Modeliranje volatilnosti financijskih vremenskih nizova zasnovano na nadziranom učenju

Aleksić, Filip

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

ZAVRŠNI RAD br. 1638

MODELIRANJE VOLATILNOSTI FINANCIJSKIH VREMENSKIH NIZOVA ZASNOVANO NA NADZIRANOM UČENJU

Filip Aleksić

Zagreb, lipanj 2024.

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

ZAVRŠNI RAD br. 1638

MODELIRANJE VOLATILNOSTI FINANCIJSKIH VREMENSKIH NIZOVA ZASNOVANO NA NADZIRANOM UČENJU

Filip Aleksić

Zagreb, lipanj 2024.

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

Zagreb, 4. ožujka 2024.

ZAVRŠNI ZADATAK br. 1638

Pristupnik:	**Filip Aleksić (0036537622)**
Studij:	Elektrotehnika i informacijska tehnologija i Računarstvo
Modul:	Računarstvo
Mentor:	doc. dr. sc. Stjepan Begušić

Zadatak: **Modeliranje volatilnosti financijskih vremenskih nizova zasnovano na nadziranom učenju**

Opis zadatka:

U sklopu ovog završnog rada potrebno je razmotriti problem procjene i predviđanja volatilnosti (standardne devijacije) vremenskih nizova povrata financijskih imovina, te standardne procjenitelje kao šro su uzoračka standardna devijacija ili eksponencijalno otežani procjenitelji. Potom je potrebno razviti model zasnovan na nadziranom učenju koji bi predviđa buduću volatilnost iz značajki izvedenih iz povijesnih vremenskih prozora. Poseban naglasak potrebno je staviti na integraciju predviđenih vektora volatilnosti u matrice kovarijance financijskih vremenskih nizova, te razviti okvir za usporedbu tako sastavljenih matrica kovarijance. Razvijene metode potrebno je primijeniti na povijesnim tržišnim podatcima i provesti statističku analizu rezultata.

Rok za predaju rada: 14. lipnja 2024.

Uvod

U svijetu financija i ulaganja, razumijevanje i predviđanje volatilnosti financijskih instrumenata od ključne je važnosti za investitore, menadžere portfelja i financijske analitičare. Volatilnost, kao mjera varijabilnosti povrata financijskog instrumenta, predstavlja vrlo bitnu komponentu u procjeni rizika i donošenju financijskih odluka.

Točno predviđanje volatilnosti ključno je za investitore i upravitelje rizicima. Volatilnost, često mjerena kao standardna devijacija povrata imovine, odražava stupanj nesigurnosti i rizika na tržištu. Strojno učenje, sa svojim pristupom temeljenim na podacima i sposobnošću otkrivanja složenih uzoraka, pokazalo se kao moćan alat za predviđanje volatilnosti. U ovom radu analizirat ćemo učinkovitost različitih procjenitelja volatilnosti u multivarijatnim financijskim vremenskim nizovima, ulazimo u svijet predikcije volatilnosti pomoću strojnog učenja, istražujući njegove metode, primjene i dubok utjecaj na financijsku industriju.

Glavni cilj ovog rada je istražiti može li se ovom matematičkom problemu predikcije pristupiti kao zadatak nadziranog strojnog učenja te kako korištenje dodatnih značajki i naprednijih modela utječe na naše predikcije volatilnosti. Također, cilj je analizirati kakav utjecaj predikcija volatilnosti ima na procjenu matrice kovarijance koja će nam služiti za kreiranje optimalnih portfelja.

1.1. Pristup predikciji volatilnosti i matrici kovarijance

Tradicionalne metode za predviđanje volatilnosti, poput jednostavnih statističkih modela (npr. povijesna volatilnost) i GARCH modela, često ne uspijevaju prepoznati složene i dinamične uzorke volatilnosti u financijskim vremenskim nizovima. Ovi modeli pretpostavljaju linearne odnose i stacionarnost, što nije uvijek slučaj u stvarnim tržišnim uvjetima. Osim toga, volatilnost može biti podložna naglim promjenama zbog neočekivanih tržišnih događaja, što dodatno komplicira njezinu predikciju.

Zbog toga su u fokusu ovog rada naprednih modela strojnog učenja koji provode predikciju volatilnosti financijskih instrumenata. Strojno učenje nudi fleksibilnost i sposobnost hvatanja nelinearnih odnosa i složenih uzoraka u podacima, što može značajno poboljšati točnost predikcija volatilnosti [1].

Implementiran je pristup koji kombinira strojno učenje za predikciju volatilnosti i korištenje predviđenih volatilnosti s beta koeficijentima za izračun matrice kovarijance. Matrica kovarijance igra ključnu ulogu u kreiranju portfelja jer omogućava kvantifikaciju rizika, identifikaciju mogućnosti za diversifikaciju, optimizaciju portfelja prema ciljanom povratu i riziku, evaluaciju performansi portfelja i kreiranje hedging strategija. Korištenje precizno procijenjenih matrica kovarijance, posebno onih dobivenih naprednim metodama strojnog učenja, može značajno poboljšati učinkovitost upravljanja portfeljem i omogućiti donošenje boljih investicijskih odluka.

Koristit ćemo napredne modele strojnog učenja, uključujući linearnu regresiju, slučajne šume, XGBoost i unaprijednu neuronsku mrežu, kako bismo predvidjeli volatilnost pojedinačnih financijskih instrumenata. Ovi modeli mogu integrirati različite značajke, uključujući povijesne povrate, obujam trgovanja, ekonomske pokazatelje i sentiment vijesti, kako bi uhvatili složene uzorke i dinamične promjene na tržištu.

Jednom kada imamo predviđene volatilnosti, koristit ćemo ih s beta koeficijentima za izračun matrice kovarijance [2]. Beta koeficijent mjeri volatilnost ili sistematski rizik imovine u odnosu na tržište. Kombinacijom ovih informacija možemo izgraditi matrica kovarijance koja bolje odražava trenutne tržišne uvjete i omogućava preciznije procjene rizika.

Iako primarno provodimo analizu kako bismo dobili uvid u sposobnost raznih metoda u predviđanju volatilnosti, ovo istraživanje može pružiti vrijedne informacije za investitore i stručnjake u financijskom sektoru. Na temelju rezultata, možemo bolje razumjeti je li moguće precizno predvidjeti volatilnost dionica i kako ta predviđanja mogu utjecati na strategije ulaganja.

1.2. Struktura rada

Rad je organiziran u nekoliko ključnih poglavlja. Za početak fokusirat ćemo se na teorijsku osnovu i objašnjavanje pojmova koje ćemo koristiti kroz ovo istraživanje. Nakon toga, opisat ćemo metodologiju, pojedine metode strojnog učenja koje koristimo. Također kreirat ćemo referentne modele koji će nam poslužiti kao točke za usporedbu rezultata. Četvrto poglavlje posvećeno je testiranju različitih algoritama i metoda strojnog učenja koje želimo istražiti. Konačno, istraživanje zaključujemo kreiranjem i evaluacijom portfelja temeljenih na dobivenim predikcijama volatilnosti.

2. Teorijska osnova

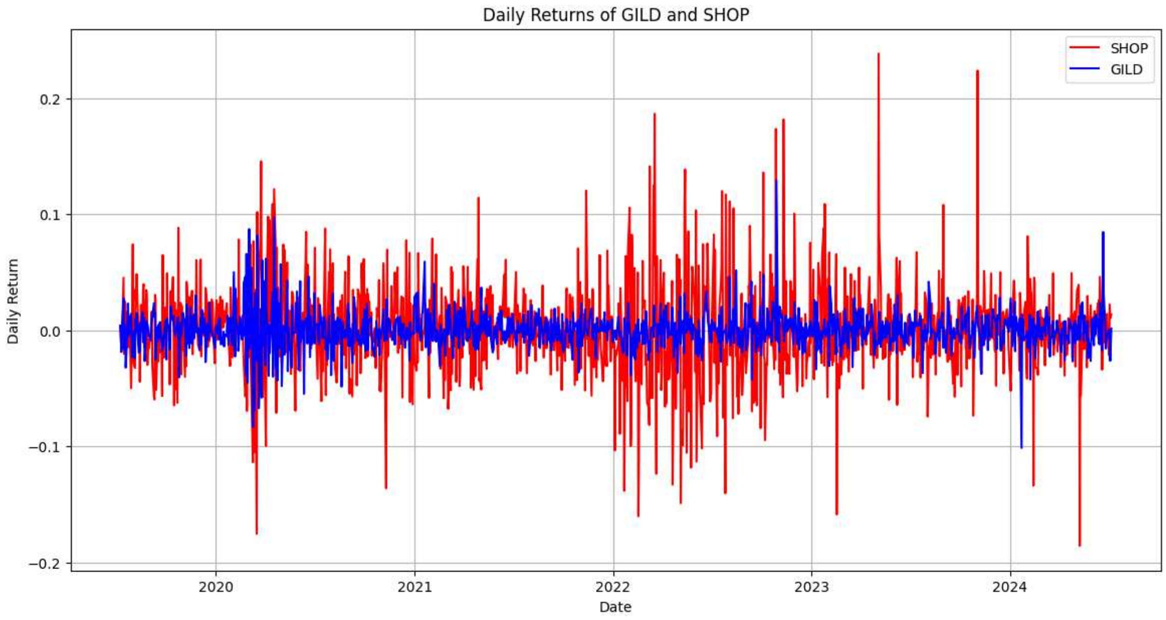
2.1. Volatilnost i njezina važnost na tržištima

Volatilnost je mjera koliko cijena ili vrijednost nekog financijskog instrumenta (npr. dionica, obveznica, roba) varira ili oscilira tijekom vremena. U jednostavnim terminima, volatilnost pokazuje koliko se cijena može promijeniti unutar određenog vremenskog razdoblja. Što je volatilnost veća, to su promjene u cijeni veće i češće.

