

# Predviđanje cijene kriptovaluta korištenjem podataka s internetskih foruma

---

**Saganić, Gabriel**

**Master's thesis / Diplomski rad**

**2021**

*Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj:* **University of Rijeka / Sveučilište u Rijeci**

*Permanent link / Trajna poveznica:* <https://um.nsk.hr/um:nbn:hr:195:547729>

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

*Download date / Datum preuzimanja:* **2024-09-13**



*Repository / Repozitorij:*

[Repository of the University of Rijeka, Faculty of Informatics and Digital Technologies - INFORI Repository](#)



Sveučilište u Rijeci – Odjel za informatiku

Sveučilišni diplomski studiji Informatika  
modul Informacijski i komunikacijski sustavi

Gabriel Saganić

# Predviđanje cijene kriptovaluta korištenjem podataka s internetskih foruma

Diplomski rad

Mentor: doc. dr. sc., Miran Pobar

Rijeka, srpanj 2021.

## SADRŽAJ

|      |  |    |
|------|--|----|
| 1.   | Zadatak diplomskog rada .....                        | 1  |
| 2.   | Uvod.....  | 2  |
| 2.1. | Definicija problema .....                            | 3  |
| 2.2. | Struktura rada .....                                 | 4  |
| 3.   | Kriptovalute.....                                    | 5  |
| 3.1. | Blockchain .....                                     | 5  |
| 3.1. | Hash funkcije i Hash .....                           | 6  |
| 3.2. | Kripto novčanici.....                                | 7  |
| 3.3. | Rudarenje .....                                      | 8  |
| 3.4. | Proof of work (PoW) algoritam .....                  | 9  |
| 3.5. | Distribuirani sustav i decentralizirani sustav ..... | 10 |
| 3.6. | Najznačajnije kriptovalute .....                     | 13 |
| 3.7. | Bitcoin (BTC).....                                   | 14 |
| 3.8. | Ethereum (ETH).....                                  | 15 |
| 3.9. | Trgovanje kriptovalutama .....                       | 15 |
| 4.   | Predviđanje cijene kriptovaluta .....                | 17 |
| 4.1. | Analiza sentimenta teksta.....                       | 17 |
| 5.   | Eksperiment .....                                    | 19 |
| 5.1. | Podaci.....  | 20 |
| 5.2. | Analiza podataka.....                                | 23 |
| 6.   | Modeli predviđanja.....                              | 29 |
| 6.1. | Vektorska autoregresija .....                        | 29 |
| 6.2. | Neuronske mreže .....                                | 30 |
| 6.3. | Algoritam K najbližih susjeda .....                  | 32 |
| 6.4. | Stablo odlučivanja .....                             | 33 |

|      |                                     |    |
|------|-------------------------------------|----|
| 7.   | Rezultati .....                     | 35 |
| 7.1. | Vektorska autoregresija .....       | 35 |
| 7.2. | Neuronske mreže .....               | 39 |
| 7.3. | Algoritam k najbližih susjeda ..... | 41 |
| 7.4. | Stablo odlučivanja .....            | 42 |
| 7.5. | Analiza rezultata.....              | 43 |
| 8.   | Zaključak.....                      | 44 |
| 9.   | Popis slika .....                   | 45 |
| 10.  | Popis tablica .....                 | 46 |
| 11.  | Popis priloga.....                  | 47 |
| 12.  | Literatura.....                     | 48 |



# 1. ZADATAK DIPLOMSKOG RADA



Rijeka, 1.6.2021.

## Zadatak za diplomski rad

Pristupnik: Gabriel Saganić

Naziv diplomskog rada: Predviđanje cijene kriptovaluta korištenjem podataka s internetskih foruma

Naziv diplomskog rada na eng. jeziku: Predicting cryptocurrency prices using internet forums data

Sadržaj zadatka: Prikupiti objave o odabranim kriptovalutama sa internetskih foruma. Izraditi skup podataka za učenje modela strojnog učenja na osnovu povijesnih podataka o cijenama kriptovalute te prethodno prikupljenih objava. Opisati korištene postupke. Naučiti, testirati i opisati modele strojnog učenja koji na osnovu podataka o objavama na internetskim forumima predviđaju kretanje cijene odabrane kriptovalute.

Mentor:  
Doc. dr. sc. Miran Pobar

Voditeljica za diplomske radove:  
Izv. prof. dr. sc. Ana Meštrović

Komentor:

Zadatak preuzet: 1.6.2021.

(potpis pristupnika)

## 2. UVOD

Kriptovalute su posljednja novost u svijetu financija. Povijesno gledajući, ljudi su se oduvijek bavili razmjenom dobara što je rezultiralo razvojem oblika plaćanja usluga. Roba za robu, plemeniti metali te u konačnici novac. Svakoj je valuti dodijeljena središnja banka i vrijednost na temelju zlata koju je ista posjedovala. Trgovanje novčanicama generalno je dobro osmišljen sustav s jednom manom – ovisi o središnjoj banci koja kontrolira transakcije i ukoliko se nađe u opasnosti, ovlaštena je printati više novčanica nego što ima zlata u banci. Izdavanjem novčanica bez pokrića vrijednost valute pada i dolazi do inflacije. Kako bi se izbjegao navedeni problem i omogućile privatne transakcije stvoreni su digitalni ekvivalenti fizičkom novcu, kriptovalute, koje nemaju pokriće u zlatu ni institucije povezane s njima već se zasnivaju na matematici i algoritmima. U prošlosti su se trgovanjem kriptovalutama bavili pojedinci, dok se danas sve veći postotak populacije upušta u rudarenje kriptovaluta koje ima mnoge sličnosti trgovanju dionicama. Kupovina i prodaja kriptovaluta slična je kupovini i prodaji dionica, a vrlo se često oba tržišta za vrijeme krize kreću u istom smjeru. Cijene dionica ovise o poslovanju tvrtke čija je dionica kupljena, dok je kod kriptovaluta volatilnost cijene veća te je moguće ostvariti puno veći dobitak ili gubitak u kratkom vremenskom periodu. Mnogi ljudi sada povezuju kriptovalute s dionicama. Dionice i kriptovalute obje su vrste ulaganja, ali postoji nekoliko ključnih razlika među njima. Prvotni cilj kreatora kriptovalute bio je da kriptovaluta jednog dana postane valuta koja je opće prihvaćena kao gotovina. Za usporedbu, tvrtke nude dionice kao udjel ili vlasništvo u toj tvrtki. Vrijednost kriptovaluta i dionica može rasti i padati, pa kupnja bilo koje od njih uključuje rizik. Međutim, kriptovaluta je stekla reputaciju naglih i drastičnih promjena u vrijednosti koje se mogu dogoditi bez upozorenja. Dionice su, s druge strane, izravno povezane s tvrtkama koje moraju javno i redovito objavljivati kako rade i kako očekuju rad u budućnosti. Kada kupujete dionice, ulažete u određenu tvrtku za koju vjerujete da će nastaviti rasti u budućnosti. S kriptovalutama možete kupiti određene valute kako biste ih koristili kao oblik plaćanja, ili ih jednostavno možete držati u nadi da će povećati vrijednost. Kriptovalute također ne trguju na tradicionalnim burzama. Za ulaganje u kriptovalutu morate koristiti „crypto exchange“, internetsku uslugu koja pruža alate za kupnju i prodaju digitalnih valuta, te valutu spremate u svoji kripto novčanik. Kripto novčanik je opširnije opisani dalje u radu.. Drugim riječima, kripto ne možete kupiti i prodati kao dionice. Federalne agencije, poput američke Komisije za vrijednosne papire i burze (SEC), imaju ovlast nad cijelim tržištem dionica da štite poštenu trgovinu. Nasuprot tome, do sada ne postoji središnje tijelo koje regulira kripto tržište. Za svaku jedinstvenu kriptovalutu upravljanje se distribuira među onima koji se bave rastom i održavanjem njegove tehnologije. Međunarodno istraživanje ING pokazalo je da bi oko 30% ljudi u dobi između 18 i 34 godine radije uložilo 1000 dolara u Bitcoin nego u državne dionice i obveznice. (Ipsos, 2018) Popularnost ulaganja privlači sve veći broj kripto znatiželjnika, no isti vrlo često nisu dovoljno upućeni u ponudu tržišta već prate trendove.

