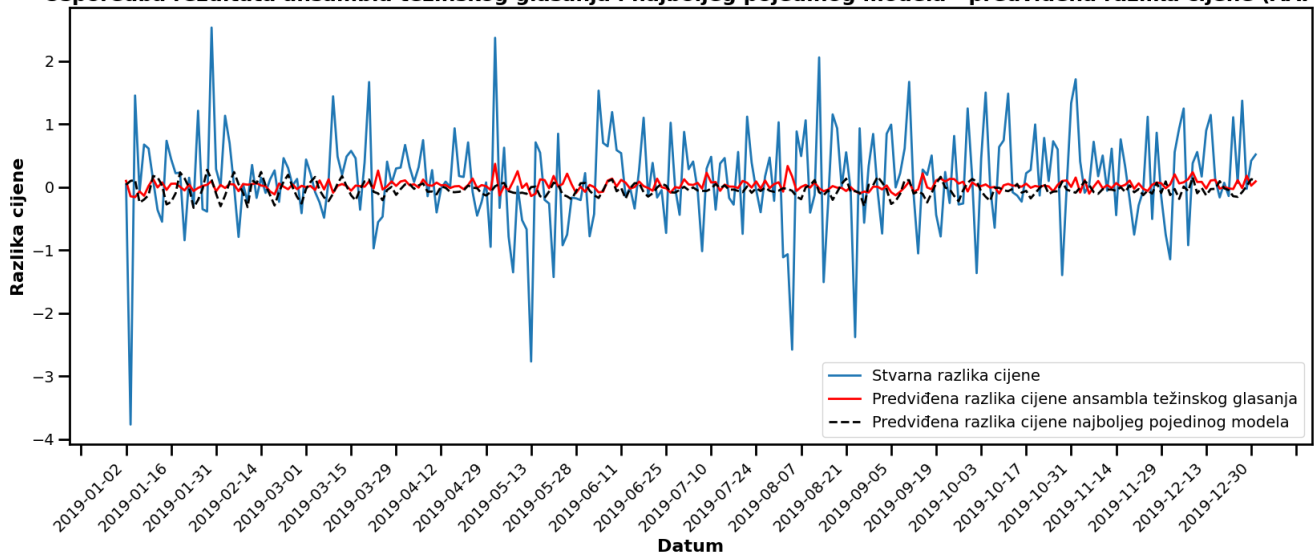
### 5.3. Usporedba pojedinih modela s predloženim ansamblom

Procjene ansambla težinskog glasanja nadmašile su pojedine modele u slučaju procjene dionica Apple i Amazon, dok je model ARIMA-GARCH ostvario bolje rezultate samo pri procjeni dionica Google. Ansambl težinskog glasanja pokazuje prednosti u usporedbi s ostalim modelima, s neznatno boljim prosječnim RMSE u procjeni dionica Applea i Amazona. Usporedba prosječnih RMSE prikazana je u tablici 5.11, dok su rezultati ansambla i najboljeg pojedinog modela prikazani na grafovima u slici 5.16. Međutim, treba uzeti u obzir da je svaki pojedini model koristio validaciju s vremenskim pomicanjem unaprijed, dok je u ansamblu težinskog glasanja ovaj korak skraćen kako bi se smanjili potrebni računalni resursi. Bitno je napomenuti da je sam ansambl težinskog glasanja koristio metodu validacije s vremenskim pomicanjem unaprijed, no pojedini modeli nisu implementirali ovu metodu već su unaprijed istrenirani s povijesnim podacima. Ovaj proces mogao je utjecati na preciznost procjena ansambla težinskog glasanja, što sugerira da bi rezultati mogli biti nešto bolji i da bi ansambl težinskog glasanja mogao imati još veću prednost u odnosu na pojedine modele. Detaljniji prikazi rezultata mogu se pronaći u priložima P.4, P.5 i P.6.

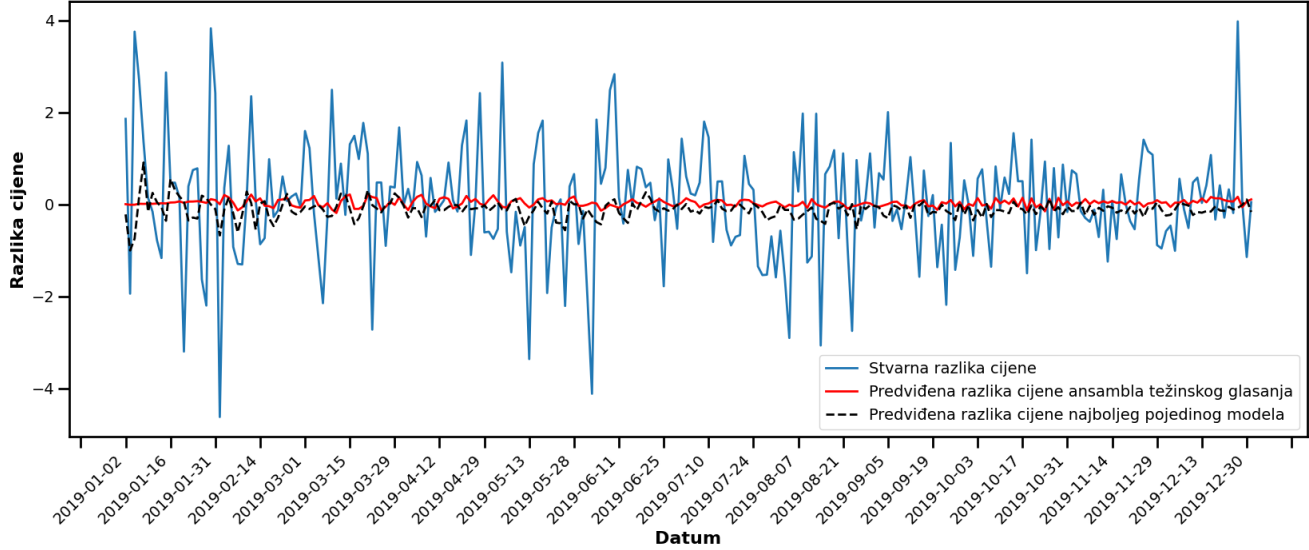
Tablica 5.11: Usporedba rezultata ansambla težinskog glasanja s najboljim modelima iz svake analize

Dionica	Ansambl težinskog glasanja	ARIMA-GARCH (analiza vremenskih serija)	RF (tehnička analiza)	RF (fundamentalna analiza)	RF (analiza sentimenta)
AAPL	0.77919	0.78248	0.80558	0.80857	0.78377
AMZN	1.25064	1.29216	1.31087	1.27951	1.2976
GOOG	0.88377	0.88201	0.97835	0.91185	0.89566

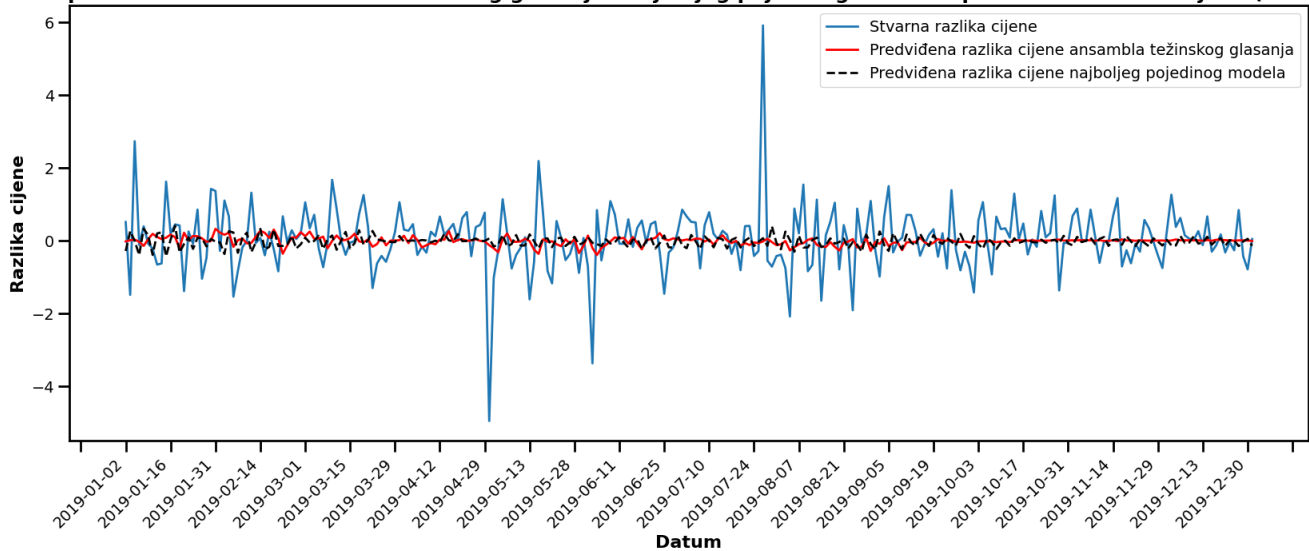
Usporedba rezultata ansambla težinskog glasanja i najboljeg pojedinog modela - predviđena razlika cijene (AAPL)



Usporedba rezultata ansambla težinskog glasanja i najboljeg pojedinog modela - predviđena razlika cijene (AMZN)



Usporedba rezultata ansambla težinskog glasanja i najboljeg pojedinog modela - predviđena razlika cijene (GOOG)



Slika 5.16: Usporedba rezultata ansambla težinskog glasanja i najboljeg pojedinog modela

## 5.4. Analiza učinka optimizacije težina

Za analizu razvijenog ansambla koji koristi težinsko glasanje i algoritam DE za optimizaciju težina, odlučeno je usporediti rezultate s jednostavnijim ansamblom koji primjenjuje prosječno glasanje. Ovaj jednostavniji ansambl izračunava prosjek procjena dobivenih od četiri modela kako bi dobio konačni rezultat. Odabir ovog ansambla kao referentnog temelji se na činjenici da pripada skupini ansambala zasnovanih na glasanju, ali je strukturno jednostavniji.

Analizom prosječnog RMSE utvrđeno je da je jednostavniji ansambl, temeljen na prosječnom glasanju, pokazao veću točnost u procjeni za tvrtke Google i Apple, dok se ansambl težinskog glasanja pokazao uspješnijim za tvrtku Amazon, što je prikazano u tablici 5.12. Iako su rezultati u

bliskom rasponu pogreške, nije uočena značajna prednost ansambla težinskog glasanja u odnosu na ansamblu prosječnog glasanja.

Tablica 5.12: Usporedba rezultata ansambla zasnovanom na prosječnom glasanju i ansambla zasnovanom na težinskom glasanju

Model	Prosječni RMSE
<b>Apple</b>	
<b>Ansambl zasnovan na prosječnom glasanju</b>	<b>0.76899</b>
Ansambl zasnovan na težinskom glasanju	0.77919
<b>Amazon</b>	
Ansambl zasnovan na prosječnom glasanju	1.25691
<b>Ansambl zasnovan na težinskom glasanju</b>	<b>1.25064</b>
<b>Google</b>	
<b>Ansambl zasnovan na prosječnom glasanju</b>	<b>0.84799</b>
Ansambl zasnovan na težinskom glasanju	0.88377

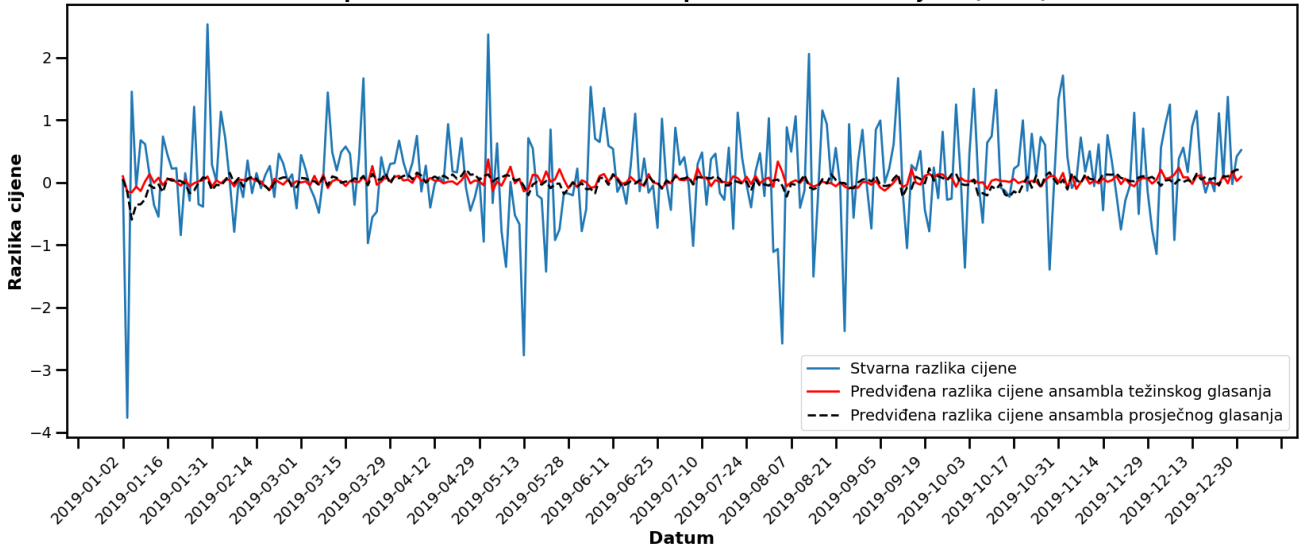
Međutim, važno je napomenuti da je jednostavniji ansambl koristio rezultate modela koji su primijenili metodu vremenske unakrsne validacije s pomicanjem unaprijed, što je detaljno opisano u poglavlju 5.3. Kako bi se osigurala točnija usporedba rezultata, korištena je ista metoda za ansambl težinskog glasanja i ansambl prosječnog glasanja, pri čemu su oba ansambla primijenila unaprijed istrenirane modele radi smanjenja vremena izvođenja. Rezultati su prikazani u tablici 5.13, gdje se jasno uočava prednost ansambla težinskog glasanja u odnosu na prosječni ansambl. Performanse ansambla težinskog glasanja značajno su bolje kada se primjenjuje ista metodologija kao i kod jednostavnijeg ansambla.

Tablica 5.13: Usporedba rezultata ansambla zasnovanom na prosječnom glasanju i ansambla zasnovanom na težinskom glasanju gdje ansamblu primjenjuju istu metodu treniranje modela

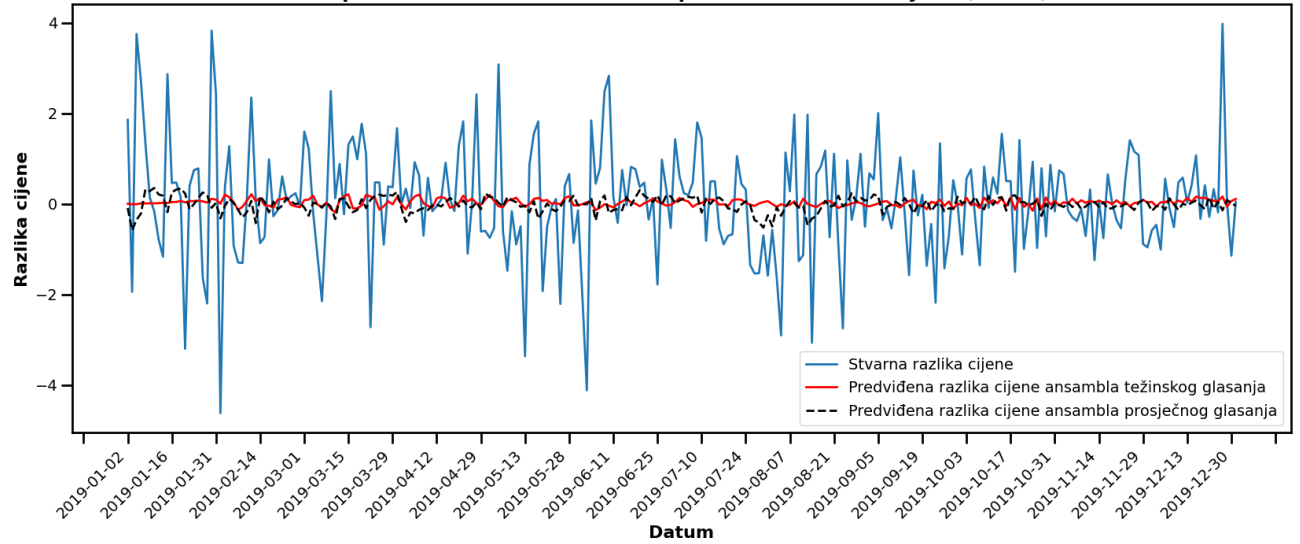
Model	Prosječni RMSE
<b>Apple</b>	
Ansambl zasnovan na prosječnom glasanju	0.79271
<b>Ansambl zasnovan na težinskom glasanju</b>	<b>0.77919</b>
<b>Amazon</b>	
Ansambl zasnovan na prosječnom glasanju	1.26756
<b>Ansambl zasnovan na težinskom glasanju</b>	<b>1.25064</b>
<b>Google</b>	
Ansambl zasnovan na prosječnom glasanju	0.97272
<b>Ansambl zasnovan na težinskom glasanju</b>	<b>0.88377</b>

Jedna od sličnosti prisutna u svim procjenama jest tendencija da rezultati osciliraju oko nule, što implicira da ansambl često predviđa minimalne promjene cijene u odnosu na trenutni dan. Ovaj fenomen proizlazi iz karakteristika korištenih modela, koji sami po sebi ne generiraju širok raspon varijacija u predviđanjima. Nasuprot tome, prirodno je očekivati da će DE algoritam optimizirati težine tako da rezultati gravitiraju prema nuli, čime se minimizira prosječna pogreška. Sama volatilitnost cijena dionica, uz varijabilnost rezultata modela, dodatno otežava kombinaciju predikcija. Zbog toga ansambl često procjenjuje da će cijena ostati gotovo nepromijenjena, jer na taj način osigurava dugoročno najmanju moguću pogrešku u predviđanjima. Rezultati su prikazani na slici 5.17.

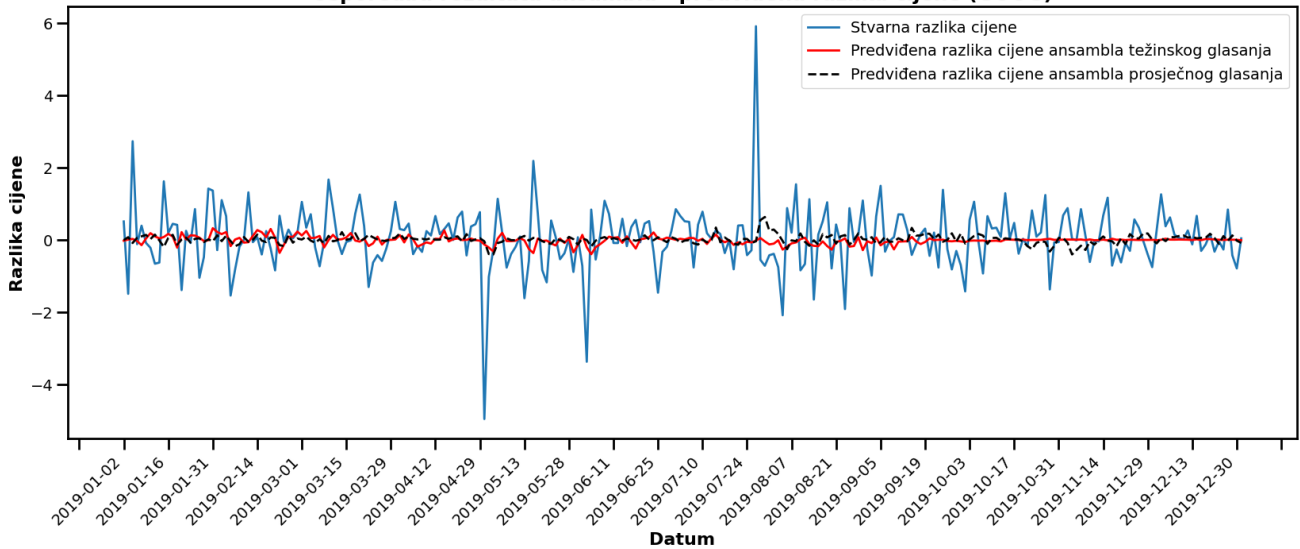
Usporedba rezultata ansambla - predviđena razlika cijene (AAPL)



Usporedba rezultata ansambla - predviđena razlika cijene (AMZN)



Usporedba rezultata ansambla - predviđena razlika cijene (GOOG)



Slika 5.17: Usporedba rezultata ansambla zasnovanom na prosječnom glasanju i ansambla zasnovanom na težinskom glasanju

## 6. ZAKLJUČAK

Zaključci ovog istraživanja jasno ilustriraju kompleksnost financijskih tržišta i polemike koje postoje prilikom procjene cijena dionica. U kontekstu nepredvidivih čimbenika koji utječu na tržišne vrijednosti, istraživanje je odgovorilo na temeljno pitanje: može li kombinacija različitih analitičkih metoda donijeti bolje rezultate od korištenja pojedinih modela? Dobiveni rezultati ukazuju da takva kombinacija zaista može pružiti pouzdanije procjene, no ujedno i naglašavaju važnost razumijevanja specifičnih okolnosti u kojima različiti modeli djeluju.

Analizirane metode, uključujući vremenske serije, tehničku analizu, fundamentalnu analizu i analizu sentimenta, pokazale su se korisnima u prepoznavanju različitih aspekata tržišta. Svaka od njih ima jedinstveni doprinos ukupnoj procjeni, no proces njihova usklađivanja i optimizacije zahtijeva pažljivo prilagođavanje parametrima. Suprotno očekivanjima, jednostavniji modeli poput RF i ARIMA-GARCH istaknuli su se svojom učinkovitošću u smanjenju pogreške u predikcijama, što otvara pitanja o optimalnoj razini složenosti modela u različitim tržišnim uvjetima.

Posebno upečatljiv ishod istraživanja jest uspjeh ansambla težinskog glasanja, gdje se težine optimiziraju algoritmom DE. Ovaj ansambl ne samo da je pokazao veću preciznost od prosječnog glasanja, već je i bolje odgovarao na promjene u tržišnim uvjetima. Ovo istraživanje potvrđuje hipotezu da se značaj različitih analitičkih pristupa mijenja u skladu s promjenama na tržištu. Ansambl težinskog glasanja uspio je integrirati te promjene na način koji pojedini modeli nisu mogli, čime je omogućio dublje razumijevanje tržišnih dinamika.

Kroz istraživanje se također otkrilo da sofisticirani modeli, iako teoretski sposobni za prepoznavanje skrivenih obrazaca, često ne ostvaruju svoju punu funkcionalnost u kraćim vremenskim okvirima, kao što je procjena na dnevnoj razini. Navedeno sugerira da jednostavniji pristupi, osobito u dinamičnim tržišnim uvjetima, mogu biti učinkovitiji zbog svoje prilagodljivosti i brzine izvođenja. Ovo saznanje otvara vrata za daljnja istraživanja primjene tih metoda na duljim vremenskim intervalima, poput tjednih ili mjesečnih procjena, gdje bi složeniji modeli možda pokazali veće prednosti.

Zaključno, iako su korišteni modeli i metode sofisticirani, precizna procjena kretanja cijena dionica ostaje izazovan zadatak zbog inherentne nepredvidivosti tržišta. Rezultati ovog rada ukazuju na potrebu za pristupom u kojem jednostavnost i složenost modela moraju biti prilagođeni



specifičnim tržišnim uvjetima. Buduća istraživanja mogla bi se usmjeriti na analizu performansi modela na duljim vremenskim okvirima i širim uzorcima podataka, čime bi se dodatno unaprijedilo razumijevanje prednosti različitih pristupa u prognoziranju cijena dionica.

## LITERATURA

- [1] P. Gratton, What Is the Stock Market and How Does It Work?: The bottom line [online], Investopedia, USA, 2024, dostupno na: <https://www.investopedia.com/terms/s/stockmarket.asp#:~:text=The%20stock%20market%20is%20where,perceived%20value%20and%20market%20conditions.> [26.6.2024.]
- [2] D. Harper, Forces That Move Stock Prices: Key takeaways [online], Investopedia, USA, 2024, dostupno na: <https://www.investopedia.com/articles/basics/04/100804.asp> [26.6.2024.]
- [3] M. Lam, "Neural network techniques for financial performance prediction: integrating fundamental and technical analysis", ScienceDirect, sv. 37(4), str. 567-582, 2004.
- [4] G. Kumar, U.S. Singh, S. Jain, "Stock Market Forecasting Using Computational Intelligence: A Survey", Springer, str. 6, 2020.
- [5] T. Yell, "4 Effective Trading Indicators Every Trader Should Know: TRADING WITH RSI" [online], DailyFX, USA, 2018, dostupno na: <https://community.ig.com/blogs/entry/842-4-effective-trading-indicators-every-trader-should-know/> [26.6.2024.]
- [6] I. Kof Nti, A.F. Adekoya, B.A Weyori, "A systematic review of fundamental and technical analysis of stock market predictions", "Springer", str. 5, 2019.
- [7] U.S. Securities and Exchange Commission (SEC) [online], USA, dostupno na: <https://www.sec.gov/> [26.6.2024.]
- [8] U.S. Securities and Exchange Commission (SEC): Apple [online], USA, dostupno na: <https://www.sec.gov/Archives/edgar/data/320193/000032019324000069/aapl-20240330.htm> [26.6.2024.]
- [9] G. Kumar, U.S. Singh, S. Jain, "Stock Market Forecasting Using Computational Intelligence: A Survey", Springer, str. 3, 2020.
- [10] G. Kumar, U.S. Singh, S. Jain, "Stock Market Forecasting Using Computational Intelligence: A Survey", Springer, str. 26, 2020.
- [11] What is Sentiment Analysis? [online], AWS, USA, dostupno na: <https://aws.amazon.com/what-is/sentiment-analysis/#:~:text=Sentiment%20analysis%20is%20the%20process,social%20media%20comments%2C%20and%20reviews.> [26.6.2024.]
- [12] P. Dixit, Sentiment Analysis: Building from the Ground Up [online], Medium, USA, 2020, dostupno na: <https://medium.com/analytics-vidhya/sentiment-analysis-building-from-the-ground-up-e12e9195fac4> [26.6.2024.]
- [13] I. Kof Nti, A.F. Adekoya, B.A Weyori, "A systematic review of fundamental and technical analysis of stock market predictions", "Springer", str. 15, 2019.
- [14] G. Sonkavde, D. P Dharrao, A. M. Bongale, S. T. Deokate, D. Doreswamy, S. K. Bhat, "Forecasting Stock Market Prices Using Machine Learning and Deep Learning Models: A Systematic Review, Performance Analysis and Discussion of Implications", mdpj, str.5, 2023

- [15] A. Stojanović, Ž. Kovačević, D. Ivošević, "ANSAMBL METODE STROJNOG UČENJA", PD.TVZ.HR, str. 248, 2023.
- [16] I. Kof Nti, A.F. Adekoya, B.A Weyori, "A comprehensive evaluation of ensemble learning for stock market prediction", Springer, str. 8, 2020
- [17] M. Georgioudakis, V. Plevris, A Comparative Study of Differential Evolution Variants in Constrained Structural Optimization [online], Frontiers, Cipar, dostupno na: <https://www.frontiersin.org/journals/built-environment/articles/10.3389/fbuil.2020.00102/full> [18.8.2024]
- [18] R. Storn and K. Price, "Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces", J. Glob. Optim., vol. 11, no. 4, str. 341–359, 1997.
- [19] C. Lin, A. Qing, and Q. Feng, "A new differential mutation base generator for differential evolution", J. Glob. Optim., vol. 49, str. 69–90, 2011.
- [20] M. Dudjak, B. Zorić, D. Bajer, "An ensemble based framework for biomedical classification problems", FERIT, str. 1-5, 2024.
- [21] Yahoo!finance, dostupno na: <https://finance.yahoo.com/>
- [22] TA-Lib – Technical Analysis Library (2001). ta-lib-0.4.0, dostupno na: <https://ta-lib.org/>
- [23] Hugging Face (2018). RoBERTa, dostupno na: [https://huggingface.co/docs/transformers/en/model\\_doc/roberta](https://huggingface.co/docs/transformers/en/model_doc/roberta)
- [24] R. Santra, Tests for Stationarity in Time Series — Dickey Fuller Test & Augmented Dickey Fuller(ADF) Test [online], Medium, USA, 2023, dostupno na: <https://medium.com/@ritusantra/tests-for-stationarity-in-time-series-dickey-fuller-test-augmented-dickey-fuller-adf-test-d2e92e214360> [13.9.2024]
- [25] A. Kumar, AIC & BIC for Selecting Regression Models: Formula, Examples [online], Vitalflux, India, 2023, Dostupno na: <https://vitalflux.com/aic-vs-bic-for-regression-models-formula-examples/>
- [26] Scikit-learn (2007). scikit-learn 1.5, dostupno na: <https://scikit-learn.org/stable/>
- [27] TensorFlow (2015). TensorFlow 2.0, dostupno na: <https://www.tensorflow.org/>

## POPIS SLIKA

Slika 2.1: Utjecajni čimbenici na cijenu [2] .....	3
Slika 2.2: Tehnički indikatori (RSI) [5] .....	5
Slika 2.3: Dio financijskog izvještaj tvrtke Apple [8].....	7
Slika 2.4: Primjer procjene dionice Apple putem ARIMA-GARCH modela.....	8
Slika 2.5: Analiza sentimenta [12] .....	9
Slika 2.6: Različite tehnike za procjenu cijene dionica [14] .....	10
Slika 3.1: Način rada ansambla zasnovanog na glasanju .....	11
Slika 3.2: Način rada ansambla zasnovanog na slaganju .....	12
Slika 3.3: Način rada ansambla zasnovanog na pakiranju .....	13
Slika 3.4: Način rada ansambla zasnovanog na pojačavanju .....	14
Slika 4.1: Ansambl za procjenu cijene dionice koristeći četiri vrste analiza .....	16
Slika 4.2: Dijagram algoritma DE.....	19
Slika 4.3: Pseudokod algoritma DE .....	21
Slika 4.4: Dijagram primjene algoritma DE za optimizaciju težina .....	22
Slika 5.1: Način rada validacije uz vremensko pomicanje unaprijed .....	31
Slika 5.2: Način treniranja i vrednovanja modela .....	32
Slika 5.3: Način treniranja i vrednovanja ansambla.....	34
Slika 5.4: Predviđena razlika cijene modela vremenskih serija za tvrtku Apple .....	37
Slika 5.5: Predviđena razlika cijene modela vremenskih serija za tvrtku Amazon .....	38
Slika 5.6: Predviđena razlika cijene modela vremenskih serija za tvrtku Google .....	39
Slika 5.7: Predviđena razlika cijene modela za tehničku analizu za tvrtku Apple.....	42
Slika 5.8: Predviđena razlika cijene modela za tehničku analizu za tvrtku Amazon.....	44
Slika 5.9: Predviđena razlika cijene modela za tehničku analizu za tvrtku Google.....	45
Slika 5.10: Predviđena razlika cijene modela za fundamentalnu analizu za tvrtku Apple .....	48
Slika 5.11: Predviđena razlika cijene modela za fundamentalnu analizu za tvrtku Amazon.....	49
Slika 5.12: Predviđena razlika cijene modela za fundamentalnu analizu za tvrtku Google .....	51
Slika 5.13: Predviđena razlika cijene modela za analizu sentimenta za tvrtku Apple .....	54
Slika 5.14: Predviđena razlika cijene modela za analizu sentimenta za tvrtku Amazon .....	55
Slika 5.15: Predviđena razlika cijene modela za analizu sentimenta za tvrtku Google .....	56
Slika 5.16: Usporedba rezultata ansambla težinskog glasanja i najboljeg pojedinog modela .....	58
Slika 5.17: Usporedba rezultata ansambla zasnovanom na prosječnom glasanju i ansambla zasnovanom na težinskom glasanju .....	61

## POPIS TABLICA

Tablica 5.1: Modeli za analizu vremenskih serija.....	26
Tablica 5.2: Upotrebljeni parametri i skup podataka za SVM modele u sklopu svake analize ....	27
Tablica 5.3: Upotrebljeni parametri i skup podataka za RF modele u sklopu svake analize.....	28
Tablica 5.4: Upotrebljeni parametri i skup podataka za ANN modele u sklopu svake analize ....	29
Tablica 5.5: Upotrebljeni parametri i skup podataka za LSTM modele u sklopu svake analize ..	30
Tablica 5.6: Upotrebljeni parametri i skup podataka za algoritam DE.....	31
Tablica 5.7: Prosječni RMSE za analizu vremenskih serija.....	36
Tablica 5.8: Prosječni RMSE modela za tehničku analizu .....	40
Tablica 5.9: Prosječni RMSE modela za fundamentalnu analizu .....	46
Tablica 5.10: Prosječni RMSE modela za analizu sentimenta.....	52
Tablica 5.11: Usporedba rezultata ansambla težinskog glasanja s najboljim modelima iz svake analize.....	57
Tablica 5.12: Usporedba rezultata ansambla zasnovanom na prosječnom glasanju i ansambla zasnovanom na težinskom glasanju .....	59
Tablica 5.13: Usporedba rezultata ansambla zasnovanom na prosječnom glasanju i ansambla zasnovanom na težinskom glasanju gdje ansambli primjenjuju istu metodu treniranje modela ..	60

## SAŽETAK

U kontekstu globalnih burzi, precizno prognoziranje cijena dionica postalo je presudno za velike tvrtke, banke i globalnu ekonomiju, gdje svaka prednost može značiti razliku između uspjeha i neuspjeha. Ovaj rad se bavi kompleksnošću procjene cijena dionica, koja je izazvana brojnim nepredvidivim čimbenicima poput informacija s društvenih mreža, financijskih izvještaja, tehničkih indikatora i samoj cijeni dionice.

