

Modeliranje financijskih tržišta pomoću tehnika strojnog učenja

Mikulić, Mateo

Undergraduate thesis / Završni rad

2022

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Rijeka, Faculty of Engineering / Sveučilište u Rijeci, Tehnički fakultet**

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:190:242907>

Rights / Prava: [Attribution 4.0 International](#)/[Imenovanje 4.0 međunarodna](#)

Download date / Datum preuzimanja: **2024-11-30**



Repository / Repozitorij:

[Repository of the University of Rijeka, Faculty of Engineering](#)



SVEUČILIŠTE U RIJECI
TEHNIČKI FAKULTET
Preddiplomski sveučilišni studij računarstva

Završni rad

**Modeliranje financijskih tržišta pomoću
tehnika strojnog učenja**

Rijeka, srpanj 2022.

Mateo Mikulić
0069087917

SVEUČILIŠTE U RIJECI
TEHNIČKI FAKULTET
Preddiplomski sveučilišni studij računarstva

Završni rad

**Modeliranje financijskih tržišta pomoću
tehnika strojnog učenja**

Mentor: doc. dr. sc. Goran Mauša

Rijeka, srpanj 2022.

Mateo Mikulić
0069087917

Rijeka, 14. ožujka 2022.

Zavod: **Zavod za računarstvo**
Predmet: **Uvod u objektno orijentirano programiranje**
Grana: **2.09.04 umjetna inteligencija**

ZADATAK ZA ZAVRŠNI RAD

Pristupnik: **Mateo Mikulić (0069087917)**
Studij: **Preddiplomski sveučilišni studij računarstva**

Zadatak: **Modeliranje financijskih tržišta pomoću tehnika strojnog učenja / Financial market modelling based on machine learning techniques**

Opis zadatka:

Proučiti upotrebu tehnika strojnog učenja u modeliranju financijskih tržišta. Prikupiti informacije o relevantnim skupovima podataka koji mogu poslužiti za treniranje modela. Predložiti vlastitu metodologiju za modeliranje tržišta. Definirati metrike vrednovanja, te ocijeniti kvalitetu predloženih modela. Konačni model usporediti sa značajnim modelima iz literature, te ocijeniti mogućnost primjene istoga u realnom scenariju.

Rad mora biti napisan prema Uputama za pisanje diplomskih / završnih radova koje su objavljene na mrežnim stranicama studija.



Zadatak uručen pristupniku: 14. ožujka 2022.

Mentor:



Doc. Goran Mauša, dipl. ing.

Predsjednik povjerenstva za
završni ispit:



Prof. dr. sc. Kristijan Lenac

Izjava o samostalnoj izradi rada

Izjavljujem da sam samostalno izradio ovaj rad.

Rijeka, srpanj 2022.



Mateo Mikulić

Zahvala

Zahvaljujem doc. dr. sc. Goranu Mauši na omogućavanju izrade rada u grani koja je meni interesantna. Zahvaljujem mag. ing. comp. Marku Njirjaku na podršci tijekom izrade i pisanja ovoga rada kao i na korisnim raspravama i savjetima. Također zahvaljujem svim profesorima i asistentima na prenesenom znanju. Zahvaljujem roditeljima i svima ostalima na podršci tijekom preddiplomskog studija.

Sadržaj

Popis slika	ix
Popis tablica	x
1 Uvod	1
2 Strojno učenje	3
2.1 Što je strojno učenje?	3
2.2 Umjetne neuronske mreže	3
2.2.1 Što su umjetne neuronske mreže?	3
2.2.2 Vrste umjetnih neurona	6
2.2.3 Slojevi neuronske mreže	7
2.2.4 Učenje umjetne neuronske mreže	7
2.3 Aplikacijski okvir TensorFlow	10
2.3.1 Strojno učenje u TensorFlowu	10
2.3.2 TensorFlow Data podatkovni cjevovod (<i>engl. pipeline</i>)	10
3 Pregled strojnog učenja u modeliranju financijskih tržišta	11
4 Skupovi podataka	15
4.1 Porijeklo podataka	15

Sadržaj

4.2	Podatci korišteni u treniranju	16
4.2.1	Cijena	16
4.2.2	RSI	16
4.2.3	SMA	17
4.2.4	EMA	17
4.2.5	Volumen	17
4.3	Formatiranje podataka	17
4.4	Distribucija podataka	18
4.5	Normalizacija podataka	23
5	Prijedlog modeliranja financijskih tržišta	24
5.1	Podatci	24
5.1.1	Ulazni podatci	24
5.1.2	Izlazni podatci	25
5.2	Arhitektura predloženog modela	26
5.3	Rezultati predloženog modela	29
5.3.1	Metrike	29
5.3.2	Rezultati cijelog skupa podataka za testiranje	30
5.3.3	Rezultati na podacima dionice OXY	33
5.3.4	Rezultati na podacima dionice IBM	34
5.3.5	Rezultati na podacima dionice NFLX	37
6	Zaključak	38
6.1	Ocjena modela	38
	Bibliografija	39
	Pojmovnik	42

Sadržaj

Sažetak

43

Popis slika

2.1	Prikaz perceptrona, preuzeto iz [4] i preuređeno	4
2.2	Prikaz višeslojne neuronske mreže	5
4.1	Distribucija Simple Moving Average (SMA) i Exponential Moving Average (EMA) podataka u cijelom skupu podataka	18
4.2	Distribucija cijena prilikom otvaranja/zatvaranja u cijelom skupu podataka	19
4.3	Distribucija krajnjih dnevnih cijena u cijelom skupu podataka	20
4.4	Distribucija Relative Strenght Index (RSI) podataka u cijelom skupu podataka	21
4.5	Distribucija podataka o dnevnom volumenu u cijelom skupu podataka	22
5.1	Prikaz metode pomičnog prozora	25
5.2	Prikaz arhitekture predloženog modela	27
5.3	Prikaz svih predviđanja modela	31
5.4	Pregled raznih scenarija predviđanja	32
5.5	Prikaz svih predviđanja modela za dionicu OXY	33
5.6	Prikaz svih predviđanja modela za dionicu IBM	34
5.7	Pregled uzastopnih predviđanja na IBM dionici	35
5.7	Pregled uzastopnih predviđanja na IBM dionici	36
5.8	Prikaz svih predviđanja modela za dionicu NFLX	37

Popis tablica

4.1	Primjer podataka za dionicu IBM na dan 10. lipnja 2022.	16
5.1	Prikaz metrika za vrednovanje skupa podataka za testiranje	30
5.2	Prikaz metrika za vrednovanje podataka na dionici OXY	33
5.3	Prikaz metrika za vrednovanje podataka na dionici IBM	34
5.4	Prikaz metrika za vrednovanje podataka na dionici NFLX	37

Poglavlje 1

Uvod

Tvrtke svoje dionice prodaju na financijskim tržištima u svrhu stjecanja gotovog novca u zamjenu za prodaju udjela vlasništva.

S obzirom na sve veću dostupnost financijskih tržišta široj populaciji raste interes za ulaganjem u ista. Veća dostupnost financijskih tržišta omogućena je putem raznih internetskih brokera i brokerskih mobilnih aplikacija. Prije pojave ovakvih rješenja, ulaganja su bila moguća putem brokerskih kuća koje imaju veće provizije zbog zaposlenika (*brokera*) koje moraju plaćati.

Ulaganja opće populacije često završe sa gubitkom zbog neupućenosti u tržišta ili naglih promjena na tržištima.

Osim opće populacije u financijska tržišta ulažu i tvrtke koje za svoje uloge zbog velikih suma koje ulažu često vrše duboke financijske analize poslovnih subjekata pa su gubitci ipak rjeđi. Ulagачi u financijska tržišta nastoje se financijski okoristiti za kratkoročnu korist ili za dugoročnu korist. Također, ulaganjem u financijska tržišta nastoje se zaštititi od inflacije.

S obzirom na veliki interes ulagača u financijska tržišta počele su se koristiti razne metode koje pokušavaju predvidjeti ponašanje tržišta. Postoje razni tehnički pokazatelji koji se računaju iz prethodnog ponašanja cijene dionice kroz neki određeni vremenski period. Također, izrađuju se statistički i matematički modeli.

Razvojem grana umjetne inteligencije i strojnog učenja pojavljuju se i modeli strojnog učenja čime će se ovaj rad baviti. Za primjenu umjetne inteligencije u

Poglavlje 1. Uvod

predviđanju finansijskih tržišta postoji veliki interes zbog mogućnosti rada s velikim skupovima podataka koji su čovjeku nemogući za pojmiti i procesuirati na relevantan način.

Ovim radom predložiti će se model strojnog učenja koji će pokušati predviđati cijene dionica na finansijskim tržištima.

Poglavlje 2

Strojno učenje

2.1 Što je strojno učenje?

Strojno učenje (*engl. machine learning*) je grana umjetne inteligencije (*engl. artificial intelligence*) i računalnih znanosti koja nastoji dati računalima mogućnost učenja bez eksplicitnog programiranja. Nastoji se omogućiti računalima da uče na način sličan ljudima. Algoritmi nadziranog strojnog učenja se pretežno koriste za predviđanje, pri čemu razlikujemo predviđanje u diskretne kategorije koje nazivamo klasifikacijom i predviđanje kontinuiranih veličina koje nazivamo regresijom.

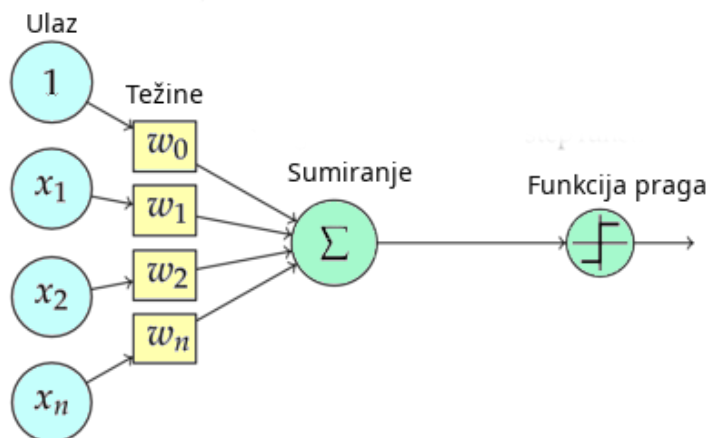
2.2 Umjetne neuronske mreže

2.2.1 Što su umjetne neuronske mreže?