Ako cijena dionica ovisi o dobrom vođenju neke tvrtke također i cijena kriptovalute ovisi o dobroj tehnologiji koja stoji iza te kriptovalute. Sigurnost kriptovalute je jako važan čimbenik na cijenu kriptovalute. Većom sigurnosti dolazi do većeg povjerenja korisnika i samim time većom popularnosti valute. Broj čvorova dobar je pokazatelj vrijednosti kriptovalute. Broj čvorova je broja aktivnih novčanika na mreži koji se mogu pronaći na internetu ili na početnoj stranici neke kriptovalute. Ako cijena fiatovog novca <sup>1</sup>padne, tada bi cijena Bitcoina porasla u odnosu na tu valutu. To je zato što ćete moći dobiti više tog novca za istu cijenu kriptovalute. Dugoročna vrijednost digitalne valute bit će određena mnogim faktorima, poput pravog masovnog usvajanja i stvarne vrijednosti koju takve valute donose našem svakodnevnom životu, u usporedbi s drugim tradicionalnim načinima plaćanja. Drugi kritični faktor bit će hoće li vlade i središnje banke prihvatiti novu tehnologiju ili se boriti protiv nje. Na kraju, ali ne i najmanje važno, upotrebljivost i sigurnost ključna su komponenta koja se mora razvijati iznad današnjih standarda kako bi svima omogućila da svoje kriptovalute drže lako i sigurno dostupnima za uporabu u svakodnevnom životu. (antonylewis2015, 2020)

## 2.1. Definicija problema

Danas je na svijetu oko 4.5 milijardi korisnika Interneta od čega je više od 3.7 milijardi aktivnih Internet korisnika koji na dnevnoj bazi objave preko 50 milijardi GB informacija. Tako velik broj informacija utječe i na okolinu u kojoj živimo i na stvari koje radimo. Kao što više čitamo o temama s kojima smo se negdje na internetu susreli i radije ćemo kupiti predmete za koje smo već čuli, tako ćemo prije uložiti u kriptovalutu koja se često spominje na društvenim mrežama. Mogli bi pretpostaviti pojednostavljenu formulu za cijenu kriptovalute koja glasi: viša cijena valute jednako više korisnika valute jednako više objava na društvenim mrežama o toj valuti. Zanima nas da li je moguće promatranjem broja objava neke valute predvidjeti da li će cijena te valute rasti ili padati te koliko daleko u budućnost možemo predvidjeti i koliko točno? Taj problem pokušat ćemo riješiti u ovom diplomskom radu. (Smith, 2021)

Ključne riječi ( Blockchain, kriptovalute, dionice, modeli predviđanja, analiza sentimenta, vektorska autoregresija, neuronske mreže, algoritam K najbližih susjeda, stablo odlučivanja)

---

<sup>1</sup> novac je koji nema pokrića svoje vrijednosti ni u kakvom konkretnom dobru poput zlata ili srebra.



## 2.2. Struktura rada

Predmet istraživanja ovog diplomskog rada je predviđanje kretanja cijena kriptovaluta Bitcoin i Ethereum na temelju podataka preuzetih s 4Chan foruma, kategorije business & Finance. Rad se sastoji od 7 poglavlja. Prvi dio je uvod koji obuhvaća i definiciju problema te strukturu rada. U drugom je dijelu objašnjen pojam kriptovaluta te su predstavljene najpopularnije kriptovalute koje su i glavno područje proučavanja ovog rada, Bitcoin i Ethereum, opisane su tehnologije na kojima se zasniva rad kriptovaluta: blockchain, kripto novčanici, rudarenje i hash. Objašnjen je i Proof of work algoritam te način funkcioniranja distribuiranih i decentraliziranih sustava. U trećem poglavlju objasnjena je analiza sentimenta. Četvrto poglavlje govori o razgradnji projekta, prikazuje proces predviđanja za koje su objave i komentari pomoću Python skripte i BeautifulSoup biblioteke preuzeti s web mjesta, zatim su filtrirani kako bi preostali samo oni u kojima se spominju za ovaj rad ključne teme, kriptovalute. Tekstovi su semantički analizirani koristeći TextBlob biblioteku te je na temelju dobivenih rezultata kreirana tablica za analizu. Od dobivenih rezultata vizualizirani su grafovi na temelju kojih je izrađen model predviđanja cijena kriptovaluta te napravljena analiza. U petom poglavlju objašnjavamo modela predviđanja s kojima smo radili, objašnjeni su model vektorske autoregresije, model neuronske mreže, algoritam K najbližih susjeda i stablo odlučivanja. Šesti dio objašnjava proces izrade modela te analiziramo rezultate. Posljednje, sedmo poglavlje je zaključak rada.

### 3. KRIPTOVALUTE

Kriptovaluta (eng. Cryptocurrency) je digitalno sredstvo osmišljeno za rad kao sredstvo razmjene na internetu. Pojedinačni zapisi o vlasništvu kriptovaluta pohranjuju se u obliku računalne baze podataka na internetskim stranicama koje pružaju te vrste usluga. Nastaju korištenjem softvera koji rješava matematičke zadatke i koriste jaku kriptografiju kako bi se osigurala evidencije transakcija i kontroliralo stvaranje dodatnih digitalnih kovanica te provjera prijenosa vlasništva nad digitalnim kovanicama. Ne postoji u fizičkom obliku i ne izdaje ju središnje tijelo (banka). Vrijednost kriptovalute varira na temelju ponude i potražnje te nije povezana s nikakvom materijalnom stvari.