Zbog raznolike prirode i složenosti podataka, efikasna analiza zahtijeva odabir i razvoj odgovarajućih modela. U ovom radu istražene su četiri ključne metode analize: fundamentalna analiza, tehnička analiza, analiza sentimenta i analiza vremenskih serija. Proučena je mogućnost kombiniranja ovih metoda u ansambl kako bi se unaprijedila preciznost prognoza. Ansambl omogućuje obuhvaćanje različitih izvora informacija i prilagodbu modela prema trenutnim uvjetima, čime se postiže dublje razumijevanje i točnije procjene cijena dionica.

Cilj ovog rada je procijeniti efikasnost različitih analitičkih modela za procjenu cijena dionica i usporediti njihove performanse s ansamblom koji kombinira rezultate više pojedinih modela. Konkretno, istraživanje će obuhvatiti analizu uspješnosti svakog modela, uključujući fundamentalnu analizu, tehničku analizu, analizu sentimenta i analizu vremenskih serija. Svaki od ovih modela pruža različite uvide i koristi specifične skupove podataka, stoga je važno razumjeti njihove prednosti i ograničenja u kontekstu procjene cijena dionica.

Nakon evaluacije učinkovitosti pojedinih modela, rad će se usredotočiti na razvoj i implementaciju ansambl modela koji integrira rezultate svih ispitanih modela. Ansambl model, koji koristi različite pristupe i metode, omogućuje integraciju između pojedinih modela i može pružiti cjelovitiju i precizniju procjenu cijena dionica. Cilj je utvrditi može li kombinacija različitih modela poboljšati točnost prognoza u odnosu na upotrebu pojedinih modela, te procijeniti koristi od ove integracije u stvarnim tržišnim uvjetima.

Kroz ovu usporedbu, rad će nastojati identificirati najučinkovitije pristupe i tehnike za analizu cijena dionica, te pružiti preporuke za primjenu ovih metoda u praksi.

**Ključne riječi:** ansambl, diferencijalna evolucija, procjena dionice, težinsko glasanje

## **ABSTRACT**

### **STOCK PRICE PREDICTION USING AN ENSEMBLE OF MACHINE LEARNING ALGORITHMS**

In the context of global stock exchanges, accurately forecasting stock prices has become crucial for large companies, banks, and the global economy, where every advantage can mean the difference between success and failure. This study addresses the complexity of stock price prediction, which is influenced by numerous unpredictable factors such as social media information, financial reports, technical indicators, and the stock price itself.

Due to the diverse nature and complexity of the data, effective analysis requires selecting appropriate methods and developing suitable models. This research examines four key analytical methods: fundamental analysis, technical analysis, sentiment analysis, and time series analysis. The study explores the potential of combining these methods into an ensemble to enhance forecasting accuracy. An ensemble allows for the integration of various information sources and adapts models according to current conditions, leading to a deeper understanding and more accurate stock price predictions.

The aim of this work is to evaluate the effectiveness of different analytical models for stock price prediction and compare their performance with an ensemble that combines the results of multiple individual models. Specifically, the research will cover the performance analysis of each model, including fundamental analysis, technical analysis, sentiment analysis, and time series analysis. Each of these models provides different insights and uses specific data sets, making it crucial to understand their strengths and limitations in the context of stock price prediction.

After evaluating the effectiveness of individual models, the study will focus on developing and implementing an ensemble model that integrates the results of all examined models. The ensemble model, which employs various approaches and methods, facilitates the integration of individual models and can provide a more comprehensive and precise stock price assessment. The goal is to determine whether combining different models can improve forecasting accuracy compared to using individual models and to assess the benefits of this integration in real-world market conditions.

Through this comparison, the study aims to identify the most effective approaches and techniques for stock price analysis and provide recommendations for applying these methods in practice.

**Keywords:** ensemble, differential evolution, stock prediction, weighted voting

## **ŽIVOTOPIS**

Antonio Žugec rođen je 1. travnja 1999. godine u Hamiltonu, Kanada. Nakon završetka II. jezične gimnazije u Osijeku 2018. godine, upisao je studij računarstva na fakultetu elektrotehnike, računarstva i informacijskih tehnologija u Osijeku. Godine 2022. završio je preddiplomski studij računarstva s temom završnog rada “Upravljanje pomoću Mitsubishi FX5UC programabilnog logičkog kontrolera“.

---

Potpis autora

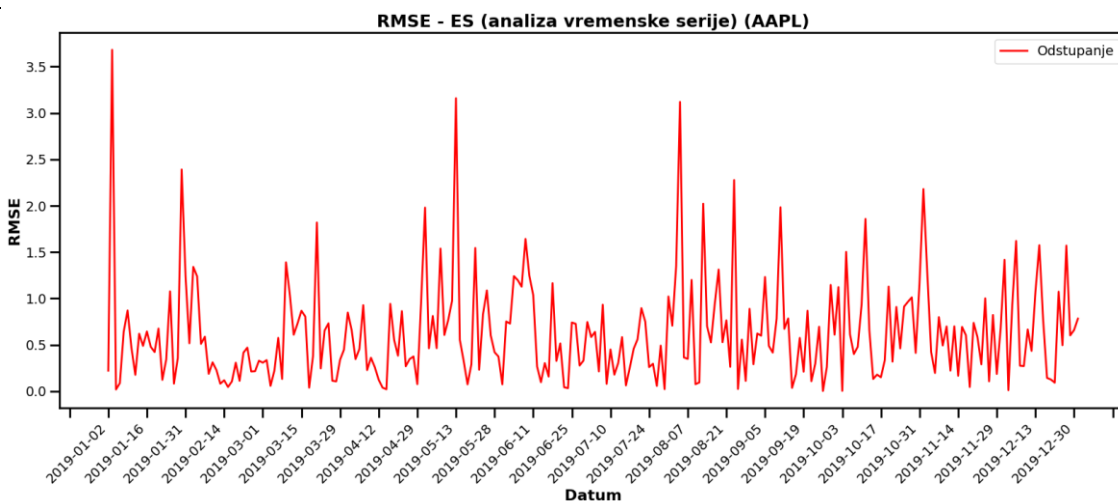
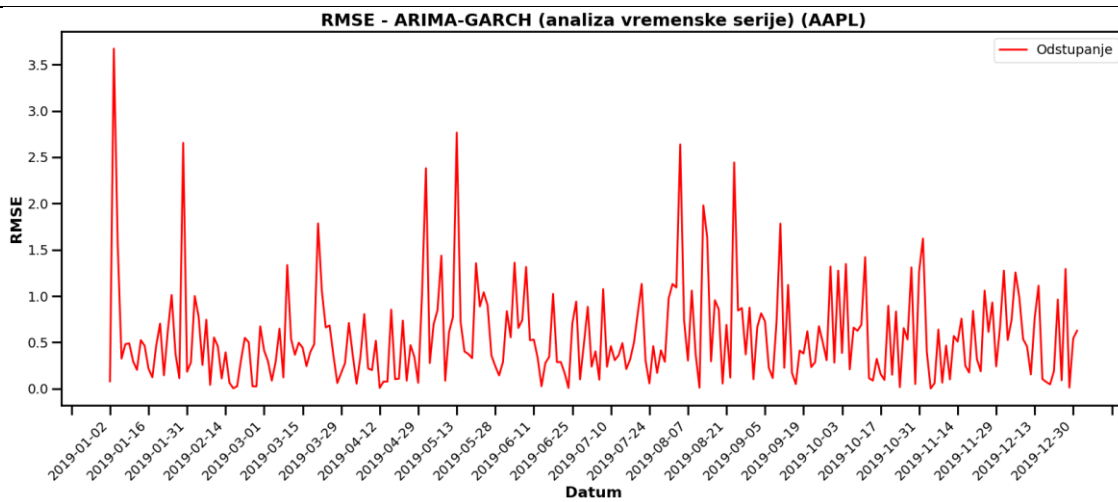
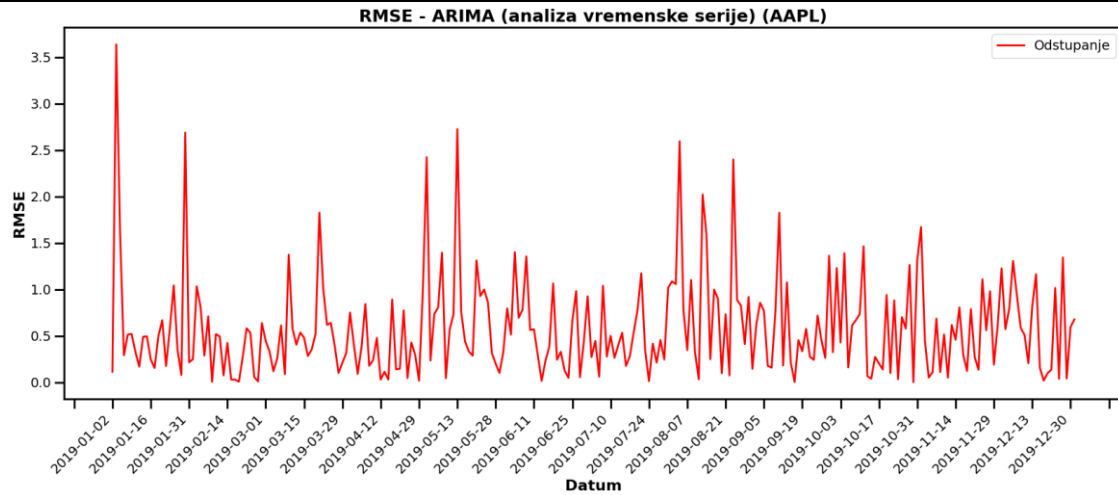


# PRILOZI

## Prilog P.1. Grafovi s prikazom rezultata modela vremenskih serija za tvrtku Apple

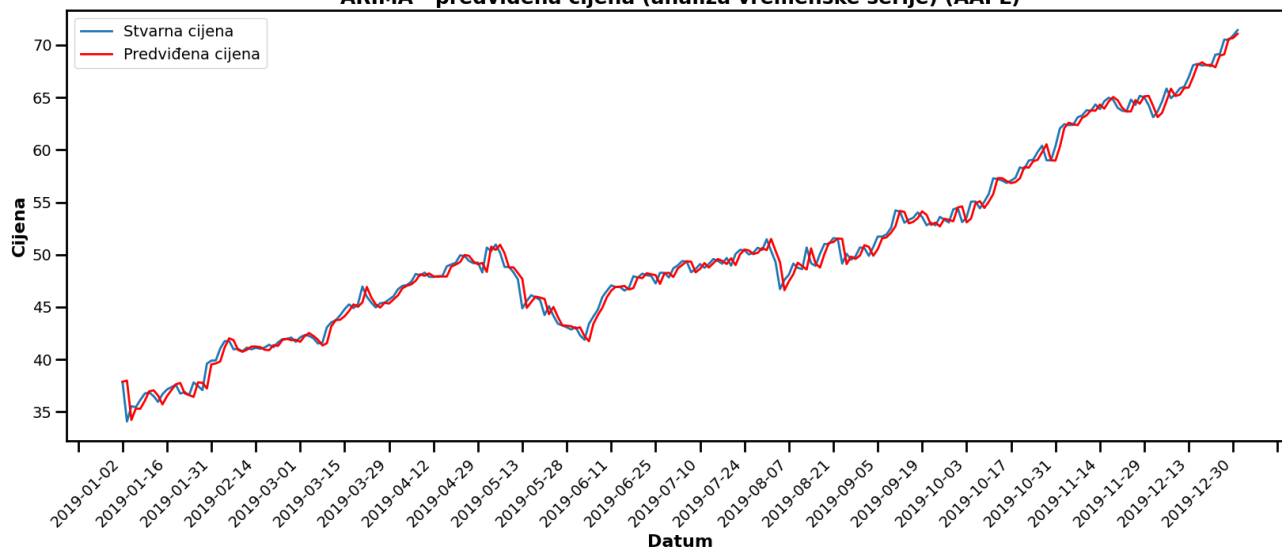
### Apple

#### RMSE

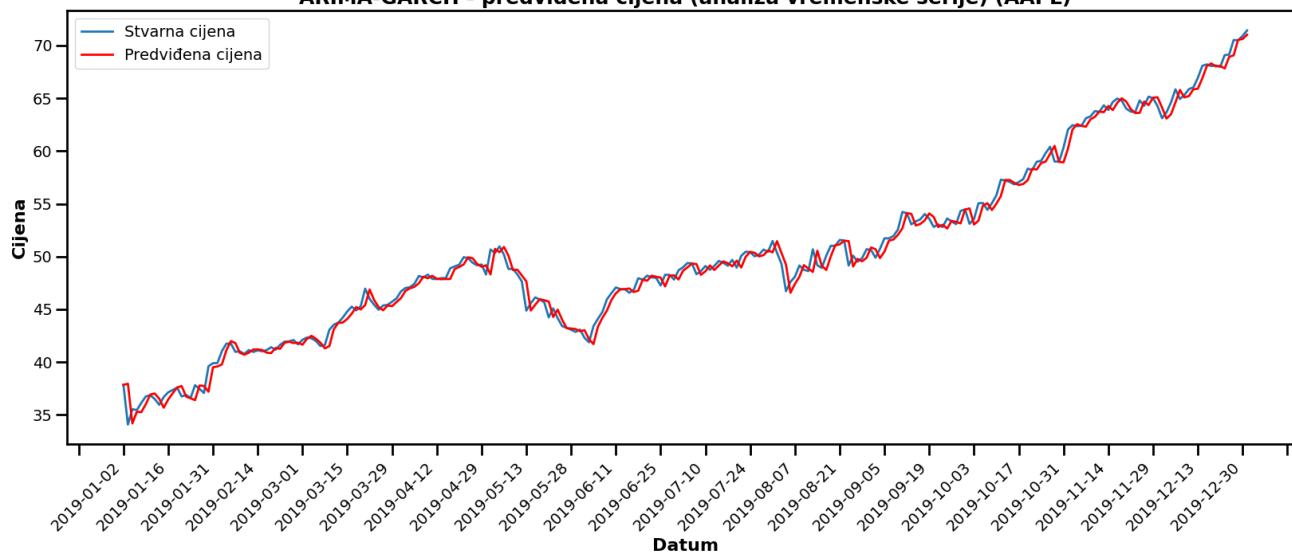


## Predviđena cijena

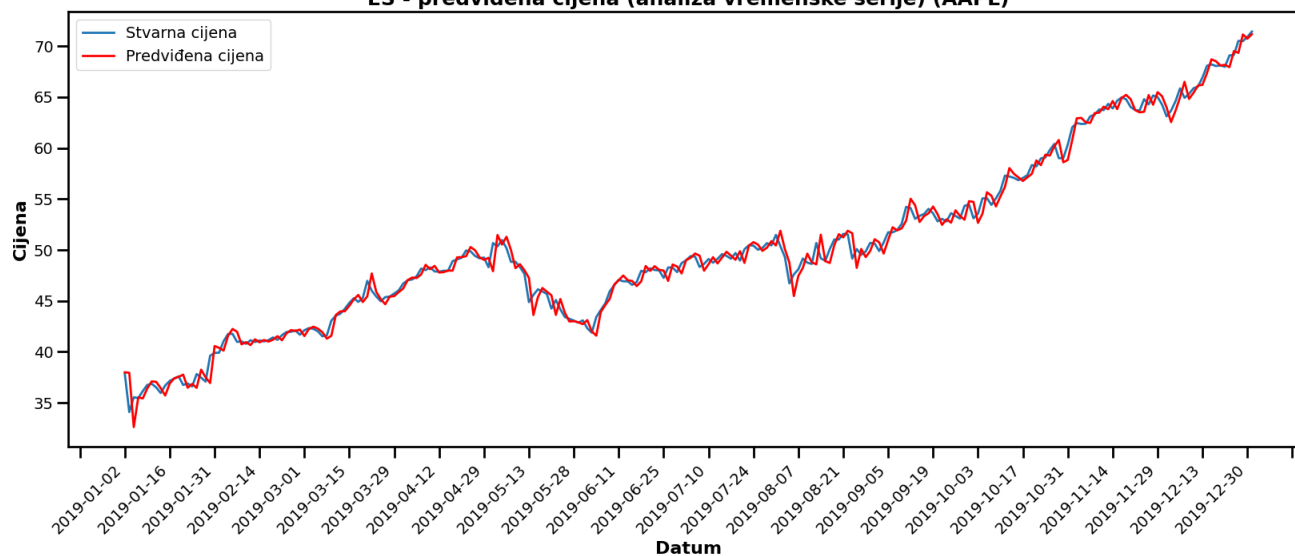
### ARIMA - predviđena cijena (analiza vremenske serije) (AAPL)



### ARIMA-GARCH - predviđena cijena (analiza vremenske serije) (AAPL)



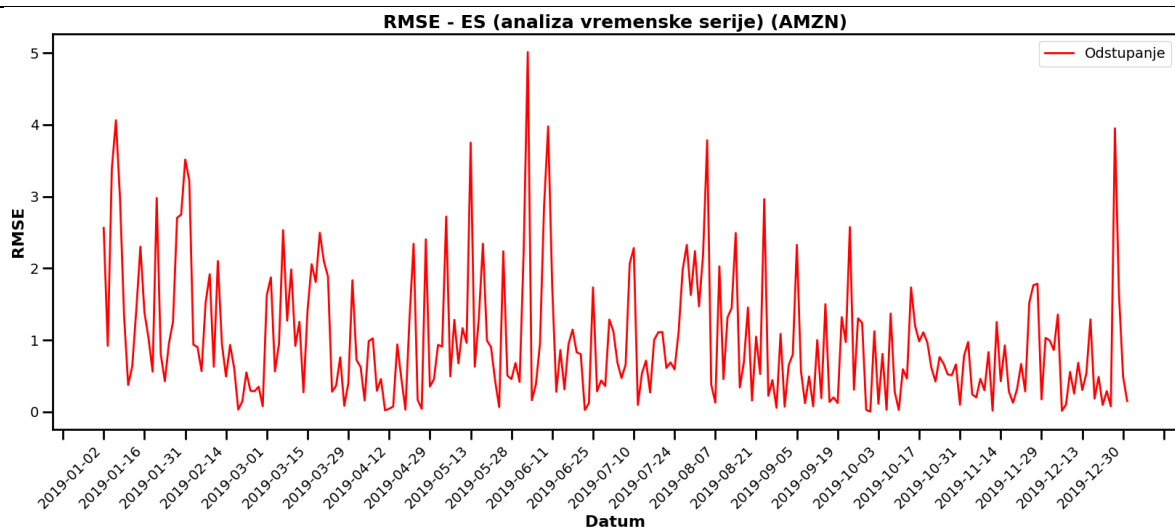
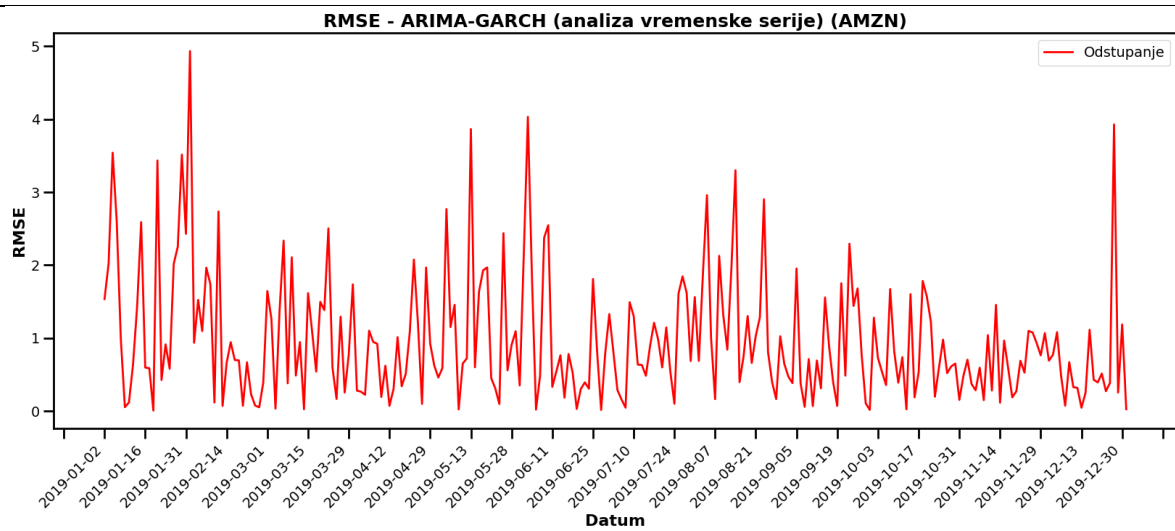
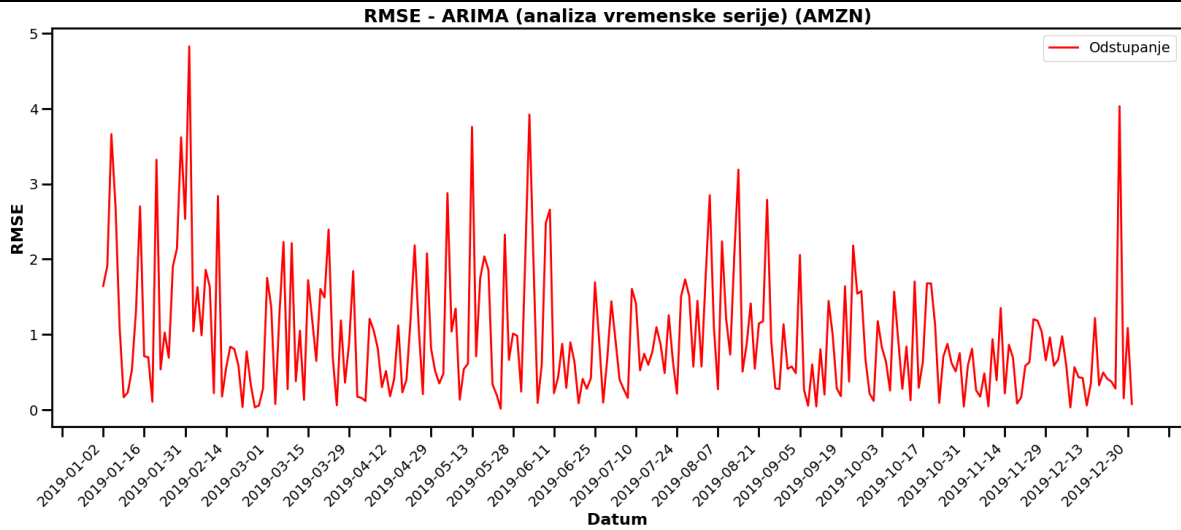
### ES - predviđena cijena (analiza vremenske serije) (AAPL)



**Prilog P.2.** Grafovi s prikazom rezultata modela vremenskih serija za tvrtku Amazon

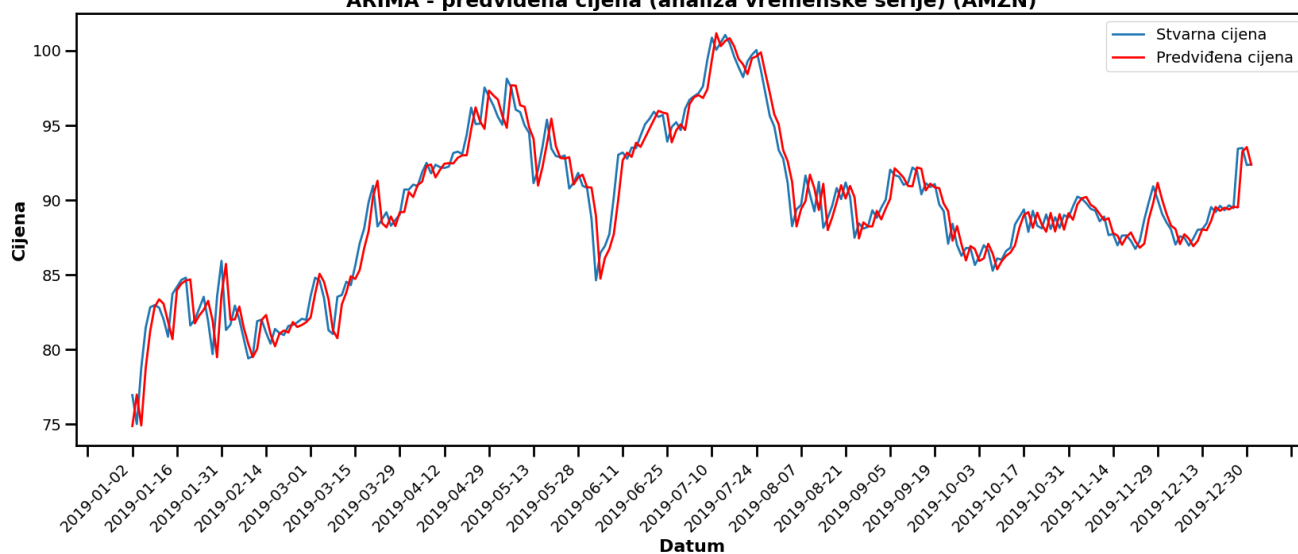
**Amazon**

**RMSE**



## Predviđena cijena

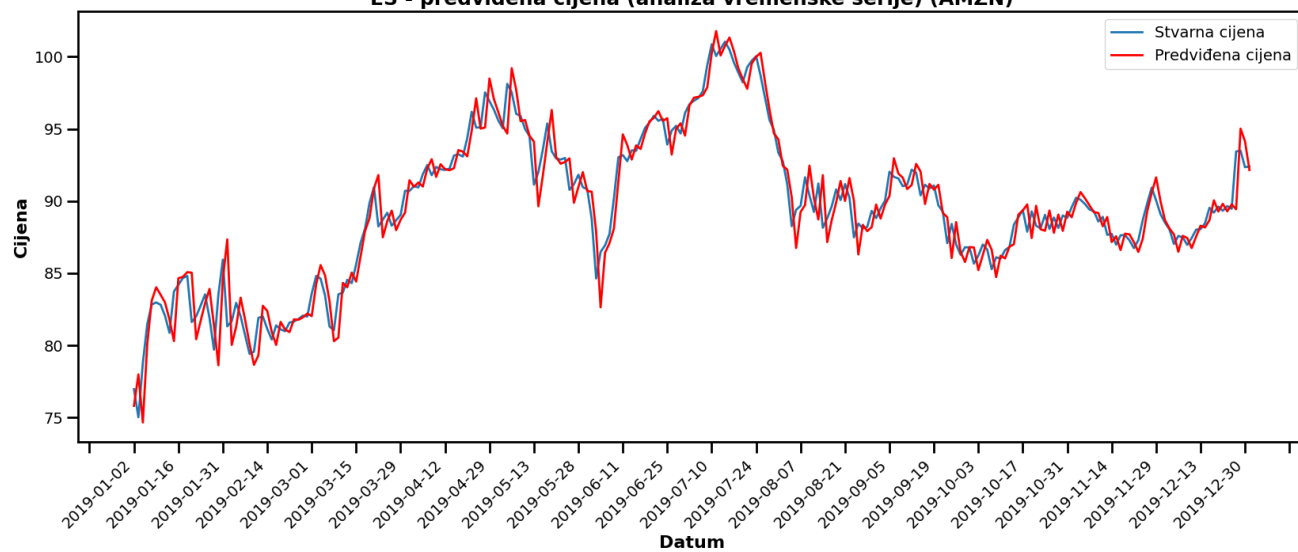
### ARIMA - predviđena cijena (analiza vremenske serije) (AMZN)