Umjetne neuronske mreže su umjetne strukture koje nastoje imitirati biološke neuronske mreže s ciljem stvaranja umjetne inteligencije. Umjetna neuronska mreža uči iskustvenim putem, a međusobne veze između neurona koriste se za pohranjivanje znanja. Umjetne neuronske mreže nisu nov koncept, prvi rad o njima je objavljen 1943. godine od strane McCulloha i Pittsa [1]. Koristili su vrlo jednostavan model neurona - perceptron. Perceptron je model koji signale x_i opisane numeričkim iznosom na ulazu množi težinskim faktorima w_i , a u tijelu neurona se ti signali sumiraju

Poglavlje 2. Strojno učenje

i uspoređuju sa pragom osjetljivosti neurona θ . Ako je suma signala veća od praga, perceptron daje izlaz kroz prijenosnu funkciju. [2][3] Prikaz perceptrona na slici 2.1.



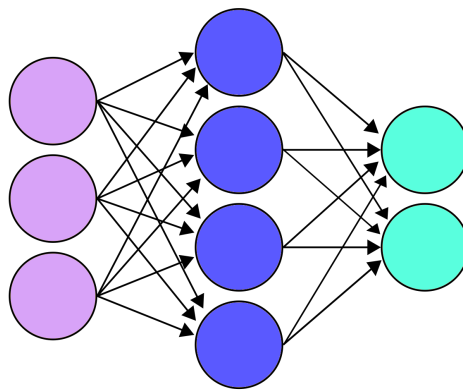
Slika 2.1 Prikaz perceptrona, preuzeto iz [4] i preuređeno

Zbog kompaktnosti se u literaturi često dogovorno pretpostavlja da je vrijednost praga $\theta = -\omega_0$ i dodaje se ulazni signal x_0 s fiksnom vrijednosti 1 pa se izlaz iz prijenosne funkcije, a time izlaz iz neurona y kako je već objašnjeno u ovom odlomku može prikazati pomoću jednadžbe:

$$y = f\left(\sum_{i=1}^n x_i w_i\right) \quad (2.1)$$

Umjetne neuronske mreže u primjeni većinom se sastoje od većeg broja čvorova, podatke obrađuju paralelno. Obradbeni moć neurona u umjetnim neuronskim mrežama je pohranjena u njihovim međusobnim vezama, odnosno njihovim težinama. Težine se prilagođavaju učenjem iz skupa podataka za treniranje. U implementaciji se koriste višeslojne mreže koje se sastoje od ulaznog sloja, skrivenog sloja (ili slojeva) i izlaznog sloja. Više o slojevima u podsekciji 2.2.3. Primjer višeslojne mreže možemo vidjeti na slici 2.2. [2]

Umjetne neuronske mreže odlično rješavaju sve probleme kod kojih postoji zavisnost između ulaznih i izlaznih varijabli, neovisno o njihovoj linearnosti. Primjeri takvih problema bi bili problemi klasifikacije i predviđanja.



Slika 2.2 Prikaz višeslojne neuronske mreže

2.2.2 Vrste umjetnih neurona

Neuroni se mogu razlikovati prema ugrađenoj prijenosnoj funkciji. S obzirom na veliki broj različitih uvriježenih prijenosnih funkcija, opisati ću još samo neurone koji su korišteni pri izradi ovoga rada.

Najjednostavniji primjer neurona je model naziva *Adaptive Linear Element* čiji je izlaz težinska suma njegovih ulaza. Prijenosna funkcija ovoga neurona još se naziva i funkcija identiteta ili linearna funkcija.[2] Jednadžba izlaza ovog neurona može se opisati kao:

$$y = \sum_{i=1}^n x_i w_i \quad (2.2)$$

Rectified Linear Unit (ReLU) funkcija za dovedenu težinsku sumu kao izlaz daje upravo tu sumu ako je veća od 0. [5] Izlaz iz neurona koji koristi ReLU funkciju može se prikazati jednadžbom:

$$y = \begin{cases} 0, & \text{za } \sum_{i=1}^n x_i w_i < 0 \\ \sum_{i=1}^n x_i w_i, & \text{za } \sum_{i=1}^n x_i w_i \geq 0 \end{cases} \quad (2.3)$$

SoftPlus funkcija je glađa aproksimacija ReLU funkcije. Može se koristiti kada se želi izlaz ograničiti na pozitivne vrijednosti. [6] Izlaz neurona koji koristi *SoftPlus* prijenosnu funkciju može se opisati jednadžbom:

$$y = \ln(1 + e^{\sum_{i=1}^n x_i w_i}) \quad (2.4)$$

2.2.3 Slojevi neuronske mreže

Slojevi neuronske mreže su skupine neurona koji paralelno djeluju.

Potpuno povezani sloj (*engl. Dense*) sloj jest sloj koji je duboko povezan sa prethodnim slojem što znači da je svaki neuron u potpuno povezanom sloju povezan sa svakim neuronom u prethodnom sloju. Potpuno povezani slojevi su najčešće korišteni u implementaciji umjetnih neuronskih mreža.

Sloj izbacivanja (*engl. Dropout*) sloj nasumično ulazne vrijednosti postavlja u nulu kako bi se izbjegao prevelik utjecaj istih neurona ili ponavljanje istih grešaka kroz više slojeva. Sloj izbacivanja se koristi kao alat protiv pretreniranja mreže.

2.2.4 Učenje umjetne neuronske mreže

Razlikujemo tri načina učenja ovisno o dostupnim podacima i metodama:

- Nadzirano učenje (*engl. supervised learning*) - kada je poznat točan izlaz, učenje se provodi sa skupovima u obliku para (ulaz, izlaz)
- Ojačano učenje (*engl. reinforcement learning*) - pokušaji i pogreške uz ocjenu sudca
- Nenadzirano učenje (*engl. unsupervised learning*) - nije poznat točan izlaz, mreža uči bez poznavanja izlaza

Ovaj rad baviti će se nadziranim učenjem neuronske mreže. Skup podataka za učenje često se dijeli na tri odvojena skupa: trening, validacija, testiranje. Skup podataka za treniranje služi za učenje mreže, podešavanje težinskih faktora. Validacijski podatci služe za provjeru i korekciju/zaustavljanje treniranja mreže kako se mreža ne bi pretrenirala (*engl. overfitting*). Pretreniranje se događa kada mreža prestane posjedovati svojstvo generalizacije i počinje dobro obrađivati samo podatke iz skupa podataka za treniranje (specijalizira se za skup podataka za treniranje). Skup podataka za testiranje služi za mjerenje točnosti i preciznosti obrade podataka. Važno je još istaknuti pojmove iteracija i epoha. Za iteraciju se podrazumijeva korak u algoritmu postupka za učenje, podešavanju težinskih faktora. Za epohu podrazumijevamo jedno predstavljanje cijelog skupa podataka. U iteraciji se još javlja pojam

Poglavlje 2. Strojno učenje

pojedinačnog učenja (*engl. on-line training*) i grupnog učenja (*engl. batch training*). Pojedinačno učenje podrazumijeva samo jedan par ulaza i izlaza, dok se u grupnom učenju mreži daje više parova odjednom.

Kako bi mreža mogla učiti potrebno je podešavati njene težinske faktore. Parametri mreže podešavaju se kroz algoritme gradijentnog spusta (*engl. gradient descent*) i propagacije pogreške unatrag (*engl. backpropagation*). [2] Za razumijevanje navedenih algoritama potrebno je i poznavati funkcije pogreške. Postoje razne funkcije pogreške, a navedne su isključivo one bitne za metodologiju ovoga rada:

Mean Squared Error (MSE) je funkcija koja mjeri prosjek kvadrata razlike između predviđenih i stvarnih vrijednosti. S obzirom na kvadriranje predviđene vrijednosti koje više odstupaju od stvarnih vrijednosti su više penalizirane s obzirom na one a manjim odstupanjem. [7] MSE možemo opisati jednadžbom:

$$MSE = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n} \quad (2.5)$$

gdje je \hat{y}_i predviđena vrijednost, a y_i stvarna vrijednost.

Mean Squared Logarithmic Error (MSLE) je funkcija pogreške koja mjeri relativne razliku između stvarnih i predviđenih vrijednosti na logaritamskoj skali. Ideja ove funkcije je ne previše penalizirati velika odstupanja koja su moguća ako se predviđaju velike vrijednosti koje nisu skalirane. [7] MSLE možemo opisati jednadžbom:

$$MSLE = \frac{\sum_{i=1}^n (\log(y_i + 1) - \log(\hat{y}_i + 1))^2}{n} \quad (2.6)$$

gdje je \hat{y}_i predviđena vrijednost, a y_i stvarna vrijednost.

Mean Absolute Error (MAE) je funkcija pogreške koja mjeri apsolutni prosjek razlike između stvarnih i predviđenih vrijednosti. Ne penalizira velika odstupanja više. [7] MAE možemo prikazati jednadžbom:

$$MAE = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n} \quad (2.7)$$

gdje je \hat{y}_i predviđena vrijednost, a y_i stvarna vrijednost.

Poglavlje 2. Strojno učenje

Root Mean Squared Error (RMSE) je standardna devijacija grešaka predviđanja. Mjeri raširenost grešaka od idealne regresijske linije. RMSE možemo opisati jednadžbom:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}} \quad (2.8)$$

gdje je \hat{y}_i predviđena vrijednost, a y_i stvarna vrijednost.

Gradijentni spust je optimizacijski algoritam koji se koristi za minimiziranje funkcije iterativnim spuštanjem u smjeru najstrmijeg pada. Gradijentni spust može se primjeniti samo na konveksne i diferencijabilne funkcije.[8] Gradijentni spust matematički može biti zapisan kao:

$$P_{n+1} = P_n - \eta \nabla f(P_n) \quad (2.9)$$

gdje P_n predstavlja trenutne parametre neuronske mreže, a P_{n+1} parametre u sljedećoj iteraciji, $\nabla f(P_n)$ predstavlja gradijent za n-dimenzionalnu funkciju pogreške u točki P_n , η je stopa učenja.

Gradijent u točki P za n-dimenzionalnu funkciju pogreške se može prikazati kao:

$$\nabla f(P) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(P) \\ \cdot \\ \cdot \\ \cdot \\ \frac{\partial f}{\partial x_n}(P) \end{bmatrix} \quad (2.10)$$

Algoritam propagacije pogreške unatrag služi za promjenu težinskih faktora kroz neuronsku mrežu s ciljem minimizacije funkcije pogreške u sljedećoj iteraciji, odnosno računanje gradijentnog spusta u neuronskim mrežama.