Prema Janu Lanskyu, istraživaču kriptovaluta, kriptovaluta je sustav koji udovoljava šest uvjeta (Lanasky, 2018) Sustav ne zahtijeva središnje tijelo, njegovo se stanje održava distribuiranim konsenzusom.

- Sustav zadržava pregled jedinica kriptovalute i njihovog vlasništva.
- Sustav definira mogu li se stvoriti nove jedinice kriptovaluta. Ako se mogu stvoriti nove jedinice kriptovalute, sustav definira okolnosti njihovog podrijetla i način utvrđivanja vlasništva nad tim novim kovanicama.
- Vlasništvo nad kovanicama kriptovaluta može se dokazati isključivo kriptografski.
- Sustav omogućava obavljanje transakcija u kojima se mijenja vlasništvo nad kriptografskim kovanicama. Izvod transakcija može izdati samo subjekt koji dokazuje trenutno vlasništvo nad tim kovanicama.
- Ako se istovremeno unose dvije različite upute za promjenu vlasništva istih kriptografskih jedinica, sustav izvodi najviše jedan od njih. (Lanasky, 2018)

Kako bi razumjeli način funkcioniranja kriptovaluta, prvotno se treba objasniti tehnologija koja stoji iza nje. Rad kriptovaluta zasniva se na blockchain tehnologiji čiju sigurnost osigurava kriptografija (kriptografski potpis), proof of work i distribuirani sustavi. U ovom je poglavlju objašnjen i pojam rudarenja te što su to novčanici za kriptovalute..

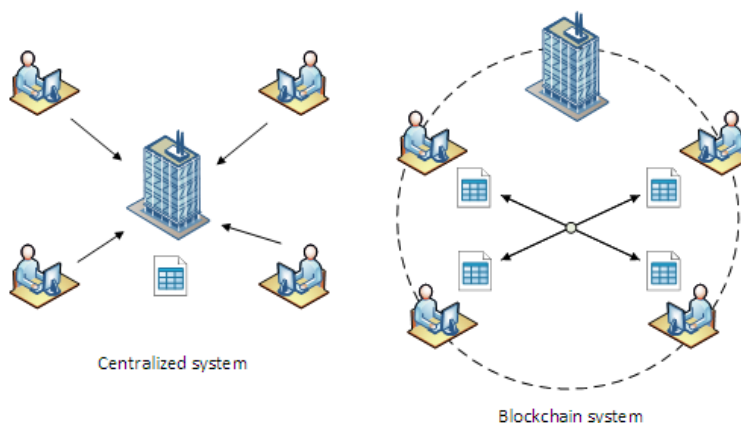
#### 3.1. Blockchain

Blockchain koncept teoretski je zamišljen još 90-ih godina ali za praktičnu primjenu čekalo se sljedećih 20 godina. U najosnovnijem slučaju, Blockchain je složena baza podataka u čijoj su glavnoj knjizi zapisane sve verificirane transakcije. Lanac blokova je jednostavni opis potvrđenih blokova gdje je svaki blok povezan sa svojim prethodnikom sve do prvoga bloka u lancu koji se koristi za inicijalizaciju podataka. Blockchain je često vizualiziran kao vertikalni stog. Blokovi su slojevito poredani jedan na drugi dok prvi blok služi kao temelj tog stoga. Prilikom ovakve vizualizacije blokova poredanih jedan na drugog pojavljuju nam se pojmovi kao što je “visina” koja se odnosi na udaljenost od prvoga bloka i pojam “vrh” koji se

odnosi na zadnji dodani blok na stog. U blockchain mreži pohranjen je veliki broj zapisa. Održavanje ovih zapisa pomaže korisnicima pratiti sve informacije od početka zadanog blockchaina. Blockchain se sastoji od tri glavna dijela:

- Blok - lista sastavljena od raznih podataka
- Lanac - hash koji povezuje jedan blok s drugim
- Mreža - mreža se sastoji od čvorova

Svaki čvor sadrži cjeloviti zapis o svim podacima koji su ikad bili u tom blockchainu. Čvorovi se mogu nalaziti diljem svijeta i mogu biti upravljani od bilo koga.



Slika 1: centralizirani i blockchain sistem

Koncept blockchaina temelj je postojanja kriptovaluta. Izbacivanjem centralne institucije koja bilježi transakcije (banke), a korištenjem blockchain sustava riješen je problem stvaranja distribuirane baze podataka neovisne o trećoj, centraliziranoj strani kojoj korisnici trebaju vjerovati. Sastavni dio te sigurnosti su privatni i javni ključevi.

### 3.1. Hash funkcije i Hash

Hash funkcija je temeljni dio Blockchain tehnologije i da bi razumjeli Blockchain prvo moramo dobro poznavati hashiranje. Hash funkcija je bilo koja funkcija koja može „mapirati“ podatak proizvoljne veličine u podatak fiksne veličine. Preciznije rečeno, hash funkcija uzima poruku fiksne duljine i za izlaz daje šifriranu poruku. Hash funkcije možemo podijeliti na dvije vrste: hash funkcije bez ključa, koje za ulaz imaju samo poruku i hash funkcije s ključem koje uz poruku za ulaz imaju i privatni ključ. Kod hash funkcije s ključem uz poruku se šalje i MAC oznaka koju dobivamo upotrebom MAC algoritma i privatnog ključa. Kada primatelj, koji ima isti privatni ključ, primi poruku može MAC algoritmom i privatnim ključem dobiti MAC oznaku i tako potvrditi autentičnost poruke. Vrijednost dobivena hash funkcijom

nazivamo hash. Hash je najčešće string nasumičnih slova i brojeva. Najčešće korištena hash funkcija je SHA-256 (Bashir, 2017) koju koristi Bitcoin. Na primjer: za ulaz „Blockchain“ izlaz SHA-256 funkcije je