Usporedili smo dvije dionice: Gilead Sciences Inc. (GILD) kao primjer dionice s niskom volatilnošću i Shopify Inc. (SHOP) kao primjer dionice s visokom volatilnošću. Na (Slika 2.1) vidimo promjenu cijena dionica GILD i SHOP kroz vremenski period od oko 5 godina. GILD pokazuje manje promjene i stabilnije kretanje cijena, dok SHOP pokazuje veće i češće promjene. Može se zaključiti da je SHOP dionica više volatilna i više rizična od dionice GILD zbog većih promjena u cijenama.

Slika 2.1 Prikaz cijena dionica GILD i SHOP

Prikazan je odnos dnevnih postotnih povrata ove dvije dionice (Slika 2.2). Dionica SHOP, prikazana crvenom bojom, ostvaruje veće promjene cijena od dionice GILD koja je prikazana plavom bojom.



Slika 2.2 Odnos dnevnih povrata dionica GILD i SHOP

2.1.1. Procjenjivanje volatilnosti

Volatilnost u financijama prikazuje se kao mjera varijabilnosti ili raspršenosti povrata financijskog instrumenta tijekom vremena [1]. Uzorak povrata korišten je za procjenu volatilnosti. U financijama, nepristrani procjenitelji koriste se kako bi se osiguralo da procjene budu što bliže stvarnim vrijednostima u populaciji. Najčešće korištene metode za mjerenje volatilnosti su varijanca i standardna devijacija.

Povrati financijskog instrumenta između dva vremenska trenutka t i t+1 računaju se po formuli (1) gdje je gdje je P_t cijena dionice u trenutku t, a R_t povrat u periodu t.

$$R_t = \frac{P_{t+1} - P_t}{P_t} \tag{1}$$

Prosječni povrat na nekom prozoru veličine n dana računamo po formuli (2) gdje X_i predstavlja dnevni povrat.

$$\overline{\mathbf{X}} = \frac{1}{n} \sum_{i=1}^{n} X_i \tag{2}$$

Varijanca je statistička mjera koja pokazuje koliko su povrati raspršeni od njihove prosječne vrijednosti. Varijanca se koristi za procjenu koliko se vrijednosti pojedinačnih podataka razlikuju od prosjeka. Nepristrani procjenitelj za varijancu uzorka definira se formulom (2).

$$s^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{\mathbf{X}})^2 \tag{2}$$

Varijanca je izražena u kvadratnim jedinicama varijable čiju varijancu računamo i to nekad predstavlja problem kod interpretacije. Standardna devijacija rješava ovaj problem jer se računa formulom (3). Standardna devijacija se procjenjuje na uzorku po formuli (4)

$$\sigma=\sqrt{\mathrm{Var}(\mathbf{X})} \tag{3}$$

$$s=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{\mathbf{X}})^2} \tag{4}$$

2.1.2. Volatilnost u financijskim instrumentima

Volatilnost je vrlo važan koncept u financijama jer utječe na procjenu rizika i donošenje investicijskih odluka. Investitori koriste volatilnost kako bi procijenili rizik povezan s određenim financijskim instrumentom. Visoka volatilnost ukazuje na veći rizik, jer su cijene sklonije velikim promjenama. Investitori koji ne vole rizik obično preferiraju instrumente s nižom volatilnošću.

U tržištima opcija, volatilnost je ključna komponenta u određivanju cijene opcija [6]. Viša volatilnost obično znači višu cijenu opcije, jer postoji veća vjerojatnost da će cijena osnovnog instrumenta značajno porasti ili pasti.

Menadžeri portfelja koriste volatilnost kako bi kreirali diversificirane portfelje. Kombiniranjem imovina s različitim razinama volatilnosti mogu smanjiti ukupni rizik portfelja. Investitori koriste volatilnost za razvoj trgovačkih strategija. Na primjer, tijekom visoke volatilnosti mogu primijeniti strategije koje iskorištavaju velike promjene cijena.

Volatilnost je temeljni koncept u financijama koji pomaže investitorima i analitičarima da razumiju rizike i potencijalne nagrade povezane s različitim financijskim instrumentima. Razumijevanje volatilnosti omogućava bolje donošenje odluka, bilo da se radi o procjeni rizika, određivanju cijena opcija, upravljanju portfeljem ili bilo kojem ulaganju.

2.2. Kovarijanca

Kovarijanca je mjera koja pokazuje koliko dvije varijable variraju zajedno. U kontekstu financija, kovarijanca se koristi za mjerenje zajedničkog kretanja povrata dviju dionica. Pozitivna kovarijanca ukazuje na to da se povrati dionica kreću u istom smjeru, dok negativna kovarijanca znači da se kreću u suprotnim smjerovima.

Kovarijanca se koristi za razumijevanje odnosa između različitih dionica unutar portfelja. Ako dvije dionice imaju visoku pozitivnu kovarijancu, to znači da će njihove cijene vjerojatno rasti i padati zajedno, što može povećati rizik portfelja. S druge strane, negativna kovarijanca može pomoći u smanjenju rizika portfelja kroz diversifikaciju, jer kada jedna dionica gubi vrijednost, druga može dobivati na vrijednosti, čime se ublažavaju ukupni gubici [2].

Kovarijanca između dvije dionice $\boldsymbol{X_1}$ i $\boldsymbol{X_2}$ može se izračunati pomoću formule (5) gdje su $\boldsymbol{X_1}$ i $\boldsymbol{X_2}$ dnevni povrati dionica, a μ_{X_1} i μ_{X_2} predstavljaju srednju vrijednost povrata na prozoru veličine *n* gdje računamo kovarijancu.

$$\mathrm{Cov}(\boldsymbol{X}_1, \boldsymbol{X}_2) = \frac{1}{n-1} \sum_{i=1}^{n} (X_{1,i} - \mu_{X_1})(X_{2,i} - \mu_{X_2}) \tag{5}$$

Matrica dimenzija *n* × *n,* gdje je n broj dionica u našem portfelju, koja sadrži međusobne kovarijance dionica zove se matrica kovarijance i izgleda ovako (7).

$$\boldsymbol{\Sigma} = \begin{bmatrix} Cov(X_1, X_1) & \cdots & Cov(X_1, X_N) \\ \vdots & \ddots & \vdots \\ Cov(X_N, X_1) & \cdots & Cov(X_N, X_N) \end{bmatrix} \tag{7}$$

Matrica kovarijance ključni je alat u financijama i investicijskom menadžmentu jer omogućava razumijevanje međusobnih odnosa između povrata različitih dionica. Ona pruža informacije o tome kako se dvije ili više dionica kreću zajedno. Matrica kovarijanci je temeljni element u modelima za izradu efikasnih portfelja, kao što je Markowitzova teorija portfelja, pomažući u određivanju optimalnih težina svake dionice u portfelju kako bi se postigao najbolji omjer rizika i povrata [2].

2.3. Korelacija

Korelacija je mjera koja pokazuje snagu i smjer linearne veze između dvije varijable. U kontekstu financija, korelacija se koristi za mjerenje koliko se povrati dviju dionica kreću zajedno. Pozitivna korelacija ukazuje na to da se povrati dionica kreću u istom smjeru, dok negativna korelacija znači da se povrati kreću u suprotnim smjerovima. Korelacijski koeficijent ρ može imati vrijednosti između -1 i 1.

Korelacija, slično kao i kovarijanca, pomaže u razumijevanju odnosa između različitih dionica unutar portfelja. Međutim, dok kovarijanca mjeri apsolutnu zajedničku varijabilnost dviju dionica, korelacija normalizira ovu mjeru, omogućujući usporedbu bez obzira na različite skale povrata dionica. Ako dvije dionice imaju visoku pozitivnu korelaciju, to znači da će se njihove cijene vjerojatno zajedno kretati, što može povećati rizik portfelja. S druge strane, negativna korelacija može pomoći u smanjenju rizika portfelja kroz diversifikaciju, jer kada jedna dionica gubi vrijednost, druga može dobivati na vrijednosti, čime se smanjuju ukupni gubici.

Korelacija između dionica $\boldsymbol{X_1}$ i $\boldsymbol{X_2}$ računa se po formuli (6) gdje su $\sigma_{X_1}, \sigma_{X_1}$ standardne devijacije dionica, a matricu korelacija prikazuje formula (9).

$$\rho_{X_1X_2} = \frac{\mathrm{Cov}(\boldsymbol{X}_1, \boldsymbol{X}_2)}{\sigma_{X_1}\sigma_{X_1}} \tag{6}$$

$$\mathbf{R} = \begin{bmatrix} \rho_{X_1X_2} & \cdots & \rho_{X_1X_n} \\ \vdots & \ddots & \vdots \\ \rho_{X_nX_2} & \cdots & \rho_{X_nX_n} \end{bmatrix} \tag{9}$$

Korelacijska matrica, kao i matrica kovarijance, ključna je u financijama i upravljanju investicijama jer omogućava razumijevanje međusobnih odnosa između povrata različitih dionica. Korelacijska matrica pomaže investitorima u procjeni rizika, omogućavajući analizu rizika povezanih s različitim kombinacijama dionica i identificiranje dionica s visokom pozitivnom ili negativnom korelacijom. Diversifikacija je strategija smanjenja rizika kombiniranjem različitih vrsta investicija, a korelacijska matrica pomaže u identificiranju dionica koje imaju nisku ili negativnu korelaciju, smanjujući ukupni rizik portfelja.

2.4. Faktorski model vremenskih nizova povrata

Faktorski modeli su statistički alati koji se koriste u financijama i ekonomiji za analizu i razumijevanje povrata na investicije. Ovi modeli rastavljaju ukupni povrat na komponente koje su povezane s određenim faktorima. Pretpostavka faktorskog modela je da povrat može biti objašnjen drugim faktorima. U ovom istraživanju koristit ćemo jednostavni jednofaktorski model koji koristi tržišni indeks kao jedini faktor i prikazan je beta koeficijentima.

Beta koeficijent predstavlja mjeru volatilnosti i rizika pojedinih dionica u usporedbi s cjelokupnim tržištem. U svijetu financija beta koeficijenti se koriste za procjenu kako će reagirati neka dionica ako se dogodi promjena u tržištu.

U konkretnom slučaju gledamo promjenu cijene dionice u odnosu na promjenu u tržišnom indeksu. Tržišni indeks može predstavljati bilo koji skup dionica koji se koristi kao reprezentativni uzorak tržišta. U ovom slučaju korišten je portfelj jednakih težina u kojem sadržimo najveće američke dionice. Udio svake dionice u portfelju jednakih težina portfelju je jednak.