2.3 Aplikacijski okvir TensorFlow

2.3.1 Strojno učenje u TensorFlowu

Za implementaciju modela strojnog učenja korišten je aplikacijski okvir TensorFlow. TensorFlow je aplikacijski okvir otvorenog koda koji služi za lakšu implementaciju modela strojnog i dubokog učenja. TensorFlow nudi nekoliko modela apstrakcije za potrebe svih vrsta korisnika. Za potrebe ovoga rada koristiti ću knjižnicu *Keras Application Programming Interface (API)* koja omogućava visoku razinu apstrakcije što uvelike olakšava implementaciju strojnog učenja za početnika. Za veću fleksibilnost i intuitivnost kod otklanjanja pogrešaka postoji *eager* način rada. Za velike sustave strojnog učenja uvriježeno je upotrebljavati *Distribution Strategy API* za distribuirano treniranje modela na različitim hardverskim konfiguracijama bez mijenjanja definicije modela. Velika prednost TensorFlow aplikacijskog okvira je što ima veliku zajednicu pa je lako dobiti odgovore na eventualna pitanja ili pronaći rješenje problema. Prednost uvriježenosti TensorFlowa su i velike mogućnosti kreiranja korisničkih aplikacija ili sustava. Postoji TensorFlow Lite za pokretanje na mobilnim uređajima, TensorFlow.js za pokretanje na JavaScript okruženjima itd.

2.3.2 TensorFlow Data podatkovni cjevovod (*engl. pipeline*)

Podatkovni cjevovod je alat koji omogućuje prelazak podataka od jedne do druge točke bez gubitaka i smetnji. TensorFlow Data podatkovni cjevovod se obično koristi na velikim skupovima podataka. Primjer problema je ako računalo ima 16 GB radne memorije, a skup podataka ima 64 GB, čitav skup podataka neće se moći učitati u radnu memoriju odjednom. TensorFlow Data podatkovni cjevovod podatke smješta u nizove (*engl. batches*) podataka kako se radna memorija ne bi preopteretila i kako bi se svi podatci mogli procesuirati.

TensorFlow Data podatkovni cjevovod nudi i razne alate za procesuiranje podataka od kojih su neke korištene i u ovome radu. Podatkovni cjevovod omogućava optimizirano učitavanje podataka uz puno bolje performanse. TensorFlow Data podatkovni cjevovod može se direktno dovesti na ulaz Keras modela strojnog učenja.

Poglavlje 3

Pregled strojnog učenja u modeliranju financijskih tržišta

S obzirom na globalnu digitalizaciju, financijska tržišta su ušla u tehnološko doba kako je spomenuto u uvodu. Strojno učenje postao je jedan od važnijih alata pri analiziranju ponavljajućih uzoraka u podacima. Proučavanje tih uzoraka je moguće zbog velikog skupa podataka koji se svakodnevno generiraju zbog frekvencije trgovanja i veličine tržišta. Pošto je opseg podataka prevelik za ljudsku analizu u nekom razumnom vremenu, ljudi su se okrenuli računalima. Algoritimi strojnog učenja imaju mogućnost brze analize kompleksnih podataka i proizvodnje točnijih izlaznih podataka. Za analizu financijskih tržišta koriste se dvije glavne vrste podataka: tekstualni (vijesti, objave na društvenim mrežama itd.) i podatci o tržištu (cijene, tehnički pokazatelji itd.). Razlog korištenja tekstualnih podataka je sve veći utjecaj oblikovanja javnog mnijenja kroz društvene mreže ili kroz ostale medije. Tekstualni podatci se većinom koriste za predviđanje trenda u kojem se pomiče cijena dionice, odnosno klasifikaciju dvije varijable rasta i pada. Podatci o tržištu mogu biti korišteni za klasifikaciju kao i tekstualni podatci, ali su većinom korišteni za direktno predviđanje cijene dionice.

Tehnike strojnog učenja korištene u modeliranju financijskih tržišta iz literature [15]:

- **Artificial Neural Networks (ANN)**

Umjetne neuronske mreže koje su već opisane u sekciji 2.2. Razne verzije ANN su korištene u literaturi zbog svoje svestranosti. Frekventno su korištene. Naveden je primjer studije gdje su uspoređivani MLP (mreža perceptrona sa ulaznim, jednim ili više srednjih i izlaznim slojem), DAN2 i GARCH za predviđanje cijena NASDAQ-a. Modeli su uspoređivani na temelju MSE i MAD parametara. Rezultati su pokazali da je MLP model nadmašio rezultate DAN2 i GARCH-MLP.

- **Support Vector Machines (SVM)**

Tehnika koja se bazira na učenju s učiteljem. SVM ograničava pogreške i mijenja geometrijske margine. SVM je regresijski algoritam i klasifikator uzoraka. Na primjer je dana studija gdje su uspoređivani algoritmi logističke regresije, ojačanih stabla odlučivanja, i SVM. Metrika preciznosti je korištena za ocjenu performansi. Korišteni su numerički podatci povijesnih cijena i semantički podatci iz vijesti i sa Twittera. Ojačana stabla odlučivanja su nadmašila i logističku regresiju i SVM. Ojačana stabla odlučivanja su postigla preciznosti od 54.8% za bankarski sektor, 76% za rudarski sektor i 76.9% za naftni sektor. Logistička regresija je postigla preciznosti 65.4%, 61% i 44.2%. SVM su postigle 51%, 59% i 44.2% za navedene sektore. Studija je predložila uzimanje u obzir unutar-dnevne pomake cijena za preciznije određivanje cijene sljedećeg dana.

- **Naive Bayes (NB)**

Tehnika klasificiranja koja klasificira točke podataka prema Bayesovom Teoremu vjerojatnosti. Ova metoda klasificiranja je vrlo brza pa se koristi na velikim skupovima podataka. Korištena za klasifikaciju tekstualnih podataka za određivanje trendova rasta ili pada.

- **Genetic Algorithms (GA)**

Heuristički pristup rješavanju problema koji oponaša evolucijske procese u prirodi. Koristi se za fino podešavanje parametara za automatsko trgovanje dionicama.

- **Fuzzy Algorithms (FA)**

FA jest logika bazirana na ljudskom odlučivanju gdje se vrijednosti između 0 i 1 pridružuju klasama kao srednje mogućnosti, na čemu se temelji odlučivanje. U literaturi se ova vrsta algoritama koristi za klasifikaciju trendova dionice.

- **Deep Neural Networks (DNN)**

DNN su nadogradnja nad uobičajenim umjetnim neuronskim mrežama. U dubokim neuronskim mrežama se nalazi više skrivenih slojeva. Neuronima imaju zadatak automatski izvlačiti ulazne parametre i raditi transformacije. Često su korištene s numeričkim i s tekstualnim ulaznim podacima. U literaturi je naveden primjer studije gdje su se uspoređivali 4 modela za predviđanje. Korišteni su ARIMA, VAR, LSTM, NARX. Performanse modela su uspoređene metrikom preciznosti. Modeli su analizirani na cijena zatvaranja tržišta NASDAQ-a. NARX je imao dobra kratkotrajna predviđanja, a na dugoročnima nije bio uspješan. Zaključeno je da LSTM bolje uče dugotrajne uzorke.

- **Regression Algorithms (RA)**

Regresijom se pokušava modelirati veza između zavisne i nezavisne varijable. U literaturi su korištene različite vrste regresijskih metoda. Naveden je primjer gdje se je model koristio za dionicu tvrtke po korisnikovoj želji nekoliko dana unaprijed što je dalo zadovoljavajuće rezultate. Model je unaprijeđen različitim algoritmima strojnog učenja i poboljšana preciznost od 85% je postignuta.

- **Hybrid Approaches (HA)**

Hibridni pristup je mješavina različitih tehnika kako bi se poboljšale performanse modela. U literaturi je navedeno da razni hibridni pristupi poboljšavaju modele predviđanja. Naveden je primjer korištenja linearnih (eksponencijalno zaglađivanje) i nelinearnih modela (autoregresijska pomična referentna neuronska mreža). Linearni model je korišten za previđanje. Pogreške predviđanja su izračunate i proslijeđene u nelinearni model. Nelinearni model je minimizirao pogreške. Rezultati i novi pristup su zabilježeni kao obećavajući.

Poglavlje 4

Skupovi podataka

4.1 Porijeklo podataka

Podatci koji su korišteni za treniranje, validaciju i testiranje modela su stvarni javno dostupni podatci. Za dohvaćanje podataka koristio sam *AlphaVantage API* koji zbog partnerstva sa mnogim financijskim institucijama i tržištima ima pristup velikom skupu podataka. Za besplatno korištenje API-a potrebno se je registrirati za dobivanje vlastitog ključa koji se šalje u URL-u svakog zahtjeva. Ograničenje za besplatne korisnike je pet zahtjeva po minuti što ne predstavlja problem uz ispravno korištenje skripte koju sam napravio upravo za te potrebe. [10]

4.2 Podatci korišteni u treniranju

U tablici 4.1 prikazan je uzorak podataka za jedan dan otvorenog tržišta iz skupa podataka koji se prikupljaju na dnevnoj razini. U tablici 4.1 stupac *Otvaranje* reprezentira cijenu prilikom otvaranja tržišta, stupac *Max* reprezentira maksimalnu dnevnu cijenu, a *Min* minimalnu dnevnu cijenu. Stupac *Zatvaranje* predstavlja cijenu prilikom zatvaranja tržišta za taj dan. Objašnjenja svakog korištenog pokazatelja slijede u nastavku.

Tablica 4.1 Primjer podataka za dionicu IBM na dan 10. lipnja 2022.

Otvaranje	Max	Min	Zatvaranje	Volumen	RSI	SMA	EMA
137.80	137.80	135.25	136.19	4661716	52.48	135.88	136.60

4.2.1 Cijena

Glavni podatak koji određuje vrijednost dionice jest njezina cijena. Cijena se određuje ponudom i potražnjom dionice. Za potrebe treniranja modela korištene su dnevne cijene prilikom otvaranja tržišta, prilikom zatvaranja tržišta, maksimalne vrijednosti, minimalne vrijednosti.

4.2.2 RSI

Relative Strenght Index (RSI) jest oscilator indeks relativne snage. RSI mjeri magnitudu oscilacija cijene u cilju detektiranja podcijenjenosti ili precijenjenosti, odnosno previše prodane ili previše kupljene dionice. RSI vrijednosti se kreću u rasponu od 0 do 100. Vrijednosti RSI-a iznad 70 označavaju dionicu koja je previše kupljena ili precijenjena, odnosno da je dionica sklona skoroj korekciji cijene. Vrijednosti ispod 30 označavaju premalo kupljenu ili podcijenjenu dionicu. [11] Period na kojem je izračunat dnevni RSI jest 25 dana.