*625da44e4eaf58d61cf048d168aa6f5e492dea166d8bb54ec06c30de07db57e1,*

a za ulaz „Blockchain2“ dobijemo izlaz

*4f07e031df167abda5623314c19c38e11358853f1c876b892a1d2ae494cc1058.*

Iz ovog primjera vidimo da je gotovo nemoguće otkriti ulaz promatranjem samo hash-a, te da samo najmanja promjena u ulazu kompletno mijenja izlaz. Hash funkcije su jednosmjerne funkcije, nemaju inverznu funkciju. Jedini način da se dobiju ulazni podaci hash funkcije iz izlaza je bruteforce algoritmom (pokušavanje svih mogućih kombinacija), ali za takvo nešto bi trebala sva računala na svijetu da rade od početka svijeta do danas. Bitno svojstvo hash funkcije je kompresija, veliki ulaz je komprimiran u vrlo kratki niz tog unosa. U blockchain tehnologiji je bitno da sto više podataka ulazi u hash funkciju da je teže ili gotovo nemoguće otkriti ulaz. Iz bitcoin bloka vrlo je jednostavno proizvesti unikatni hash ali je, gledajući samo stvoren hash, gotovo nemoguće otkriti koji su to podaci. Ukoliko se promjeni samo jedna brojka ili slovo, hash se u potpunosti mijenja. Kako bi hashevi bili legitimni svaki od njih pohranjuje podatke o hashu zadnjeg bloka pohranjenog u lancu blokova čim se kreira digitalni pečat. Ukoliko bi netko pokušao krivotvoriti transakciju mijenjajući postojeći blok, vrlo bi jednostavno bilo dokazati da je blok krivotvoren pokretanjem hash funkcije koja bi se razlikovala od one pohranjene kraj njega u lancu. Mijenjanje jednog bloka izazvalo bi lančanu reakciju i promjenu svih slijedećih blokova do kraja lanca. Kod bitcoina, svaki put kad je hash uspješno kreiran, rudari kao poticaj nastavka rudarenja dobivaju 25 bitcoinova. (Crobotcoin, 2020)

Kako bi spriječili kreiranje hasheva od podataka, bitcoin protokol otežava situaciju korištenjem dokaza o radu (eng. Proof of work).

### 3.2. Kripto novčanici

Kripto novčanici čuvaju privatne ključeve - lozinke koje omogućuju pristup kriptovalutama. Za razliku od klasičnog novčanika s gotovinom, kripto novčanici ne pohranjuju kriptovalute već ključeve potrebne za pristup blockchainovima na kojima su pohranjene.

Proces kriptiranja:

Novčanik generira privatni ključ → privatni ključ generira javni ključ → javni ključ generira hash javnog ključa → hash javnog ključa generira adresu → adresa se koristi za prihvaćanje Bitcoina

U kriptografiji kriptografski algoritam uvijek ima dva ili više ključeva povezanih s njim. To su privatni i javni ključ. Privatni ključ tajni je ključ koji se koristi za šifriranje i dešifriranje poruka i koristi se s javnim ključem. Treba ga držati privatnim u svakom trenutku i ne ga dijeliti s drugima. Gubitkom privatnog ključa onemogućava se pristup novčaniku i sredstvima. S druge strane, javni ključ koristi se samo za šifriranje poruka i može se slati drugima na šifriranje. Nakon što pošiljatelj šifrira poruku javnim ključem on se može otključati samo privatnim ključem. Javni ključ i adresa mogu se generirati iz privatnog ključa, međutim obrnuta funkcija nije moguća. To znači da ako imate nečiju adresu ili čak javni ključ, ne možete generirati privatni ključ. Zatim javni ključ prolazi kroz jednosmjernu hash funkciju koja zauzvrat generira javnu adresu. Javna adresa vidljiva je svima i može se podijeliti sa svima radi slanja kovanica ili sredstava.

Novčanik je softver koji pohranjuje privatne ključeve, javne ključeve i javne adrese, omogućuje slanje i primanje kovanica, a također djeluje kao osobna knjiga stanja i transakcija. Laički rečeno, novčanik poput mrežne platforme bankovnog računa, adresa je poput broja računa, blockchain je poput glavne knjige banke, a kod skrbničkih novčanika skrbnik ima funkciju bankara dok je rudar korisnik banke.

### 3.3. Rudarenje

Za razliku od tradicionalnih monetarnih sistema u kojima središnje banke izdaju novac kad im je to potrebno, kod kriptovaluta novac se ne izdaje već „otkriva“ postupkom rudarenja. Proces rudarenja podrazumijeva traženje broja pomoću kojeg će rudar, zajedno s ostalim podacima iz bloka, izračunati hash koji zadovoljava određene kriterije. Kriteriji se mijenjaju s obzirom na vrijeme koje je bilo potrebno da se riješi block kako bi se postigao jednak razmak između dodavanja novih blokova. Dobiveno rješenje zovemo Proof of Work. Rješenje potvrđuje da je rudar potrošio mnogo vremena i resursa na pronalaženje broja. U svakom se trenutku odvijaju transakcije i da nitko ne vodi evidenciju o njima ne bi se znalo koliko tko novaca ima. Mreže to sprječavaju stavljanjem obavljenih transakcija u lance blokova, blockchain, a rudari ih potvrđuju i zapisuju u glavnu knjigu za što dobivaju isplatu u kriptovalutama.

Dugi lanac blokova čini glavnu knjigu koja se koristi za provjeru transakcija. Nakon izvršenja transakcije, novi transakcijski blok dodaje se u lanac blokova što stvara dugu listu svih napravljenih transakcija i svi sudionici dobivaju novu kopiju glavne knjige kako bi znali što se događa. Kako bi osigurali glavnu knjigu, odmah nakon stvaranja transakcijskog bloka rudari ga stavljaju u proces obrade. Informacije pohranjene u bloku korištenjem hash funkcije rudari pretvaraju u naizgled nasumični niz brojeva i slova nazvan hash. Hash se pohranjuje skupa s blokom na kraju lanca blokova.

### 3.4. Proof of work (PoW) algoritam

Dokaz rada (proof of work, PoW) je konsenzusni algoritam u kojem je stvaranje i prikupljanje podataka skupo i dugotrajno, ali drugima je lako provjeriti jesu li podaci točni. Najpopularnija kriptovaluta Bitcoin koristi sustav dokaza o radu Hashcash. Iako je početna ideja Hashcash-a bila borba protiv neželjene pošte, Satoshi je tu ideju primijenio na bitcoin transakcije.

Cynthia Dward i Moni Naor objavile su 1993. godine znanstveni rad *Pricing via Processing or Combatting Junk Mail* u kojem su pokušale predstaviti novu metodu borbe protiv neželjene e-pošte koja se šalje pomoću računala za svaku e-poštu (ili transakciju). Navode: "Ako vas ne poznajem i želite mi poslati poruku, morate dokazati da ste potrošili, recimo, deset sekundi CPU vremena, samo za mene i samo za ovu poruku." Objasnile su da je lako i jeftino slati neželjenu poštu i predložile rješenje prema kojem se od nepoznatog pošiljatelja koji nam po prvi put šalje poruku zahtjeva da uloži 10 sekundi procesorskog vremena računala prije slanja poruke pri čemu slanje postaje skuplje. U radu je funkcija koju je teško izračunati, a lako provjeriti nazvana funkcijom određivanja cijena.

Kod transakcija, da bi mreža prihvatila blok, rudari trebaju popuniti dokaz o radu kako bi se sve transakcije u bloku provjerile. Rudari se međusobno natječu u rješavanju funkcije sažimanja za čije je rješavanje zbog kompleksnosti potrebno puno računalne snage.