Značenje vrijednosti beta koeficijenta

- Beta = 1: Dionica ima istu volatilnost kao i tržište. Ako tržište poraste ili padne za 1%, očekuje se da će i dionica porasti ili pasti za 1%.
- Beta > 1: Dionica je volatilnija od tržišta. Ako tržište poraste ili padne za 1%, dionica će se kretati više od 1% u istom smjeru. Na primjer, beta od 1.5 znači da će dionica porasti za 1.5% ako tržište poraste za 1%.

- Beta < 1: Dionica je manje volatilna od tržišta. Ako tržište poraste ili padne za 1%, dionica će se kretati manje od 1% u istom smjeru. Na primjer, beta od 0.5 znači da će dionica porasti za 0.5% ako tržište poraste za 1%.
- Beta < 0: Dionica se kreće u suprotnom smjeru od tržišta. Ako tržište poraste za 1%, dionica će pasti.

Beta koeficijenti za pojedinu dionicu u ovom radu su izračunati primjenom linearne regresije između dnevnih povrata dionice i dnevnih povrata portfelja jednakih težina, koji je korišten kao prikaz tržišta. U linearnoj regresiji, beta koeficijent je nagib regresijske linije koja prikazuje odnos između povrata dionica (x) i povrata tržišta (y).

Beta koeficijent (β) je izračunat kao nagib regresijske linije uz varijablu x u formuli linearne regresije (10).

$$Y = \alpha + \beta X + e \tag{10}$$

gdje je:

- Y zavisna varijabla,
- α slobodni član (intercept),
- β nagib (slope),
- X nezavisna varijabla,
- e slučajna pogreška.

Izračunati beta koeficijenti koristit će se za značajke kod treniranja modela te izračun korelacija između dionica koje služe za izgradnju optimalnih portfelja minimalne varijance.

2.5. Predikcija matrice kovarijance

Matrica kovarijance ima jako bitnu ulogu kod prikazivanja odnosa među dionicama portfelja. Ako bi mogli precizno predvidjeti matricu kovarijance konstruirali bi portfelj koji bi imao optimalna svojstva. To je ideja ovog rada, da uz pomoć strojnog učenja i predviđanja volatilnosti na kraju predvidimo i matricu kovarijance s kojom ćemo kreirati portfelje. Matrica kovarijance može se prikazati i formulom (11). $\boldsymbol{D}$ je matrica prikazana formulom (12) koja na dijagonalama ima standardne devijacije pojedinih dionica, a ostali elementi su

0. Korelacijska matrica $\boldsymbol{R}$, formula (9), prikazuje korelacije dionica i na mjestima i i j sadrži $\rho_{X_i X_j}$ objašnjenu formulom (6).

$$\boldsymbol{\Sigma}=\boldsymbol{DRD} \tag{11}$$

$$\boldsymbol{D} = \begin{bmatrix} \sigma_{X_1} & 0 & \cdots & 0 \\ 0 & \sigma_{X_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{X_n} \end{bmatrix} \tag{12}$$

Korelacijsku matricu konstruiramo pomoću beta koeficijenata dobivene linearnom regresijom (7). Množenjem beta koeficijenta dviju dionica dobili bi kovarijancu između njih, ali za izračun beta koeficijenata korišteni su standardizirani povrati pa množenjem dobijemo korelaciju između njih (13). Matrica korelacije na dijagonalama treba imati 1 jer je svaka dionica savršeno korelirana sama sa sobom [3]. Iz tog razloga matrici (13) dodajemo matricu (14) kako bi na dijagonalama imali 1, tim zbrojem dobijemo korelacijsku matricu (15).

$$\beta^{\mathrm{T}}\beta = \begin{bmatrix} \beta_1\beta_1 & \beta_1\beta_2 & \cdots & \beta_1\beta_n \\ \beta_2\beta_1 & \beta_2\beta_2 & \cdots & \beta_2\beta_n \\ \vdots & \vdots & \ddots & \vdots \\ \beta_n\beta_1 & \beta_n\beta_2 & \cdots & \beta_n\beta_n \end{bmatrix} \tag{13}$$

$$\Psi = \begin{bmatrix} 1-\beta_1\beta_1 & 0 & \cdots & 0 \\ 0 & 1-\beta_2\beta_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1-\beta_n\beta_n \end{bmatrix} \tag{14}$$

$$R = \beta^{\mathrm{T}}\beta + \Psi \tag{15}$$

$$R = \begin{bmatrix} 1 & \beta_1\beta_2 & \cdots & \beta_1\beta_n \\ \beta_2\beta_1 & 1 & \cdots & \beta_2\beta_n \\ \vdots & \vdots & \ddots & \vdots \\ \beta_n\beta_1 & \beta_n\beta_2 & \cdots & 1 \end{bmatrix} \tag{15}$$

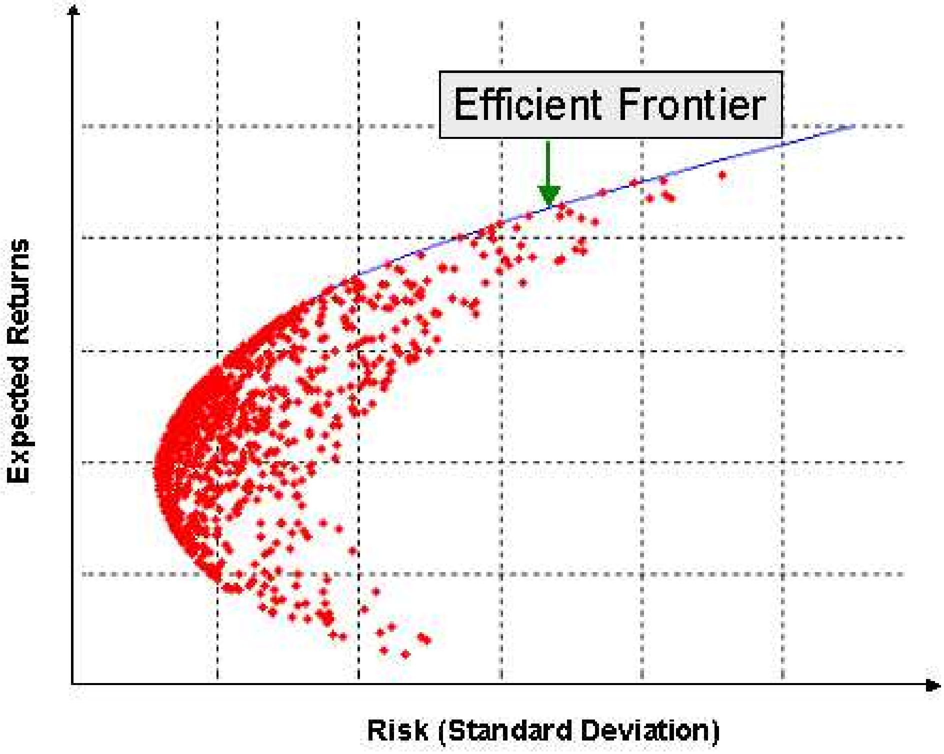
Problem kod predikcije matrice kovarijance je što budući beta koeficijenti i buduća volatilnost nisu poznati i moramo ih procijeniti. Ovaj rad se fokusira na predviđanje volatilnosti modelima strojnog učenja pa ćemo standardnu devijaciju predviđati modelima o kojima će se kasnije pričati više. Beta koeficijente ćemo „predviđati“ uzoračkom procjenom na povijesnom prozoru od 126 dana, pretpostavljamo da će beta koeficijent neke dionice u sljedećih mjesec dana biti jednak kao i beta koeficijent prošlih 6 mjeseci.

Cilj ovog rada je vidjeti koliko dobro možemo predvidjeti volatilnost dionica i pomoću te „dobre“ procjene i uzoračke procjene beta koeficijenata kreirati nadamo se „dobru“ matricu kovarijance. Nakon kreiranja predviđene matrice kovarijance možemo iskoristiti optimizacijske algoritme i kreirati optimalne portfelje. Jedna takva teorija je Markowitzova teorija portfelja koja može kreirati portfelje minimalnog rizika.

2.6. Markowitzova teorija portfelja

Markowitzova teorija portfelja predstavlja temelj moderne teorije portfelja koju je razvio Harry Markowitz 1952. godine. Ova teorija se fokusira na maksimiziranje povrata za zadanu razinu rizika ili minimiziranje rizika za zadani povrat kroz diversifikaciju ulaganja u različite imovine [4].