4.2.3 SMA

Simple Moving Average (SMA) je aritmetički pomični prosjek izračunat dodavanjem novijih cijena i njihovim dijeljenjem sa vremenskim periodima u izračunu prosjeka. Kratkoročni pomični prosjeci brzo reagiraju na promjenu cijene dok dugoročni pomični prosjeci reagiraju sporije. Sve vrijednosti u prošlosti imaju veću težinu. [12] Period na kojem je izračunat dnevni SMA jest 25 dana.

4.2.4 EMA

Exponential Moving Average (EMA) je tip pomičnog prosjeka koji daje veću težinu bližim vremenskim točkama u prošlosti. [13] Period na kojem je izračunat dnevni EMA jest 25 dana.

4.2.5 Volumen

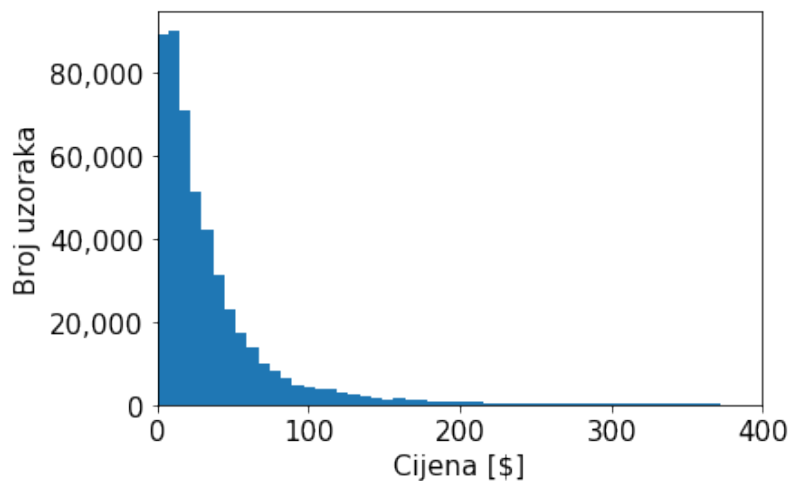
Volumen je broj trgovanja nekim financijskim entitetom u nekom vremenskom periodu. Najčešće je korišten dnevni volumen. Volumen služi kao pokazatelj aktivnosti i popularnosti određene dionice. Obično pomak cijene uz nizak volumen znači i manju važnost tog pomaka, dok pomak cijene uz visok volumen obično označava veću važnost pomaka. [14] U skupu podataka koristio sam dnevni volumen trgovanja određenom dionicom.

4.3 Formatiranje podataka

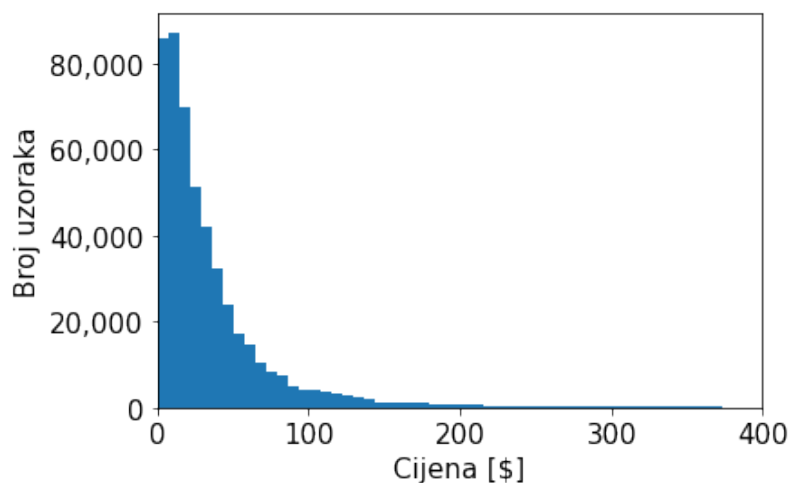
Podatci se iz API-a pomoću skripte spremaju u JSON datoteku. Nakon rasčlanjivanja iz datoteke podatci stariji od 1. siječnja 2018. godine se spremaju u poseban skup podataka za treniranje, podatci mlađi od skupa podataka za treniranje i stariji od 1. siječnja 2020. godine se spremaju u skup podataka za validaciju. Podatci nakon 1. siječnja 2020. godine spremaju se u skup podataka za testiranje.

4.4 Distribucija podataka

Na slici 4.1 možemo vidjeti da su distribucije vrijednosti pomičnih prosjeka cijena slične pošto su brze oscilacije tržišta koje prati EMA malobrojne u odnosu na veliki broj stabilnijih vrijednosti pomičnih prosjeka. Postoje minimalne razlike između grafova 4.1a i 4.1b za koje su zaslužne upravo te brze oscilacije.



(a) SMA

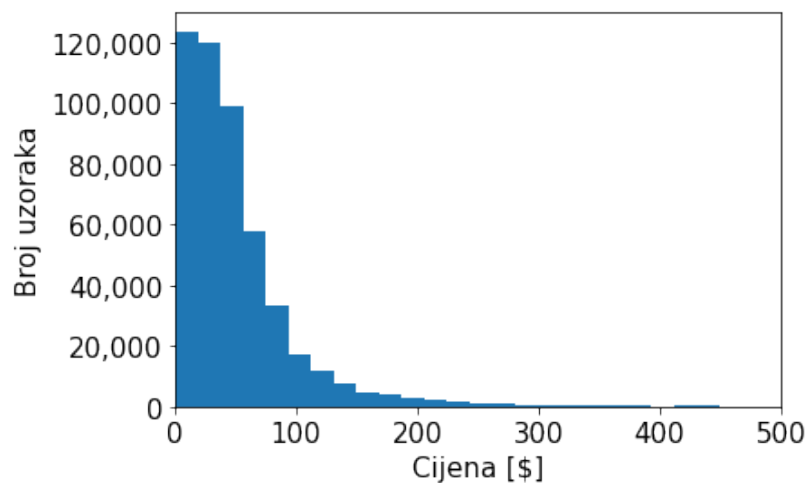


(b) EMA

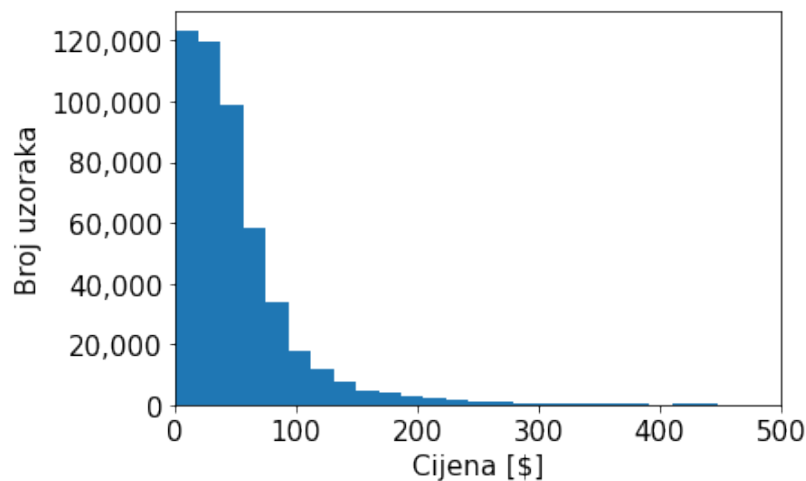
Slika 4.1 Distribucija SMA i EMA podataka u cijelom skupu podataka

Poglavlje 4. Skupovi podataka

Kako možemo vidjeti na slikama 4.2 i 4.3 cijene se u najvećem broju nalaze između 0 i \$100, a najzastupljenije su cijene u rasponu između 0 i \$50. Najzastupljeniji raspon je logičan s obzirom da većina tvrtki krene s manjim cijenama dionica, a tek nekolicina tvrtki je toliko uspješna da im vrijednost dionice na financijskom tržištu naraste na iznose koji se mjere u stotinama ili još rjeđe tisućama američkih dolara.



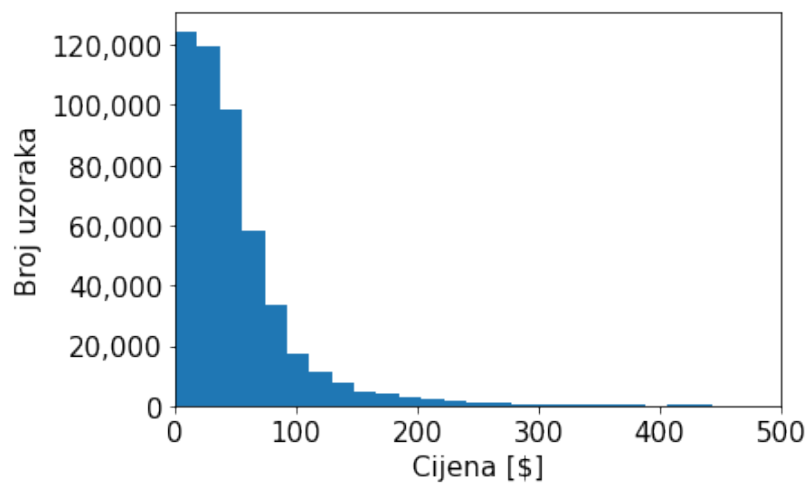
(a) Cijena prilikom otvaranja tržišta



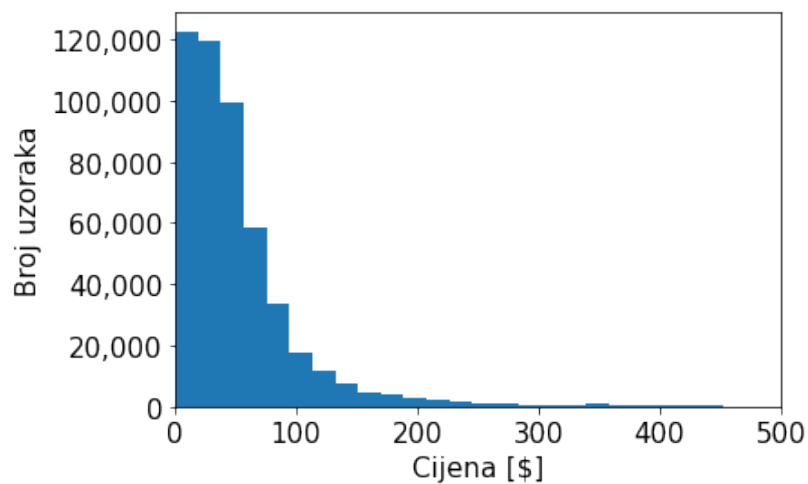
(b) Cijena prilikom zatvaranja tržišta

Slika 4.2 Distribucija cijena prilikom otvaranja/zatvaranja u cijelom skupu podataka