### ARIMA-GARCH - predviđena cijena (analiza vremenske serije) (AMZN)



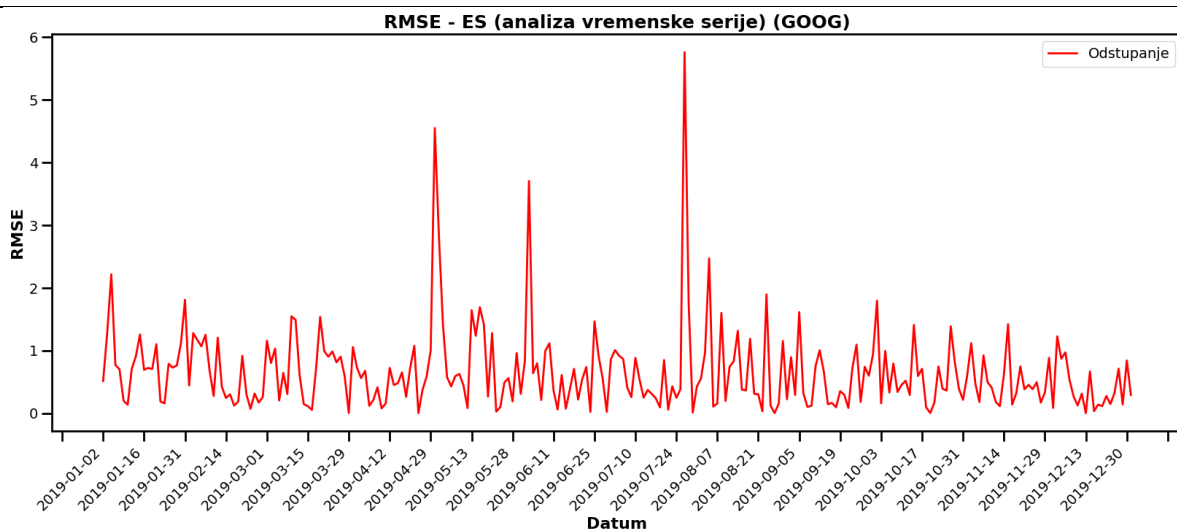
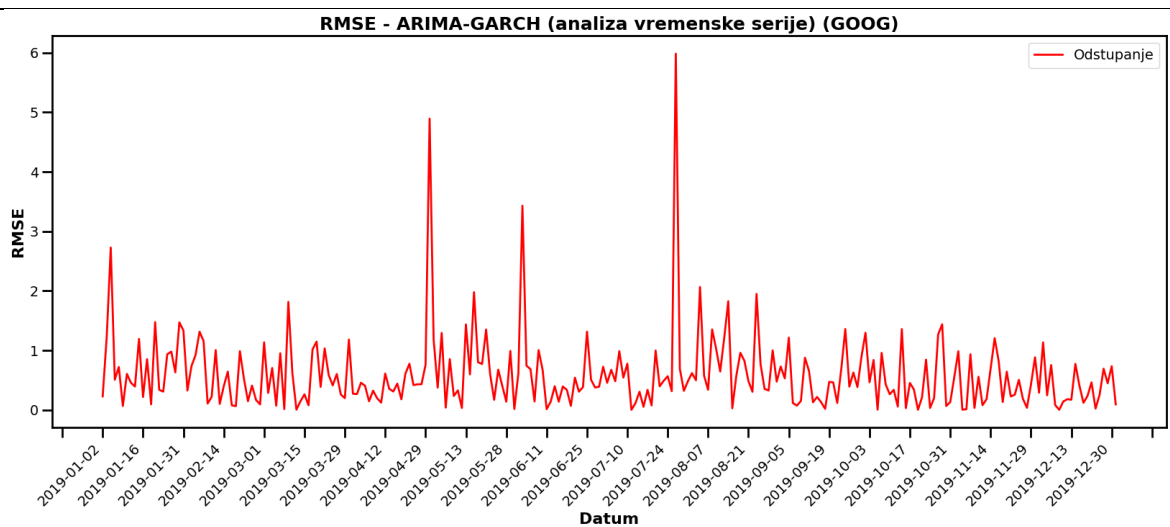
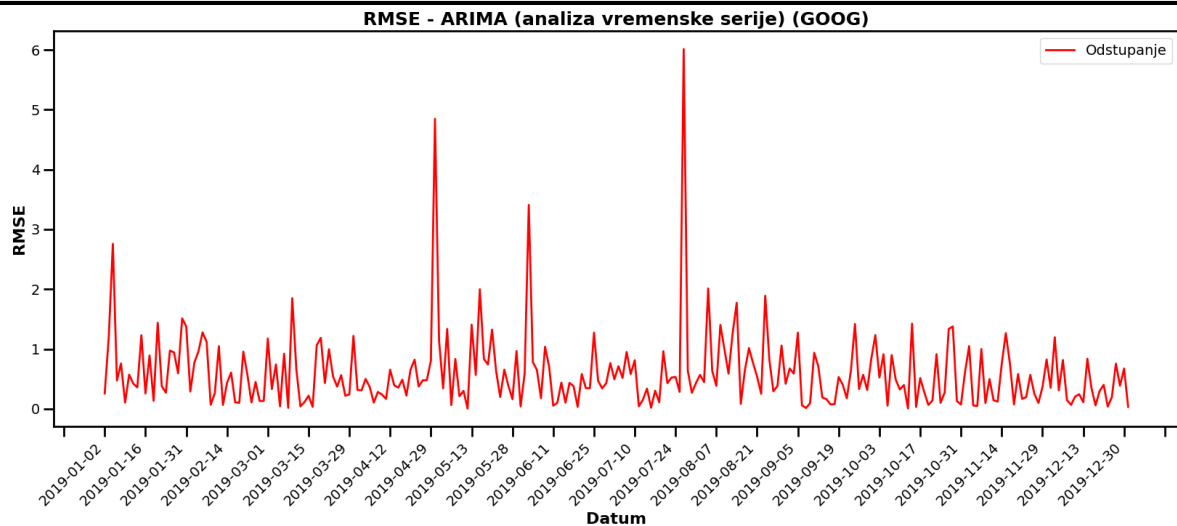
### ES - predviđena cijena (analiza vremenske serije) (AMZN)



### Prilog P.3. Grafovi s prikazom rezultata modela vremenskih serija za tvrtku Google

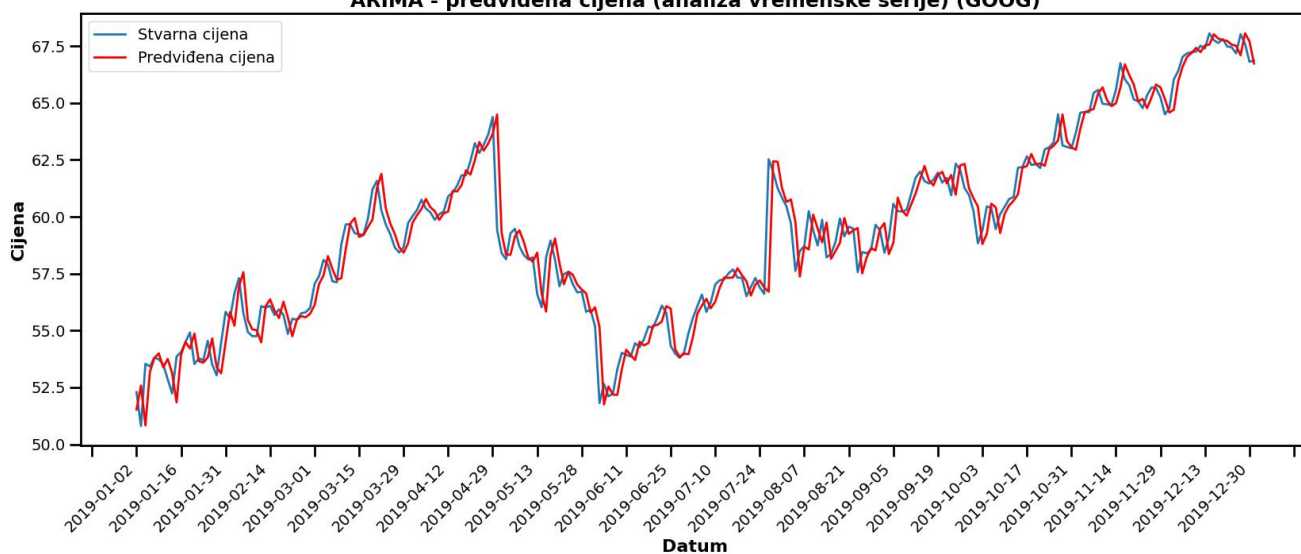
#### Google

#### RMSE

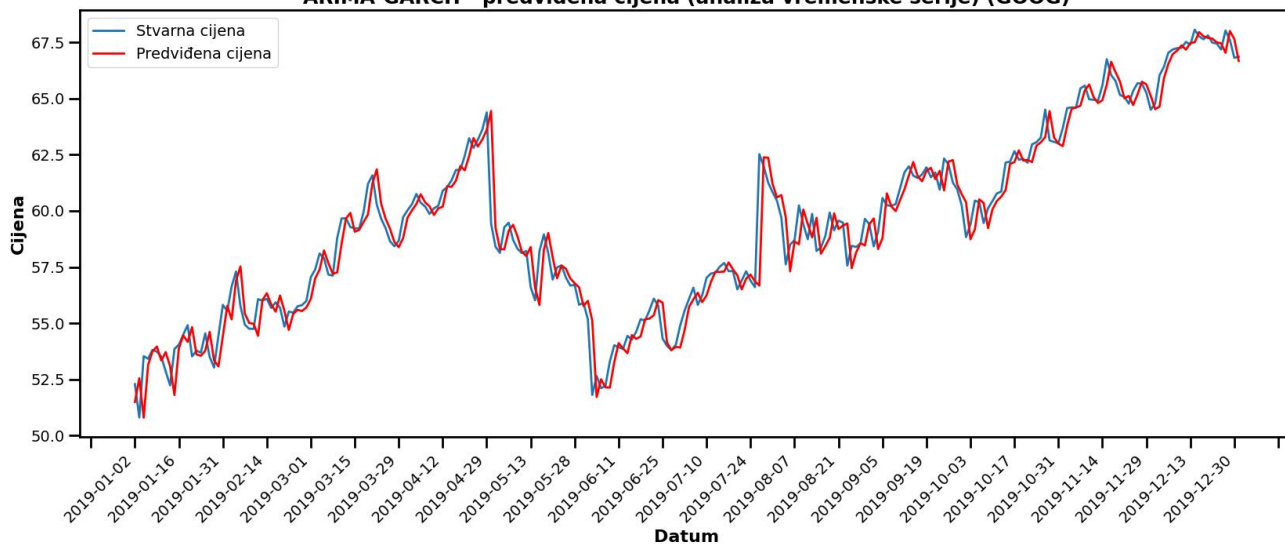


## Predviđena cijena

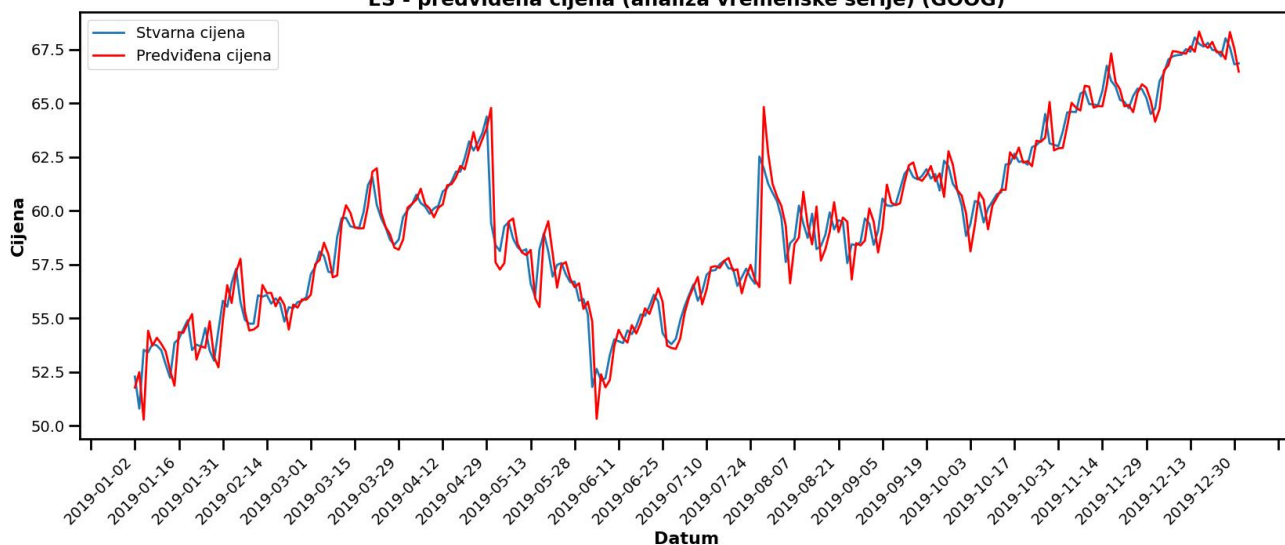
### ARIMA - predviđena cijena (analiza vremenske serije) (GOOG)



### ARIMA-GARCH - predviđena cijena (analiza vremenske serije) (GOOG)



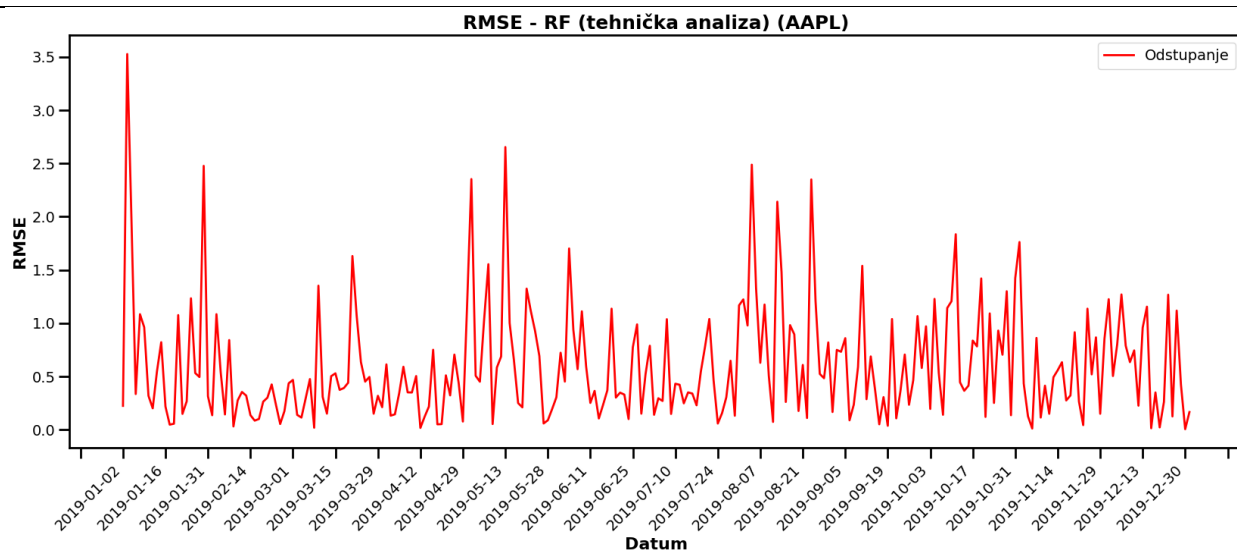
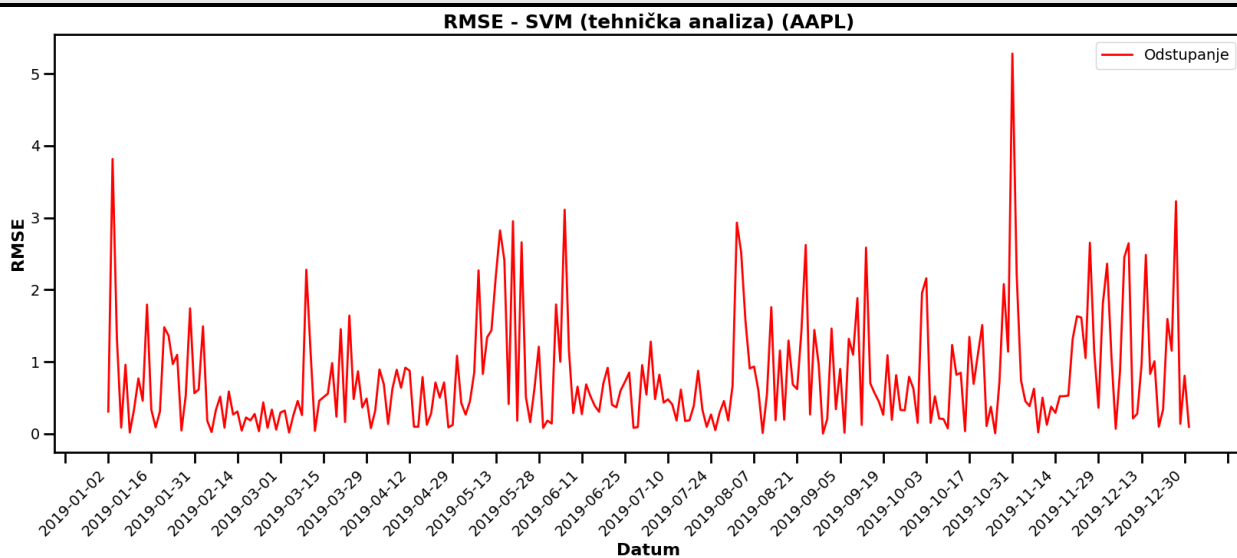
### ES - predviđena cijena (analiza vremenske serije) (GOOG)



## Prilog P.4. Grafovi s prikazom rezultata modela za tehničku analizu za tvrtku Apple

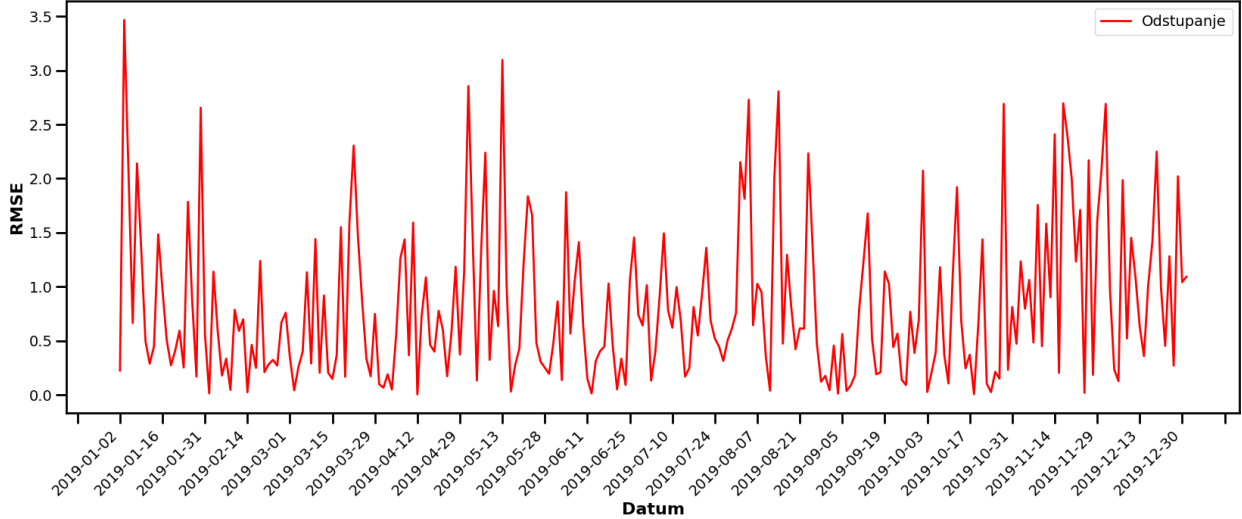
### Apple

#### RMSE

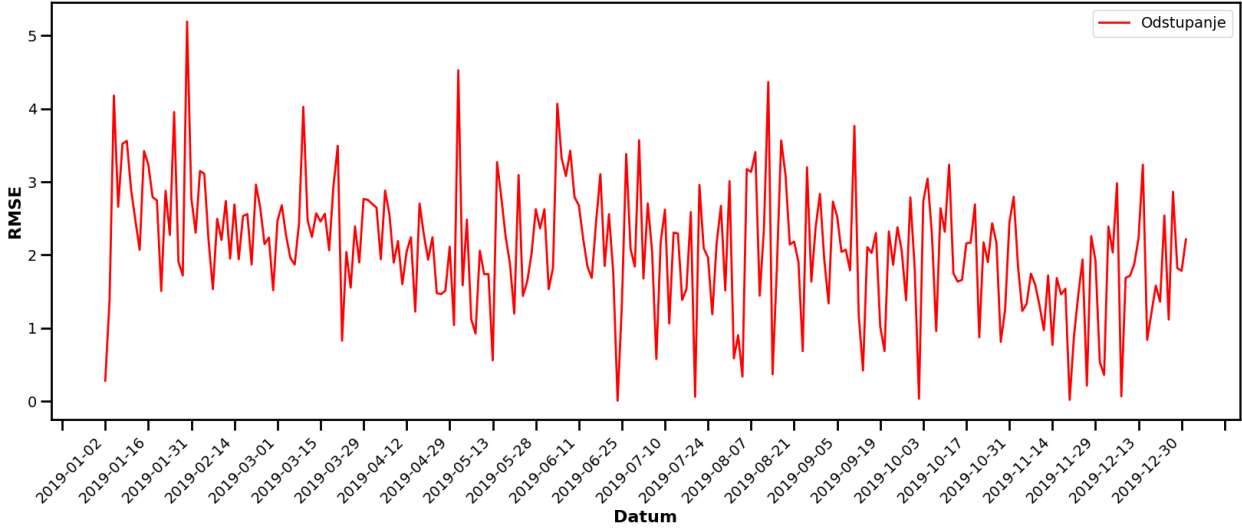




RMSE - ANN (tehnička analiza) (AAPL)

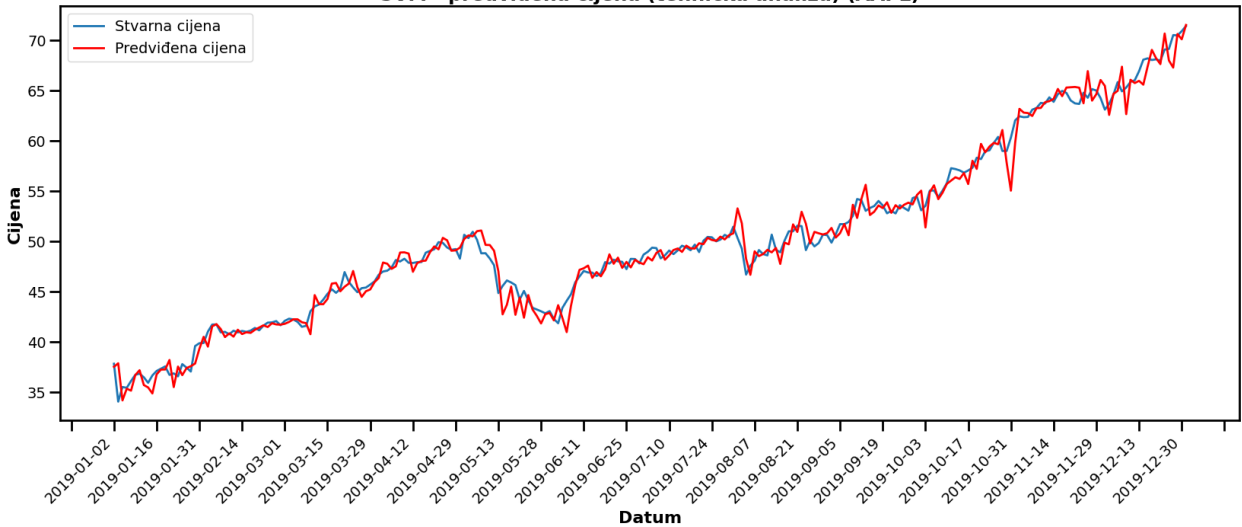


RMSE - LSTM (tehnička analiza) (AAPL)



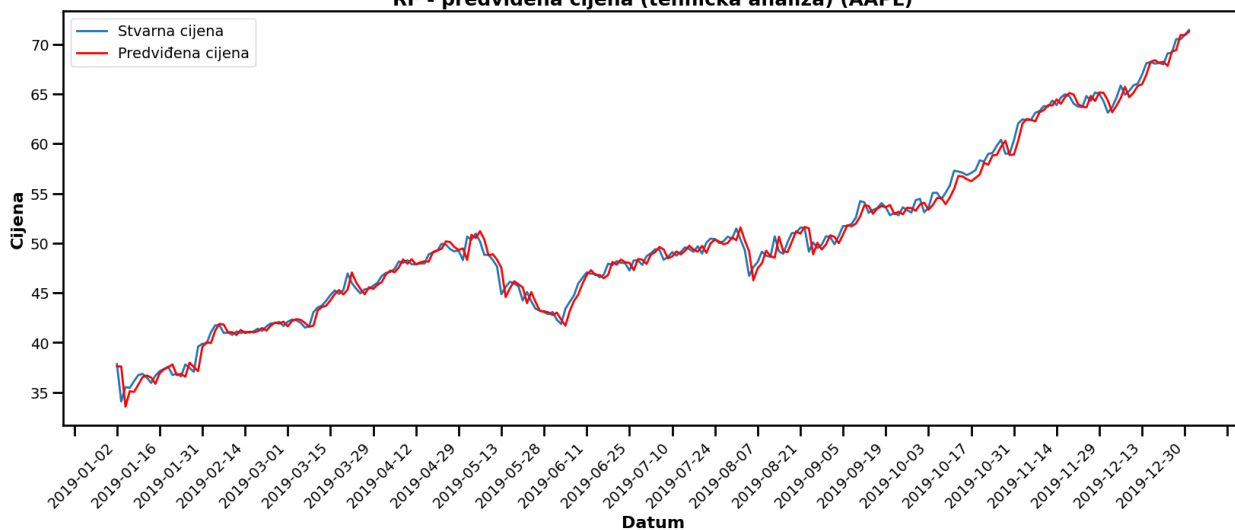
Predviđena cijena

SVM - predviđena cijena (tehnička analiza) (AAPL)

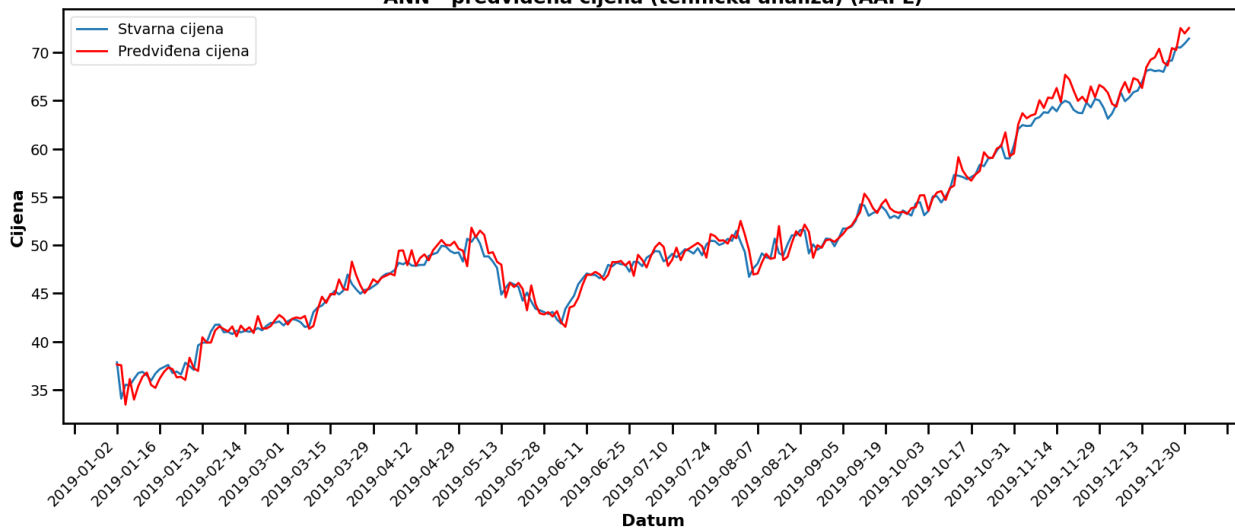




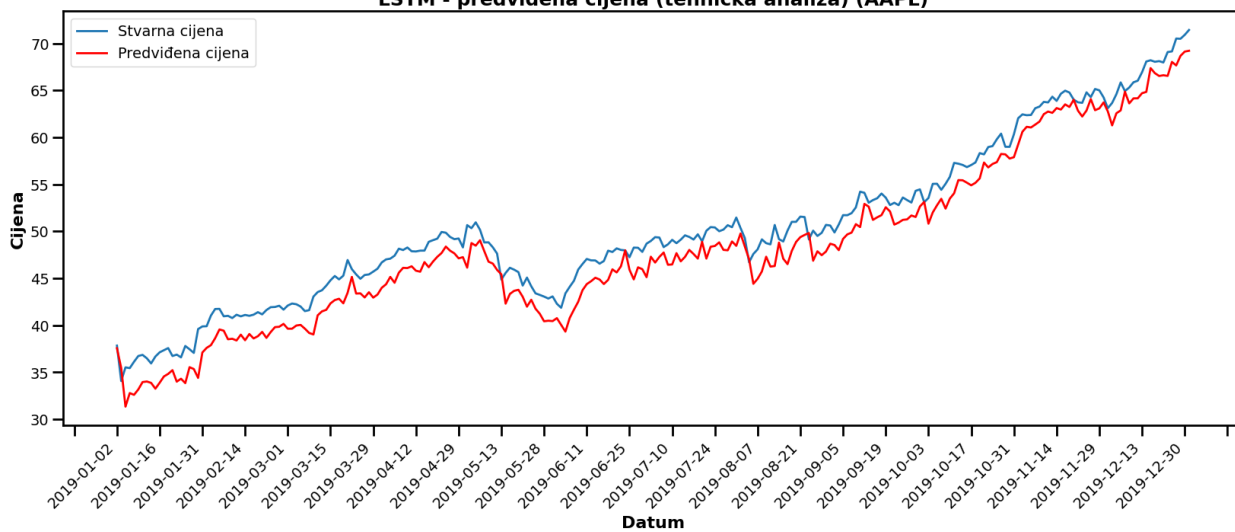
**RF - predviđena cijena (tehnička analiza) (AAPL)**



**ANN - predviđena cijena (tehnička analiza) (AAPL)**



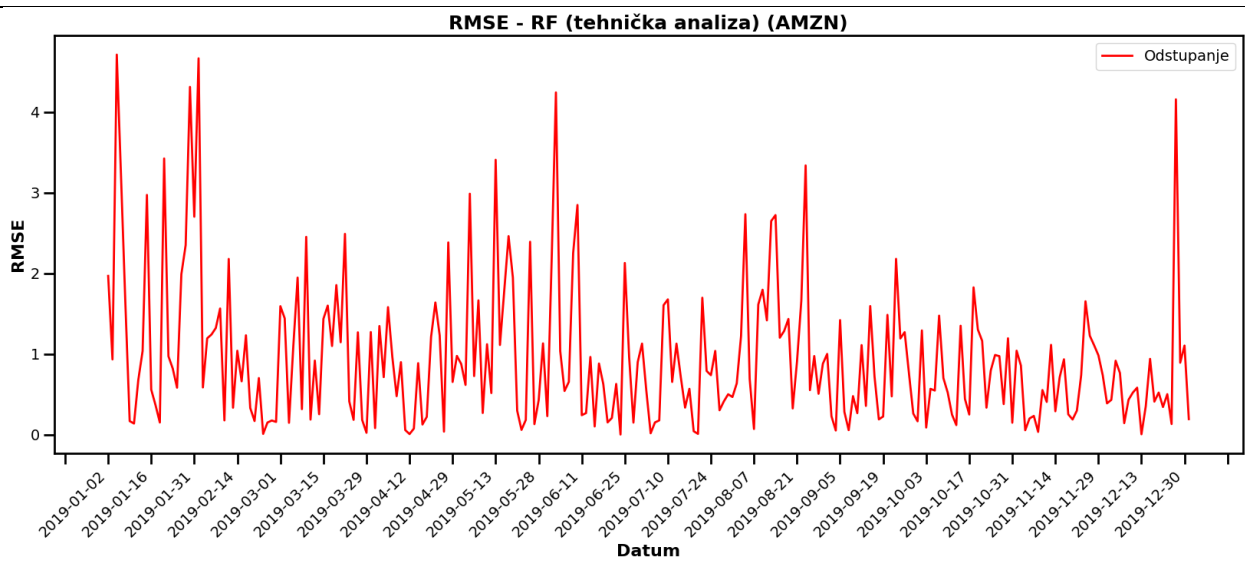
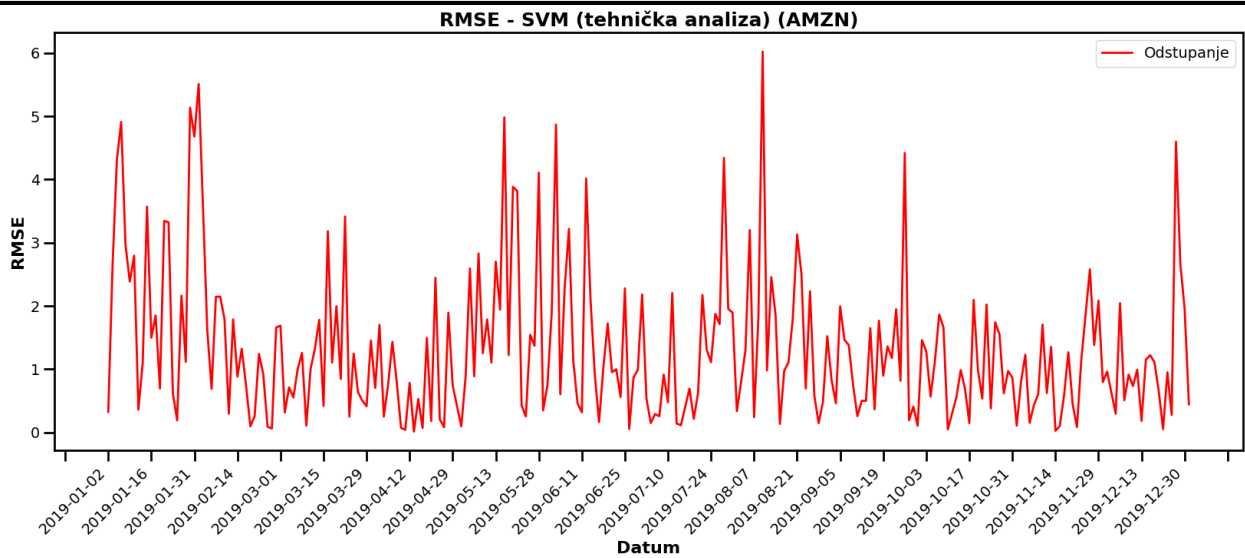
**LSTM - predviđena cijena (tehnička analiza) (AAPL)**