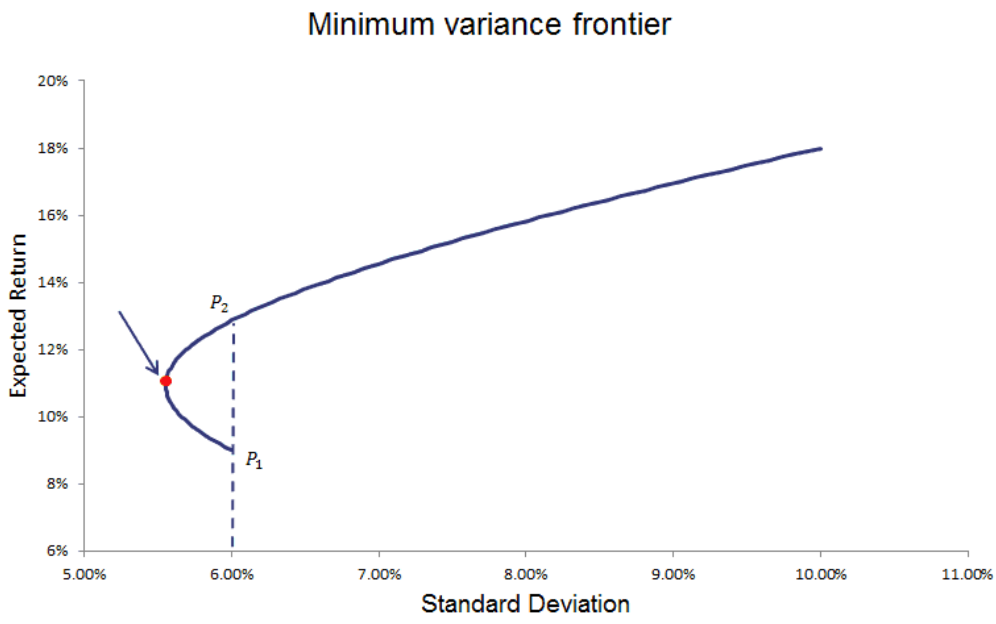
Osnovni koncepti teorije portfelja uključuju diversifikaciju, rizik i povrat te efikasnu granicu. Glavni cilj teorije je smanjiti rizik portfelja diversifikacijom, tj. kombiniranjem različitih imovina koje nisu savršeno korelirane kako bi se smanjio rizik. Svaka imovina ima očekivani povrat i rizik, mjeren standardnom devijacijom, a kombiniranjem imovina možemo stvoriti portfelj s određenim očekivanim povratom i rizikom. Efikasna granica (*Efficient Frontier*) koju prikazuje (Slika 2.3), predstavlja skup svih mogućih portfelja koji imaju maksimalni očekivani povrat za određeni nivo rizika [9]. Portfelji na efikasnoj granici su optimalni jer nude najbolje moguće povrate za određeni nivo rizika. Diversifikacija omogućuje smanjenje rizika bez proporcionalnog smanjenja povrata. Kombiniranjem imovina s niskom ili negativnom korelacijom postiže se veća stabilnost portfelja. Rizik i povrat su ključni faktori u izradi strategija upravljanja portfeljem.



Slika 2.3 Efikasna granica optimalnih portfelja

2.6.1. Portfelj minimalne varijance

Portfelj minimalne varijance je specijalan slučaj Markowitzove teorije portfelja, gdje je cilj minimizirati rizik bez obzira na očekivani povrat [10]. Drugim riječima, to je portfelj s najmanjim mogućim rizikom (varijancom) koji se može konstruirati od zadanog skupa imovina, označen je strelicom na (Slika 2.4).



Slika 2.4 Optimalni portfelj s minimalnom varijancom

Za izgradnju portfelja minimalne varijance potrebna je matrica kovarijance koja prikazuje kako pojedine dionice među djeluju [4]. Problem optimizacije postavljen je formulom (16) u kojem želimo minimizirati varijancu portfelja uz uvjet (17) gdje se sve težine sumiraju u 1. Nakon optimizacije vektor **w** predstavlja težine dionica u portfelju za koje imamo najmanju varijancu.

$$\min_{\mathbf{w}} \mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w} \tag{16}$$

$$\sum_{i=1}^{N} w_i = 1 \tag{17}$$

3. Metodologija

U ovom poglavlju detaljno su opisani koraci i metode koje su korištene za predviđanje volatilnosti dionica američkog tržišta. Fokusirat ćemo se na pripremu podataka, referentne modele, te specifične metode strojnog učenja implementirane u istraživanju.

3.1. Podaci i priprema podataka

Podaci za analizu preuzeti su s Yahoo Finance-a i obuhvaćaju povijesne cijene američkog fonda Russell 3000 za razdoblje od 2000. do 2021. godine. Podaci su prikupljani na dnevnoj bazi.

Nad podatcima je obavljeno dopunjavanje, nedostajuće vrijednosti su zamijenjene susjednim vrijednostima. Podatci su pretvoreni iz cijena dionica u dnevne povrate pomoću formule (1).

Modeli za predviđanje volatilnosti treniraju se sa podatcima izračunatih na pomičnim prozorima. Pomaci pomičnih prozora su 21 dan. Analizirani su povijesni podatci o povratima dionica i nad različitim prozorima su računate značajke pomoću kojih treniramo modele. Izlazna varijabla za treniranje modela je buduća volatilnost nad prozorom od 21 dan, sljedećih mjesec dana.

U svakom trenutku analiziraju se podaci iz četiri različita prozora:

Tjedni prozor: 5 dana

Mjesečni prozor: 21 dan

Tromjesečni prozor: 63 dana

Godišnji prozor: 126 dana

Unutar svakog prozora, računaju se značajke koje se koriste u modelima predikcije volatilnosti.

3.1.1.1 Značajke korištene u modelima

Odabir značajki za treniranje modela bio je usmjeren na značajke koje bi mogle dobro opisivati buduću volatilnost i imati utjecaj na kretanje cijena dionice. Za ovo istraživanje korištene su četiri različite značajke koje su računate na pomičnim prozorima povijesnih podataka.

Prosječni povrati

Prosječni povrati odabrani su jer predstavljaju osnovni pokazatelj dugoročnog trenda cijene dionice. Analizom povijesnih povrata može se dobiti uvid u opću izvedbu dionice tijekom vremena. Stabilni prosječni povrati mogu signalizirati nižu buduću volatilnost, dok nestabilni ili negativni prosječni povrati mogu ukazivati na povećanu volatilnost. Uključivanjem ove značajke model može razlikuje dionice s različitim rizicima temeljenim na njihovoj povijesnoj izvedbi.

Beta koeficijenti

Beta koeficijenti odabrani su jer mjere osjetljivost dionice na promjene u širem tržištu. Dionice s različitim beta vrijednostima pokazuju različite stupnjeve rizika u odnosu na tržišne promjene. Dionice s visokom beta vrijednošću obično imaju veću volatilnost, jer su osjetljivije na tržišne fluktuacije. Beta koeficijenti omogućuju modelu da procijeni kako promjene na tržištu mogu utjecati na volatilnost određene dionice.

VIX indeks

VIX je indeks koji mjeri očekivanu volatilnost tržišta u sljedećih mjesec dana. Računa se pomoću implicirane volatilnosti opcija na indeks S&P 500 indeks. Niske vrijednosti VIX indeksa upućuju na stabilno i smireno stanje na tržištu, dok visoke vrijednosti sugeriraju nesigurnost ili strah među investitorima. VIX indeks odabran je kao jedna od značajki jer mjeri buduću volatilnost i široko je prihvaćen kao pokazatelj tržišnog rizika i nesigurnosti. Odabirom VIX-a model ima informaciju o općoj tržišnoj sigurnosti uzrokovane opcijama.

Standardna devijacija

Standardna devijacija odabrana je jer izravno mjeri raspršenost povrata oko prosjeka. To je osnovna mjera volatilnosti. Visoka standardna devijacija pokazuje velike fluktuacije u povratima, što je izravni pokazatelj visoke volatilnosti. Ova značajka pomaže modelu da precizno predvidi buduće oscilacije cijena dionica na temelju povijesne volatilnosti.

Odabrane značajke računamo na četiri različita prozora, tjedni, mjesečni, tromjesečni i polugodišnji. Korištenje različitih prozora za računanje značajki omogućava modelu da hvata informacije o ponašanju dionice u različitim vremenskim okvirima, pružajući višeslojni uvid. Kratki prozori otkrivaju trenutne promjene, dok dulji prozori daju kontekst dugoročnih trendova, poboljšavajući robusnost modela i smanjujući rizik prevelike prilagodbe jednom prozoru. Ovakav pristup nam omogućava precizniju i pouzdaniju analizu podataka. Nakon svih prozora imamo 16 različitih značajki, svaka od 4 glavne značajke prikazana na 4 različita prozora.

3.2. Referentni modeli

Za evaluaciju performansi modela strojnog učenja za predikciju volatilnosti koristit ćemo tri referentna modela. Ovi modeli služe kao točke za usporedbu i procjenu učinkovitosti naših modela.

3.2.1. Oracle Metoda

Oracle metoda predstavlja idealni referentni model koji savršeno predviđa buduću volatilnost. Ova metoda koristi stvarne vrijednosti buduće volatilnosti za određeni vremenski period kao predikcije, omogućujući nam izračunavanje srednje kvadratne pogreške (MSE) i pružajući referentnu točku za procjenu koliko su dobre ostale metode u usporedbi s idealnom predikcijom. Iako je nerealno imati savršenu predikciju u stvarnim uvjetima, Oracle metoda služi kao savršeni model za usporedbu. Ona postavlja gornju granicu učinkovitosti koju ostali modeli mogu dosegnuti, pružajući nam najbolji mogući scenarij protiv kojeg možemo mjeriti performanse naših modela strojnog učenja.

3.2.2. Uzorački procjenitelj

Uzorački procjenitelj predviđa buduću volatilnost na temelju pretpostavke da će volatilnost iz prethodnog mjeseca ostati ista u budućem prozoru. Konkretnije, ova metoda izračunava volatilnost za prethodni mjesec, konkretno standardnu devijaciju dnevnih povrata za prethodnih 21 dan, i koristi tu volatilnost kao predikciju za budući mjesec. Iako je ovaj pristup jednostavan za implementaciju i koristi nedavne podatke za predikciju, njegova glavna slabost leži u pretpostavci da se volatilnost neće mijenjati. To ignorira promjene u tržišnim uvjetima i dinamičnu prirodu financijskih tržišta, čime se može pokazati kao neadekvatan u nestabilnim ili promjenjivim tržišnim uvjetima. Ipak, uzorački procjenitelj pruža osnovni model za usporedbu, omogućujući nam da procijenimo koliko su sofisticiraniji modeli bolji u odnosu na jednostavne pristupe.

3.2.3. Portfelj jednakih težina

Portfelj jednakih težina pretpostavlja da će sve dionice u portfelju imati jednaku težinu i koristi tu pretpostavku za predikciju volatilnosti portfelja. Ova metoda izračunava volatilnost portfelja koristeći jednake težine svih dionica, bez obzira na njihove individualne karakteristike i rizike. Iako je pristup jednostavan i ne zahtijeva složene izračune ili modele, njegova glavna mana je ignoriranje specifičnih karakteristika i rizika pojedinačnih dionica te promjena u tržišnim uvjetima ili korelacija između dionica. Portfelj jednakih težina često se koristi u praksi kao početni model zbog svoje jednostavnosti, ali služi kao referentna točka za usporedbu s naprednijim metodama. Korištenjem ove metode možemo procijeniti koliko su naši sofisticiraniji modeli učinkoviti u odnosu na jednostavan pristup ravnomjernog ponderiranja.