Poglavlje 4. Skupovi podataka



(a) Maksimalna dnevna cijena dionice na tržištu

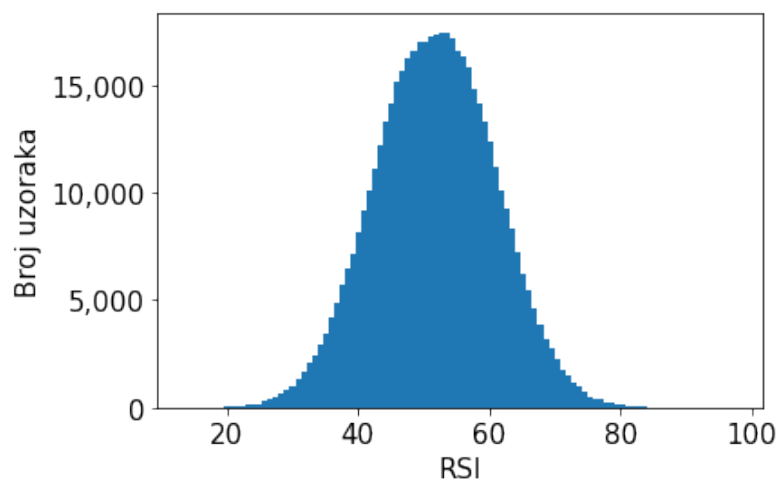


(b) Minimalna dnevna cijena dionice na tržištu

Slika 4.3 Distribucija krajnjih dnevnih cijena u cijelom skupu podataka

Poglavlje 4. Skupovi podataka

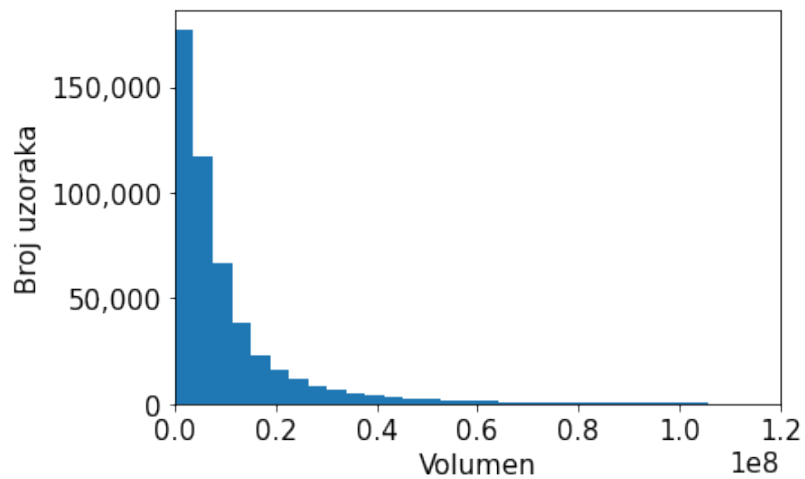
Distribucija RSI vrijednosti prati normalnu raspodjelu odnosno Gaussovu krivulju što možemo vidjeti na slici 4.4 . RSI vrijednosti dionica se većinom nalaze u rasponu gdje nisu niti precijenjene niti podcijenjene, a oko vrijednosti indeksa 50.



Slika 4.4 Distribucija RSI podataka u cijelom skupu podataka

Poglavlje 4. Skupovi podataka

Distribucija dnevnog volumena trgovanja sa dionicom proteže se od 0 do vrijednosti 10^8 , a najviše vrijednosti se nalazi u rasponu do 10^7 .



Slika 4.5 Distribucija podataka o dnevnom volumenu u cijelom skupu podataka

4.5 Normalizacija podataka

Mnogi algoritmi strojnog učenja trendove u podacima dobivaju iz uspoređivanja točki. Problem u tome je ako su podatci na različitim skalama, npr. kako je prikazano u slikama 4.5 i 4.4 gdje je je volumen na skali i do 10^8 a RSI u rasponu vrijednosti 0 do 100. Cilj normalizacije je te podatke svesti na istu skalu. Za normaliziranje korištena je *Min-Max normalizacija*. Kod *Min-Max normalizacije* podataka cilj je najveću vrijednost skupa podataka postaviti kao 1, a najmanju vrijednost na 0. Navedena normalizacija primijenjena je na cijelom skupu podataka. Normalizacija je primijenjena na način da se za svaki parametar u skupu trening podataka izlučuju minimalne i maksimalne vrijednosti. Na temelju minimalnih i maksimalnih vrijednosti parametara izračuna se normalizirana vrijednost parametra za svaki parametar u danu. Formula za *Min-Max normalizaciju* izgleda ovako:

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (4.1)$$

gdje x_{norm} predstavlja normaliziranu vrijednost, x vrijednost koju treba normalizirati, x_{min} minimalnu vrijednost parametra u cijelom skupu podataka, x_{max} maksimalnu vrijednost parametra u cijelom skupu podataka.

Tako iz formule 4.1 za npr. $x_{min} = 10$, $x_{max} = 100$, $x = 50$ dobijemo:

$$x_{norm} = \frac{50 - 10}{100 - 10} = \frac{4}{9}$$

$$x_{norm} = 0.4444$$

Poglavlje 5

Prijedlog modeliranja financijskih tržišta

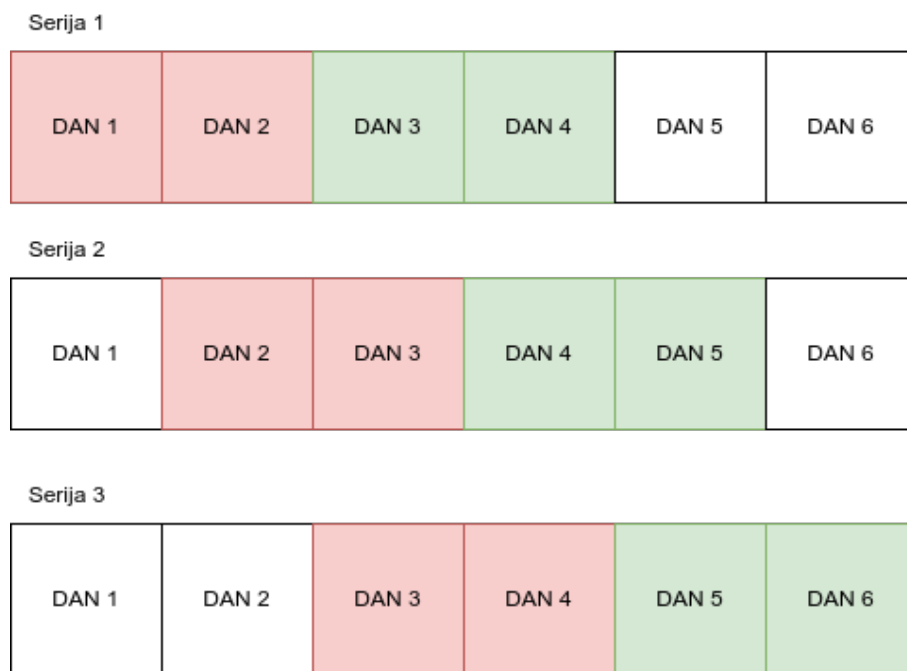
U ovom poglavlju biti će iznesen prijedlog modela strojnog učenja za predviđanje prosječne dnevne cijene dionice na kojemu ovaj rad počiva.

5.1 Podatci

5.1.1 Ulazni podatci

Ulazne značajke (*engl. features*) u predloženi model su podatkovni nizovi od 60 dana, pomak *pomičnog prozora* po jednoj ulaznoj značajki je 1 dan. Primjer pomičnog prozora za uzorkovanje jednog niza ulaznih značajki od 2 dana dan je na slici 5.1, a ulazne značajke označene su crvenom bojom. Na slikovnom primjeru pomak pomičnog prozora također je 1 dan po značajki. Kako je već navedeno u sekciji 4.2 za svaki dan ulazni podatci su cijena dionice kod otvaranja tržišta, najniža cijena, najviša cijena, cijena kod zatvaranja tržišta, RSI, SMA, EMA i dnevni volumen tržišta. Prema navedenom, kada se ti podatci konkatenuju u vektor, jedan niz podataka sadržavati će vektor od 480 vrijednosti. Za jednu iteraciju u epohi u model se šalje 32 podatakovna niza. Korištenje nizova podataka objašnjeno je u podsekciji 2.3.2.

Poglavlje 5. Prijedlog modeliranja financijskih tržišta



Slika 5.1 Prikaz metode pomičnog prozora

5.1.2 Izlazni podatci

Izlazni podatak (*engl. labels*) iz predloženog modela je vektor koji sadrži srednje vrijednosti cijene dionice svakog dana u razdoblju od 30 dana. Primjerice, uzevši vektor podataka za prethodnih 60 dana, izlaz je vektor podataka za narednih 30 dana. Prosječnu cijenu (*engl. typical price*) računamo prema formuli:

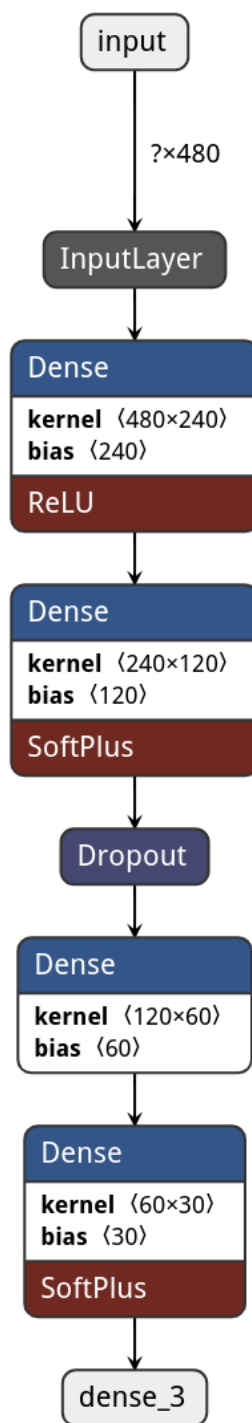
$$TP = \frac{(x_h + x_l + x_c)}{3} \quad (5.1)$$

gdje je TP prosječna cijena dionice, x_h najviša cijena dana, x_l najniža cijena dana i x_c cijena kod zatvaranja tržišta.