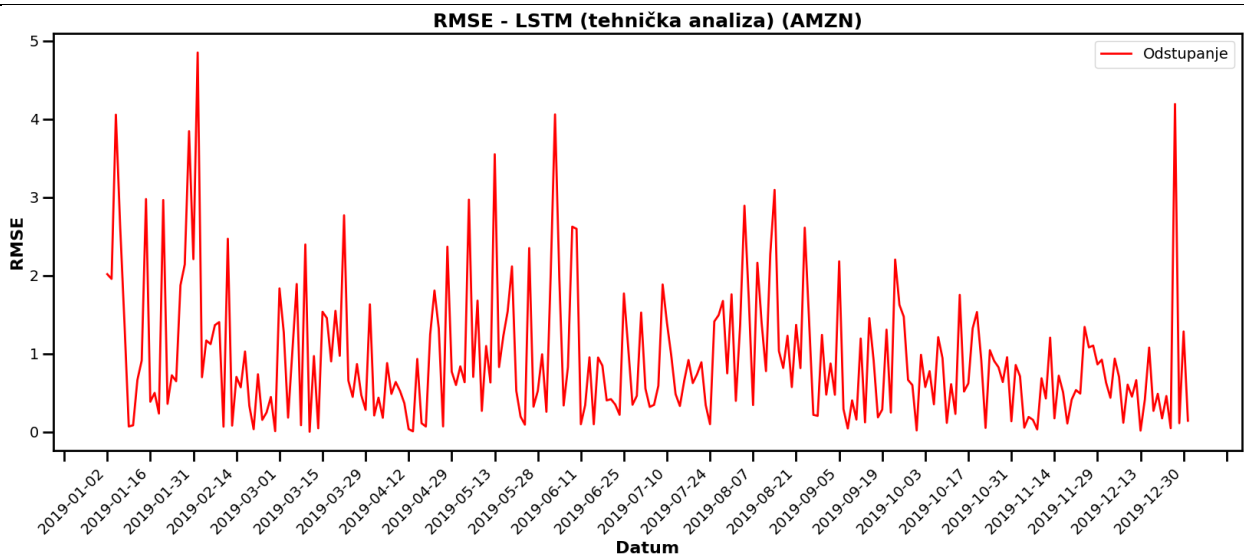
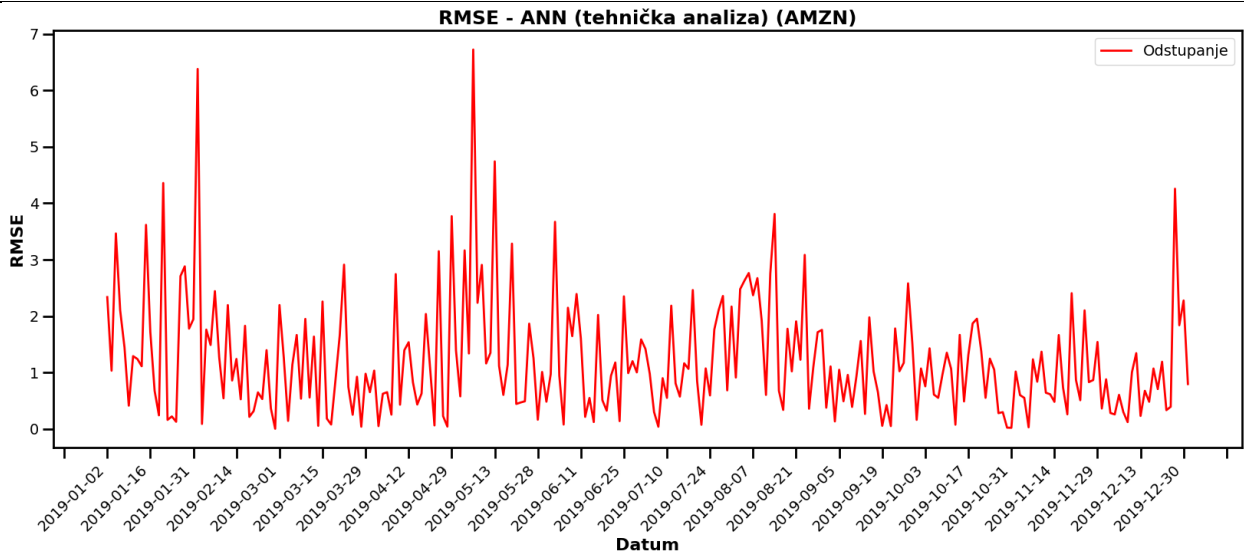


**Prilog P.5.** Grafovi s prikazom rezultata modela za tehničku analizu za tvrtku Amazon

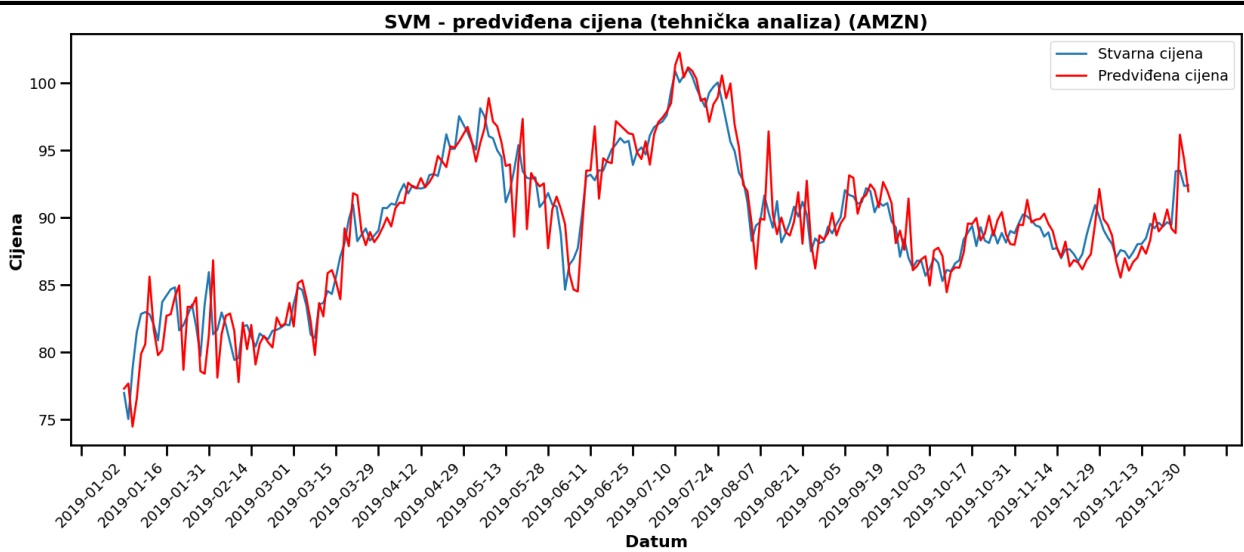
**Amazon**

**RMSE**

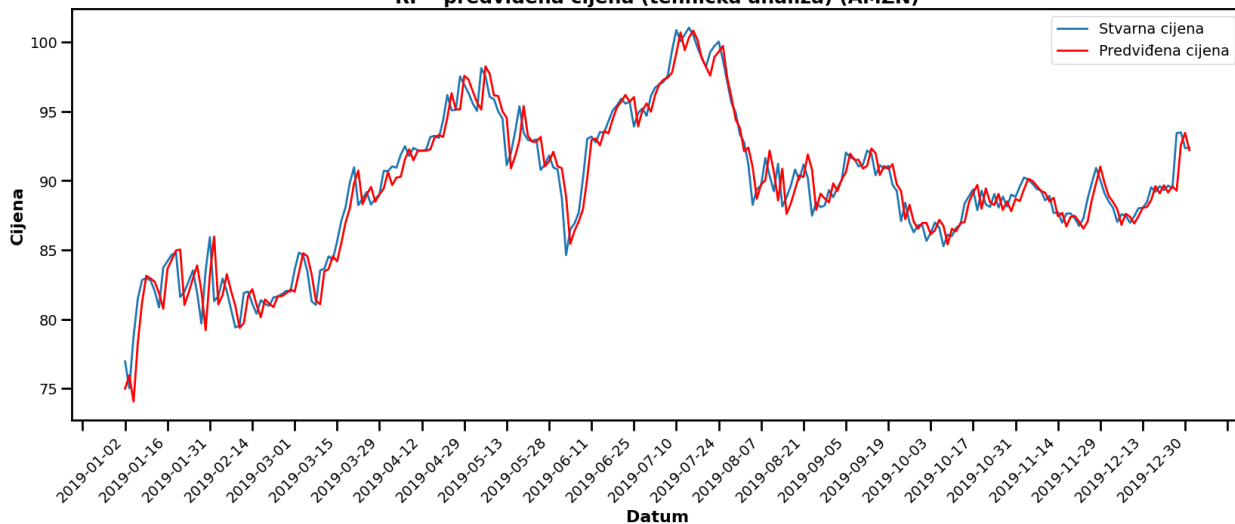




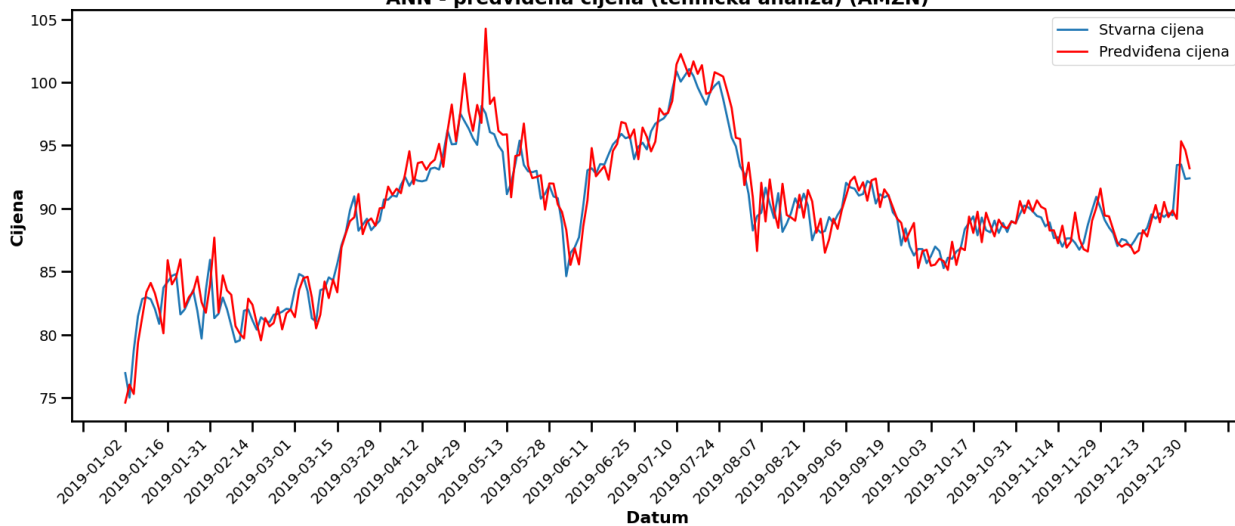
**Predviđena cijena**



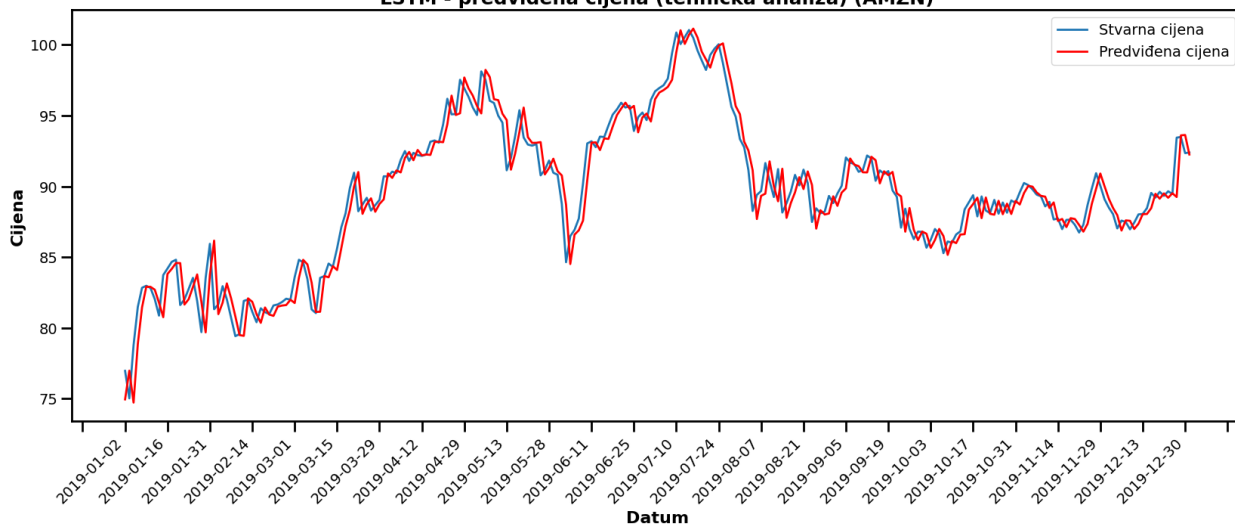
**RF - predviđena cijena (tehnička analiza) (AMZN)**



**ANN - predviđena cijena (tehnička analiza) (AMZN)**



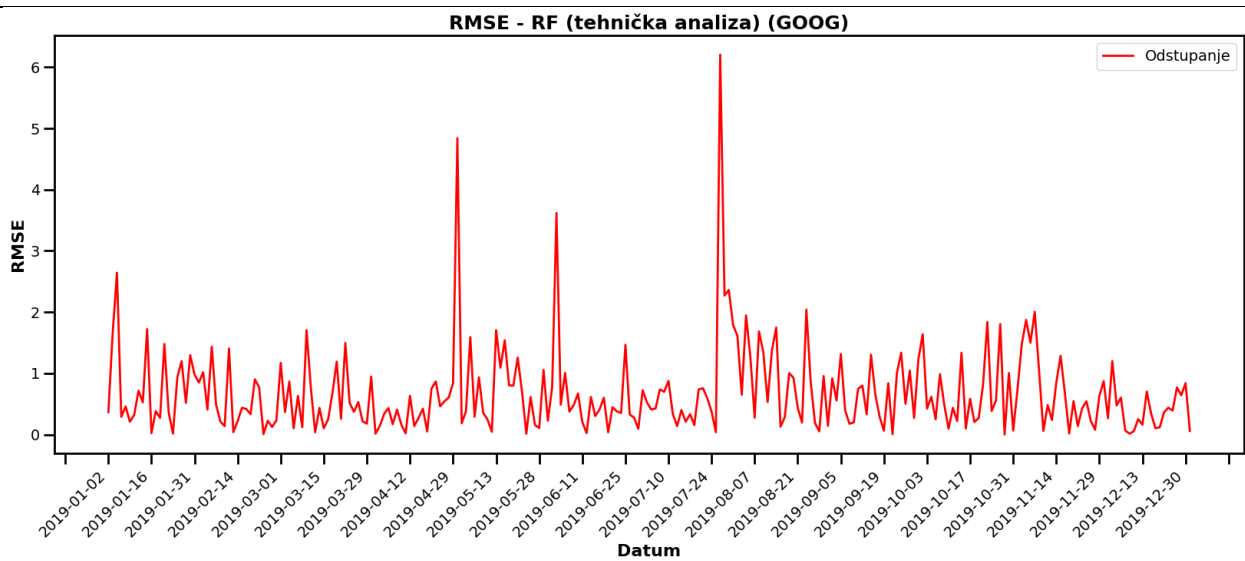
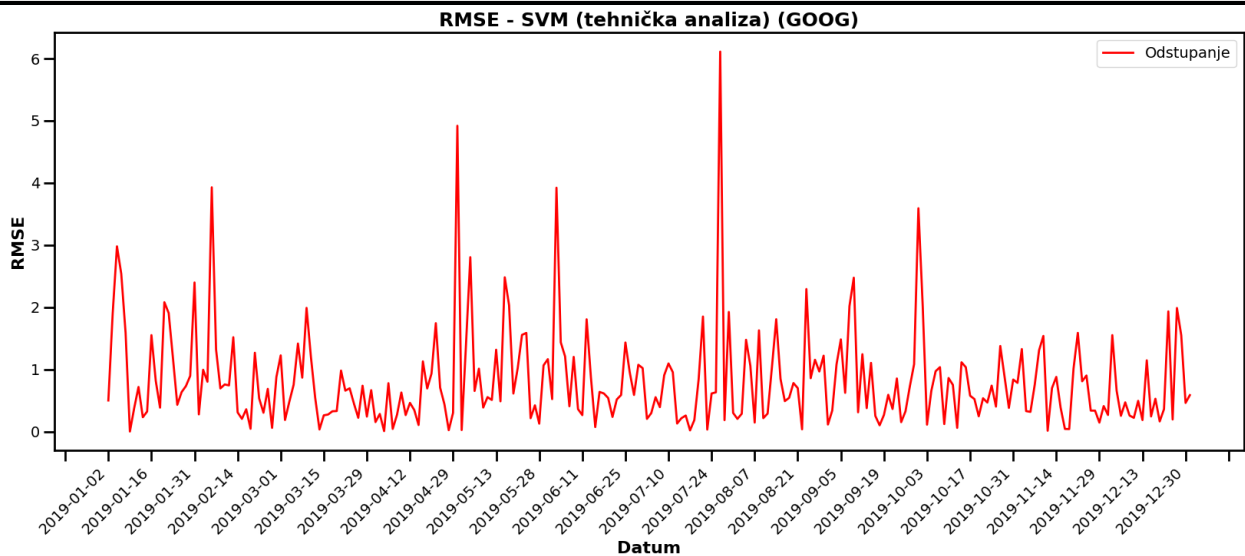
**LSTM - predviđena cijena (tehnička analiza) (AMZN)**



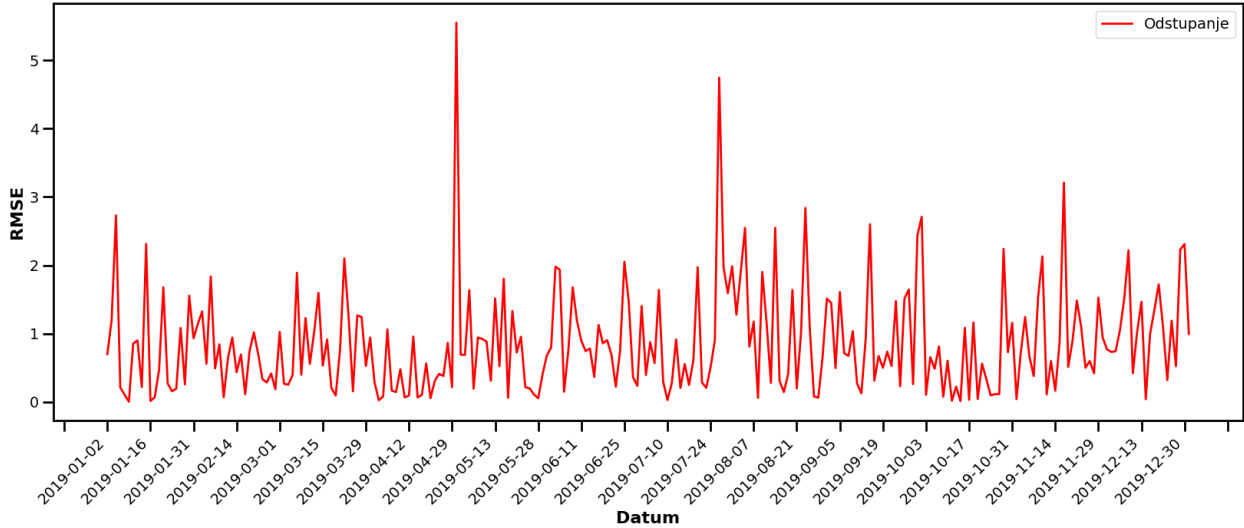
**Prilog P.6.** Grafovi s prikazom rezultata modela za tehničku analizu za tvrtku Apple

**Google**

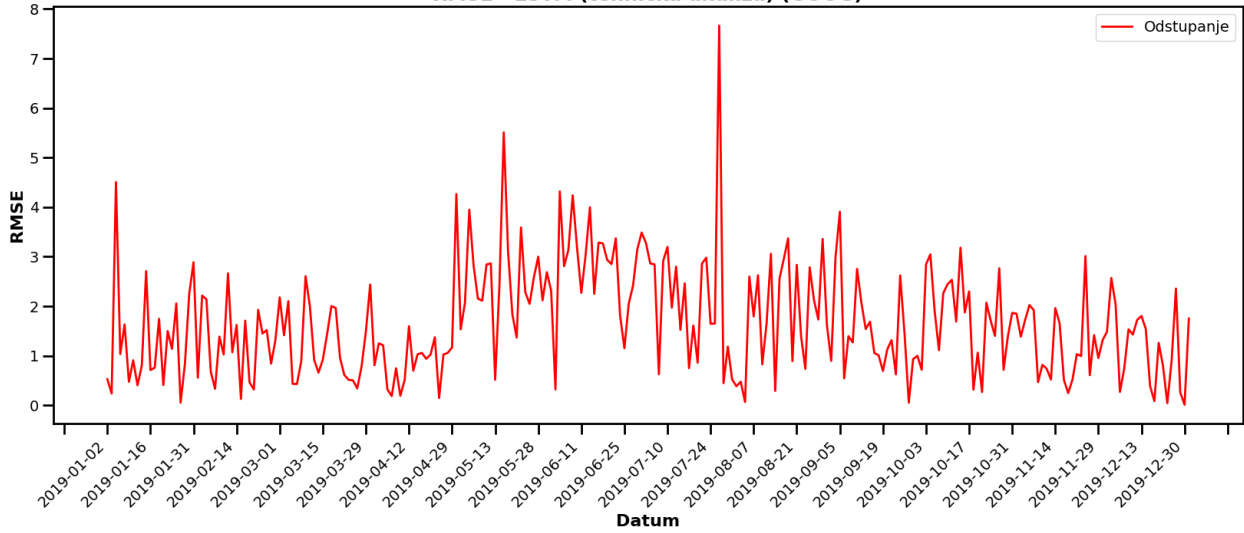
**RMSE**



RMSE - ANN (tehnička analiza) (GOOG)

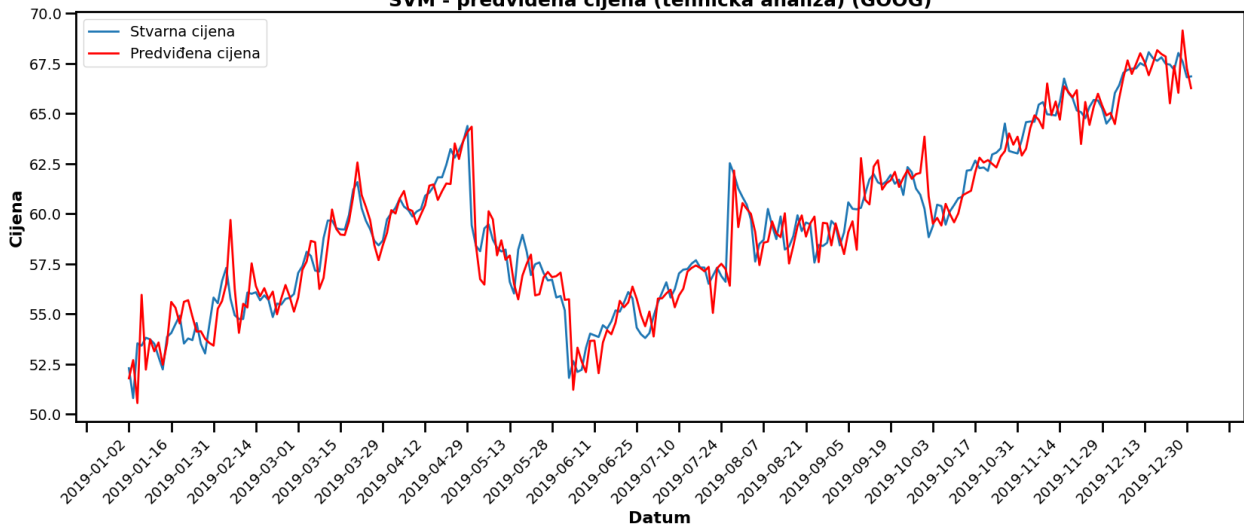


RMSE - LSTM (tehnička analiza) (GOOG)

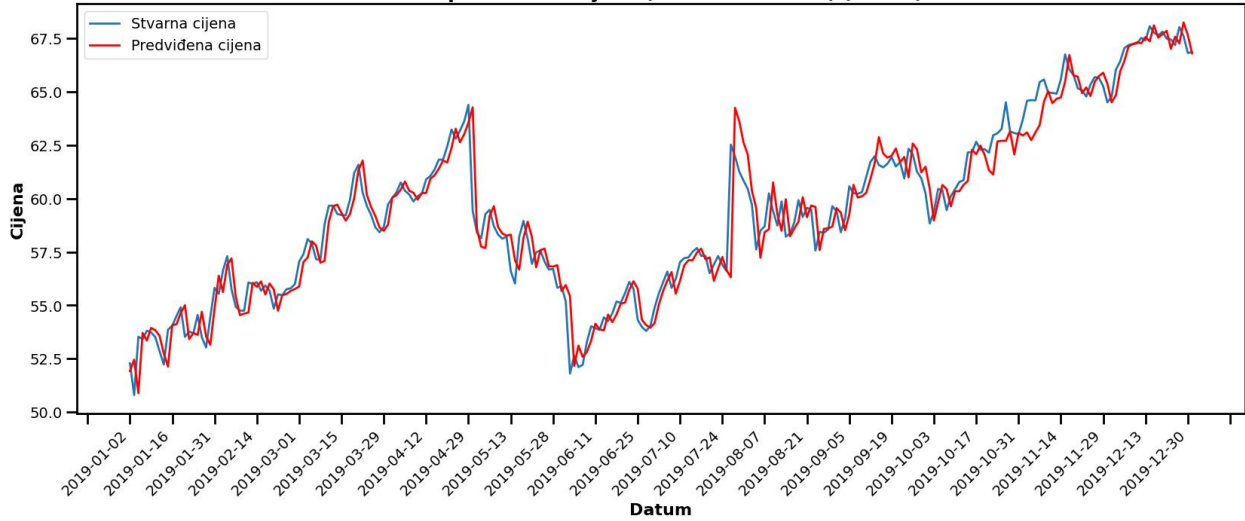


Predviđena cijena

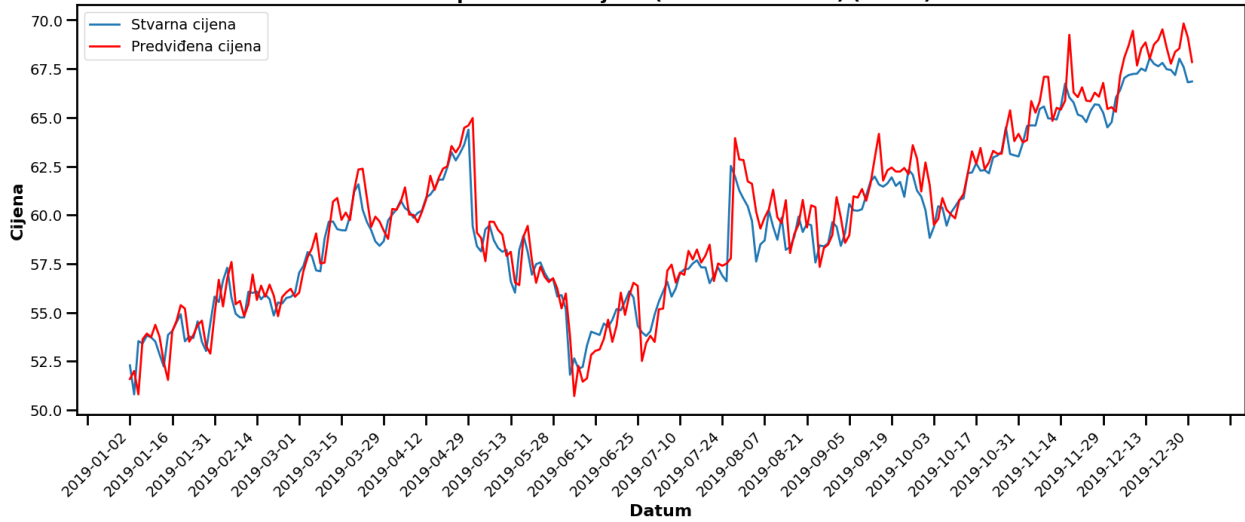
SVM - predviđena cijena (tehnička analiza) (GOOG)



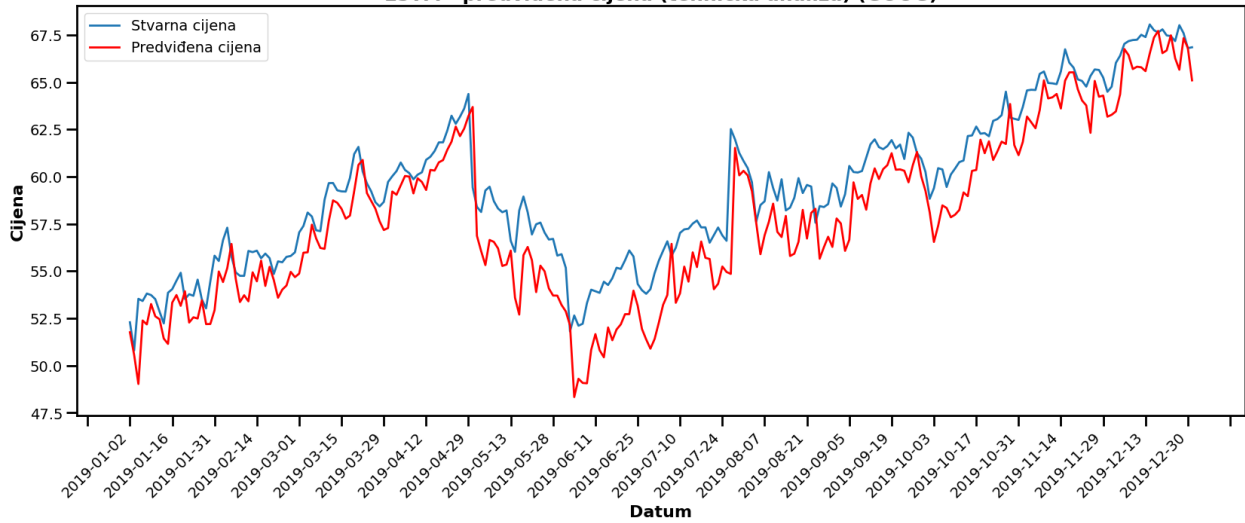
**RF - predviđena cijena (tehnička analiza) (GOOG)**