Kao što je već objašnjeno, svaki blok u blockchainu ima svoj hash (id). Kako bi provjerili blok, uzimamo hash njegovog prethodnika i dodajemo trenutni broj transakcija. Slijedeći korak je dodavanje slučajnog broja na kraj bloka teksta. Za početak izračuna, upotrebljava se hash funkcija i mijenja slučajni broj dok se ne dobije niz koji ispred sebe ima određeni broj nula za što računalo mora izvršiti  $10n21$  kalkulacija kako bi pronašlo pravi broj te je za navedeni izračun potrebno oko 10 minuta.

Ukratko, dokaz rada je sustav koji osigurava sigurnost i konsenzus u cijeloj blockchain mreži. Očito je da je sudionik, koji validira blok, uložio značajnu računalnu snagu da to učini. Za rad PoW algoritma potrebna je velika količina električne energije i računarske snage. Osim gubitka resursa, problem su i centralizacija i rudarski bazeni koji su protivni ideji decentralizacije, no nerealno je za očekivati velike rudarske bazene u državama u kojima je to skupo. Više od 70% rudarske snage dolazi iz Kine u kojoj su troškovi električne energije povoljniji nego u drugim zemljama.

### 3.5. Distribuirani sustav i decentralizirani sustav

Za razmjenu kriptovaluta mogu se koristiti dvije vrste blockchaina: centraliziran i decentraliziran. Centralizirane kriptovalute su u nečijem vlasništvu i nadzire ih središnji entitet. S obzirom da platformu kontrolira središnji entitet korisnici nemaju potpunu kontrolu nad svojim sredstvima i dužni su plaćati naknade za korištenje te kriptovalute. Osim neanonimnosti i naknada postoji i veći rizik od hakerskog napada zbog centralizirane ulazne točke.

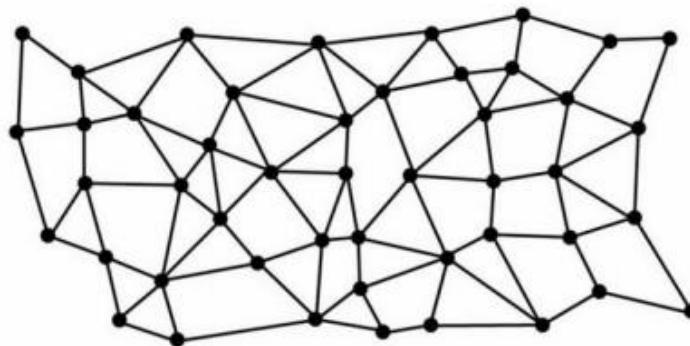
Decentralizirana razmjena kriptovaluta (dEX) je platforma za trgovanje kriptovalutama neovisna o trećoj strani. Rješavanjem propusta centralizirane razmjene poput ograničenja kontrole, sigurnosti, brzine i isplativosti korisnici dEX-a imaju potpunu kontrolu nad svojim sredstvima i anonimnošću. Naknade su svedene na najmanju moguću mjeru ili su u potpunosti besplatne, a rizici zastoja izbjegavaju se decentraliziranom razmjenom što onemogućava zlonamjerne napade. Za razliku od centralizirane mreže gdje transakcije među korisnicima mogu izvršavati samo vjerodostojni i ugledni sudionici čije se transakcije mogu revidirati, u decentraliziranoj razmjeni kriptovaluta svatko može sudjelovati i izvršavati transakcije na glavnoj knjizi.

Blockchain može biti centraliziran ili decentraliziran no to se ne smije mijenjati s distribuiranim. Iako je blockchain inherentno distribuiran (što znači da mnoge stranke drže kopije glavne knjige), on u osnovi nije decentraliziran već se to odnosi na prava sudionika na glavnoj knjizi.

Tablica 1 Usporedba centraliziranih i decentraliziranih kriptovaluta

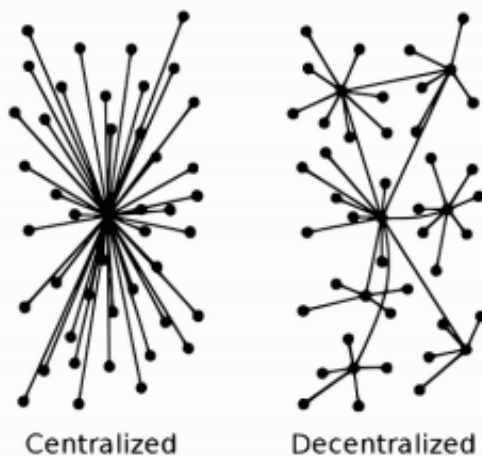
|                            | <b>Centralizirana razmjena kriptovaluta</b>          | <b>Decentralizirana razmjena kriptovaluta</b>  |
|----------------------------|--|--|
| <b>Kontrola</b>            | kontrolira središnji poslužitelj                     | Pojednostavljeno peer-to-peer decentraliziranim protokolom razmjene                    |
| <b>Upotreba</b>            | Jednostavno i za početnike                           | Nije prilagođeno korisnicima, previše tehnički čak i za srednje ili napredne korisnike |
| <b>Značajke</b>            | Nudi napredne značajke trgovanja                     | Nudi osnovne značajke trgovanja  |
| <b>Naknade</b>             | Potrebne, veće naknade za transakcije                | Gotovo nepostojeće   |
| <b>Identitet korisnika</b> | Nije anonimno  | Anonimno   |
| <b>Brzina</b>              | Sporija, rizik zastoja poslužitelja                  | Brže, manji rizik zastoja poslužitelja   |
| <b>Sigurnost</b>           | Jedna ulazna točka, osjetljiva na zlonamjerne napade | Više ulaznih točki, manja šansa zlonamjernog napada                                    |

Distribuirani sustav je mreža koja se sastoji od autonomnih čvorova koja su spojena pomoću distribucijske mreže (Bashir, 2017). Pomažu u dijeljenju različitih resursa i mogućnosti kako bi korisnicima pružili jedinstvenu mrežu. U distribuiranim sustavima nema središnjeg autoriteta. Svaki čvor je povezan sa svakim drugim čvorom i ima isti autoritet.



Slika 2: Distribuirani sustav (Izvor: Usman M. Sheikh, *Virtual currency: current regulatory and civil litigation trends*, 2018)

Distribuirani sustavi sastoje se od velikog broja istovrsnih procesa, takozvanih čvorova, Slika 2. Čvorovi obavljaju zadane zadaće prema potrebama svojih korisnika. Kada je čvoru pri obavljanju neke zadaće potrebna pomoć on stupa u komunikaciju sa susjednim čvorom, a ti susjedi sa svojim susjedima i tako se komunikacija odvija na razini cijelog sustava.