5.2 Arhitektura predloženog modela

Na slici 5.2 prikazana je arhitektura predloženog modela. Arhitektura se čita odozgo prema dolje. Na ulazni sloj dovode se serije podataka od 480 vrijednosti kako je objašnjeno u sekciji 5.1.1. Broj serija podataka je označen sa znakom "?" pošto se broj vrlo lako može promijeniti ovisno o potrebama, zahtjevima i specifikacijama sustava na kojem se model trenira. Vrijednosti se nakon toga propagiraju u prvi skriveni *dense* sloj sa aktivacijskom funkcijom ReLU čija je jednažba prikazana u 2.3. Više o *dense* sloju u podsekciji 2.2.3. Prvi sloj ima izlaz od 240 vrijednosti. Sljedeći skriveni *dense* sloj ima aktivacijsku funkciju *SoftPlus* čija jednažba je prikazana u 2.4. Izlaz se sastoji od 120 vrijednosti. *Dropout* sloj objašnjen u podsekciji 2.2.3 je sljedeći. Pretposljednji skriveni *dense* sloj jest onaj sa linearnom aktivacijskom funkcijom čija jednažba je prikazana u 2.2. Izlaz pretposljednjeg skrivenog sloja jest 60 vrijednosti. Zadnji skriveni sloj sadrži *SoftPlus* aktivacijsku funkciju kako bi vrijednosti na izlazu uvijek bile pozitivne (cijena dionice nikada neće biti negativna). Zadnji skriveni sloj 30 vrijednosti (objašnjeno u podsekciji 5.1.2) propagira izlaznom sloju označenom sa *dense_3* na slici arhitekture predloženog modela.

Poglavlje 5. Prijedlog modeliranja financijskih tržišta



Slika 5.2 Prikaz arhitekture predloženog modela

Poglavlje 5. Prijedlog modeliranja financijskih tržišta

Kao optimizator korišten je ADAM sa početnom stopom učenja od 0.0010, koja se pri tri prestanka napretka funkcije pogreške skupa podataka za validaciju dijeli sa 2, a najmanje iznosi 10^{-6} . Funkcija pogreške je MSLE. Model je treniran 50 epoha, pri čemu je dosegno $MSLE = 0.0217$, $MSE = 150.1958$, $MAE = 4.9031$ na trening skupu podataka. Na skupu podataka za validaciju dostignuo je $MSLE = 0.0150$, $MSE = 737.8923$, $MAE = 6.0451$.

5.3 Rezultati predloženog modela

5.3.1 Metrike

Za vrednovanje rezultata modela korištene su funkcije pogreške iz podsekcije 2.2.4 kao i Pearsonovim i Spearmanovim korelacijskim koeficijentima.

Korelacija je statistički postupak za izračunavanje povezanosti podataka. Korišteni korelacijski koeficijenti se definiraju:

- **Pearsonov korelacijski koeficijent**

Vrijednost Pearsonovog koeficijenta $r = 1$ označava jaku pozitivnu vezu između dvije varijable, odnosno kada vrijednost jedne varijable raste, raste i vrijednost druge. Vrijednost $r = -1$ označava jaku negativnu vezu između dvije varijable, rastom vrijednosti jedne varijable, vrijednost druge se smanjuje. Kada je vrijednost Pearsonovog koeficijenta $r = 0$, dvije varijable nisu linearno povezane.[16]

Pearsonov koeficijent izračunava se jednadžbom:

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (5.2)$$

gdje je n broj uzoraka, x_i, y_i indeksirane točke uzoraka, \bar{x} i \bar{y} srednje vrijednosti uzoraka.

- **Spearmanov korelacijski koeficijent**

Spearmanov korelacijski koeficijent označava se sa ρ , ako je vrijednost $\rho = 0$ može se reći da nema nikakve povezanosti dviju varijabli. Za $\rho = 1$ vrijedi da je jedna varijabla savršena pozitivna monotona funkcija druge, a za $\rho = -1$ vrijedi da je jedna varijabla savršena negativna funkcija druge.

Spearmanov korelacijski koeficijent izračunava se pomoću:

$$\rho = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)} \quad (5.3)$$

gdje je d_i razlika između dva uzorka, a n broj uzoraka.

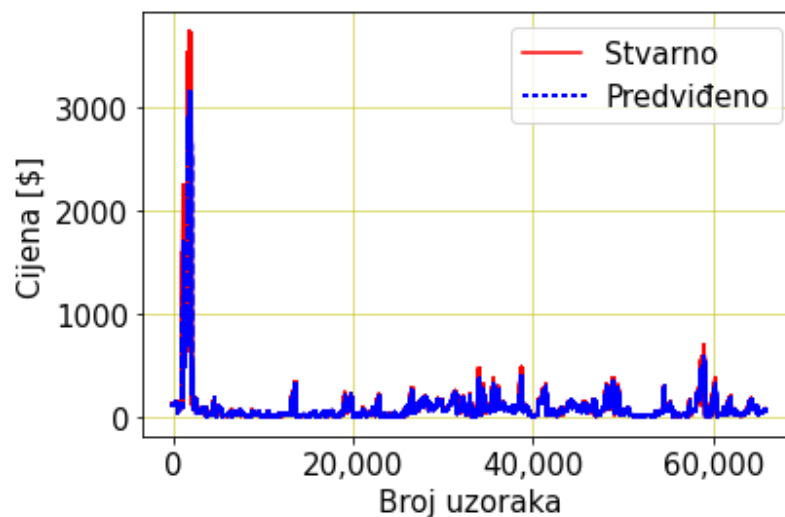
5.3.2 Rezultati cijelog skupa podataka za testiranje

Rezultati na cijelom skupu podataka za testiranje kako je opisano u sekciji 4.3 imaju vrijednosti prikazane u tablici 5.1. Predstavljene vrijednosti funkcija grešaka imaju velika odstupanja što bi dalo naslutiti da predloženi model nije dobar. Međutim, predviđanja velikih vrijednosti manje utječu na funkciju pogreške MSLE u odnosu na druge funkcije pogreške poput MSE pa je MSLE relativno niske vrijednosti što je dobro. Pearsonov i Spearmanov korelacijski koeficijent se mogu aproksimirati kao 1, što daje jaku pozitivnu korelaciju stvarnih i predviđenih vrijednosti. Odnosno, predviđanja modela jako prate trendove rasta i pada stvarnih vrijednosti.

Tablica 5.1 Prikaz metrika za vrednovanje skupa podataka za testiranje

MAE	MSE	RMSE	MSLE	Pearson kor. koef.	Spearman kor. koef.
12.5982	3594.4387	59.9536	0.0222	0.9998	0.9927

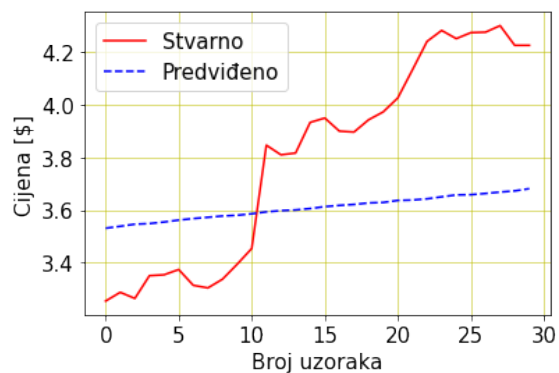
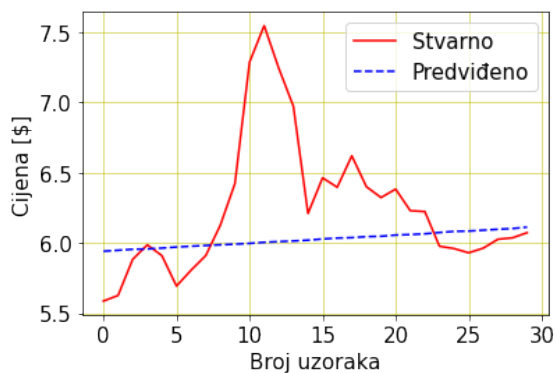
Poglavlje 5. Prijedlog modeliranja financijskih tržišta



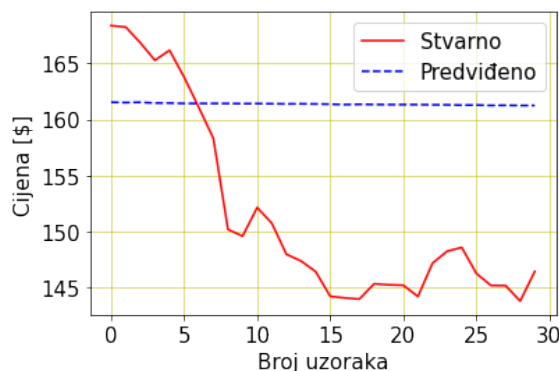
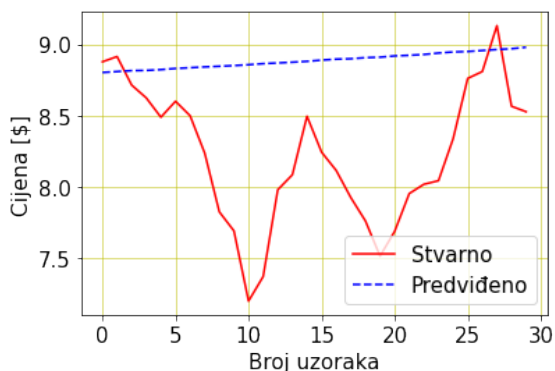
Slika 5.3 Prikaz svih predviđanja modela

Iako se na slici 5.3 svih predviđanja cijelog skupa podataka čini da predviđene vrijednosti (plava boja) skoro savršeno prate stvarne vrijednosti (crvena boja), to nije slučaj već se tako čini s obzirom na rezoluciju promatranja i broj predviđanja. Dokaz te teze prikazan na recentnim podacima dionice IBM na slici 5.7 gdje se po grafu promatra jedno predviđanje u rezoluciji od 30 dana. Na slici 5.4 prikazano je kako model prati stvarne vrijednosti skupa podataka za testiranje.