**ANN - predviđena cijena (tehnička analiza) (GOOG)**



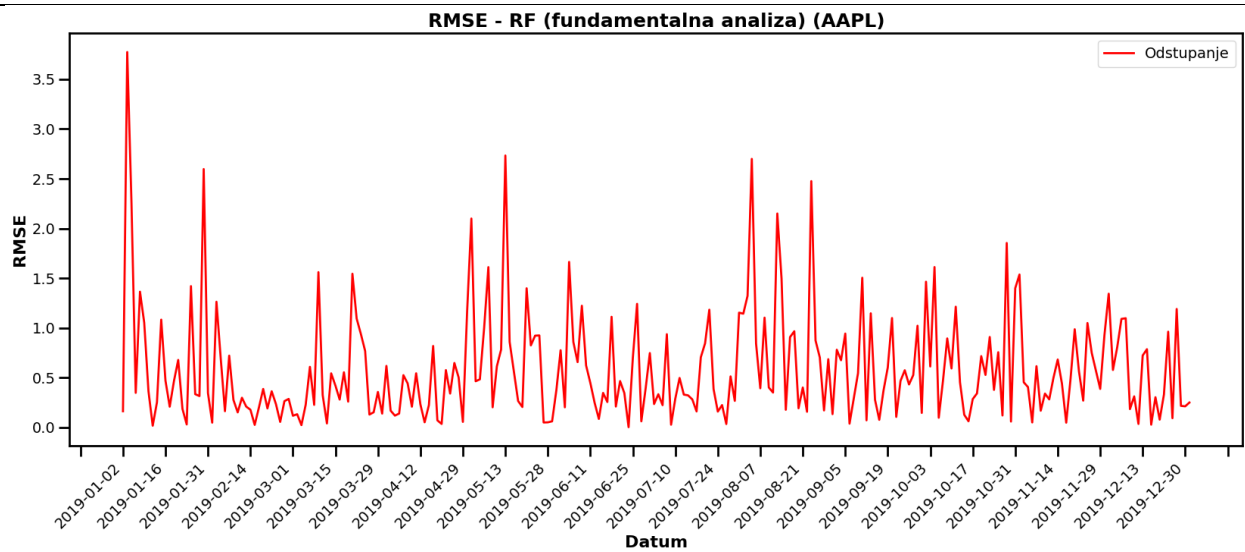
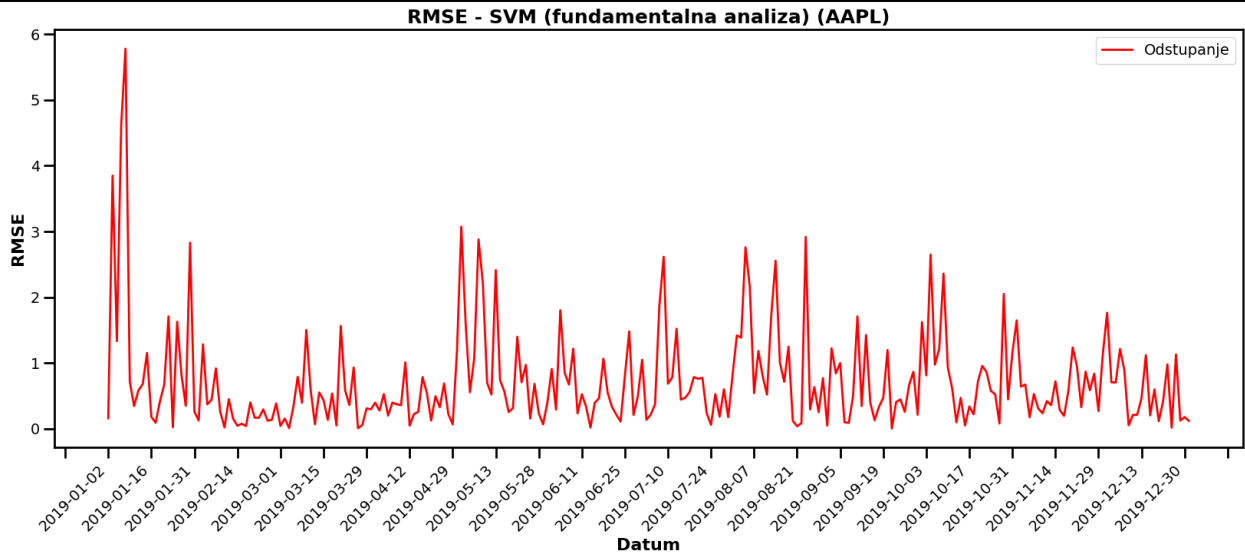
**LSTM - predviđena cijena (tehnička analiza) (GOOG)**



**Prilog P.7.** Grafovi s prikazom rezultata modela za fundamentalnu analizu za tvrtku Apple

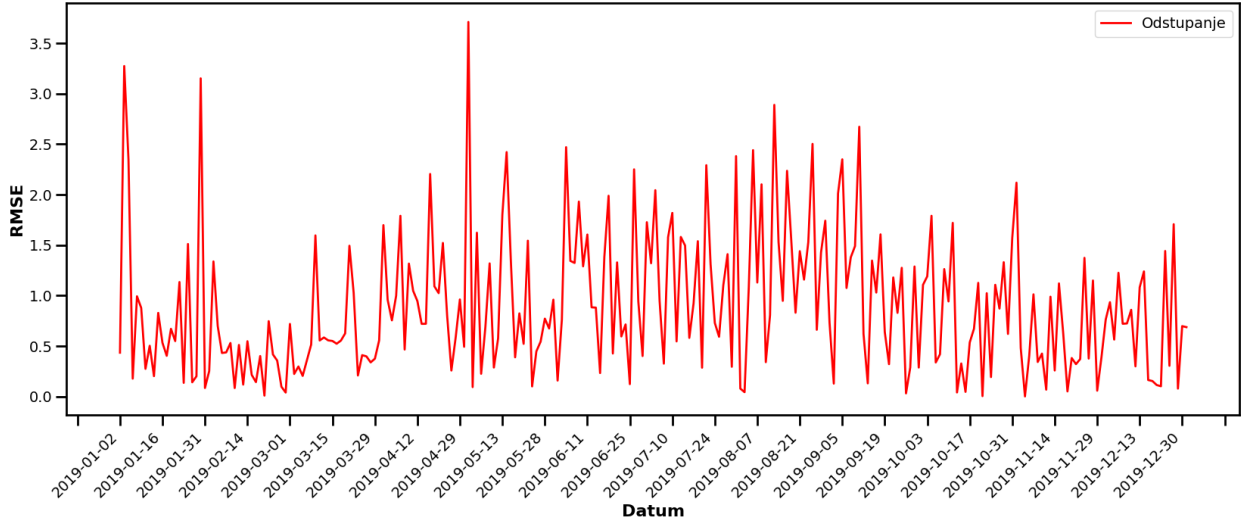
**Apple**

**RMSE**

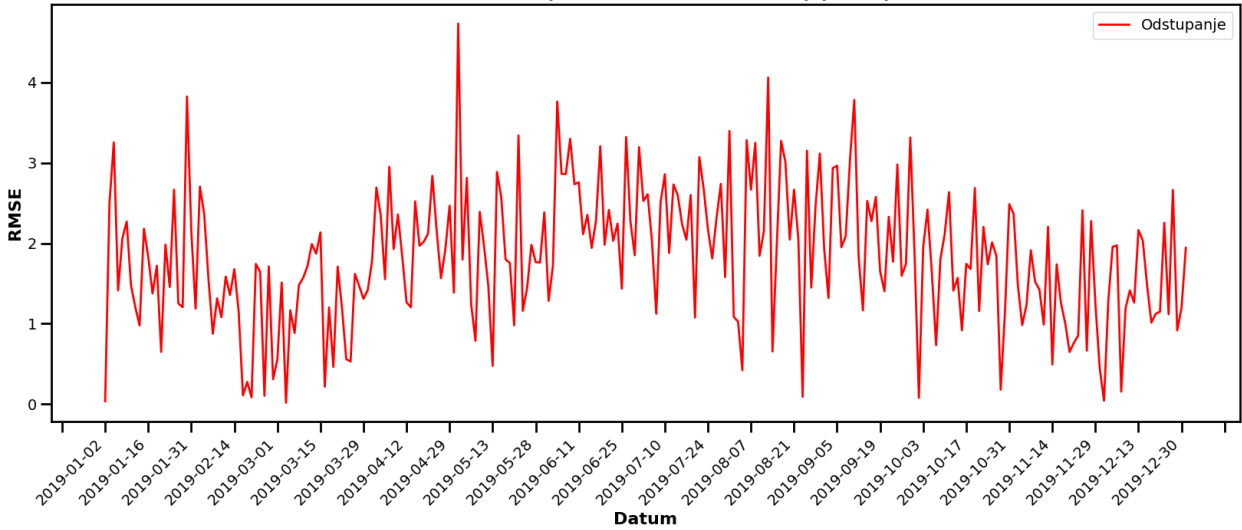




RMSE - ANN (fundamentalna analiza) (AAPL)

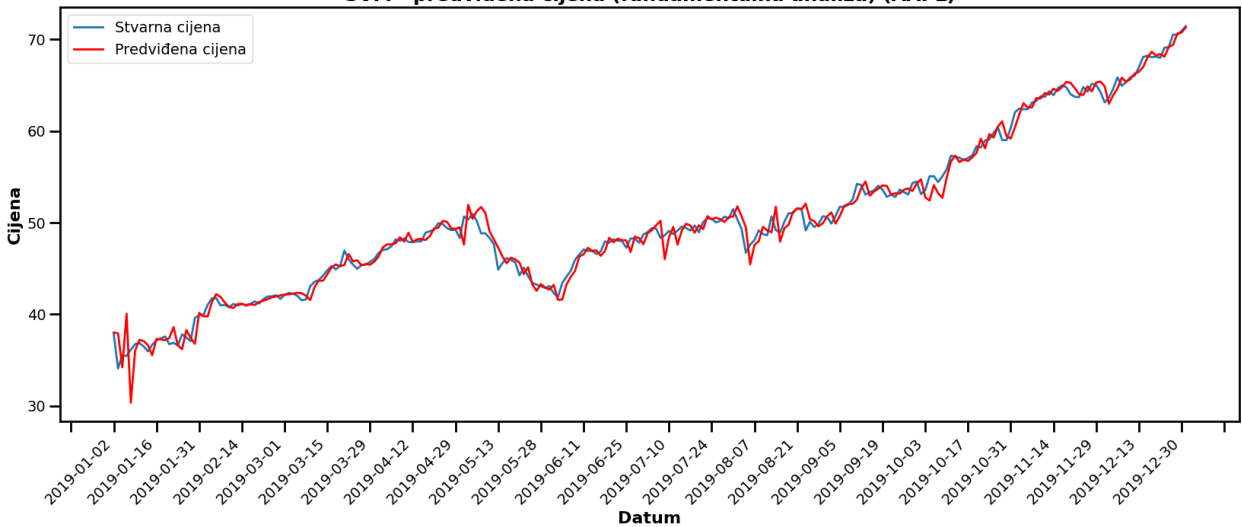


RMSE - LSTM (fundamentalna analiza) (AAPL)

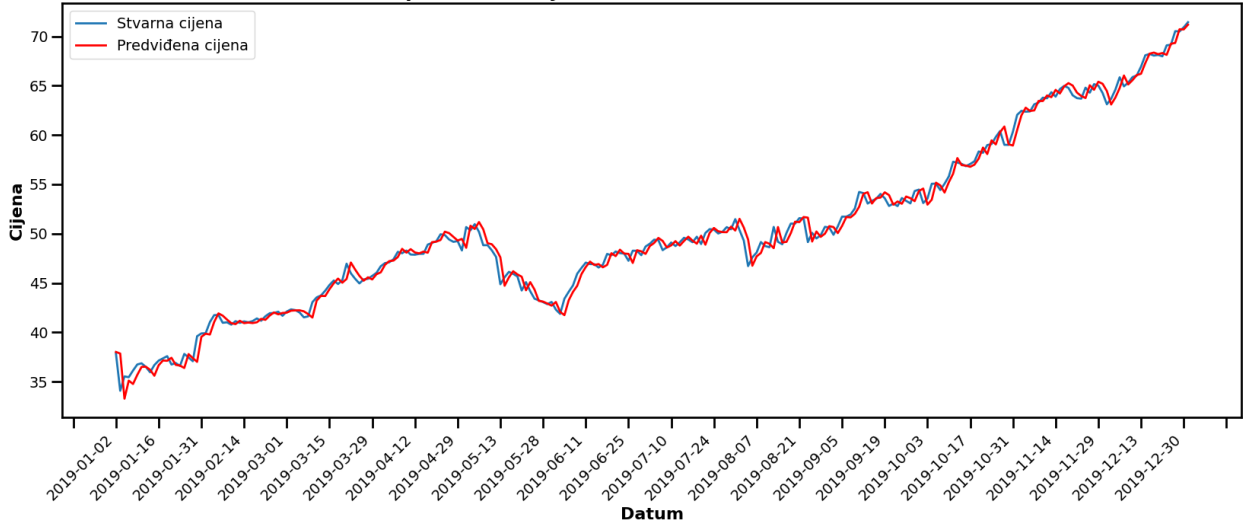


Predviđena cijena

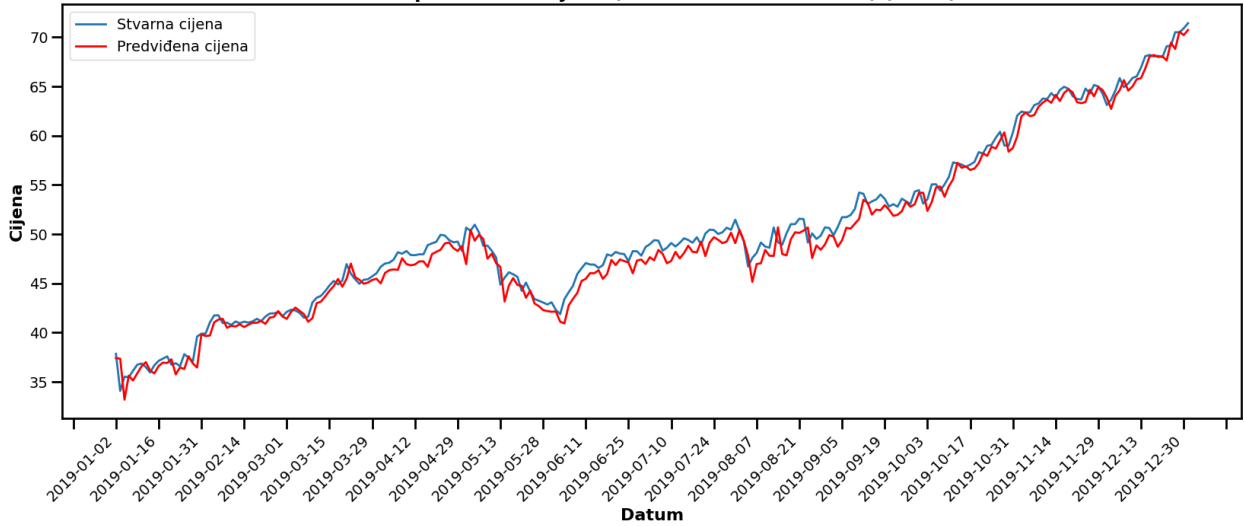
SVM - predviđena cijena (fundamentalna analiza) (AAPL)



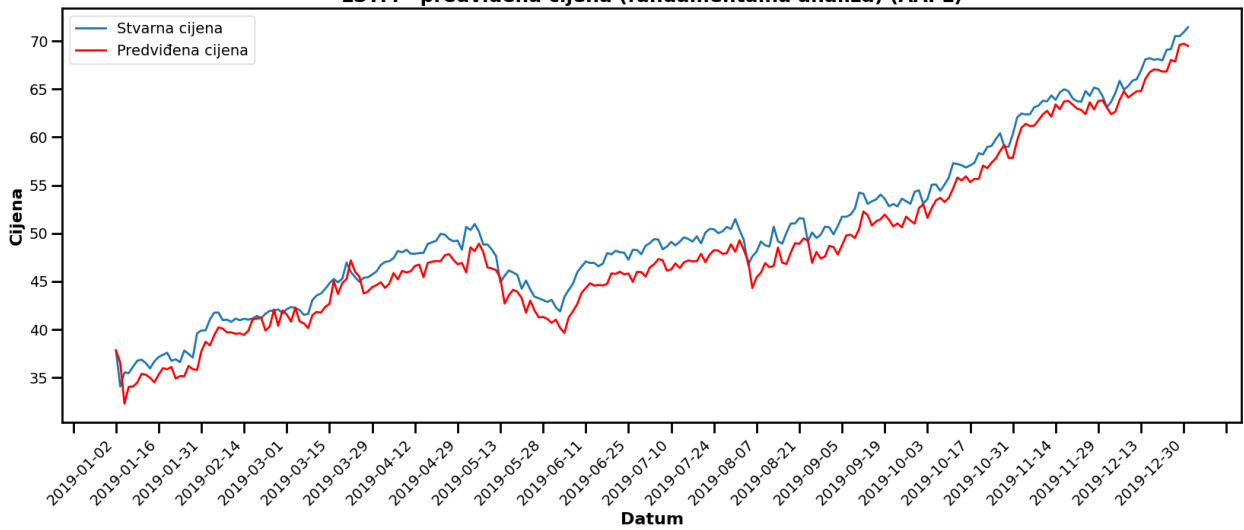
**RF - predviđena cijena (fundamentalna analiza) (AAPL)**



**ANN - predviđena cijena (fundamentalna analiza) (AAPL)**



**LSTM - predviđena cijena (fundamentalna analiza) (AAPL)**

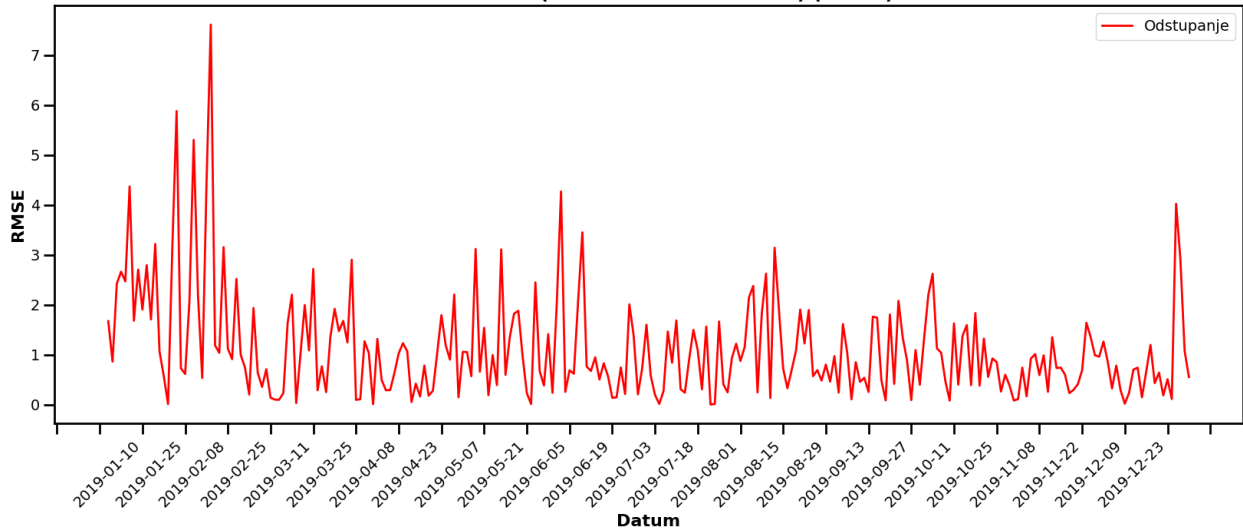


**Prilog P.8.** Grafovi s prikazom rezultata modela za fundamentalnu analizu za tvrtku Amazon

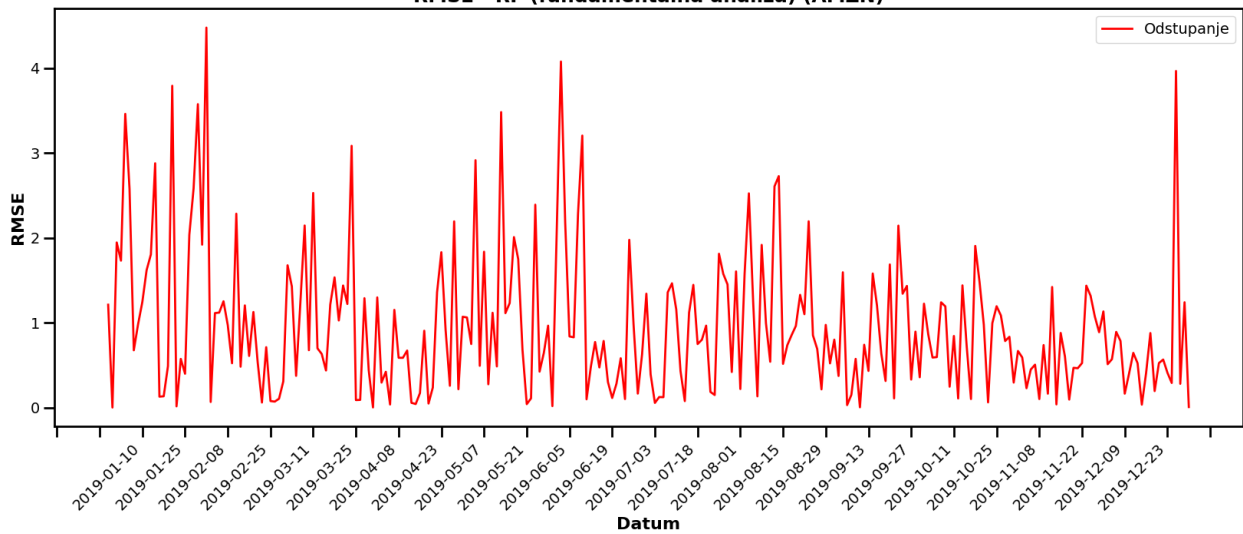
**Amazon**

**RMSE**

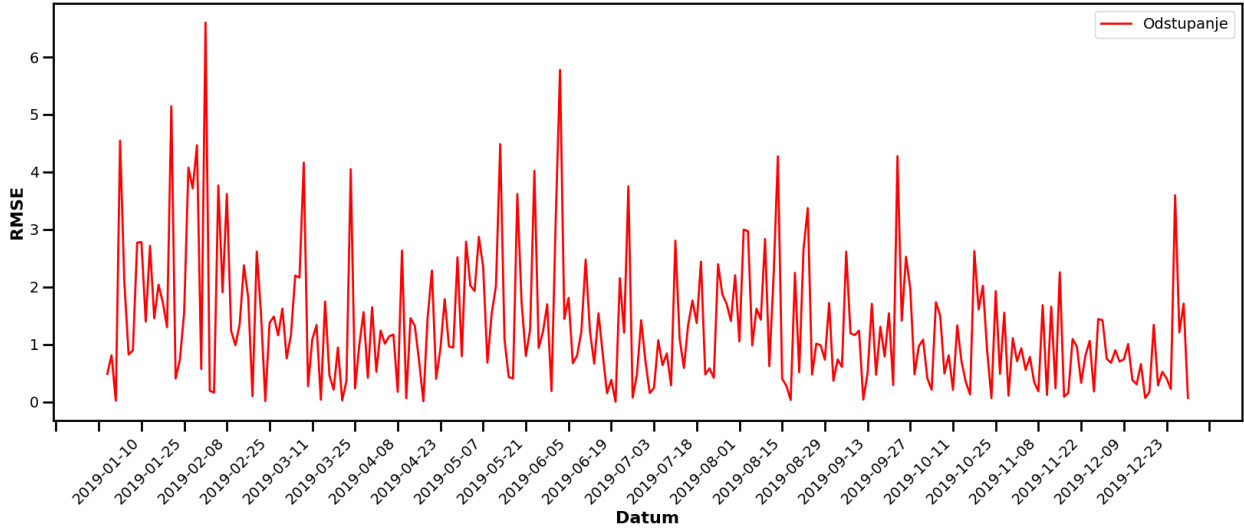
**RMSE - SVM (fundamentalna analiza) (AMZN)**



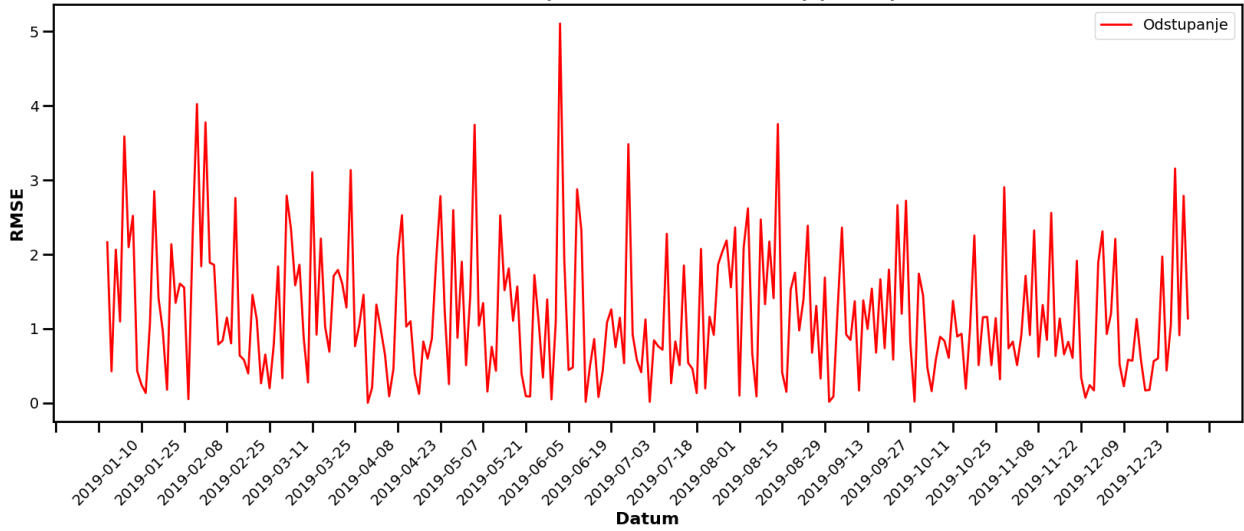
**RMSE - RF (fundamentalna analiza) (AMZN)**



**RMSE - ANN (fundamentalna analiza) (AMZN)**

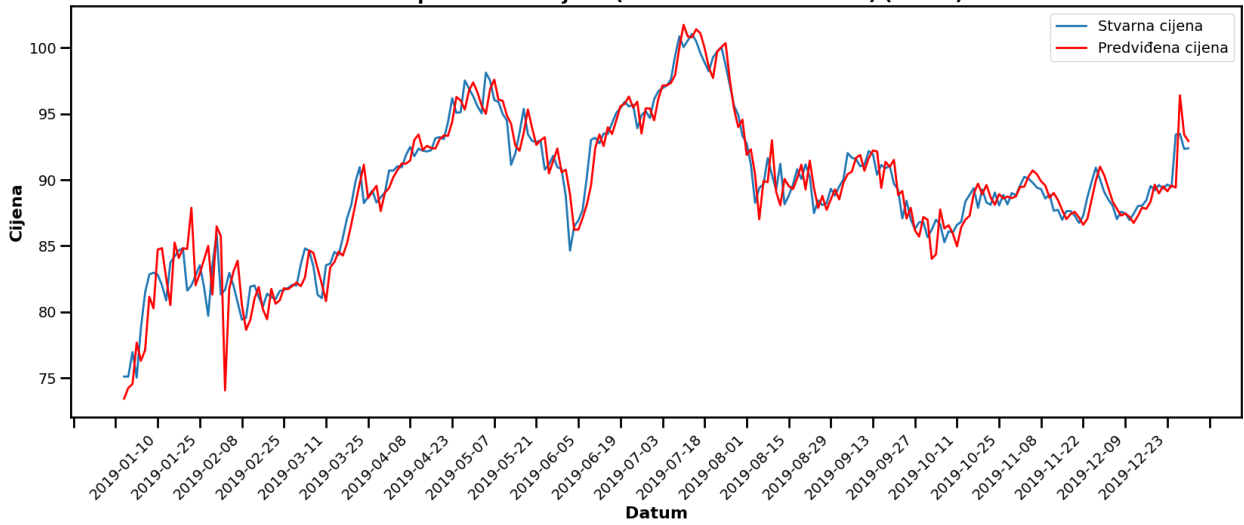


**RMSE - LSTM (fundamentalna analiza) (AMZN)**

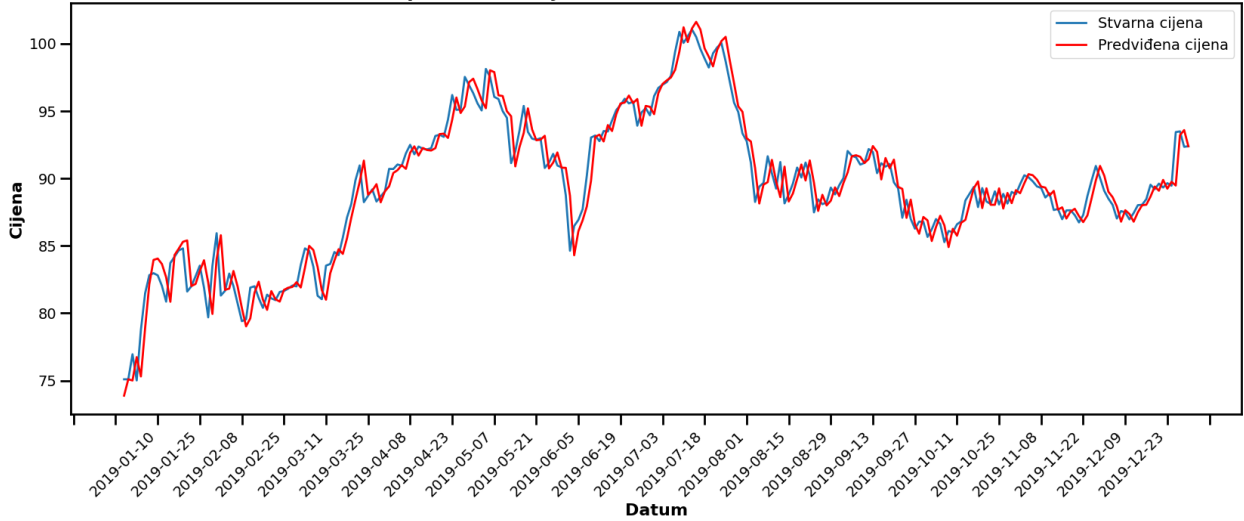


**Predviđena cijena**

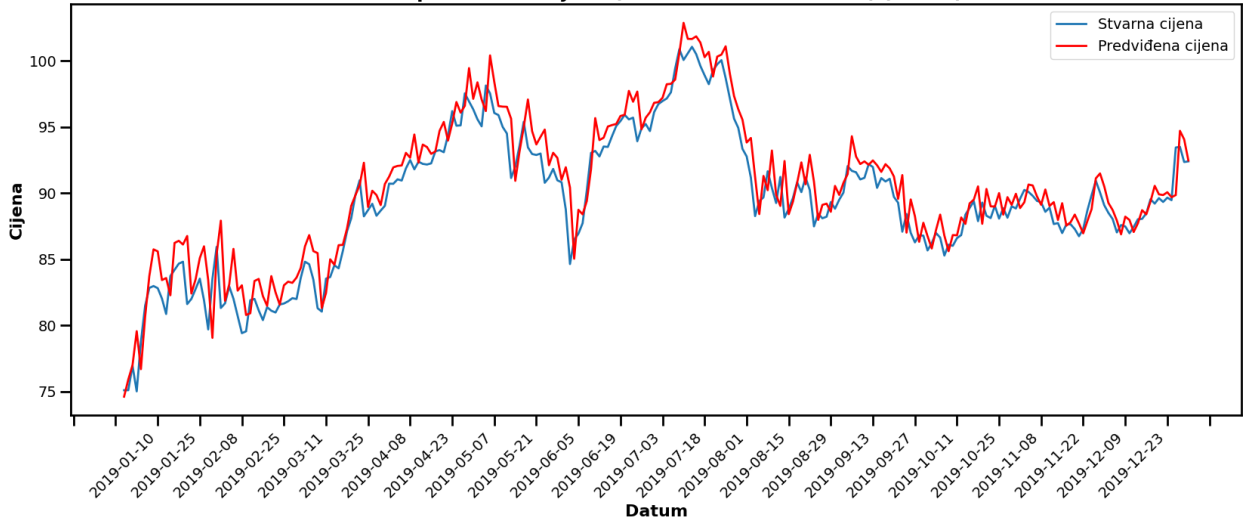
**SVM - predviđena cijena (fundamentalna analiza) (AMZN)**