Slika 3: Centraliziran i decentraliziran sustav (Izvor : Stewart Martins, *Understanding Blockchain: A Simple Approach*)



Distribuirani sustavi mogu se podijeliti prema strukturi na centralizirane i decentralizirane, Slika 3. Kod centraliziranih sustava postoji poslužitelj koji osigurava zadovoljavajuću razinu performansi i kvalitete, kod decentraliziranih sustava poslužitelja nema, te sami korisnici i dobro promišljeni sustavi obavljaju poslove poslužitelja. U slučaju digitalnog novca, poslužitelj se brine za sigurnost valute kako bi se onemogućilo i spriječilo krivotvorenje, osigurava primjerenu kvalitetu i sigurnosnu razinu tehničkog rješenja i brine se za njegovo održavanje te povezuje klijente kako bi mogli međusobno komunicirati. Izdavatelj valute, prije pojave digitalnih valuta baziranih na Blockchain tehnologiji, bio je određeni fizički subjekt koji je jamčio autentičnost i vrijednost valute. Nastankom Blockchaina odnosno Bitcoina, sustavi počinju koristiti decentralizirane distribuirane sustave. Time je omogućena razmjenu podataka kroz računalnu mrežu pri kojem čvorovi preuzimaju informacije jedni od drugih umjesto s jednog centralnog poslužitelja. Kad se dizajnira decentralizirani sustav mora se uzeti u obzir probleme koje rješava poslužitelj u centraliziranim sustavima. To su najčešće krivotvorini podaci i sigurnost sustava. Sigurnost se postiže konsenzusom između čvorova i CAP teoremom koji navodi da je nemoguće da distribuirana pohrana podataka istovremeno pruža više od dva od sljedeća tri jamstva: dosljednost, dostupnost i toleranciju particije. Dosljednost znači da je svaki odgovor poslan klijentu točan, svojstvo dostupnosti označava da svaki zahtjev koji je funkcionirajući server zaprimio mora rezultirati odgovorom, a tolerancija particije osigurava rad sustava i u uvjetima kada nastanu izolirane skupine računala. Što je veći broj čvorova potvrdilo transakciju, manji je rizik od krivotvorenih napada. Upotrebom POW (Proof of Work) mehanizma, čvorovi imaju veću motivaciju za održavanje sustava i snimanje transakcija.

### 3.6. Najznačajnije kriptovalute

Prema popisu koji datira iz siječnja 2021. na svijetu postoji više od 5000 kriptovaluta. Iako mnoge od njih imaju mali broj rudara i pratitelja, neke uživaju ogromnu popularnost među posvećenim zajednicama podržavatelja i investitora. U pozadini većine kriptovaluta stoje privatne tvrtke s vizijom koje stvaranjem kriptovaluta pronalaze investitore. Valute po uzoru na Bitcoin zajednički se nazivaju digitalnim novcem ili novčićima (altcoins), a u nekim slučajevima i "shitcoins", a često su se pokušavale predstaviti kao modificirane ili poboljšane verzije Bitcoina. Ukupna vrijednost svih kriptovaluta 27. svibnja 2021. bila je više od 1,7 bilijuna dolara. Dvije valute kojima se, prema tržišnoj kapitalizaciji, trenutno najviše trguje su Bitcoin (\$735.3 milijarda) i Ethereum (\$324.2 milijarda) te će fokus ovog rada biti usmjeren na predviđanje kretanja njihovih vrijednosti. (Coin Market, 2020)

Preostalih 8 najpopularnijih kriptovaluta (Tether, Binance Coin, Cardano, Dogecoin, XRP, USD Coin, Polkadot, Uniswap) i njihova vrijednost prikazane su na Slici 1.

| #  | Name                              | Price       | 24h %   | 7d %    | Market Cap        | Volume(24h)                             | Circulating Supply   | Last 7 Days |
|----|-----------------------------------|-------------|---------|---------|-------------------|---|----------------------|-------------|
| 1  | Bitcoin BTC <span>Buy</span>      | \$36,275.69 | -6.12%  | -11.80% | \$679,975,567,122 | \$37,330,799,218<br>1,029,086 BTC       | 18,744,662 BTC       |             |
| 2  | Ethereum ETH <span>Buy</span>     | \$2,226.60  | +6.86%  | -17.11% | \$259,377,278,098 | \$25,569,406,016<br>11,483,622 ETH      | 116,490,413 ETH      |             |
| 3  | Tether USDT <span>Buy</span>      | \$1.00      | -0.06%  | -0.07%  | \$62,492,174,268  | \$64,246,787,314<br>64,208,082,915 USDT | 62,454,526,907 USDT  |             |
| 4  | Binance Coin BNB <span>Buy</span> | \$307.28    | -4.71%  | +15.85% | \$47,146,720,986  | \$1,966,490,117<br>6,399,687 BNB        | 153,432,897 BNB      |             |
| 5  | Cardano ADA                       | \$1.39      | +5.03%  | +15.44% | \$44,271,119,389  | \$2,711,964,896<br>1,956,971,543 ADA    | 31,946,328,269 ADA   |             |
| 6  | Dogecoin DOGE <span>Buy</span>    | \$0.2672    | +4.65%  | -37.83% | \$34,793,298,118  | \$2,204,073,724<br>8,249,662,794 DOGE   | 130,228,392,047 DOGE |             |
| 7  | XRP XRP                           | \$0.7118    | -10.36% | -23.76% | \$32,848,223,564  | \$3,277,812,019<br>4,604,844,271 XRP    | 46,146,927,647 XRP   |             |
| 8  | USD Coin USDC                     | \$1.00      | -0.02%  | -0.09%  | \$25,402,765,341  | \$2,343,580,291<br>2,343,350,251 USDC   | 25,400,271,866 USDC  |             |
| 9  | Polkadot DOT <span>Buy</span>     | \$16.78     | -10.04% | -9.22%  | \$16,046,345,862  | \$1,331,865,533<br>79,351,076 DOT       | 956,023,546 DOT      |             |
| 10 | Uniswap UNI                       | \$18.83     | -7.11%  | -13.47% | \$10,830,308,476  | \$348,088,187<br>18,488,133 UNI         | 575,234,077 UNI      |             |

Slika 4: 10 najzastupljenijih kriptovaluta (Izvor: Coin Market)