Korištenjem Oracle, uzorački procjenitelj i portfelj jednakih težina kao referentne modele, možemo sveobuhvatno procijeniti učinkovitost naših modela strojnog učenja za predikciju volatilnosti. Oracle metoda pruža idealnu referentnu točku, uzorački procjenitelj koristi jednostavan pristup temeljen na povijesnim podacima, dok portfelj jednakih težina koristi ravnomjerno ponderiranje za procjenu volatilnosti portfelja. Ovi referentni modeli omogućuju nam bolje razumijevanje performansi naših modela i identificiranje područja za poboljšanje.

3.3. Korištene Metode Strojnog Učenja

3.3.1. Linearna regresija

Linearna regresija je osnovna metoda statističke analize koja modelira odnos između jedne ovisne varijable (ciljne varijable) i jedne ili više neovisnih varijabli (prediktora ili značajki). U kontekstu višestruke linearne regresije, model koristi više značajki kako bi predvidio ciljnu varijablu. Jednadžba višestruke linearne regresije može se izraziti kao:

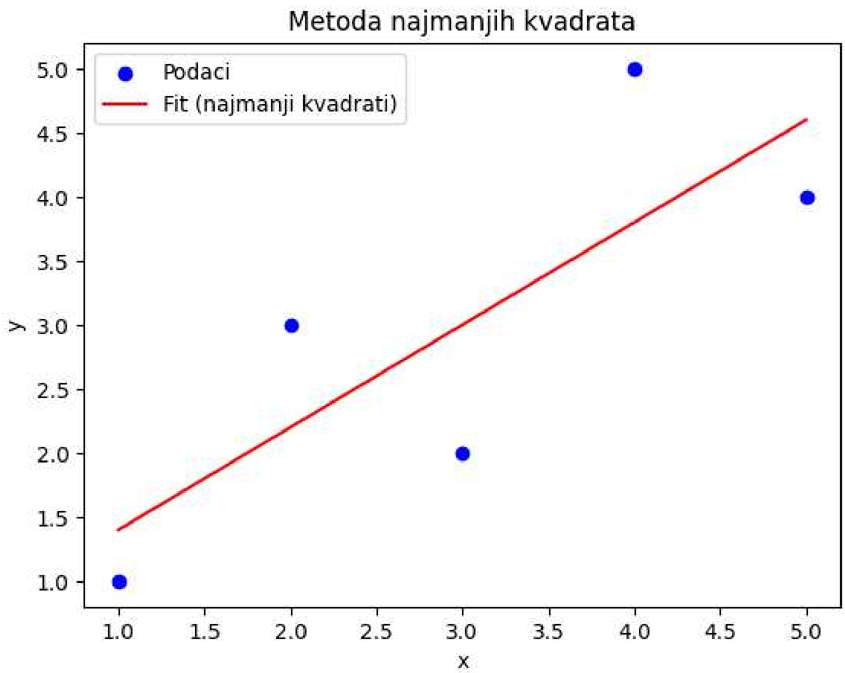
$$y = \beta_0 + x_1 \cdot \beta_1 + \ldots + x_{n-1} \cdot \beta_{n-1} + x_n \cdot \beta_n \qquad (18)$$

gdje su:

- y predviđeni koeficijent
- $x_1, x_2 \ldots x_n$ vrijednosti značajki
- $\beta_1, \beta_2 \ldots \beta_n$ težine koje opisuju određenu značajku
- β_0 slobodni član

Linearna regresija koristi metodu najmanjih kvadrata (*least squares method*) za procjenu koeficijenata (β). Cilj metode najmanjih kvadrata je minimizirati sumu kvadrata razlika između stvarnih i predviđenih vrijednosti ciljne varijable (19). (Slika 3.1) prikazuje stvarne podatke označene točkom te pravac koji smo dobili minimiziranjem kvadratnih razlika između stvarnih i predviđenih vrijednosti y.

$$S = \sum_{i=0}^{n} (y_i - (\beta_0 + \beta_1 x_i)) \qquad (19)$$



Slika 3.1 Metoda najmanjih kvadrata

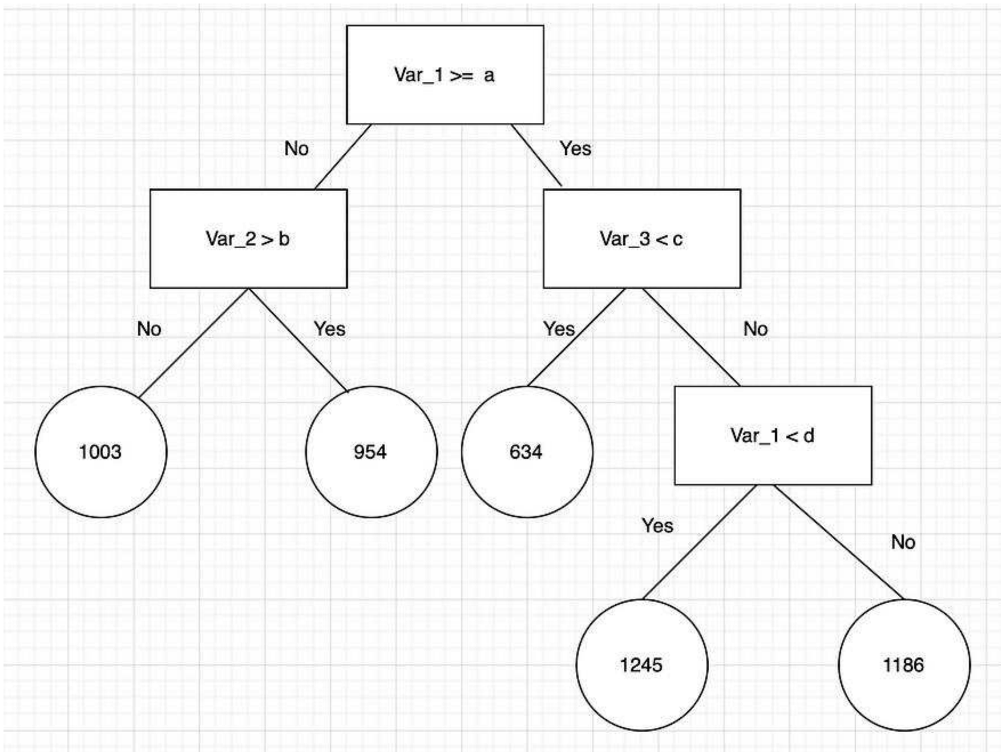
U ovom istraživanju, linearna regresija koristi se za predikciju volatilnosti na temelju skupa od 16 značajki koje uključuju povrate, beta koeficijente, VIX indeks i standardnu devijaciju, izračunate na četiri različita vremenska prozora: tjedni (5 dana), mjesečni (21 dan), tromjesečni (63 dana) i godišnji (126 dana). Predviđeni koeficijent y u našem slučaju je standardna devijacija dionice u sljedećih mjesec dana, dok su značajke x_1, x_2 … x_{16} vrijednosti naših odabranih značajki. Koeficijenti uz značajke, β_1, β_2 … β_n predstavljaju važnost samih značajki, koliko se ciljna varijabla y mijenja za jednu jedinicu promjene u značajki x_i, dok ostale značajke ostaju konstantne.

Linearna regresija je ključni alat u ovom istraživanju jer omogućuje kvantificiranje i razumijevanje utjecaja različitih značajki na volatilnost. Korištenjem linearne regresije, možemo precizno modelirati i interpretirati složene odnose između financijskih varijabli, što je ključno za učinkovito upravljanje portfeljem i donošenje informiranih investicijskih odluka.

3.3.2. Slučajne šume

Slučajne šume je algoritam strojnog učenja koji se koristi za klasifikaciju i regresiju. Radi se o ansambl metodi koja kombinira predikcije iz više odluka stabala kako bi se poboljšala točnost i robusnost modela. Svako pojedinačno stablo u šumi trenira se na različitim podskupovima podataka uz korištenje nasumičnog izbora značajki, što smanjuje problem prekomjernog učenja (overfitting) i povećava generalizacijske sposobnosti modela.

(Slika 3.2) prikazuje osnovni dijagram stabla odluke koji počinje s varijablom Var_1 i dijeli se na temelju specifičnih kriterija. Kada je odgovor 'da', stablo odluke slijedi predstavljeni put, a kada je odgovor 'ne', ide niz drugi put. Ovaj proces se ponavlja dok stablo ne dosegne list (čvor) i ne odredi rezultat. U našem istraživanju Var_1, Var_2, Var_3 i ostale varijable predstavljaju naše značajke, kroz učenje naša stabla odluke provjeravaju zadovoljavaju li vrijednosti naših značajki uvjete i po tome se određuje predviđena vrijednost volatilnosti.



Slika 3.2 Stablo odluke

Slučajne šume algoritam konstruira mnogo odluka stabala tijekom treniranja i svakom stablu dodijeli nasumični podskup podataka na kojima će se to stablo trenirati, to se naziva bootstrap metoda. Nad svakim čvorom, krećući od korijena, razmatra se određen broj značajki za koje će se provjeravat uvjet. Odabrana značajka će biti ona koja nam daje najviše informacija odnosno najbolje dijeli podatke.

Ovo su neki od bitnijih hiperparametara koji su korišteni u ovom istraživanju:

- n_estimators=300, broj različitih stabala koji se kreira
- max_features='sqrt', broj značajki koji se razmatra u svakom čvoru, u slučaju ovog istraživanja imamo 16 značajki, model slučajnih šuma će za svaki čvor razmatrati 4 nasumične značajke
- max_depth=30, maksimalna dubina stabla

Algoritam slučajnih šuma koristi srednju vrijednost (za regresiju) ili glasovanje većine (za klasifikaciju) rezultata svakog stabla za donošenje konačne predikcije.

Ovaj algoritam je korišten za predikciju volatilnosti zbog svoje otpornosti na prekomjerno učenje, sposobnost analize velikog broja podataka te pružanja važnosti pojedinih značajki.