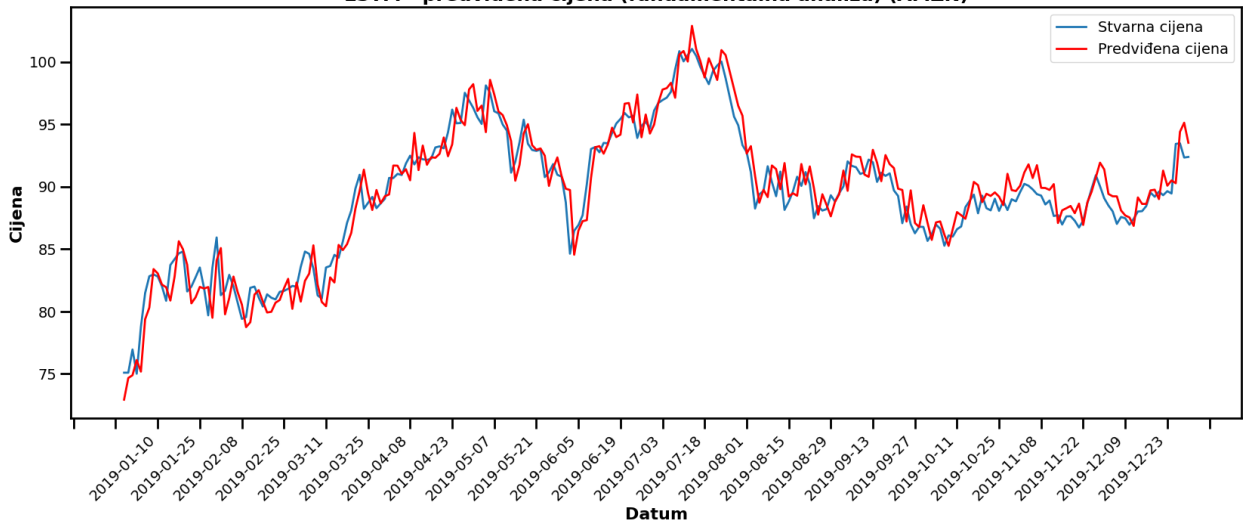
**RF - predviđena cijena (fundamentalna analiza) (AMZN)**



**ANN - predviđena cijena (fundamentalna analiza) (AMZN)**



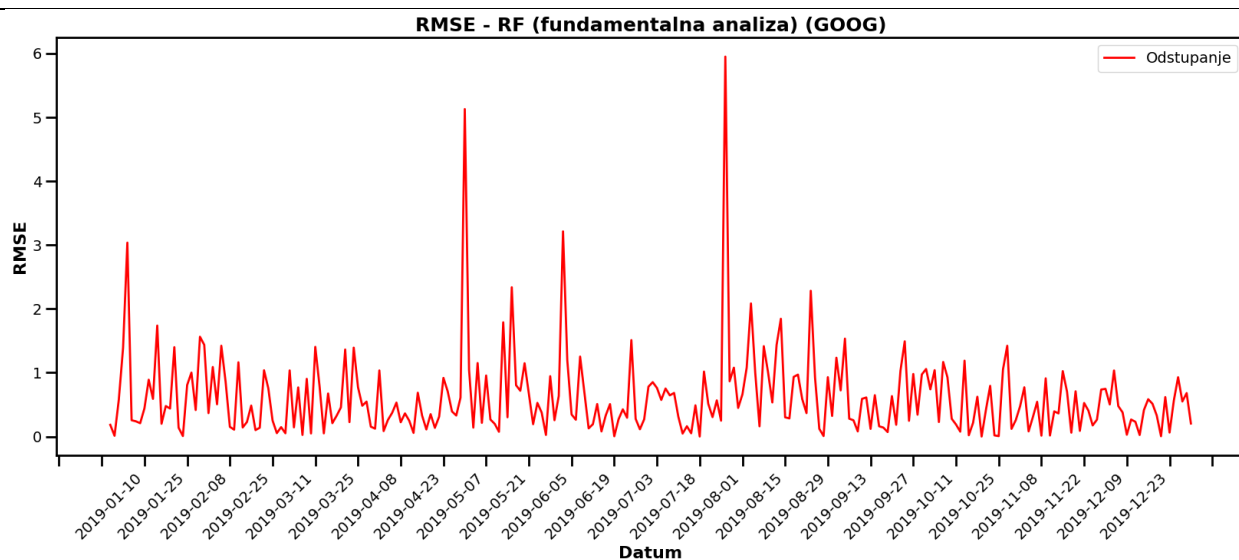
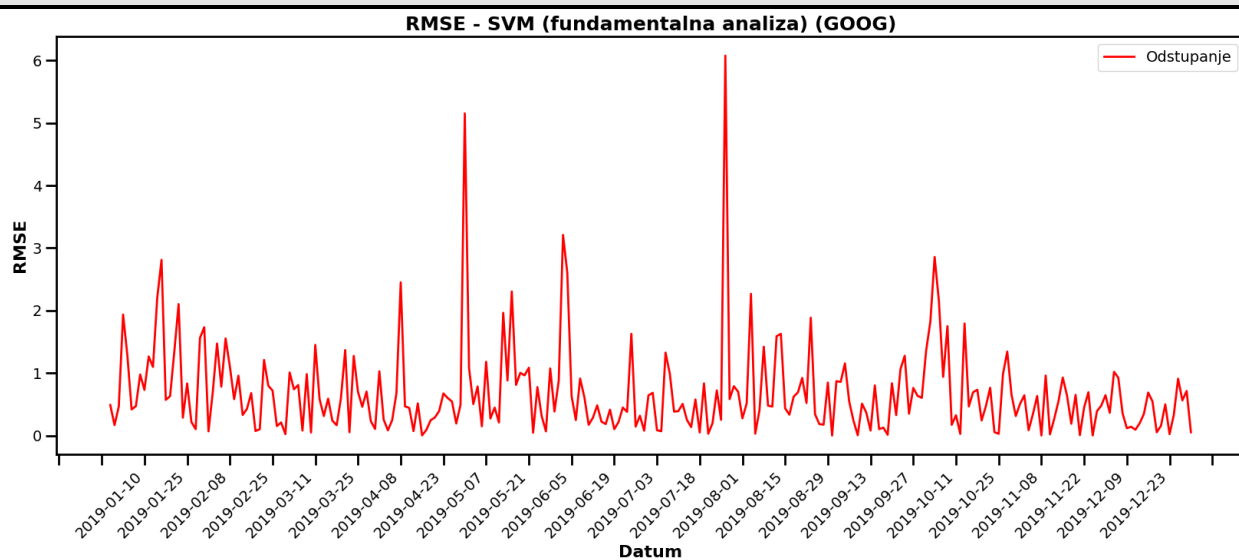
**LSTM - predviđena cijena (fundamentalna analiza) (AMZN)**



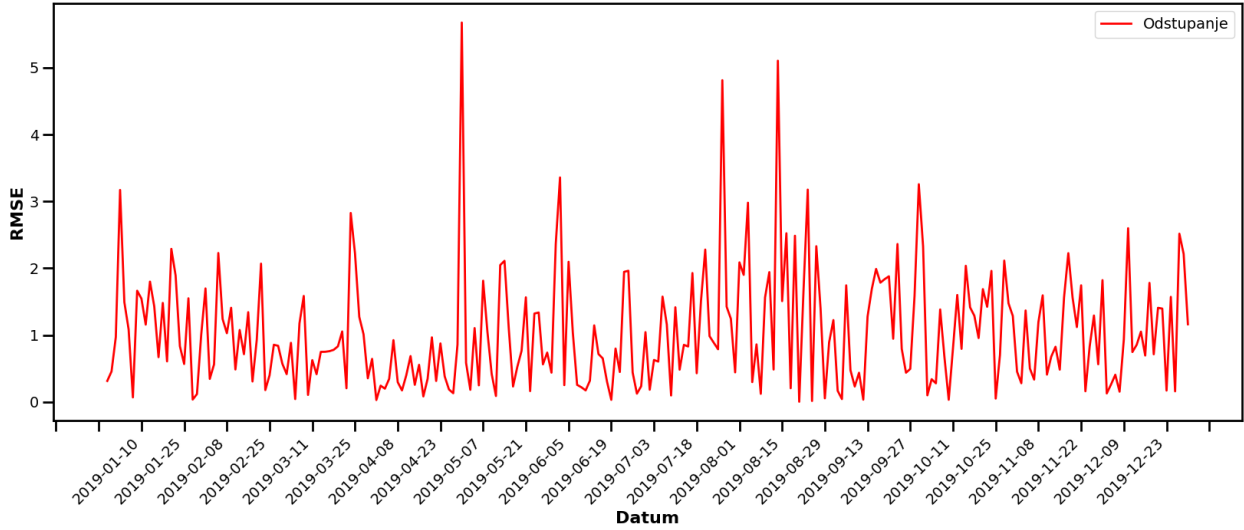
## Prilog P.9. Grafovi s prikazom rezultata modela za fundamentalnu analizu za tvrtku Google

### Google

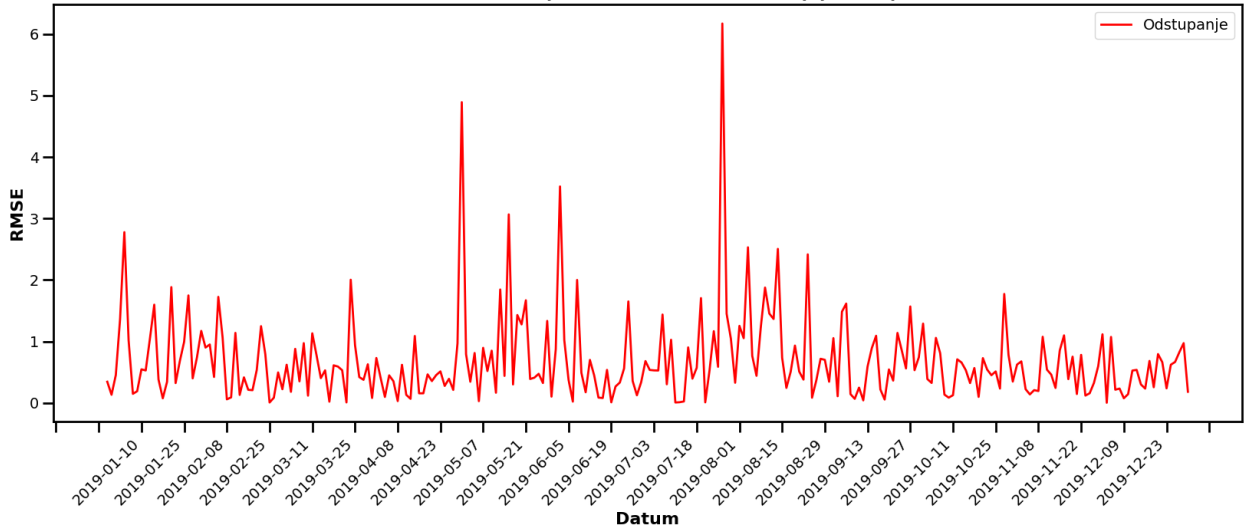
#### RMSE



RMSE - ANN (fundamentalna analiza) (GOOG)

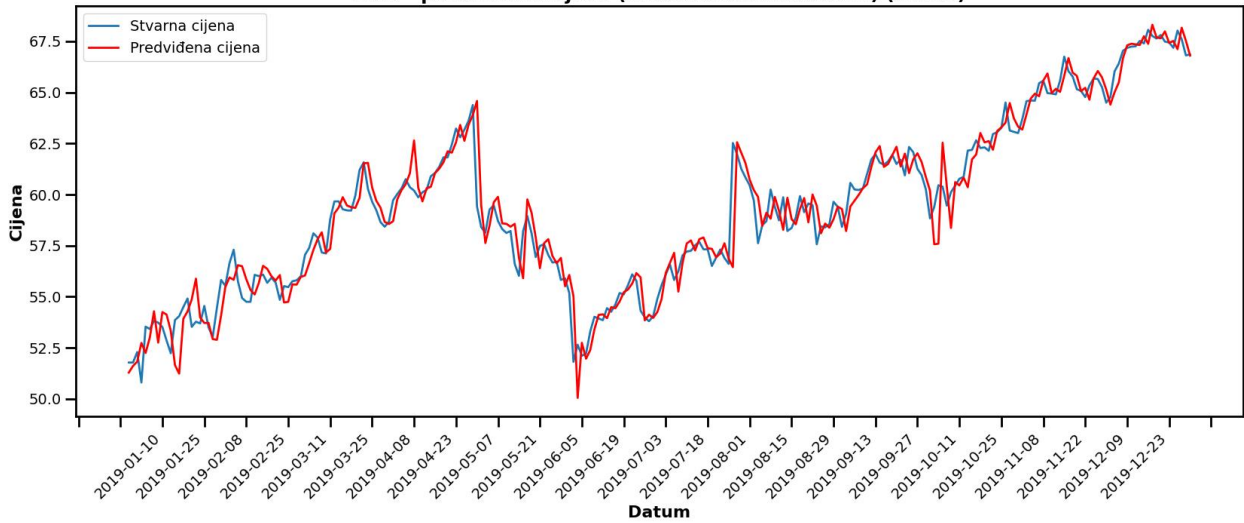


RMSE - LSTM (fundamentalna analiza) (GOOG)

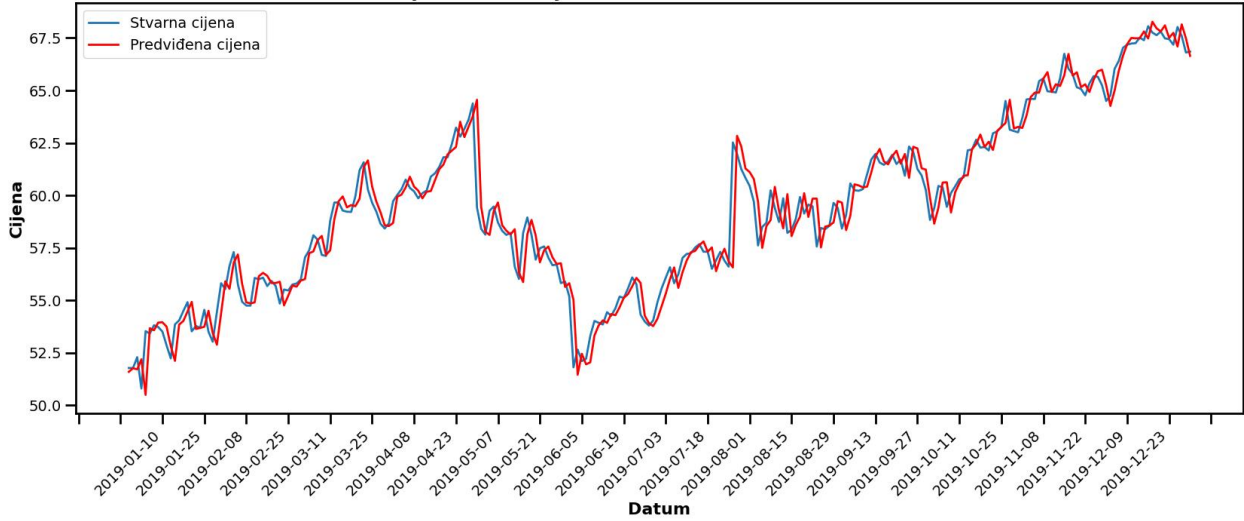


Predviđena cijena

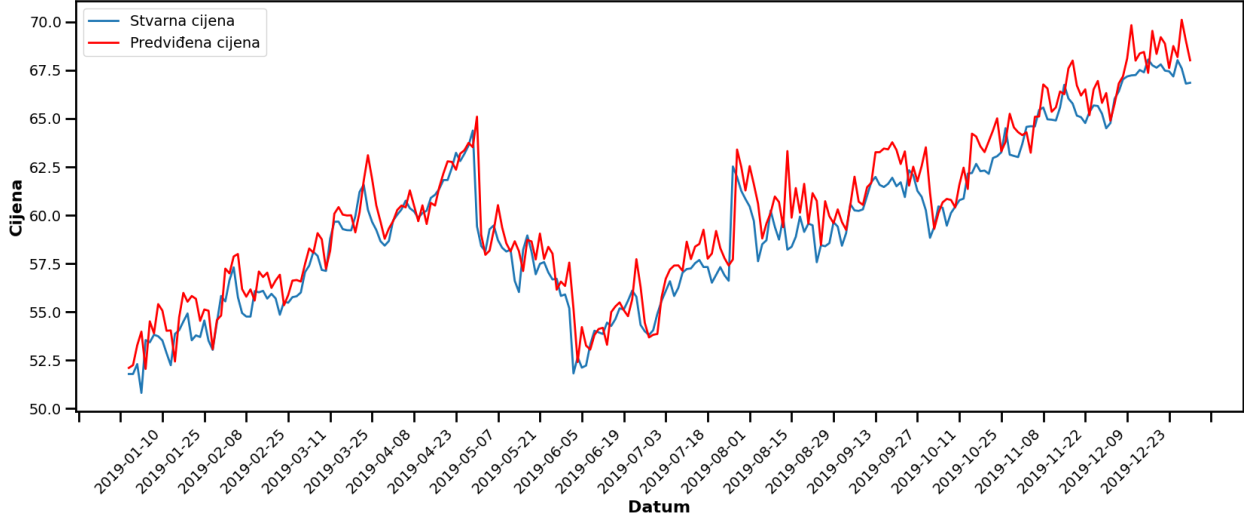
SVM - predviđena cijena (fundamentalna analiza) (GOOG)



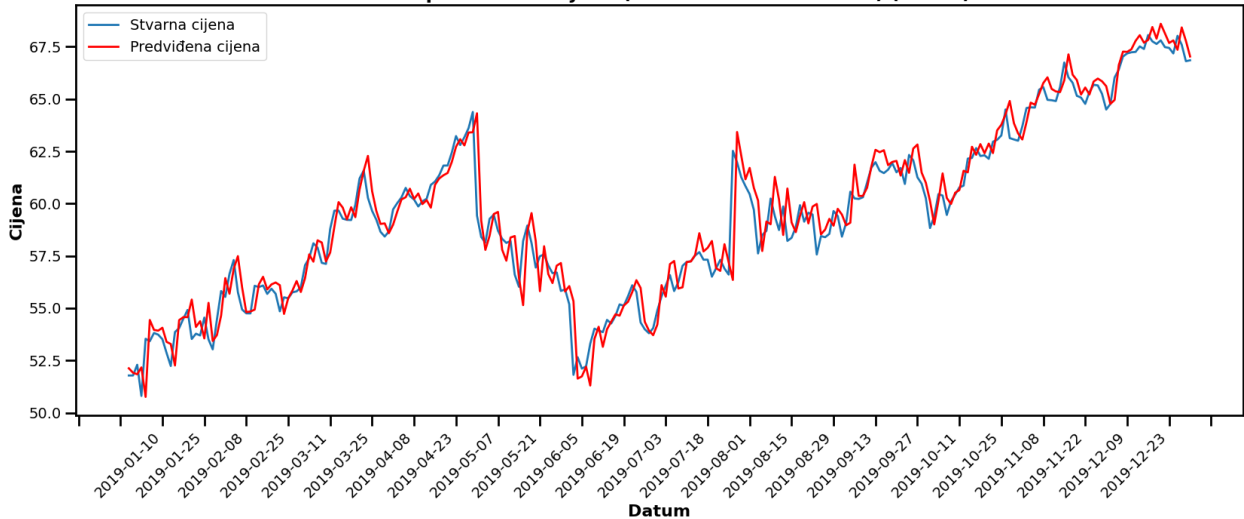
**RF - predviđena cijena (fundamentalna analiza) (GOOG)**



**ANN - predviđena cijena (fundamentalna analiza) (GOOG)**



**LSTM - predviđena cijena (fundamentalna analiza) (GOOG)**

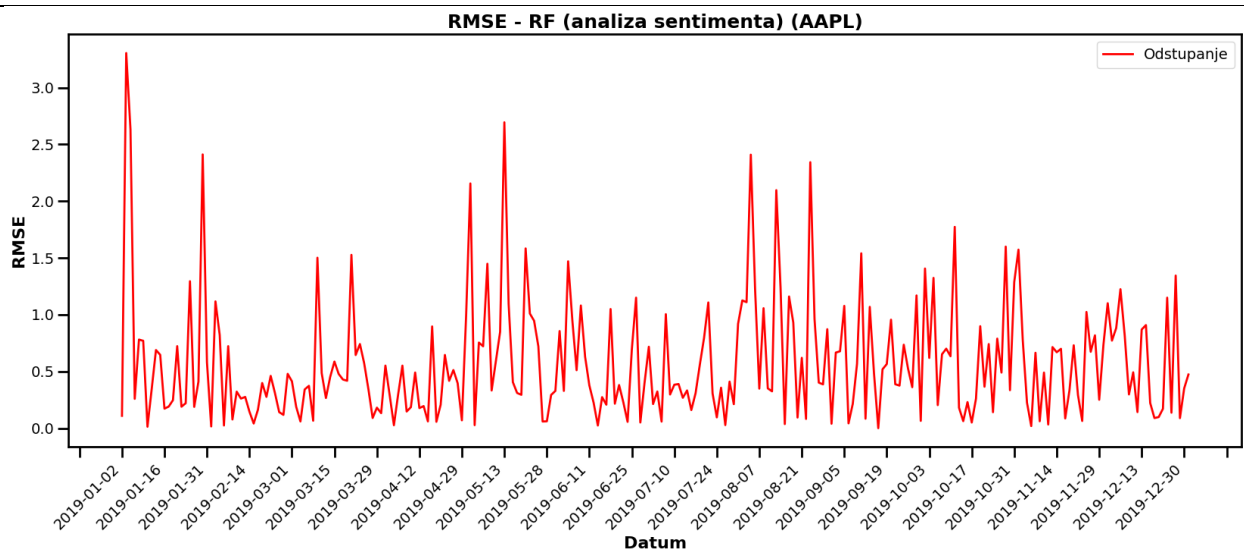
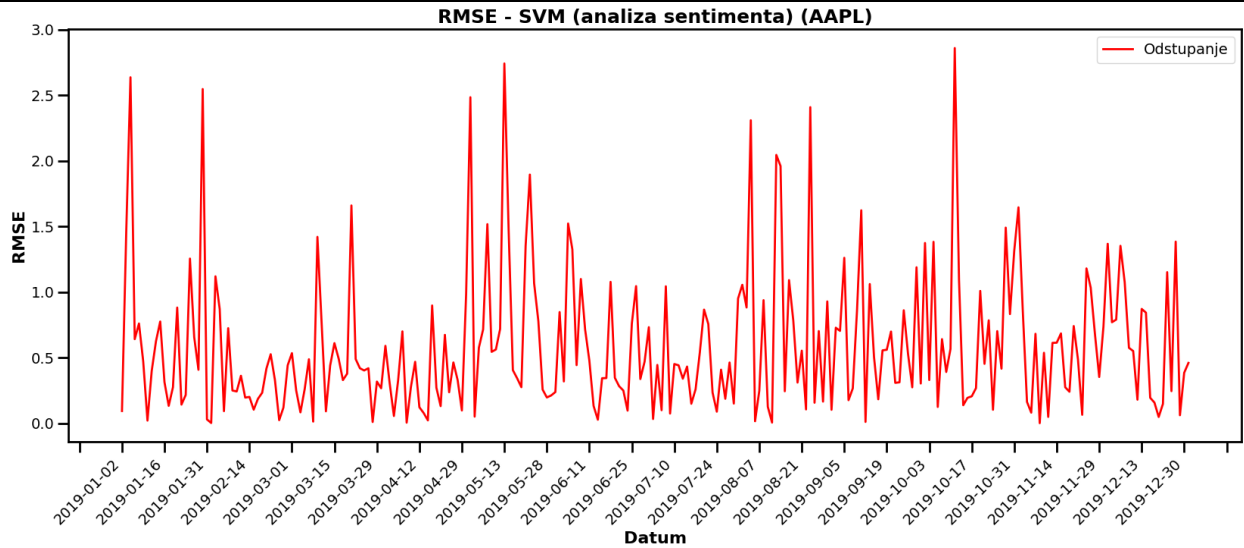




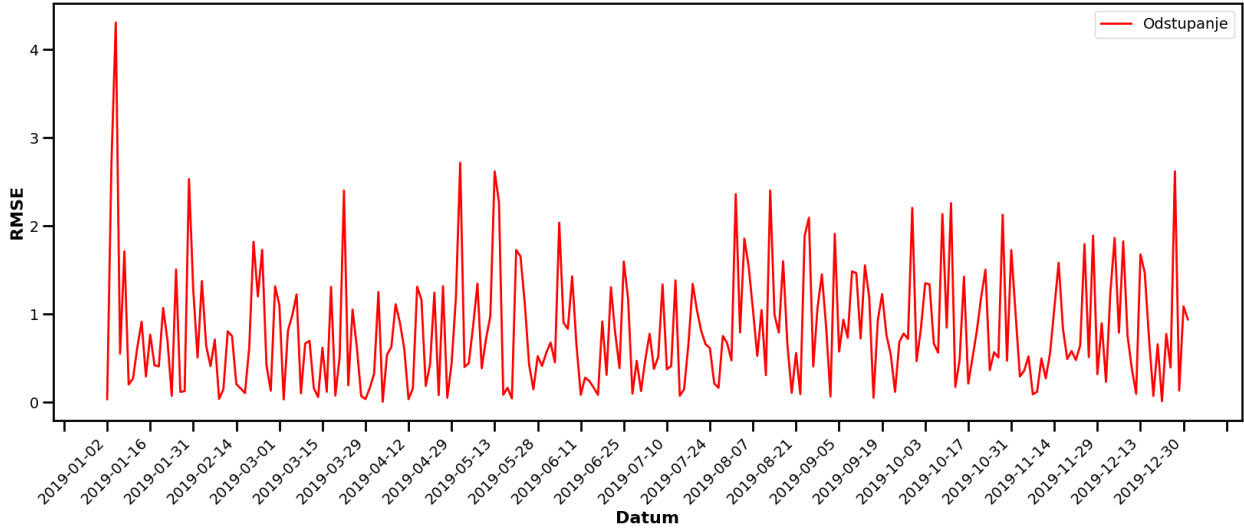
**Prilog P.10.** Grafovi s prikazom rezultata modela za analizu sentimenta za tvrtku Apple