### 3.7. Bitcoin (BTC)

Bitcoin je prva uspješna i globalno najpopularnija kriptovaluta izgrađena na decentraliziran način. Sustav elektroničkog plaćanja osmišljen od strane anonimnog pojedinca ili skupine programera koji su pod pseudonimom Satoshi Nakamoto objavili znanstveni rad naziva *Bitcoin: A Peer-to-peer Electronic Cash System* 2008. godine u kojem su izloženi konceptualni i tehnički detalji platnog sustava koji pojedincu omogućuje slanje i primanje novca bez uključivanja posredničkih financijskih institucija. Predložili su rješenje koje rješava problem duple potrošnje korištenjem peer-to-peer (svaki sa svakim) mreže. Koncept se temelji na povezivanju računala bez centralnog servera već svako računalo izravno komunicira s preostalim računalima. Mreža vremenski označuje (timestamp) transakcije hashiranjem u kontinuirani lanac dokaza o radu temeljen na hashu. Na taj se način stvara zapis kojeg je nemoguće izmijeniti bez ponovnog obavljanja dokaza o radu. Najdulji lanac ne služi samo kao dokaz o slijedu događanja, već je rezultat najvećeg skupa procesorske snage (CPU). Procesorskom snagom upravljaju računala koja su u mreži prikazana kao čvorovi i sve dok ona ne planiraju u suradnji napasti mrežu, nastaviti će rezultirati najduljim lancem i nadjačati napadače. Struktura same mreže vrlo je jednostavna, čvorovi koji su njen dio mogu napuštati i ponovo se pridruživati mreži pri čemu se prihvaća najdulji lanac dokaza o radu kao dokaz onoga što se dogodilo dok ih nije bilo. Ne moraju se identificirati jer se poruke ne preusmjeravaju na neko određeno mjesto nego samo trebaju biti prenesene u najboljem nastojanju. Bitcoin ima ograničenje da obrađuje velike količine transakcijama na svojoj platformi u kratkom vremenskom razdoblju. Taj problem zovemo problem nedovoljne skalabilnosti Bitcoina (Satoshi, 2009)

2009. godine Bitcoin je predstavljen kao opensource program.. Diljem svijeta ljudi koriste program temeljen na matematičkoj formuli za proizvodnju Bitcoina. Formula je javno dostupna svima koji se žele uvjeriti u njenu djelotvornost. Formulu popularno zovemo rudarenje (mining).

Bitcoinovi su rudareni (mined) pomoću računala, tj. korištenjem procesorskog vremena i distribuirani u bitcoin mreži. Mreža također procesira transakcije napravljene pomoću bitcoina, što rezultira stvaranjem vlastite platne mreže. Bitcoin protokol, algoritam na kojem se sustav temelji, limitiran je na izdavanje maksimalno 21 milijuna bitcoinova, no svaki bitcoin ima puno veću vrijednost nego tradicionalna valuta zbog čega je podijeljen na manje dijelove od kojih je najmanji, milijunti dio "Satoshi", nazvan po izumitelju bitcoina. (Crobotcoin, 2020)

### 3.8. Ethereum (ETH)

Od veljače 2018. godine na drugom je mjestu prema tržišnom udjelu, vrijednosti neke valute u Dolarima, najpopularnija kriptovaluta Ether. Ethereum je programibilna blockchain platforma otvorenog koda s funkcijama pametnih ugovora dizajnirana kako bi se riješio problem nedovoljne skalabilnosti Bitcoina. Pruža decentralizirani virtualni stroj EVM, *Ethereum Virtual Machine*, putem svoje namjenske kripto valute Ether.

Koncept Ethereuma kreirao je krajem 2013-te devetnaestogodišnji programer Vitalik Buterin. Razočaran činjenicom da sve izmjene koje programeri izvršavaju kako bi poboljšali Bitcoin ne rješavaju njegove osnovne nedostatke, odlučio je započeti iznova i pokrenuo crowdfunding za razvoj Ethereuma koji je 2015. godine izbačen na tržište. Ethereum pomaže u jednostavnom stvaranju pametnih ugovora, programskih kodova upućenih mreži koji bez utjecaja treće strane izvršava transakcije. Može se shvatiti kao otvorena računalna platforma temeljena na blockchain tehnologiji koja ne samo da može voditi račune, već omogućuje izgradnju i implementaciju aplikacija.

Distribuirana računalna platforma pokazuje da je Ethereum, poput Bitcoina, decentraliziran. Čvorovi distribuirani po cijelom svijetu zajednički vode račune i zajednički vode knjigu. Knjiga je otvorena i transparentna i na nju treće strane ne mogu utjecati. Za razliku od Bitcoina, Ethereum se bazira na Turingovom cjelovitom skriptnom jeziku što znači da se na Ethereumu mogu razvijati mali programi. Ether valutom plaća se izvršavanje transakcija na Ethereum Blockchainu.

Ethereum je uspostavio novu vrstu tehnologije šifriranja, a poteškoće u razvoju programa mnogo su jednostavnije od Bitcoina. Ovo je otkriće uvelike smanjilo razvojne troškove i vrijeme razvoja za programere koji primjenjuju blockchain tehnologiju. Pojava Ethereuma još je jednom ponovila nužnost decentralizacije, te izvedivost i prednosti decentraliziranih distribuiranih aplikacija, dajući postojeće ekonomsko tržište, novi smjer razvoja u financijskom sektoru i novi smjer za društvo, te potpuno nove poduzetničke ideje i mogućnosti.

### 3.9. Trgovanje kriptovalutama

Burza kriptovaluta je sustav koji omogućuje korisnicima da trguju kriptovalutama za drugu imovinu, poput fiat novca ili drugih kriptovaluta. Burza kriptovaluta mogu prihvaćati plaćanja kreditnim karticama, elektronička plaćanja ili druge oblike plaćanja u zamjenu za kriptovalute. Burza kriptovaluta je zapravo

market mejker<sup>2</sup> koji obično uzima razliku u kupovnoj i prodajnoj cijeni kao proviziju od transakcije za uslugu ili kao odgovarajuća platforma jednostavno naplaćuje naknade. Na primjer 100 dolara može se zamijeniti za bitcoin ekvivalentne vrijednosti i obrnuto. Slično tako, bitcoin vrijedan 100 dolara može se zamijeniti za Ethereum ekvivalentne vrijednosti. Isti se koncept može primijeniti na različitu imovinu na temelju ponude burze.

Centralizirane burze kriptovaluta djeluju kao treća strana između kupca i prodavatelja. Budući da njima upravlja i kontrolira tvrtka, centralizirane razmjene nude veću pouzdanost. Približno 99% svih kripto transakcija prolazi kroz centralizirane razmjene.

Decentralizirane burze kriptovaluta omogućuju korisnicima izvršavanje „peer-to-peer“ transakcija bez potrebe za trećom stranom ili posrednikom. Međutim, decentralizirane burze ne dopuštaju trgovanje fiat valuta s kriptovalutama.

---

<sup>2</sup> Investicijska tvrtka ili broker koji održava prodajne i kupovne cijene na određenoj vrijednosnici ili financijskom instrumentu tako što neprekidno spremno čeka da kupi ili proda tu istu vrijednosnicu po njezinoj javno iskazanoj cijeni.

## 4. PREDVIĐANJE CIJENE KRIPTOVALUTA

U ovom će se radu za predviđanje cijene kriptovaluta koristiti analiza upisa na društvenim mrežama, odnosno analiza sentimenta za određivanje prevladavajuće pozitivnog ili negativnog stava prema nekoj valuti, te niz algoritma za treniranje modela. Modeli koje ćemo koristiti za treniranje su vektorska autoregresija, neuronske mreže, algoritam k najbližih susjeda i stablo odlučivanja. Najprije se prikupljaju izjave s 4Chana kategorije business & Finance. Nakon toga im se određuje polaritet, tj. imaju li pozitivan ili negativan stav, a zatim različitim algoritmima učimo model te na temelju toga dobivamo rezultate za narednih par dana. U konačnici analiziramo rezultate i promatramo je li veća vjerojatnost da će cijena kriptovalute pasti ili narasti.