3.3.3. XGBoost

XGBoost (Extreme Gradient Boosting) je napredni algoritam strojnog učenja koji se koristi za regresiju i klasifikaciju. XGBoost koristi tehniku gradijentnog boostinga koja kombinira predikcije više modela (u ovom slučaju, stabala odluke) kako bi se poboljšala točnost i robusnost konačnog modela. U financijama se XGBoost široko koristi za razne zadatke, uključujući predikciju cijena dionica, procjenu kreditne sposobnosti i, kao u ovom radu, predikciju volatilnosti.

XGBoost radi kroz iterativni proces u kojem se svaki novi model trenira na greškama prethodnih modela . Algoritam počinje s jednostavnim modelom i zatim dodaje nova stabla odluke koja nastoje ispraviti preostale greške (reziduale) prethodnih modela. Svako stablo konstruira se tako da minimizira funkciju gubitka, kao što je srednja kvadratna greška (MSE), koristeći tehniku gradijentnog spusta. Stabla se kreiraju jedno za drugim, sekvencijalno treniranje, za razliku od algoritma slučajnih šuma gdje su se stabla trenirala paralelno. Konačna predikcija sastoji se od sumiranih predikcija svih stabala u ansamblu. Proces se ponavlja dok se ne dostigne zadani broj iteracija (stabala) ili dok greška ne prestane značajno opadati.

XGBoost koristi nekoliko tehnika za smanjenje greške i sprječavanje prenaučenosti (*overfittinga*), uključujući regularizaciju, kontrolu složenosti modela i metodu ranog zaustavljanja (*early stopping*).

U ovom radu koristili smo sljedeće hiperparametre za naš XGBoost model:

- learning_rate = 0.1, ovaj parametar kontrolira brzinu učenja modela.
- max_depth = 20, maksimalna dubina svakog stabla
- n_estimators = 200, broj stabala koja će se koristiti u modelu
- subsample = 0.8, udio podataka koji će se koristiti za treniranje svakog stabla
- colsample_bytree = 0.7, udio značajki koje će se koristiti pri konstruiranju svakog stabla

XGBoost algoritam je odabran za ovo istraživanje zbog robusnosti na prenaučenost modela, može obraditi veliku količinu podataka te je poznat po svojoj točnosti predikcija.

3.3.4. Unaprijedna neuronska mreža (MLP)

Umjetne neuronske mreže inspirirane su biološkim neuronskim mrežama u mozgu. Sastoje se od umjetnih neurona, koji oponašaju funkcije bioloških neurona i njihovih sinapsi [5]. Umjetni neuroni izvode matematičke operacije koje ulaznim vrijednostima dodjeljuju težine i zatim primjenjuju aktivacijsku funkciju kako bi proizveli izlaz. Iako pojedinačni neuron nije vrlo moćan, mreža neurona koja je pravilno raspoređena i međusobno surađuje može postati vrlo moćan alat za analizu i predikciju.

Neuron se sastoji od ulaza, pripadnih težina i aktivacijske funkcije. Tijekom treniranja mreže, težine se optimiziraju kako bi se minimizirala pogreška modela. Izlaz k-tog neurona može se zapisati kao:

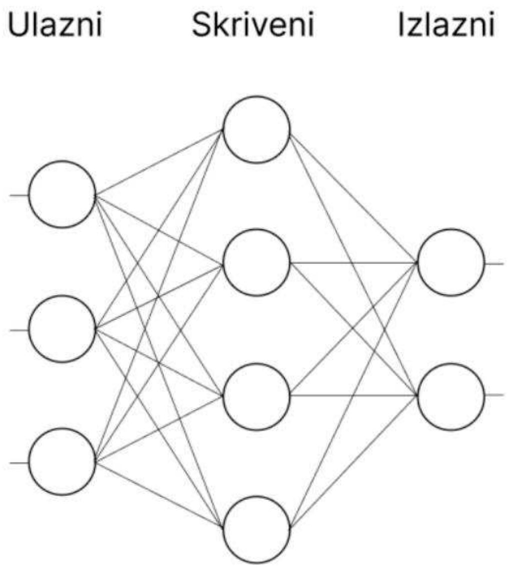
$$y_k = \phi(\sum_{j=0}^{m} w_{kj} x_j \tag{20}$$

gdje su:

- y_k izlaz k-tog neurona
- ϕ aktivacijska funkcija
- w_{kj} težine pridružene ulazima

- x_j ulazne vrijednosti

Neuronska mreža sastoji se od ulaznog sloja, jednog ili više skrivenih slojeva i izlaznog sloja. Svaki sloj može sadržavati različit broj neurona. (Slika 3.3) prikazuje strukturu jednostavne potpuno povezane neuronske mreže.



Slika 3.3 Slojevi neuronske mreže

- Ulazni sloj: prima ulazne podatke.
- Skriveni slojevi: obrađuju podatke
- Izlazni sloj: proizvodi konačnu predikciju.

Postoje različite vrste neuronskih mreža, uključujući konvolucijske neuronske mreže (CNN) za obradu slika, povratne neuronske mreže (RNN) za obradu sekvenci i MLP (Multi-Layer Perceptron) za razne primjene, uključujući financijske predikcije [8].

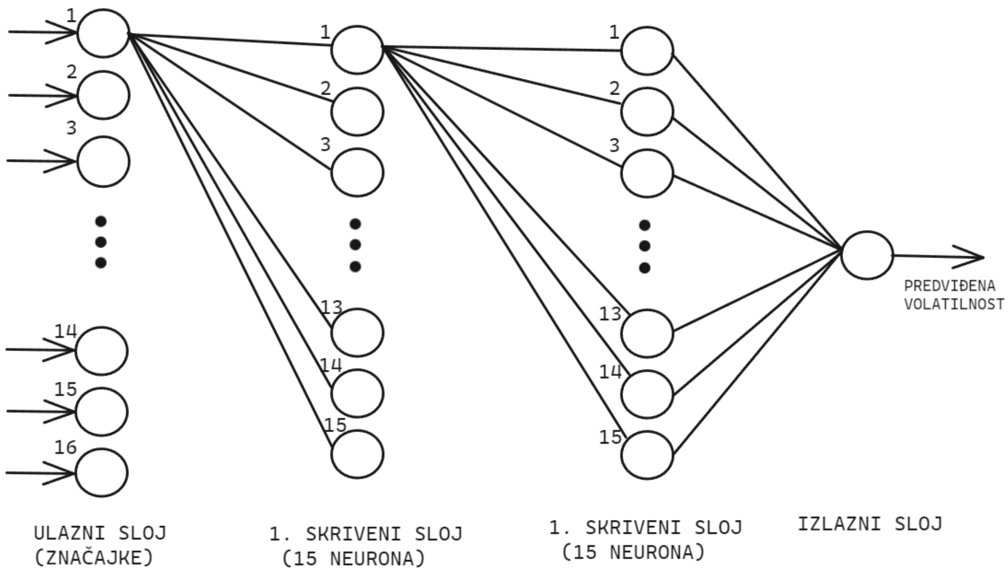
MLP je vrsta potpuno povezane umjetne neuronske mreže koja se sastoji od najmanje tri sloja neurona: ulaznog sloja, jednog ili više skrivenih slojeva i izlaznog sloja. MLP koristi naprijed propagaciju za predikciju i unatrag propagaciju za optimizaciju težina kroz algoritam gradijentnog spusta.

Propagacija unaprijed funkcionira tako da ulazne vrijednosti prolaze kroz slojeve mreže i tako kreiranju izlaznu vrijednost. Svi neuroni u slojevima primaju ulazne vrijednosti, množe ih s težinama, sumiraju sve umnoške i primjene aktivacijsku funkciju.

Aktivacijska funkcija dodaje nelinearnost modelu tako što izlaze iz neurona pretvaraju u druge vrijednosti. To omogućuje mreži da nauči složene uzorke u podatcima.

Pogreška između predviđene i stvarne vrijednosti se propagira unatrag kroz mrežu, prilagođavajući težine kako bi se minimizirala pogreška u sljedećem prolazu. Koristi se algoritam gradijentnog spusta za ažuriranje težina.

Struktura MLP mreže koju smo koristili za našu analizu sastoji se od ulaznog sloja koji prima 16 podataka koje predstavljaju naše značajke, dva skrivena sloja od 15 neurona koji koriste ReLU aktivacijsku funkciju te izlazni sloj od jednog neurona koji proizvodi predikciju buduće volatilnosti. Neuronska mreža (Slika 3.4) koju smo koristili je potpuno povezana mreža što znači da su svi neuroni iz jednog sloja povezani sa svim neuronima u sljedećem sloju.



Slika 3.4 Izgled korištene neuronske mreže

Koristi se Adam optimizator, koji kombinira prednosti AdaGrad i RMSProp optimizatora za efikasno prilagođavanje težina mreže. Funkcija gubitka je srednje kvadratna pogreška, često korištena u regresijskim problemima za kvantificiranje razlike između predviđenih i stvarnih vrijednosti. Early Stopping callback koristi se za praćenje gubitka na validacijskom skupu i zaustavljanje obuke ako se gubitak ne smanjuje nakon 10 epoha, vraćajući najbolje težine modela. Mreža se trenira kroz 50 epoha, pri čemu svaka epoha koristi grupu od 16 primjera podataka za ažuriranje gradijenata i prilagodbu težina.

MLP neuronska mreža odabrana je za ovo istraživanja zbog sposobnosti učenja složenih uzoraka, fleksibilnost i sposobnost generalizacije nad podacima.