**Apple**

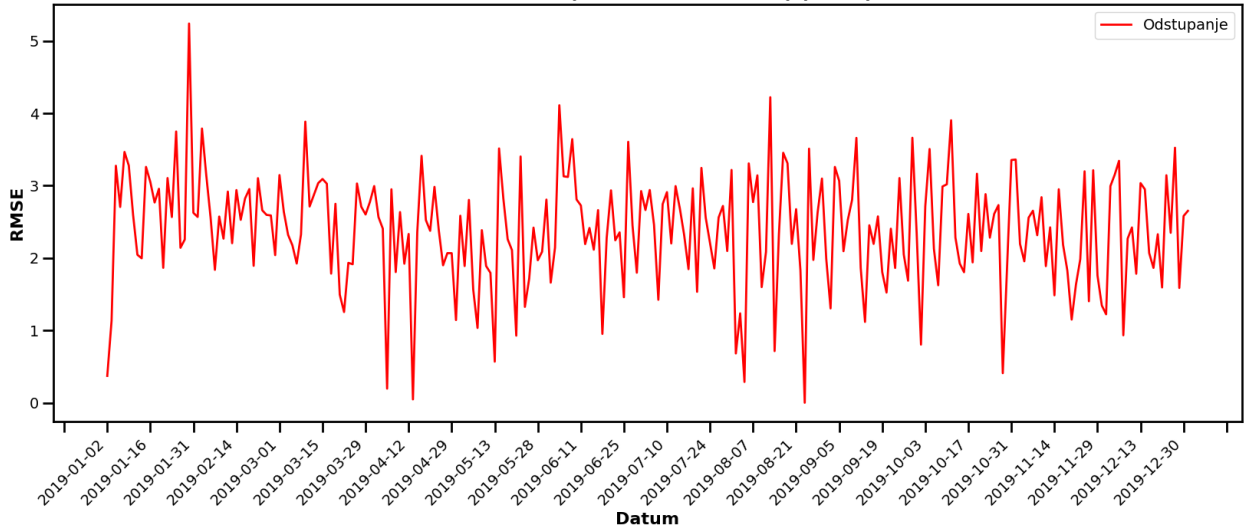
**RMSE**



**RMSE - ANN (analiza sentimenta) (AAPL)**

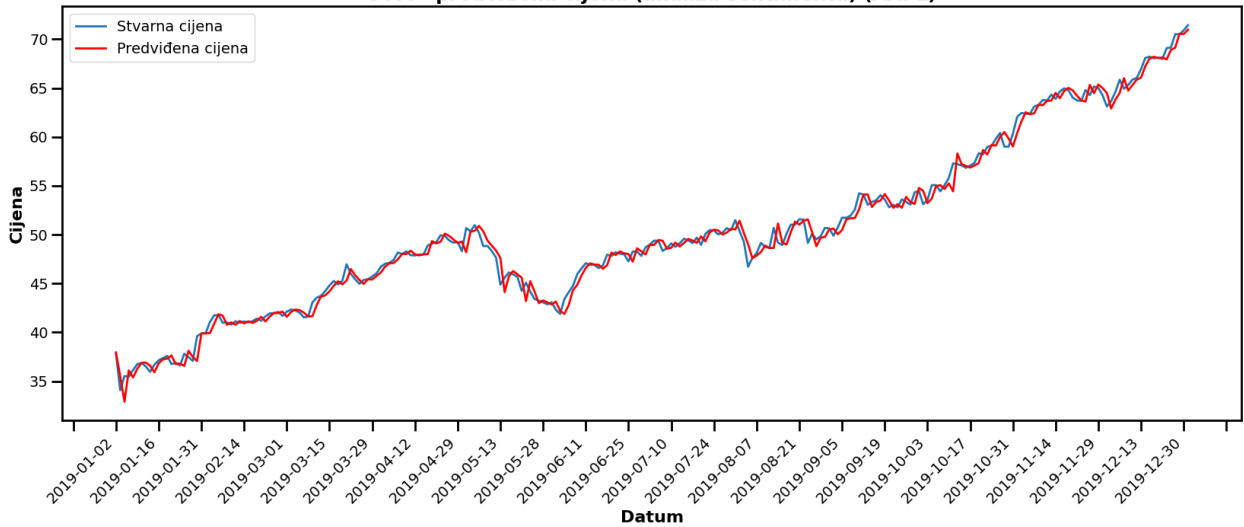


**RMSE - LSTM (analiza sentimenta) (AAPL)**

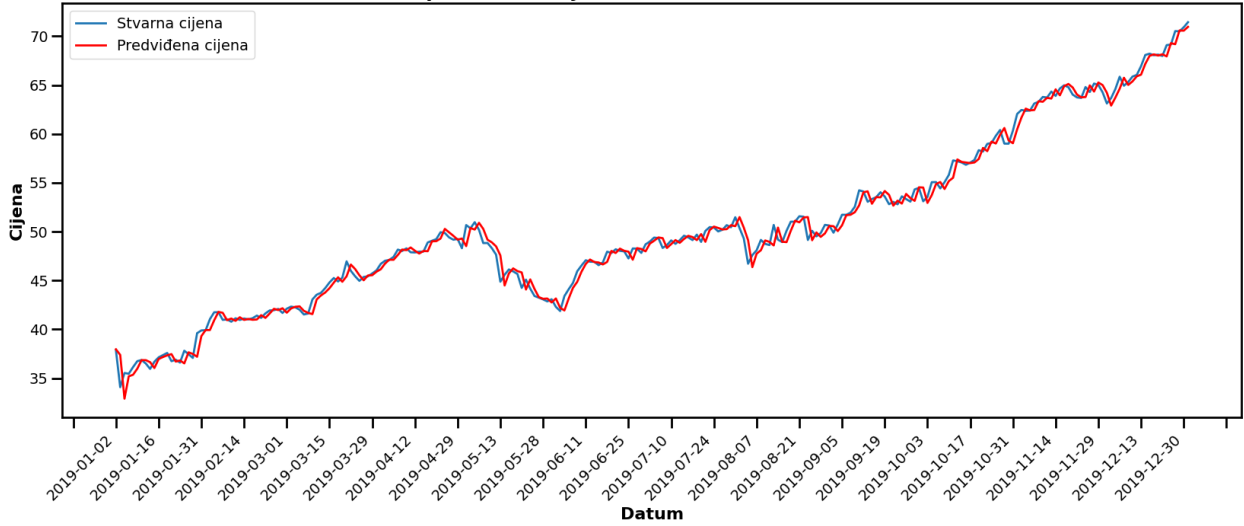


**Predviđena cijena**

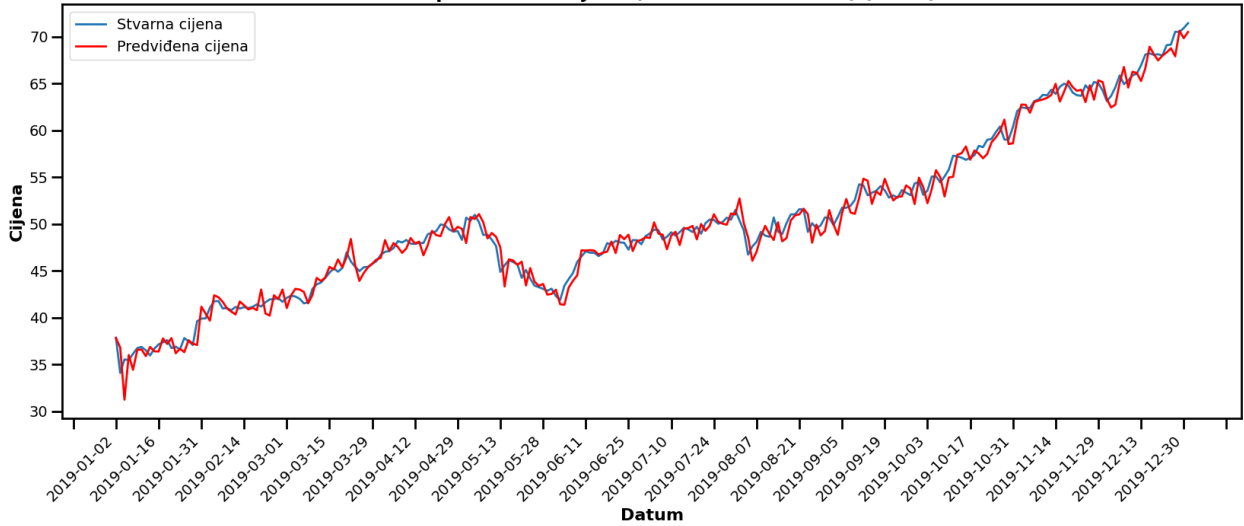
**SVM - predviđena cijena (analiza sentimenta) (AAPL)**



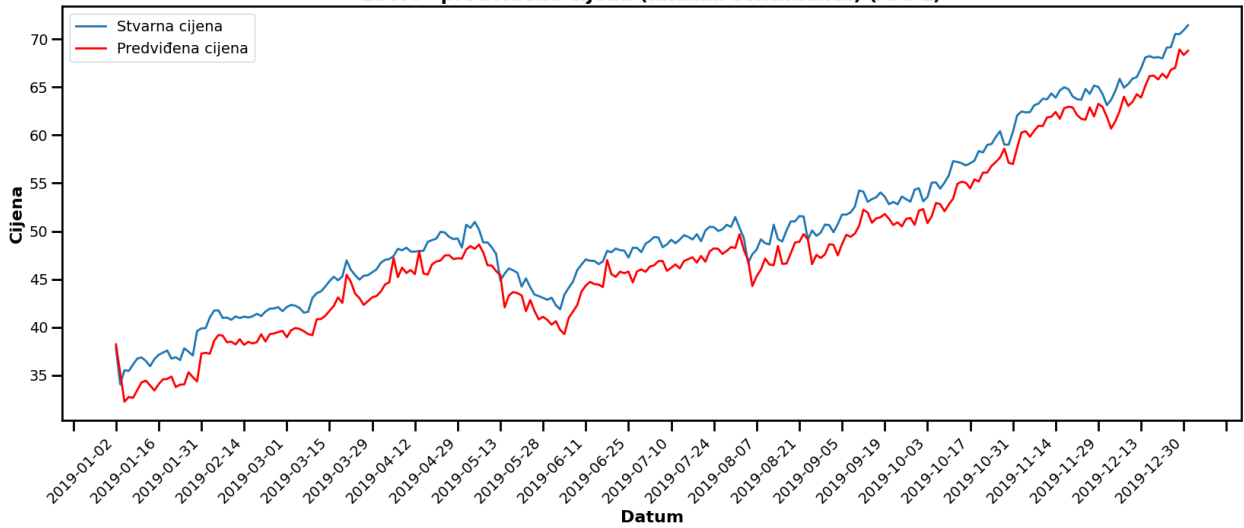
**RF - predviđena cijena (analiza sentimenata) (AAPL)**



**ANN - predviđena cijena (analiza sentimenata) (AAPL)**



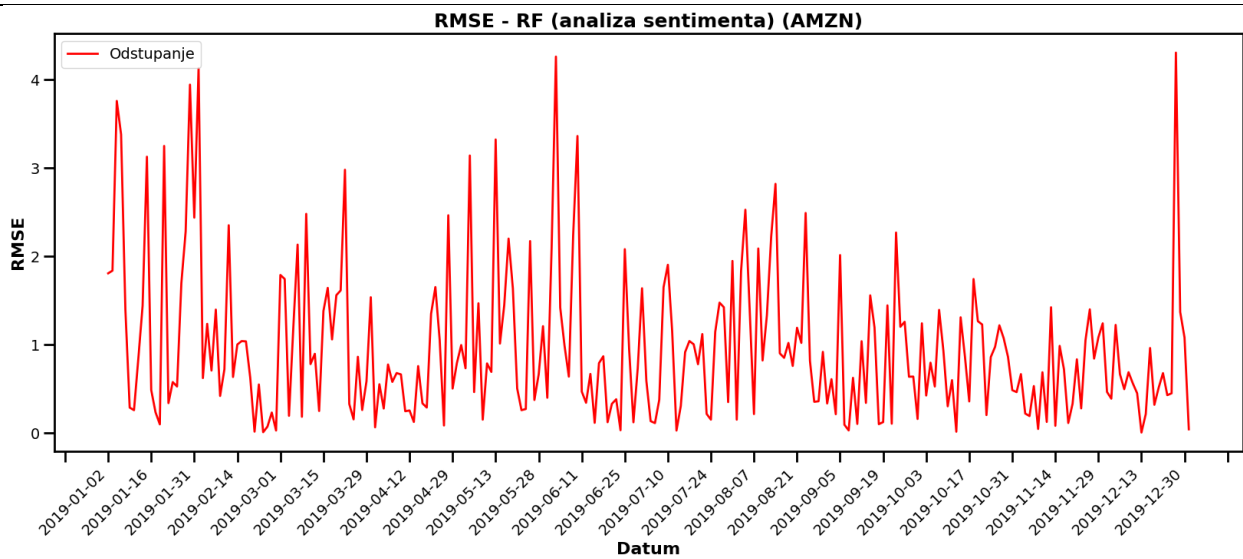
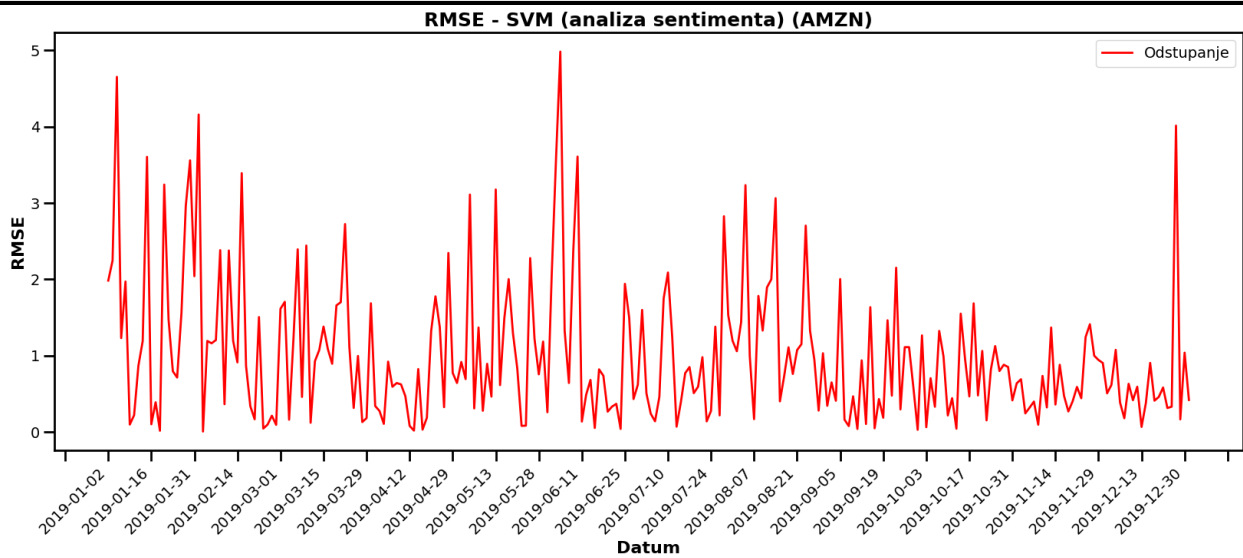
**LSTM - predviđena cijena (analiza sentimenata) (AAPL)**



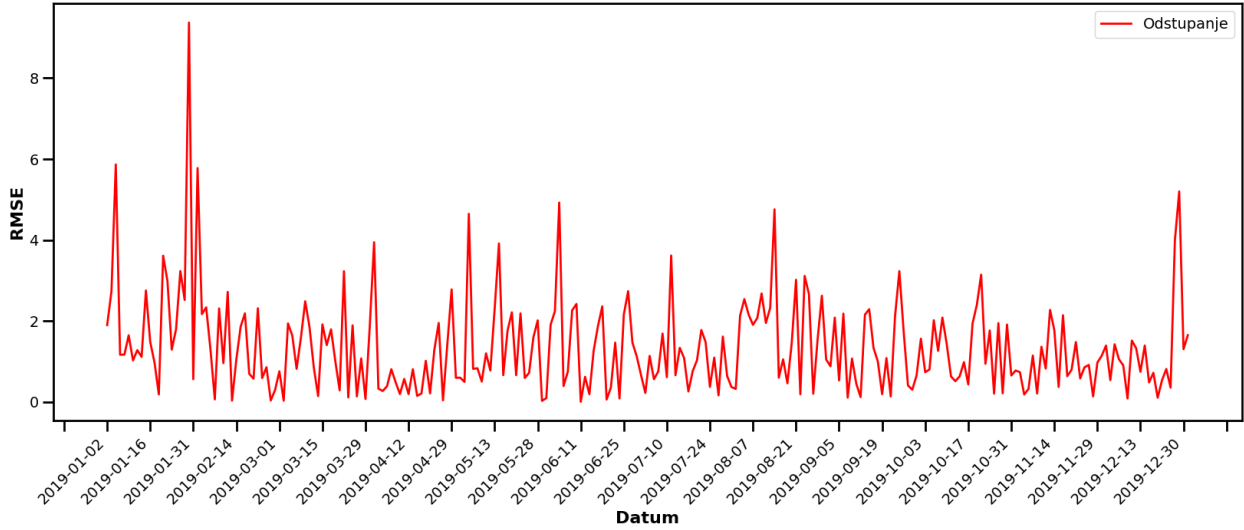
**Prilog P.11.** Grafovi s prikazom rezultata modela za analizu sentimenta za tvrtku Amazon

**Amazon**

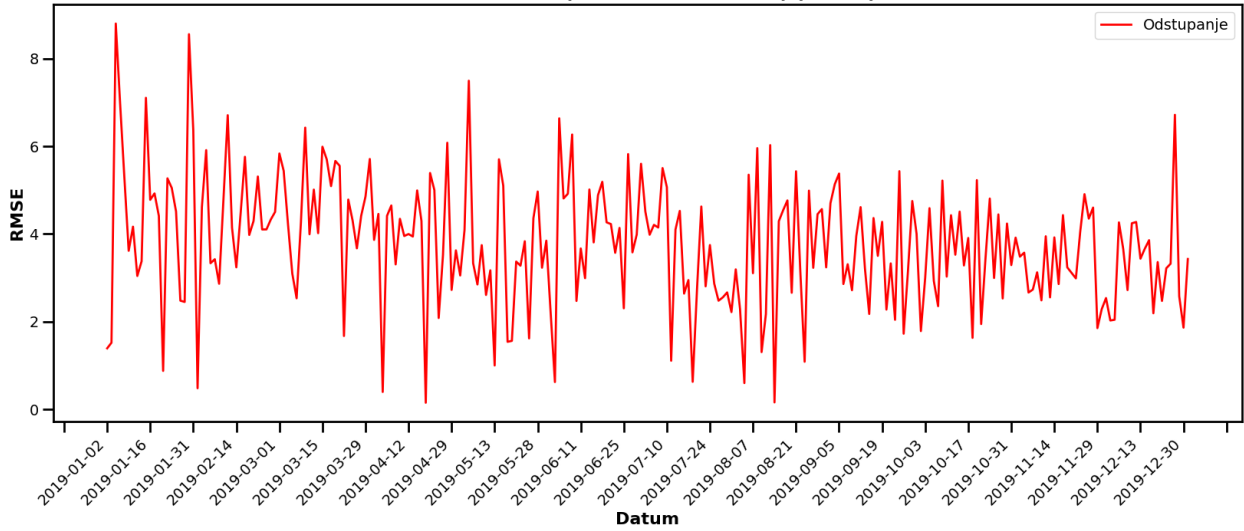
**RMSE**



**RMSE - ANN (analiza sentimenta) (AMZN)**

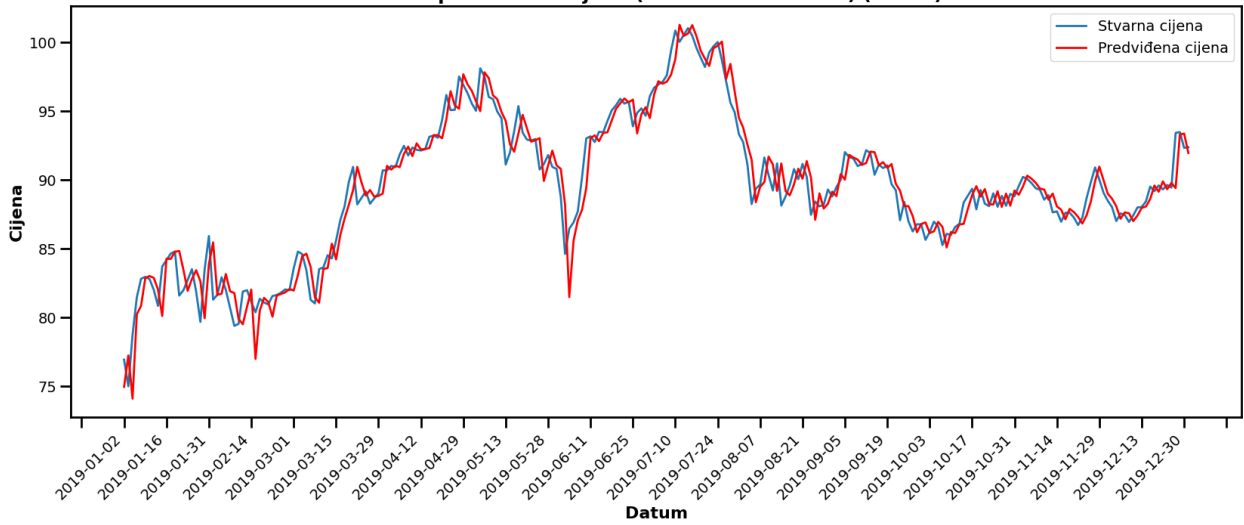


**RMSE - LSTM (analiza sentimenta) (AMZN)**

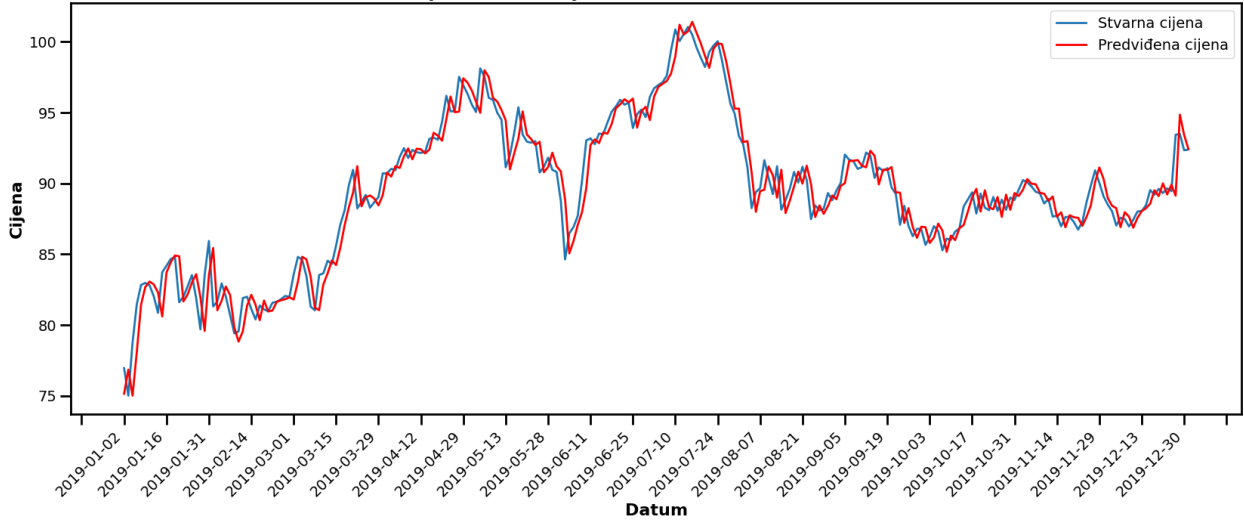


**Predviđena cijena**

**SVM - predviđena cijena (analiza sentimenta) (AMZN)**



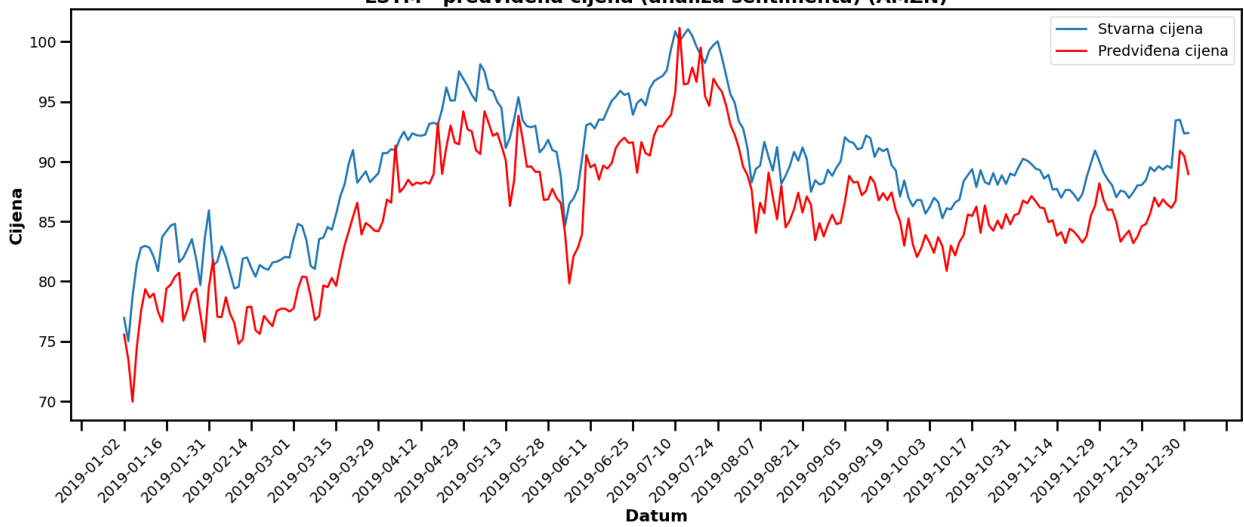
**RF - predviđena cijena (analiza sentimenta) (AMZN)**



**ANN - predviđena cijena (analiza sentimenta) (AMZN)**



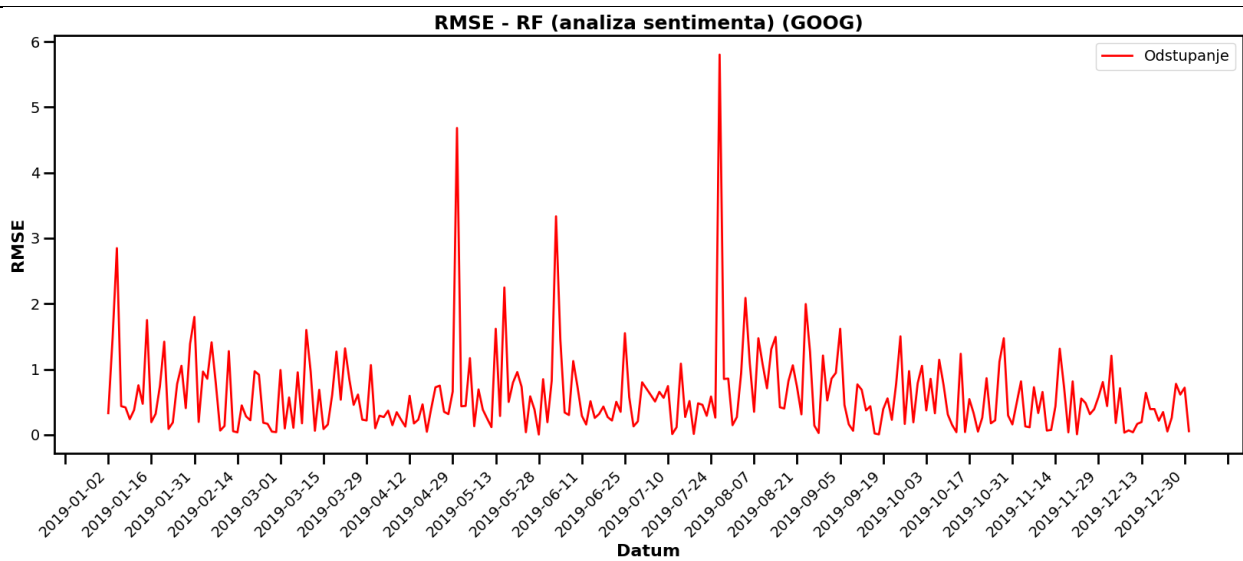
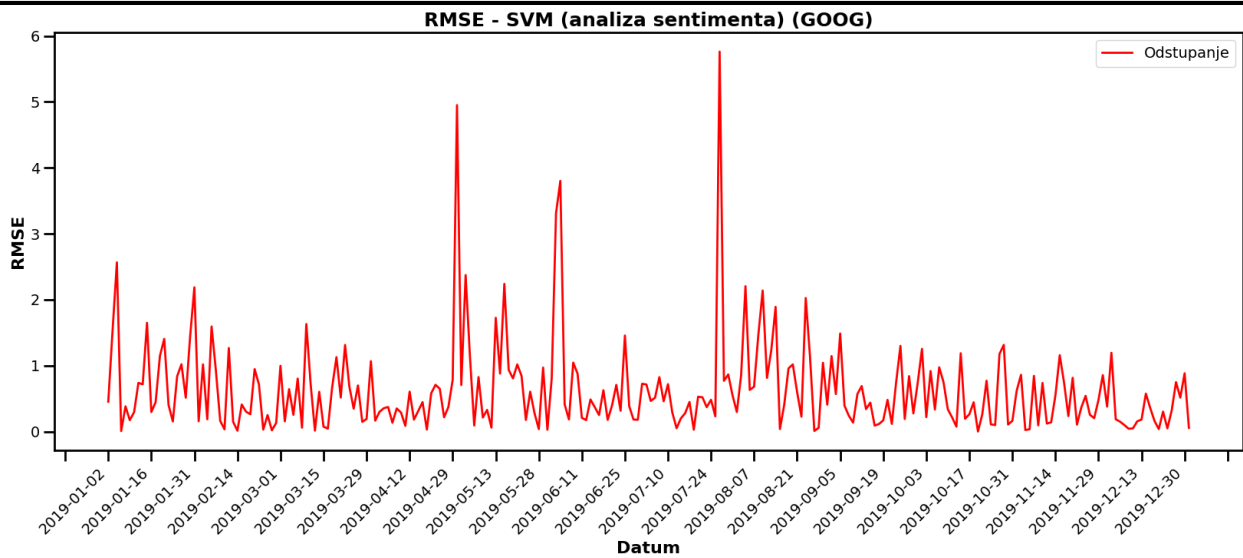
**LSTM - predviđena cijena (analiza sentimenta) (AMZN)**



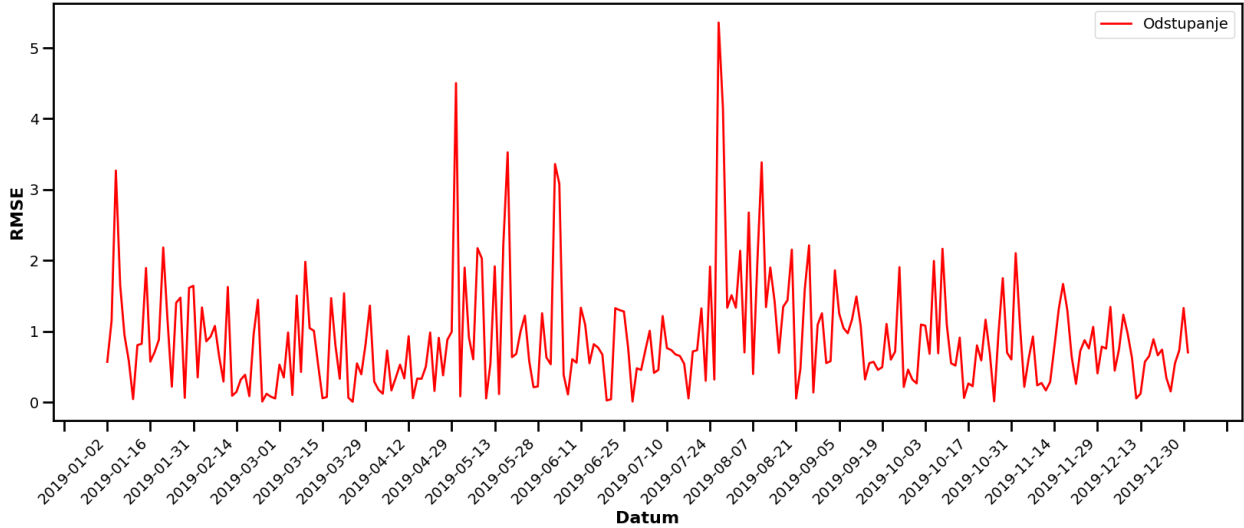
**Prilog P.12.** Grafovi s prikazom rezultata modela za analizu sentimenta za tvrtku Google

**Google**

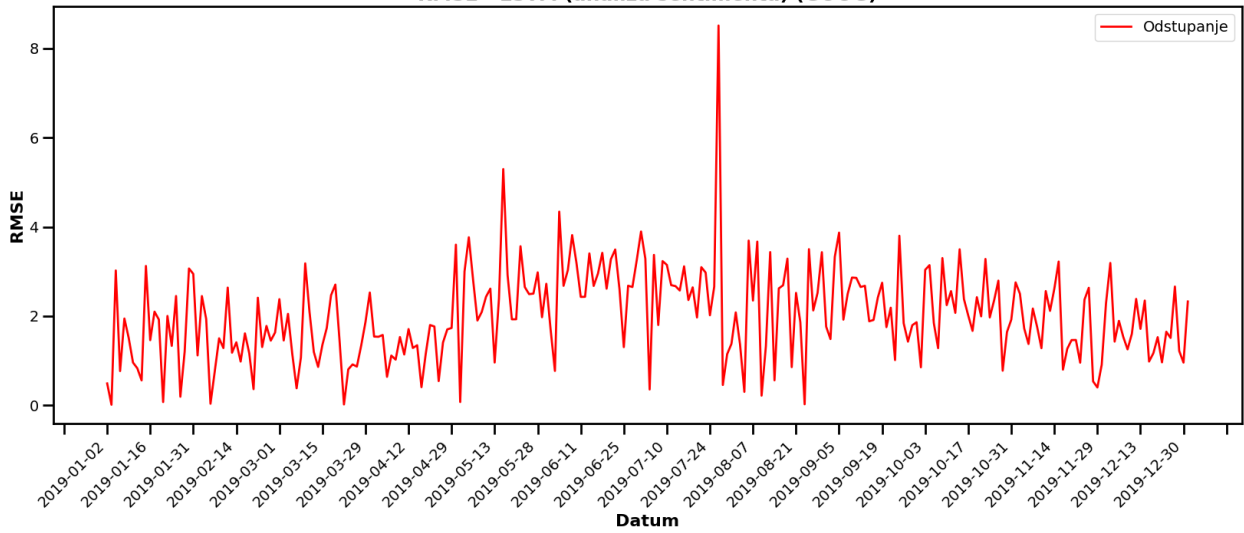
**RMSE**



RMSE - ANN (analiza sentimenta) (GOOG)