Poglavlje 5. Prijedlog modeliranja financijskih tržišta



(a) Predviđanja prate trend, ali ne i velike (b) Previđanja prate trend srednje vrijednosti
skokove



(c) Lošija predviđanja, vršne vrijednosti (d) Predviđanja ne prate velike padove

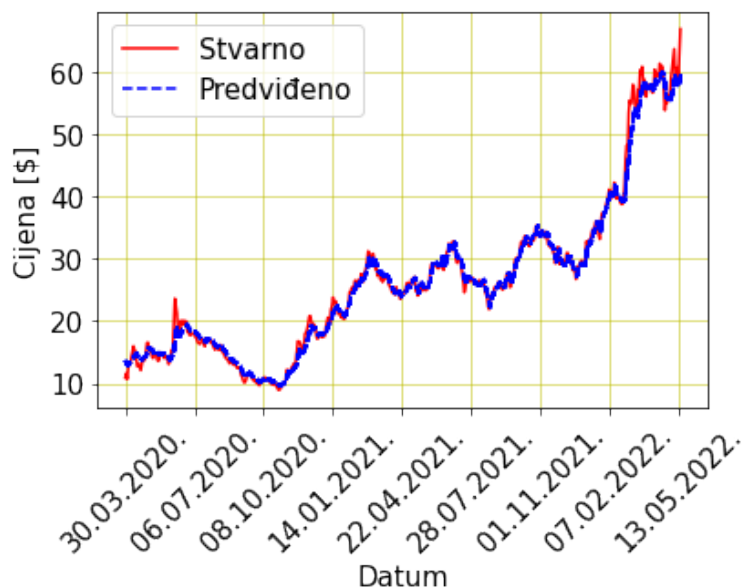
Slika 5.4 Pregled raznih scenarija predviđanja

5.3.3 Rezultati na podacima dionice OXY

Rezultati prikazani u tablici 5.2 su za dionicu tvrtke Occidental Petroleum Corporation (OXY). Usporedbom funkcija pogreške vidljivo je da su odstupanja manja. Predviđene i stvarne vrijednosti dobro koreliraju s obzirom na korelacijske koeficijente. Prikaz svih predviđanja na slici 5.5.

Tablica 5.2 Prikaz metrika za vrednovanje podataka na dionici OXY

MAE	MSE	RMSE	MSLE	Pearson kor. koef.	Spearman kor. koef.
1639.1979	22.3251	4.7250	0.0277	0.9537	0.9212



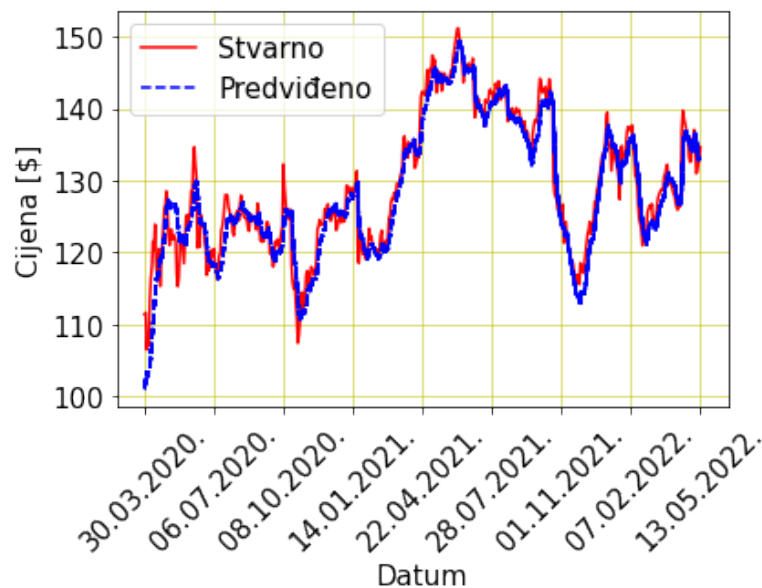
Slika 5.5 Prikaz svih predviđanja modela za dionicu OXY

5.3.4 Rezultati na podacima dionice IBM

Rezultati prikazani u tablici 5.3 su za dionicu tvrtke International Business Machines Corporation (IBM). Usporedbom funkcija pogreške vidljivo je da su odstupanja dosta velika u odnosu na stvarne vrijednosti. Predviđene i stvarne vrijednosti granično dobro koreliraju s obzirom na korelacijske koeficijente. Prikaz svih predviđanja vidljiv je na slici 5.6. Prikaz 9 uzastopnih previđanja na recentnim podacima se nalazi na slici 5.7. Na slici 5.7h vidi se i kako trend predviđanja pada, a trend stvarnih vrijednosti raste.

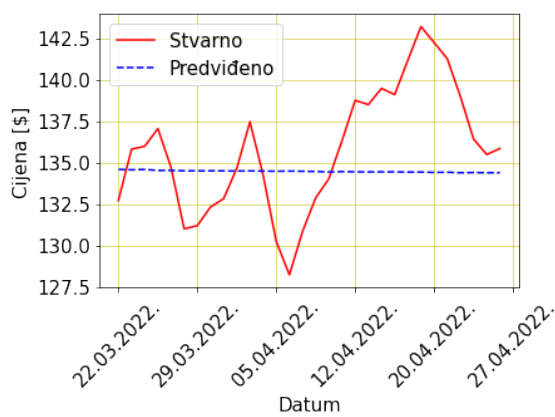
Tablica 5.3 Prikaz metrika za vrednovanje podataka na dionici IBM

MAE	MSE	RMSE	MSLE	Pearson kor. koef.	Spearman kor. koef.
1547.0221	50.1630	7.0826	0.0031	0.7127	0.6713

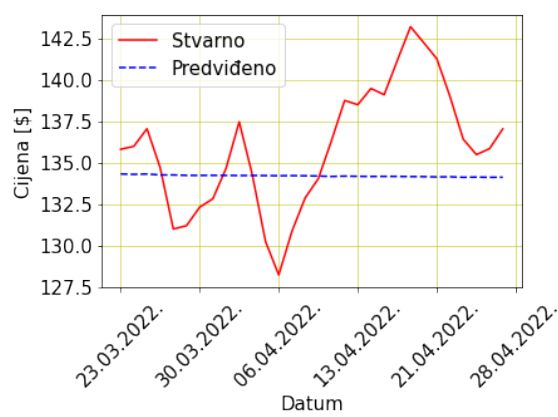


Slika 5.6 Prikaz svih predviđanja modela za dionicu IBM

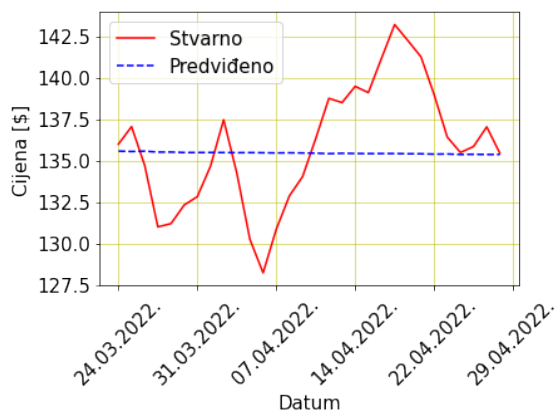
Poglavlje 5. Prijedlog modeliranja financijskih tržišta



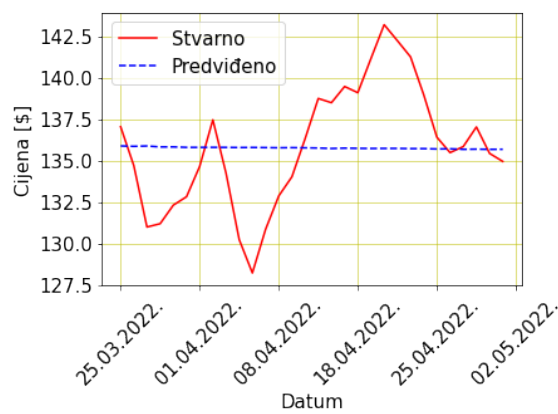
(a)



(b)



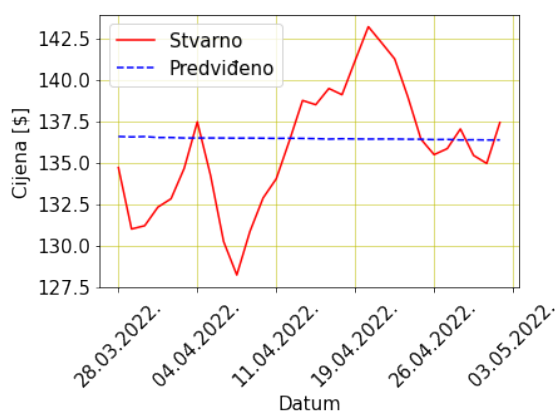
(c)



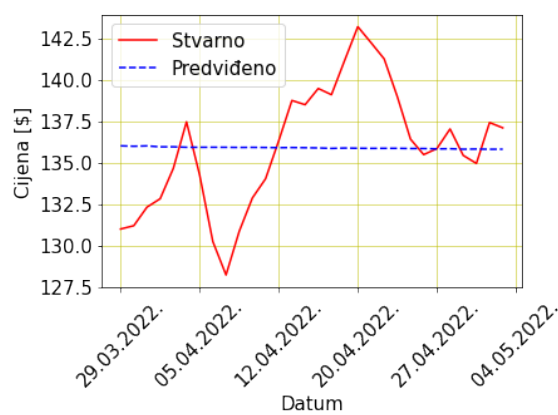
(d)

Slika 5.7 Pregled uzastopnih predviđanja na IBM dionici

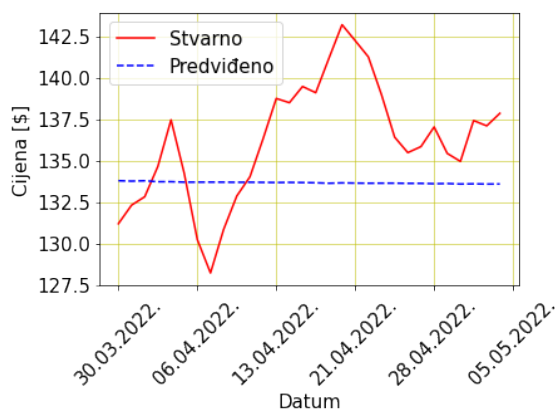
Poglavlje 5. Prijedlog modeliranja financijskih tržišta