3.4. Implementacija

Programski jezik i alati

Analiza je provedena korištenjem programskog jezika Python i nekoliko specijaliziranih biblioteka koje pružaju napredne funkcionalnosti za obradu podataka, strojno učenje i duboko učenje [7]. Python je vrlo popularan u znanstvenim krugovima i industriji zbog svojih bogatih biblioteka za znanstvene izračune i strojno učenje. Implementacija je provedena u Jupyter bilježnici, koja je popularno okruženje za interaktivno programiranje i analizu podataka. Neke od važnijih biblioteka su:

NumPy

- Rad s višedimenzionalnim nizovima i matricama
- Osnovne statističke mjere, manipulacija podacima

Pandas

- Analiza podataka u tabličnom formatu (DataFrame)
- Čitanje, čišćenje, priprema i analiza podataka

scikit-learn

- Algoritmi za strojno učenje (klasifikacija, regresija)
- Implementacija i evaluacija modela (MSE)

TensorFlow

- Izgradnja i treniranje neuronskih mreža
- Implementacija MLP modela, duboke neuronske mreže

4. Rezultati

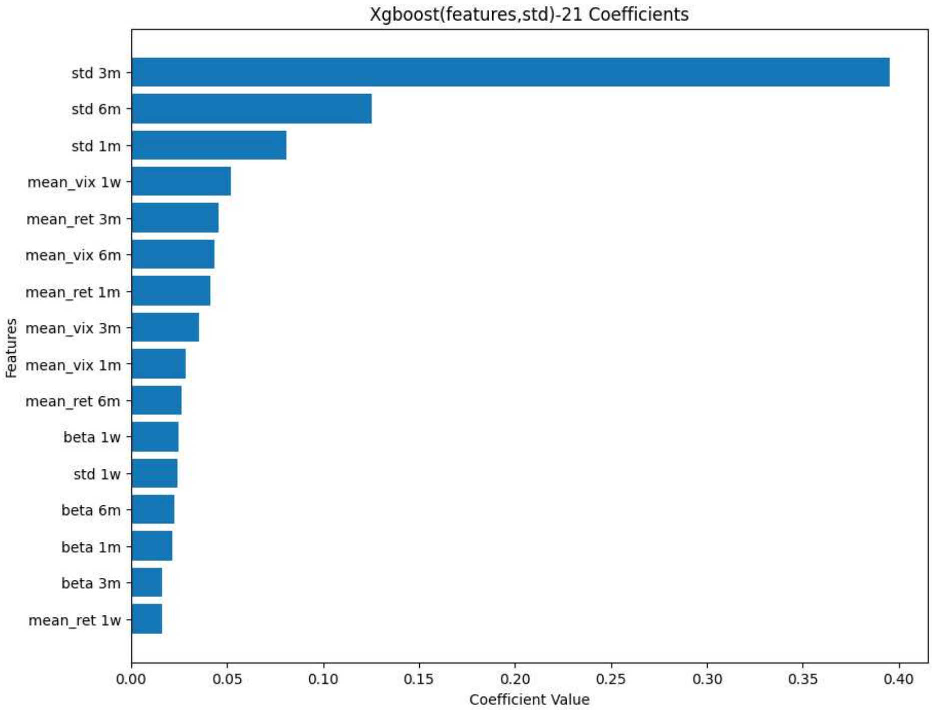
U ovom poglavlju predstavljeni su rezultati analize i predikcije volatilnosti korištenjem različitih modela strojnog učenja. Nakon predikcije volatilnosti, korištene su beta vrijednosti izračunate na prozoru od prošlih 126 dana za konstrukciju matrice korelacija. Kombinacija matrice korelacija i dijagonalne matrice volatilnosti (13) stvara matricu kovarijance koju koristimo za kreiranje optimalnih portfelja. Pozvan je *efficient frontier* algoritam kako bi se dobila dva portfelja: *CMV* (*constrained minimum variance)* portfelj gdje su težine dionica ograničene u rasponu od 0 do 1 i *GMV (global minimum variance)* portfelj, gdje težine dionica mogu ići u minus odnosno dozvoljene su kratke pozicije dionica.

Modeli su trenirani i testirani na skupu podataka koji obuhvaća 1168 dionica u razdoblju od 2000. do 2021. godine. Ukupno 75% podataka (razdoblje od 2000. do kraja 2014. godine) korišteno je za trening, dok je preostalih 25% podataka (2015. do sredine 2021. godine) korišteno za testiranje.

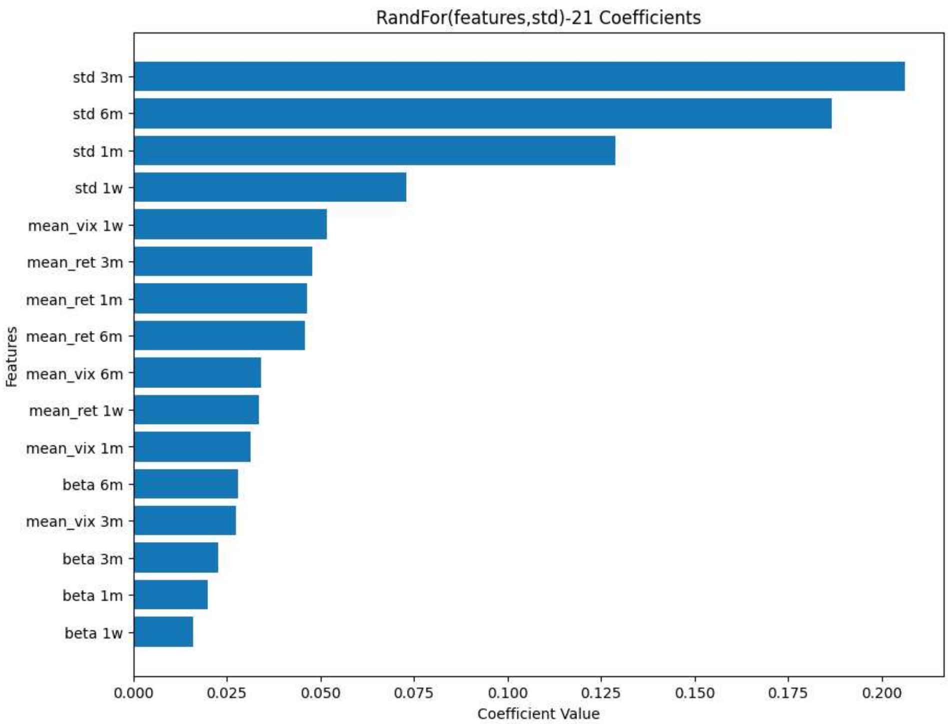
Tijekom testiranja, bilježena je godišnja volatilnost izračunata preko standardne devijacije te je računan godišnji povrat. Na kraju, izračunali smo Sharpeov omjer, u tablici prikazan slovom **S**, kako bismo procijenili odnos između povrata i rizika za svaki model i portfelj.

4.1. Važnost značajki

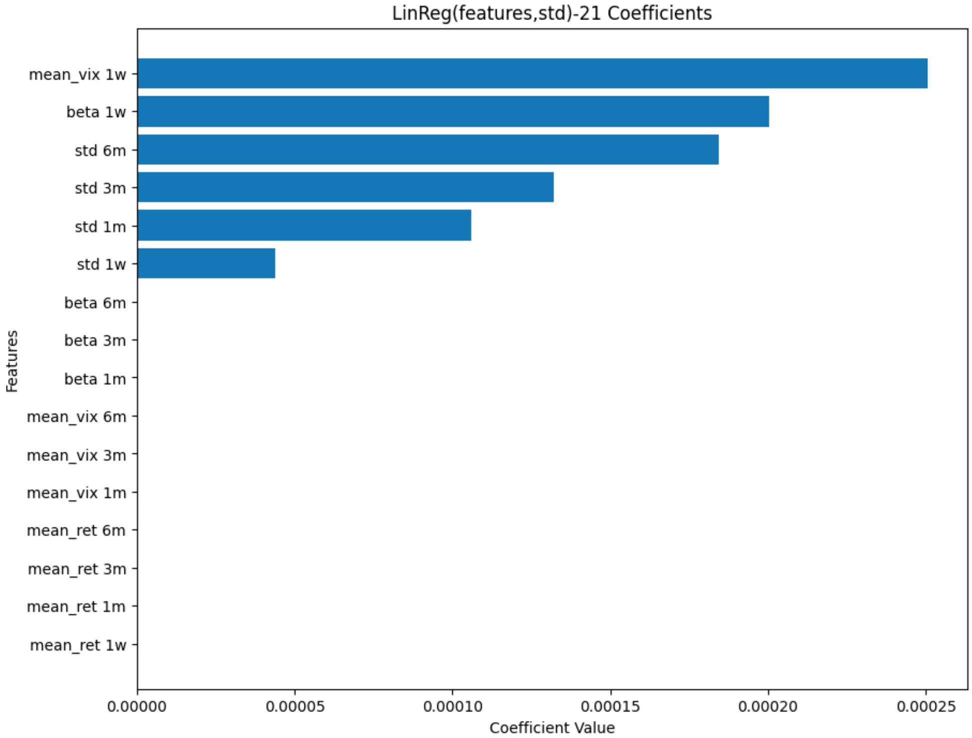
U analizama su korištene različite značajke za predikciju volatilnosti. (Slika 4.1) važnost značajki za XGBoost model, dok (Slika 4.2) i (Slika 4.3) prikazuju važnost značajki za modele slučajnih šuma i linearne regresije.



Slika 4.1 Važnost značajki XGBoost modela



Slika 4.2 Važnost značajki modela slučajnih šuma



Slika 4.3 Važnost značajki modela linearne regresije

Model XGBoost kod važnosti značajki najviše uzima u obzir standardnu devijaciju prijašnjih prozora, najviše je bitan prozor od 3 mjeseca. Prosječni povrati i VIX indeks su po sredini važnosti, beta koeficijent je najmanje bitan i nalazi se na dnu. Kod algoritma slučajnih šuma važnost značajki slična je kao kod XGBoost modela, uvjerljivo najbitnije značajke su 4 vrste standardne devijacije, VIX indeks i prosječni povrati su po sredini, a beta koeficijenti su na dnu važnosti. Linearna regresija je zanemarila sve osim 6 značajki. Nije zanemarena standardna devijacija na svim prozorima te tjedni prozori za prosječne povrate i za beta koeficijente.

4.2. Ograničeni portfelj minimalne varijance (CMV)

Ograničeni portfelj minimalne varijance (CMV) je portfelj dizajniran da minimalizira rizik pod određenim ograničenjima. Ograničenje u našem slučaju je da težine svake dionice u portfelju mogu biti između 0 i 1, pri čemu je zbroj svih težina 1. Ovo ograničenje osigurava da su sva ulaganja pozitivna, odnosno nema kratkih pozicija.