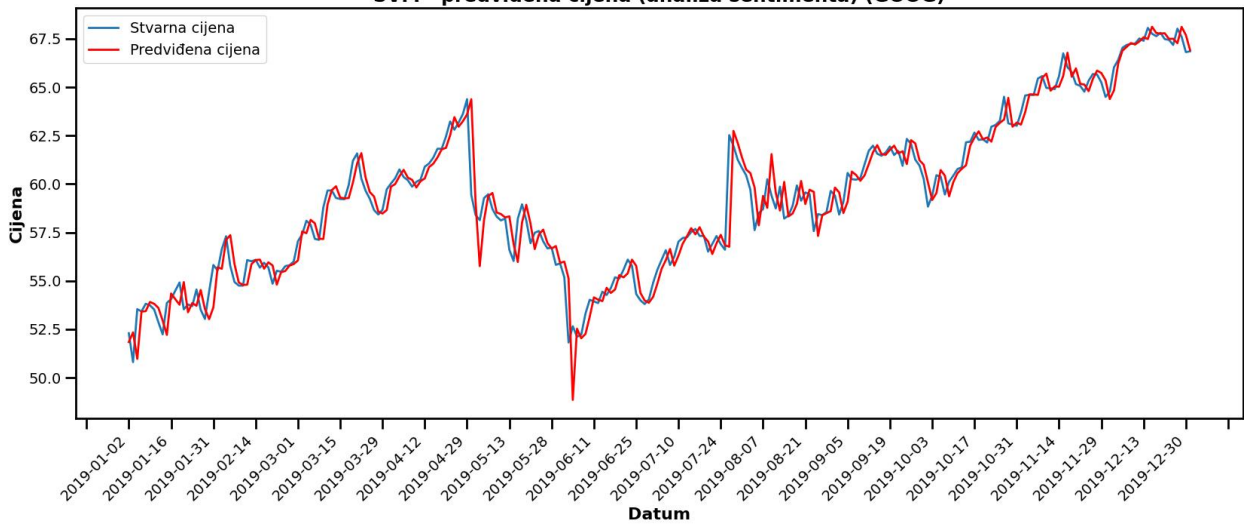


RMSE - LSTM (analiza sentimenta) (GOOG)



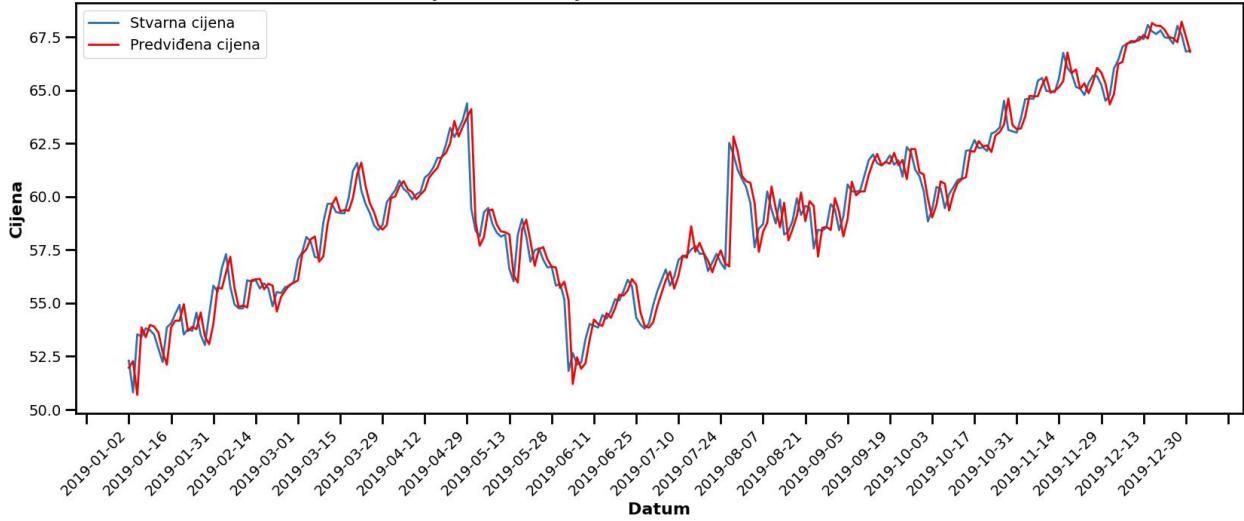
Predviđena cijena

SVM - predviđena cijena (analiza sentimenta) (GOOG)

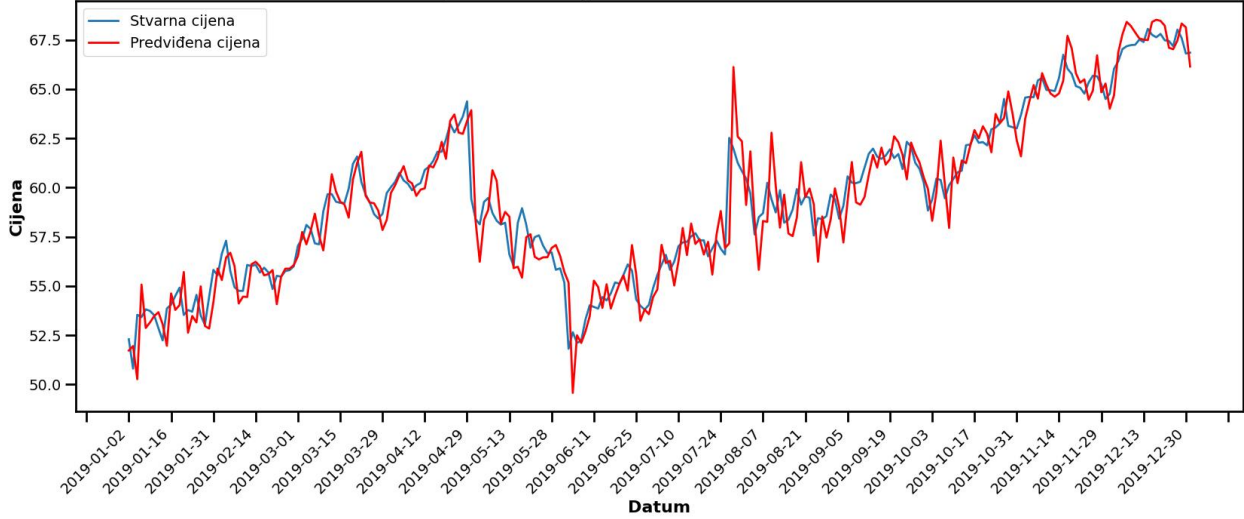




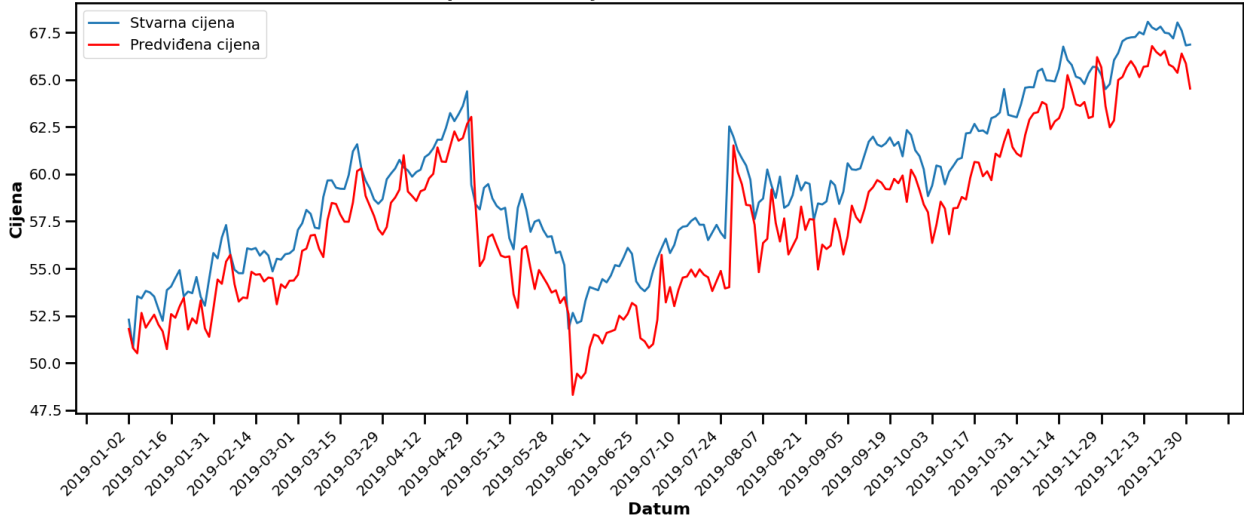
**RF - predviđena cijena (analiza sentimenta) (GOOG)**



**ANN - predviđena cijena (analiza sentimenta) (GOOG)**



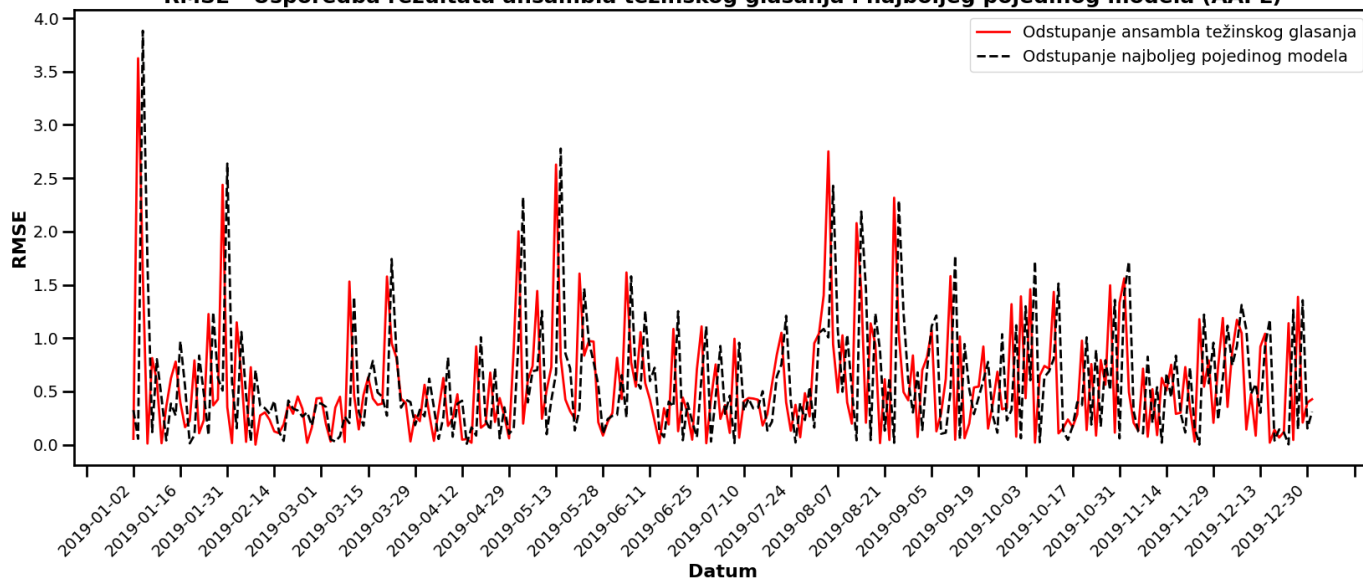
**LSTM - predviđena cijena (analiza sentimenta) (GOOG)**



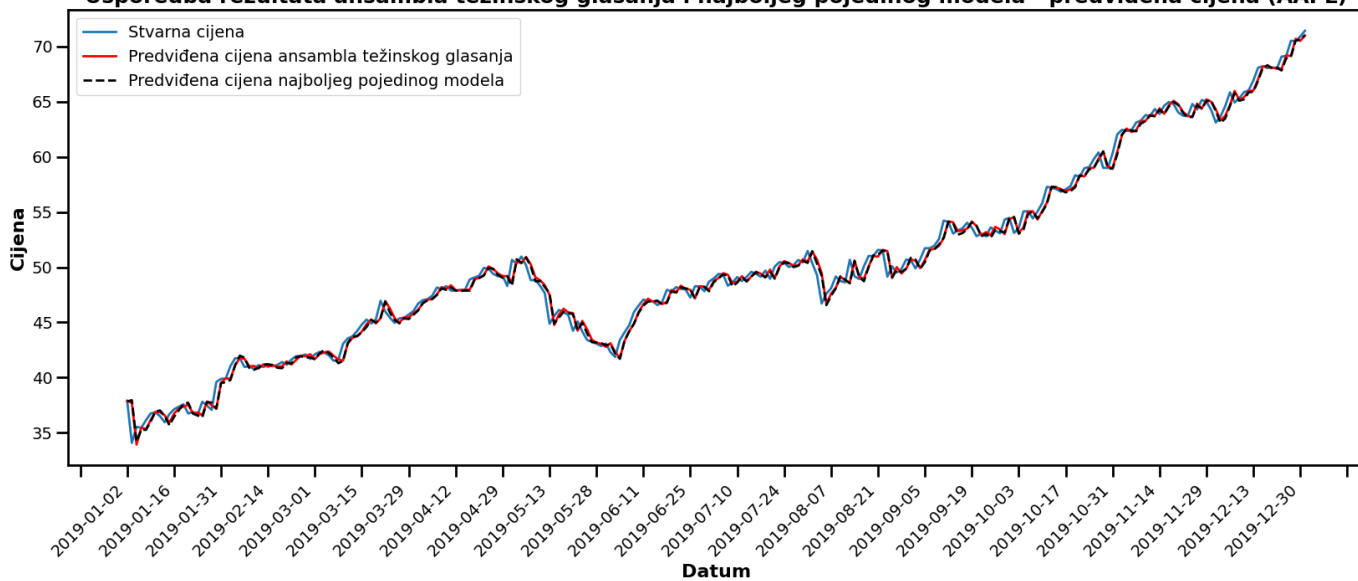
**Prilog P.13.** Usporedba rezultata ansambla zasnovanom na težinskom glasanju i najboljeg pojedinog modela

**Apple**

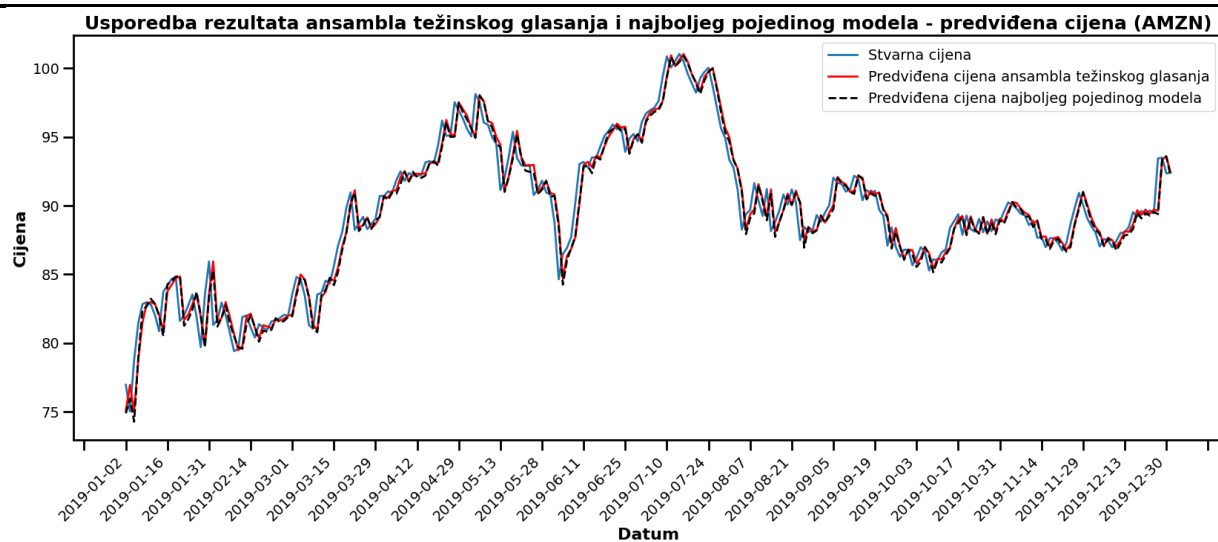
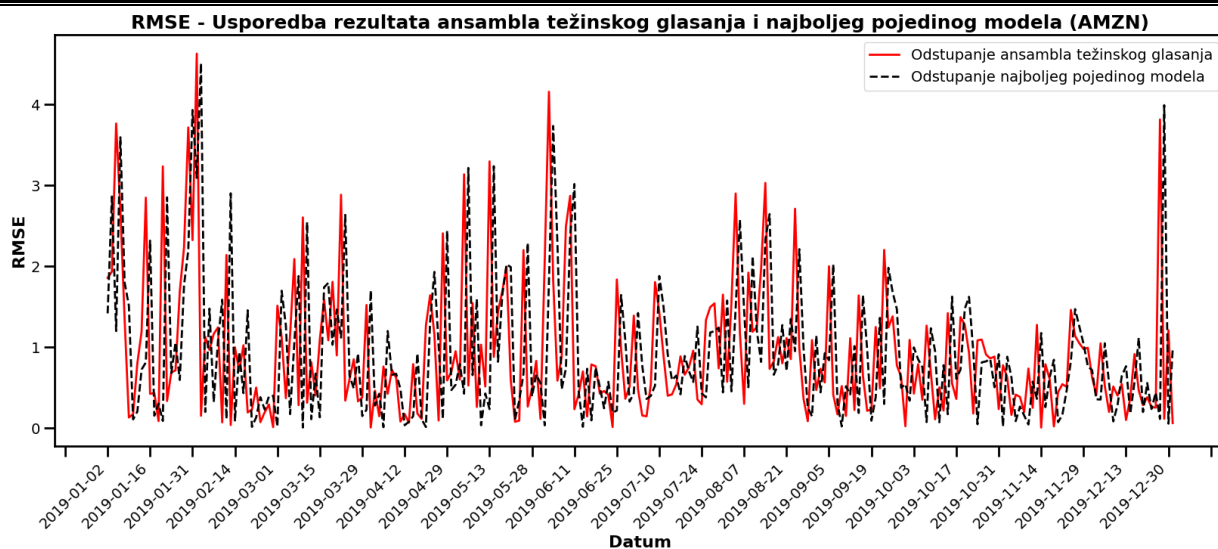
**RMSE - Usporedba rezultata ansambla težinskog glasanja i najboljeg pojedinog modela (AAPL)**



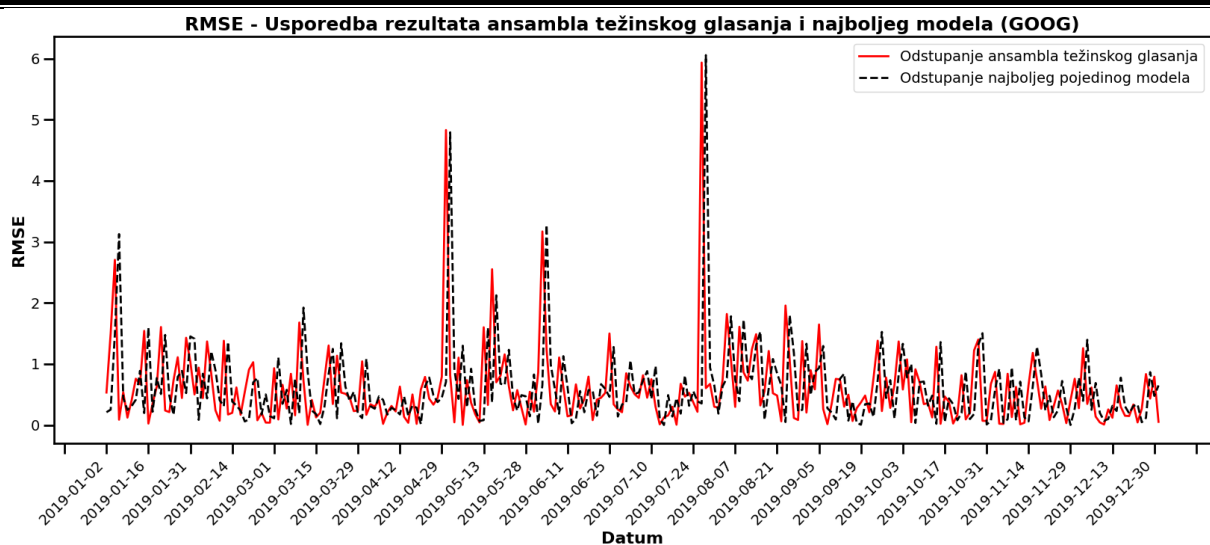
**Usporedba rezultata ansambla težinskog glasanja i najboljeg pojedinog modela - predviđena cijena (AAPL)**



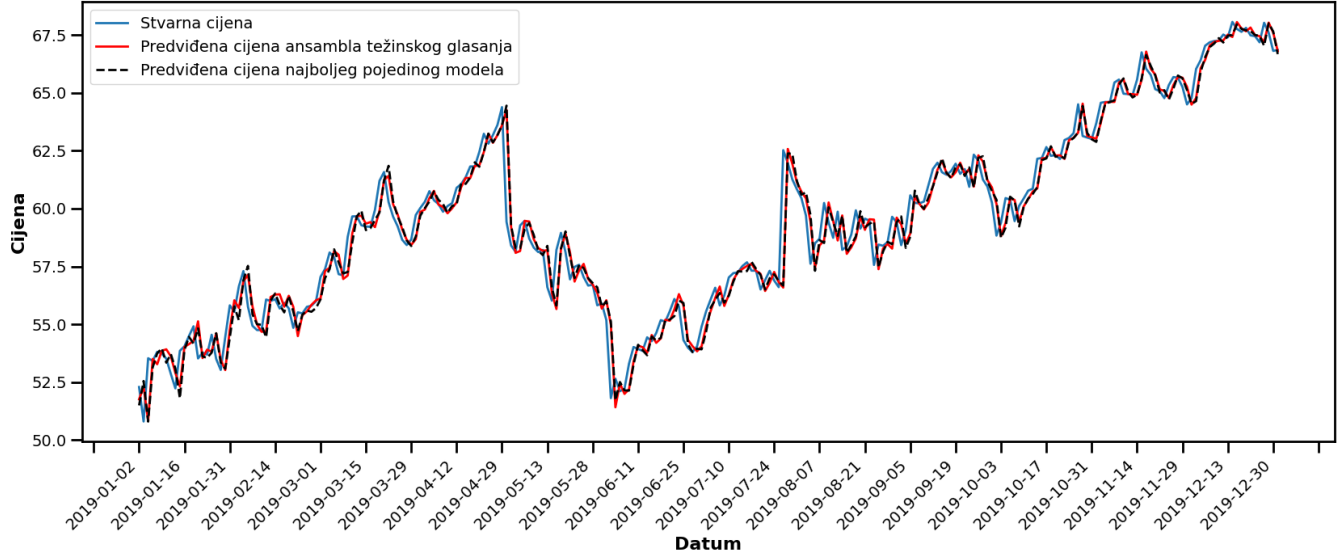
## Amazon



## Google



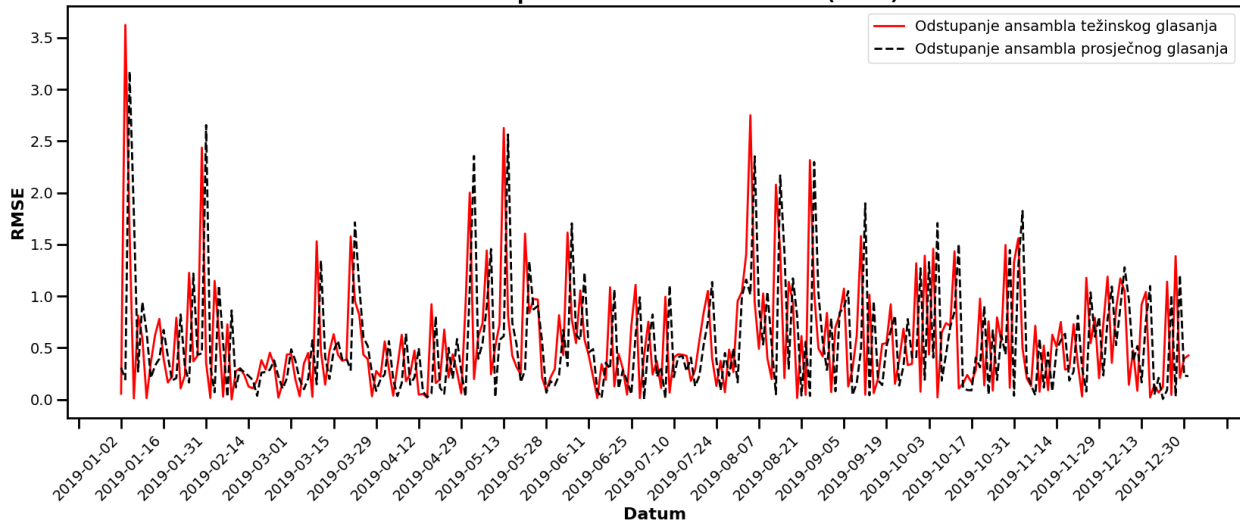
Usporedba rezultata ansambla težinskog glasanja i najboljeg pojedinog modela - predviđena cijena (GOOG)



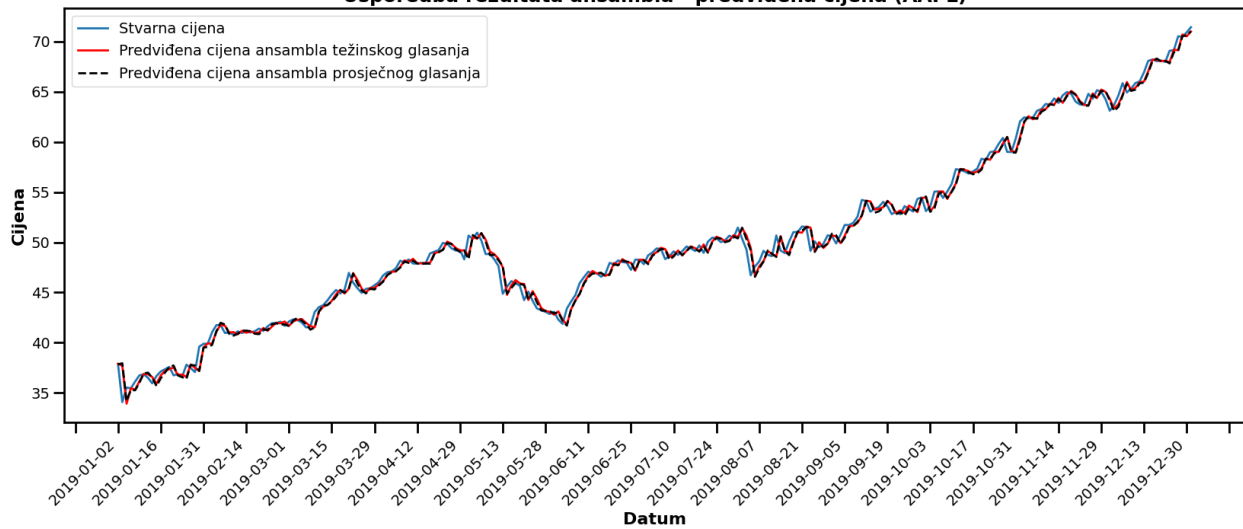
**Prilog P.14.** Usporedba rezultata ansambla zasnovanom na prosječnom glasanju i ansambla zasnovanom na težinskom glasanju

**Apple**

**RMSE - Usporedba rezultata ansambla (AAPL)**

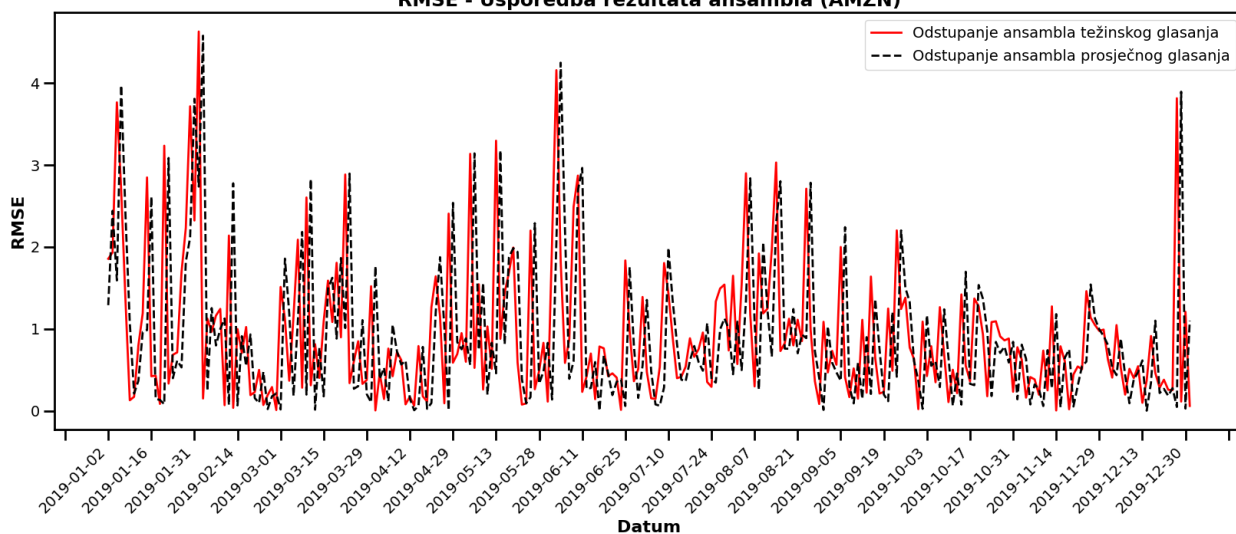


**Usporedba rezultata ansambla - predviđena cijena (AAPL)**

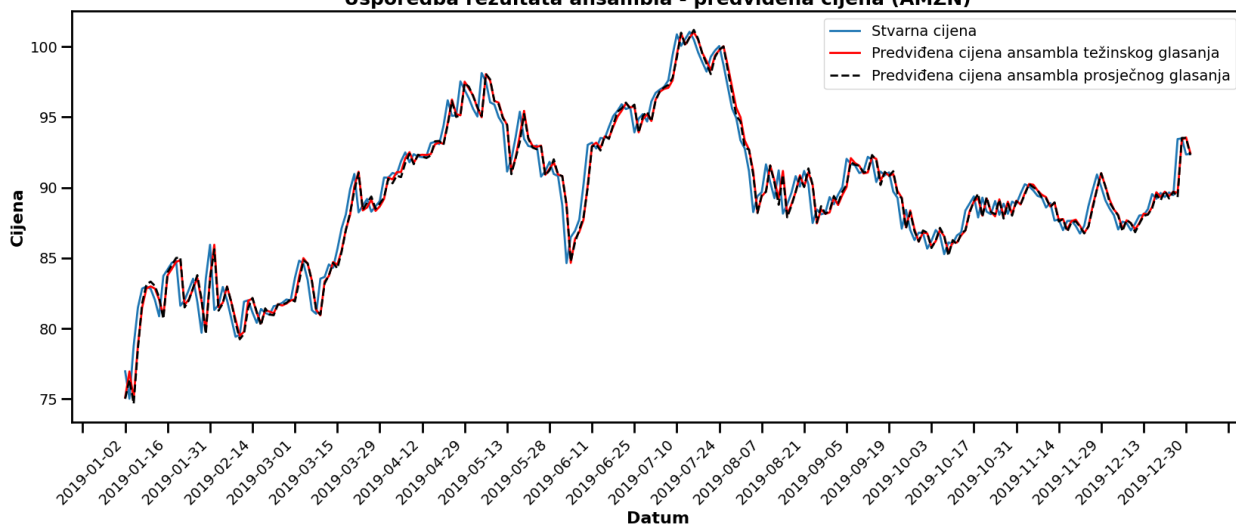


## Amazon

RMSE - Usporedba rezultata ansambla (AMZN)



Usporedba rezultata ansambla - predviđena cijena (AMZN)



## Google