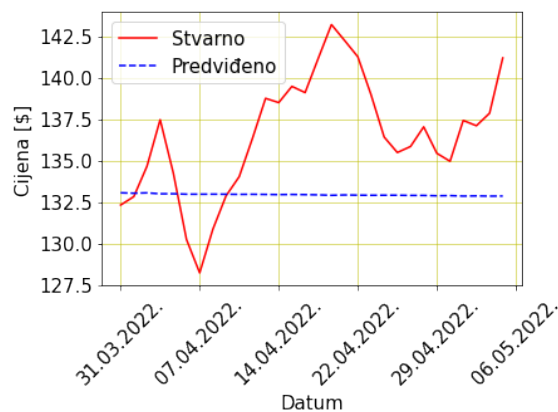
(e)



(f)



(g)



(h)

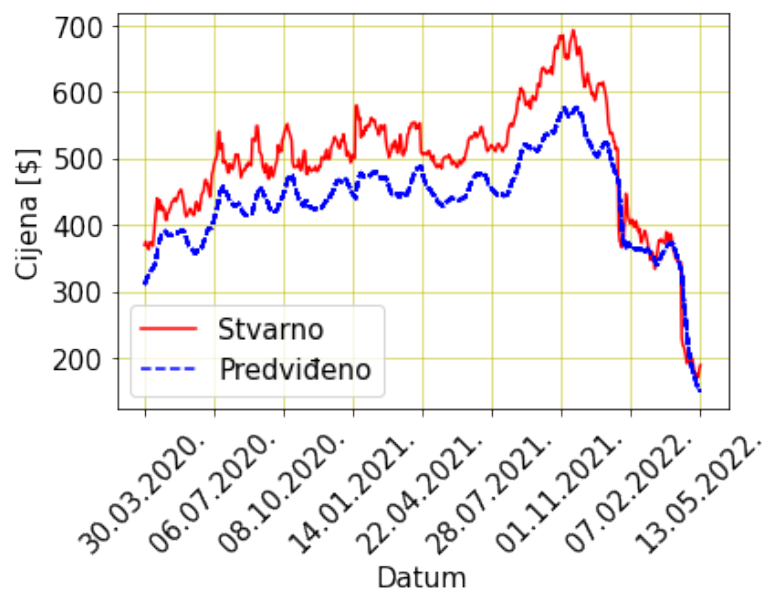
Slika 5.7 Pregled uzastopnih predviđanja na IBM dionici

5.3.5 Rezultati na podacima dionice NFLX

S obzirom na vrijednosti u tablici 5.4 za tvrtku Netflix (NFLX) može se zaključiti da model ne daje realna predviđanja što se tiče cijena (veća odstupanja funkcija pogrešaka). Predviđanja modela zadovoljavajuće koreliraju sa stvarnim vrijednostima prema Pearsonovom i Spearmanovom korelacijskom koeficijentu. Prikaz svih predviđanja nalazi se na slici 5.8.

Tablica 5.4 Prikaz metrika za vrednovanje podataka na dionici NFLX

MAE	MSE	RMSE	MSLE	Pearson kor. koef.	Spearman kor. koef.
1635.8745	5488.6455	74.0854	0.0312	0.8433	0.7621



Slika 5.8 Prikaz svih predviđanja modela za dionicu NFLX

Poglavlje 6

Zaključak

6.1 Ocjena modela

Teško je izvršiti usporedbu s drugim modelima pošto većina modela klasificira trendove cijena na *Gore* i *Dolje* na temelju čega donose zaključke o preciznosti što im daje puno veću šansu za bolju preciznost. Modeli koji predviđaju cijene poput [17] predviđaju cijenu dionice za jedan dan na temelju povijesnih podataka dužeg perioda promatranja. Predloženi model predviđa srednju dnevnu cijenu za 30 dana unaprijed i zbog toga se izdvaja od drugih.

Kao i većina modela koji modeliraju financijska tržišta, predloženi model može poslužiti za informativna predviđanja i kratku analizu financijskog tržišta s obzirom na povijesne vrijednosti i tehničke indikatore. Nudi puno više od klasičnih modela koji samo predviđaju trend rasta i pada. U većini slučajeva predloženi model solidno aproksimira srednju cijenu dionice za period od 30 dana i zadovoljavajuće korelira sa stvarnim vrijednostima. Predloženi model nije primjenjiv kao jedina, detaljna i točna okosnica za ulaganje novca u financijska tržišta.

Predloženi model je moguće unaprijediti dubljom analizom podataka koji mu se daju pri treniranju, poželjno bi bilo postaviti veće težine na brze promjene tržišta kako bi one bile bolje modelirane.

Bibliografija

- [1] McCulloch, W.S.; Pitts, W.: A logical calculus of the ideas immanent in nervous activity, *Bulletin of Mathematical Biophysics* 5, 115–133, 1943.
- [2] Dalbelo Bašić, B.; Čupić M.; Šnajder, J.: "Umjetne neuronske mreže", Fakultet elektrotehnike i računarstva, Zagreb, 2008.
- [3] Ujević Andrijić Ž. : "Umjetne neuronske mreže", *Osvježimo znanje*, 5-6, 219-220, Zagreb, 2019.
- [4] Kok-Leong, S.: Perceptron, s interneta, <https://kseow.com/perceptron>, 09.07.2022.
- [5] Nair, V.; Hinton, G.E.: "Rectified linear units improve restricted boltzmann machines", *ICML'10: Proceedings of the 27th International Conference on International Conference on Machine Learning*, 807–814, 2010.
- [6] Glorot, X.; Bordes, A.; Bengio, Y.: "Deep Sparse Rectifier Neural Networks", *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, PMLR 15, 315-323, 2011.
- [7] Brownlee, J.: "How to Choose Loss Functions When Training Deep Learning Neural Networks", s interneta, <https://machinelearningmastery.com/how-to-choose-loss-functions-when-training-deep-learning-neural-networks>, 20. lipnja 2022.
- [8] Ruder, S.: "An overview of gradient descent optimization algorithms", *arXiv:1609.04747 [cs.LG]*
- [9] Abadi, M.; Barham, P.; Chen, J.: "TensorFlow: A system for large-scale machine learning", *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)*, USENIX Association, pp. 265-283, 2016.
- [10] Alpha Vantage: "About Alpha Vantage", s interneta, <https://www.alphavantage.co/#about>, 18. lipnja 2022.

Bibliografija

- [11] Bhargavi, R.; Gumparthy, S.; Anith R.: "Relative strength index for developing effective trading strategies in constructing optimal portfolio", *International Journal of Applied Engineering Research*, 12(19), 8926-8936, 2017
- [12] Johnston F.R.; Boyland J. E.; Meadows M.; Shale E.: "Some properties of a simple moving average when applied to forecasting a time series", *Journal of the Operational Research Society*, 12, V. 50, 1267-1271, 1999.
- [13] Klinker, F: Exponential moving average versus moving exponential average, *Math Semesterber* 58, 97–107, 2011.
- [14] Gallant,ii A. R.; Rossi, P. E.; Tauchen, G.: "Stock Prices and Volume", *The Review of Financial Studies*, 2, V. 5, 199–242, 1992.
- [15] Rouf, N.; Malik, M.B.; Arif, T.; Sharma, S.; Singh, S.; Aich, S.; Kim, H.-C. "Stock Market Prediction Using Machine Learning Techniques: A Decade Survey on Methodologies, Recent Developments, and Future Directions", *Electronics*, 10(21):2717
- [16] Udovičić, M.; Baždarić K; Bilić-Zulle, L.; Petrovečki, M.: "Što treba znati kada izračunavamo koeficijent korelacije?", *Biochemia Medica*, Vol 17. No. 1, 10-15, 2007.
- [17] Mehtab, S.; Sen, J.; Dutta, A.: "Stock Price Prediction Using Machine Learning and LSTM-Based Deep Learning Models", *Machine Learning and Metaheuristic Algorithms, and Applications* , vol 1366. Springer, 88-106, Singapur, 2021

Pojmovnik

- ANN** Artificial Neural Networks. 12
- API** Application Programming Interface. 10, 15, 17
- ARIMA** Auto-Regressive Integrated Moving Average. 13
- DAN2** Dynamic Artificial Neural Network. 12
- DNN** Deep Neural Networks. 13
- EMA** Exponential Moving Average. vii, ix, 16–18, 24
- FA** Fuzzy Algorithms. 13
- GA** Genetic Algorithms. 13
- GARCH** Autoregressive conditional heteroscedasticity. 12
- GB** Giga Byte. 10
- HA** Hybrid Approaches. 14
- IBM** International Business Machines Corporation. vii, ix, x, 16, 31, 34–36
- JSON** JavaScript Object Notation. 17
- LSTM** Long-Short Term Memory. 13
- MAD** Mean Absolute Deviate. 12

MAE Mean Absolute Error. 8, 30, 33, 34, 37

MLP Multi Layer Perceptron. 12

MSE Mean Squared Error. 8, 12, 30, 33, 34, 37

MSLE Mean Squared Logarithmic Error. 8, 28, 30, 33, 34, 37

NARX Nonlinear Auto-Regressive with exogenous inputs. 13

NASDAQ National Association of Securities Dealers Automated Quotations. 12, 13

NB Naive Bayes. 12

NFLX Netflix. vii, ix, x, 37

OXY Occidental Petroleum Corporation. vii, ix, x, 33

RA Regression Algorithms. 13

ReLU Rectified Linear Unit. 6, 26

RMSE Root Mean Squared Error. 9, 30, 33, 34, 37

RSI Relative Strenght Index. vii, ix, 16, 21, 23, 24

SMA Simple Moving Average. vii, ix, 16–18, 24

SVM Support Vector Machines. 12

URL Uniform Resource Locator. 15

VAR Vector Auto Regression. 13

Sažetak

U ovome radu biti će objašnjeno što je strojno učenje i kako se ono primjenjuje u modeliranju financijskih tržišta. Predložiti će se metodologije za modeliranje financijskih tržišta. Prikazati će se relevantni skupovi podataka koji su bili korišteni za treniranje predloženog modela. Opisati će se korišteni koncepti, metodologije i tehnologije. Predloženi model usporediti će se sa modelima iz literature. Ocijeniti će se mogućnost primjene modela u realnom scenariju. Biti će predložena buduća poboljšanja predloženog modela.

Ključne riječi — **financijska tržišta, strojno učenje, dionice, predviđanje, modeliranje**

Abstract

This paper will explain what is machine learning and how it is implemented in financial market modelling. Methodologies for financial market modelling will be suggested. Relevant datasets that were used for suggested model training will be overviewed. Used concepts, methodologies and technologies will be described. Suggested model will be compared with models in literature. Realistic scenario application of suggested model will be assessed. Future improvements of model will be suggested.

Keywords — **financial markets, machine learning, stocks, forecasting, modelling**