Iz (Tablica 1) vidimo da oracle referentni model koji ima uvid u budućnost, poznaje volatilnost i beta koeficijente sljedećeg mjeseca, očekivano je ostvario najmanju volatilnost i najveći povrat. Oracle procjenitelj koji je znao buduću volatilnost a beta koeficijente uzimao iz uzoračke procjene kao i ostale metode ostvario je volatilnost 9.8% što nije prevelika razlika od XGBoosta koji je ostvario volatilnost od 11.5%. Usporedbom dva oracle procjenitelja vidimo da su dobro procijenjeni beta koeficijenti od velike važnosti kad je u pitanju volatilnost portfelja. Uzorački procjenitelj koji naivno uzima prošlu volatilnost kao buduću ostvaruje skoro pa jednaku volatilnost kao i XGBoost model, XGBoost je model strojnog učenja sa najmanjom volatilnosti kad izuzmemo referentne modele.

Tablica 1. CMV portfelji korištenih modela

CMV portfelji i njihova evaluacija			
Korišteni model	**Volatilnost**	**Povrat**	**S**
Linearna regresija	11.66	14.30	1.23
Slučajna stabla	11.64	13.09	1.12
Neuronska mreža	13.44	14.89	1.11
XGBoost	11.53	13.59	1.18
Oracle (beta, vol)	6.87	26.79	3.89
Oracle (vol)	9.85	13.91	1.41
Jednake težine	16.81	20.24	1.20
Uzorački procjenitelj	11.53	14.22	1.23

4.3. Portfelj minimalne varijance bez ograničenja (GMV)

Portfelj minimalne varijance bez ograničenja također pokušava napraviti portfelj minimalnog rizika ali bez ograničenja, sada težine dionica mogu biti negativne. Negativne težine kod dionica predstavljaju obrnutu situaciju kod kupnje neke dionice, kada kupimo dionicu nadamo se da će njena vrijednost narasti i da ćemo zaraditi. Kod negativnih težina, odnosno kratkih pozicija, kladimo se da će vrijednost dionice pasti i tako ćemo ostvariti dobit.

Rezultati portfelja minimalne varijance bez ograničenja slični su portfeljima s ograničenjima (Tablica 2). Oracle procjenitelji su i dalje najbolji po volatilnosti, iako sad je razlika puno manja nego kod CMV portfelja. Uzorački procjenitelj ima manju volatilnost od svih metoda strojnog učenja koje smo testirali.

Tablica 2. GMV portfelji korištenih modela

GMV portfelji i njihova evaluacija			
Korišteni model	**Volatilnost**	**Povrat**	**S**
Linearna regresija	13.86	10.61	0.77
Slučajna stabla	13.26	11.95	0.90
Neuronska mreža	12.36	11.92	0.96
XGBoost	12.80	12.09	0.794
Oracle (beta, vol)	7.23	29.03	4.01
Oracle (vol)	8.45	20.62	2.44
Jednake težine	16.80	20.24	1.20
Uzorački procjenitelj	11.67	14.56	1.25

4.4. MSE

Kod izračuna MSE podatke smo podijelili u tri dijela:

- sve dionice : svih 1168 dionica

- trening dionice : prvih 870 dionica (75% svih dionica)
- ostale dionice : ostalih 298 dionica (25% svih dionica)

Uzeli smo tri različita skupa podataka zbog bolje evaluacije modela, svi modeli su trenirani na skupu trening dionice, testiranje i izračun prosječne godišnje volatilnosti i prosječnog godišnjeg povrata obavljeno je na skupu sve dionice. Mjeru greške kod predviđanja volatilnosti, MSE, smo računali na sva tri skupa podataka(trening dionice, sve dionice i ostale dionice) kako bi vidjeli koliko dobro modeli predviđaju volatilnost na treniranim podatcima i podatcima koje modeli nisu nikad vidjeli.

Rezultati prikazani (Tablica 3) prikazuju da slučajne šume i XGBoost model najbolje predviđaju buduće volatilnosti na sva tri skupa podataka. Linearna regresija ima nešto lošije predikcije, greška na treniranim dionicama je jednaka kao i kod slučajnih šuma i XGBoosta ali na ostalim skupovima je veća. To nam govori da se model linearne regresije malo prenaučio na trenirane podatke. Neuronska mreža i uzorački procjenitelj imaju jako slične rezultate i najveće greške kod predviđanja volatilnosti.

Tablica 3. MSE korištenih modela

MSE			
Model	**Sve dionice**	**Trening dionice**	**Ostale dionice**
Linearna regresija	0.000326	0.000143	0.000873
Slučajna stabla	0.000291	0.000145	0.000728
Neuronska mreža	0.000585	0.000364	0.001247
XGBoost	0.000292	0.000142	0.000741
Uzorački procjenitelj	0.000478	0.000197	0.001316

5. Zaključak

U ovom radu predikcija volatilnosti u dionicama predstavljena je kao problem nadziranog strojnog učenja. Analizirane su različite metode koje su predviđale standardnu devijaciju na temelju odabranih značajki. Htjeli smo dobru predikciju volatilnosti kako bi zajedno sa beta koeficijentima koji prikazuju korelacije samih dionica i sveukupnog tržišta kreirali matricu kovarijance. Beta koeficijenti su predviđani uzoračkom procjenom iz povijesnih podataka. Kada usporedimo oracle procjenitelje koji znaju buduću volatilnost, ali jedan od njih zna buduće beta koeficijente a drugi ih procjenjuje iz uzorka, vidimo da je razlika ogromna. Procjenitelj koji zna beta koeficijente ostvaruje volatilnost 6.8% i povrate od 29% dok procjenitelj koji ne zna beta koeficijente ostvaruje volatilnost 9.8% i povrate od 14%. Ako usporedimo i uzorački procjenitelj koji prema jednostavnom principu buduću volatilnost predviđa tako da kaže da je ona jednaka prošloj vidimo jako slične rezultate kao i oracle procjenitelja, koji zna točno buduću volatilnost. Uzorački procjenitelj ostvario je 11.53% volatilnost portfelja što je 1.7% više od oracle. Model strojnog učenja koji je ostvario najmanju volatilnost je XGBoost koji ima volatilnost praktički jednaku kao uzorački procjenitelj, 11,52%. Također XGBoost ostvario je najmanji MSE između predviđene i stvarne volatilnosti koji iznosi 0.000292. Uzorački procjenitelj ostvario je MSE od 0.000492, tu još jednom možemo vidjeti da XGBoost iako bolje predviđa volatilnost opet mu sveukupna volatilnost portfelja nije bolja od uzoračkog procjenitelja. Ovakav pristup predviđanja matrice kovarijance možemo unaprijediti na više načina. Možemo koristiti kompleksnije modele strojnog učenja koji će bolje predviđati volatilnost, također možemo u modele uključiti ostale značajke koje se ne računaju iz povrata. Na posljetku kao što smo vidjeli veliku razliku oracle procjenitelja koji zna i koji ne zna beta koeficijente možemo pristupiti boljoj procjeni beta koeficijenata i samim time dobiti bolje predikcije matrice kovarijance.

Literatura

[1] Juan D. Díaz, Erwin Hansen, Gabriel Cabrera, *Machine-learning stock market volatility: Predictability, drivers, and economic value*

[2] Clifford Lam, *High-dimensional covariance matrix estimation*

[3] Jianqing Fan, Yingying Fan, Jinchi Lv, *High dimensional covariance matrix estimation using a factor model*

[4] Myles E. Mangram, *A Simplified Perspective of the Markowitz Portfolio Theory*

[5] Bojana Dalbelo Bašić, Jan Šnajder, *Strojno učenje*, Sveučilište u Zagrebu Fakultet elektrotehnike i računarstva, 2020.

[6] Poon, Ser-Huang, and Clive W.J. Granger. 2003. *Forecasting Volatility in Financial Markets: A Review*

[7] JetBrains, *Python Software Foundation*, Python Developers Survey 2018,

[8] Sven Lončarić, *Neuronske mreže*, Sveučilište u Zagrebu Fakultet elektrotehnike i računarstva, 2020.

[9] Finance Train, *Constructing an Efficient Frontier,* Constructing an Efficient Frontier - Finance Train

[10] Breaking down finance, *Minimum variance portfolio*, https://breakingdownfinance.com/finance-topics/modern-portfolio-theory/minimum-variance-portfolio/

Sažetak

Ovaj rad se bavi predikcijom volatilnosti i procjenom matrice kovarijance koristeći metode strojnog učenja. U uvodu je naglašena važnost razumijevanja i predviđanja volatilnosti financijskih instrumenata za investitore i financijske analitičare. Teorijska osnova uključuje objašnjenje volatilnosti, kovarijance, korelacije, beta koeficijenata i Markowitzove teorije portfelja. Metodologija rada opisuje prikupljanje i pripremu podataka, te primjenu različitih modela strojnog učenja, uključujući linearnu regresiju, slučajne šume, XGBoost i unaprijednu neuronsku mrežu. Rezultati pokazuju da modeli strojnog učenja, posebno XGBoost, mogu efikasno predvidjeti volatilnost i doprinijeti preciznijoj procjeni matrice kovarijance. Ovi modeli su evaluirani na osnovu njihove sposobnosti minimiziranja volatilnosti i maksimiziranja povrata portfelja. Zaključeno je da napredni modeli strojnog učenja imaju potencijal značajno poboljšati upravljanje portfeljima i donošenje investicijskih odluka.

Summary

This paper deals with the prediction of volatility and the estimation of the covariance matrix using machine learning methods. The introduction highlights the importance of understanding and predicting the volatility of financial instruments for investors and financial analysts. The theoretical foundation includes explanations of volatility, covariance, correlation, beta coefficients, and Markowitz's portfolio theory. The methodology section describes data collection and preparation, and the application of various machine learning models, including linear regression, random forests, XGBoost, and a multilayer perceptron neural network. The results show that machine learning models, especially XGBoost, can effectively predict volatility and contribute to a more accurate estimation of the covariance matrix. These models are evaluated based on their ability to minimize volatility and maximize portfolio returns. The conclusion is that advanced machine learning models have the potential to significantly improve portfolio management and investment decision-making.