### 4.1. Analiza sentimenta teksta

Analiza sentimenta korisna je za brzo sticanje uvida pomoću velikih količina tekstualnih podataka. Jedan od primjera su tvrtke za trgovanje dionicama koje pretražuju internet radi vijesti. Algoritmi sentimenta na temelju dobivenih informacija mogu otkriti određene tvrtke koje pokreću pozitivne osjećaje u vijestima što predstavlja značajnu financijsku priliku jer to može pokrenuti ljude da kupe veći dio dionica te tvrtke. Pristup ovoj vrsti podataka daje trgovcima mogućnost donošenja odluka prije nego što preostalo tržište reagira.

Model koji analizira objave na društvenim mrežama ima za cilj procijeniti raspoloženje investitora analizom spominjanja na društvenim mrežama. Osjećaj autora trebalo bi tumačiti na način da ukoliko je osjećaj pozitivan i cijena kriptovalute ne raste, to je medvjedi signal, a ako je osjećaj negativan i cijena kriptovalute ne opada, to je bikovski signal.

Moglo bi se primjerice naučiti model da pozitivni članci o kineskim ulaganjima u kriptovalute mogu imati pozitivan utjecaj na tržištu koje je bilo relativno medvjede posljednjih tjedan dana. Osnovno načelo modela analize utjecaja na tržište sentimenta bilo bi kontekstualizirati znanje o modelima sentimenta prema specifičnostima i dinamici kripto tržišta.

Za ljude je razumijevanje teksta jednostavno, prepoznajemo pojedine riječi i kontekst u kojem se koriste. Ako pročitate ovu rečenicu:

“Vaša korisnička usluga je šala! Čekam već 30 minuta i ništa se ne dešava!”

Razumijete da je kupac frustriran jer agentu za korisničke usluge predugo treba odgovor.

Međutim, strojeve prvo treba osposobiti da razumiju ljudski jezik i razumiju kontekst u kojem se riječi koriste jer bi u suprotnom riječ “šala” mogli bi pogrešno protumačiti kao pozitivnu. Pokrenuti algoritmima strojnog učenja, sustavi semantičke analize mogu razumjeti kontekst prirodnog jezika, otkriti emocije i sarkazam te izvući vrijedne informacije iz nestrukturiranih podataka, postižući točnost na ljudskoj razini. Da bi bili u mogućnosti točnije predvidjeti kretanje kriptovalute poželjno je semantički analizirati podatke te odrediti je li objava ili neki komentar bio pozitivan ili negativan. Tako imamo više parametra koji utječu na kretanje cijene valute s čime možemo preciznije odrediti putanju cijene valute.

U radu je za analizu sentimenta korištena Pythonova TextBlob biblioteka za obradu prirodnog jezika (NLP). TextBlob aktivno koristi Natural Language Toolkit (NLTK). NLTK je knjižnica koja omogućuje jednostavan pristup velikom broju leksičkih izvora i omogućuje korisnicima rad s kategorizacijom, klasifikacijom i mnogim drugim zadacima. TextBlob je jednostavna knjižnica koja podržava složenu analizu i operacije s tekstualnim podacima. Unaprijed definiranim rječnikom pojmovi se klasificiraju na negativne i pozitivne riječi.

TextBlob vraća polaritet i subjektivnost rečenice. Polaritet se kreće između  $[-1,1]$ ,  $-1$  definira negativni osjećaj, a  $1$  definira pozitivan osjećaj. Riječi negacije preokreću polaritet. TextBlob ima semantičke oznake koje pomažu u sitnoznoj analizi poput emotikona ili uskličnika. Subjektivnost je između  $[0,1]$ . Subjektivnost kvantificira količinu osobnog mišljenja i činjeničnih podataka sadržanih u tekstu. Viša subjektivnost znači da tekst sadrži osobno mišljenje, a ne činjenične podatke. Osim navedenih, TextBlob ima još jedan parametar - intenzitet. TextBlob izračunava subjektivnost promatrajući 'intenzitet'. Intenzitet određuje mijenja li riječ sljedeća riječ. Našem radu koristiti ćemo samo polaritet rečenice.

## 5. EKSPERIMENT

U radu se pokušava predvidjeti kako će se kretati cijene kriptovaluta Bitcoin i Ethereum pomoću podatka prikupljenih s 4Chan foruma business & Finance poddirektorija. 4chan je jednostavna oglasna ploča temeljena na slikama na kojoj svatko može anonimno objavljivati komentare i dijeliti slike. Za prikupljanje podataka korišten je upravo 4Chan jer je s njega najjednostavnije može preuzeti podatke te nema ograničenja za preuzimanje poput Facebooka-a, Twitter-a ili Reddita-a. Za izradu predviđanja korišten je programski jezik Python te njegove biblioteke, BeautifulSoup, Pandas, Matplotlib, Statsmodels, Sklearn, Multiprocessing. Slijed zadataka koji smo izvršili da bi dobili rezultat predviđanja je sljedeći:

1. Preuzimanje objava i komentara od 4Chan/Business & Finance pomoću Python skripte I BeautifulSoup biblioteke
2. Filtriranje dobivenih podataka gdje se spominju određene kriptovalute
3. Analizirati sentiment tekstove objava
4. Izrada tablice i priprema podataka za analizu i testiranje
5. Izrada grafova
6. Izrada modela predviđanja rasta ili pada cijene određene kriptovalute
7. Analiza dobivenih rezultata



## 5.1. Podaci

Kako bi mogli napraviti dobar predikcijski model potrebno je mnogo čistih podataka spremnih za testiranje. Pomoću Python biblioteke BeautifulSoup koja olakšava struganje podataka s web stranica s 4Chana su preuzeti i u tekstualnu datoteku 4chanAllThreadLink.csv spremljeni svi linkovi objava poddirektorija Business & Finance.

```
try:
    tryAgain = 0
    while tryAgain < 20:
        threads = []
        allLink = []
        #print("Getting all threads from page: " + str(i))

        req = requests.get(URL, headers = headers)
        soup = BeautifulSoup (req.content, features="lxml")

        article = soup.find_all("span", class_="post_controls")

        for _a in article:
            threads.append(_a.find("a", class_="btnr parent"))

        for thread in threads:
            if thread.text == 'View':
                link = thread.get('href')
                allLink.append(link)
        if not allLink:
            tryAgain = tryAgain + 1
            print("Trying again getting links No. " + str(tryAgain))
        else:
            tryAgain = 100
            q.put(allLink)
except Exception as e:
    logFile = open('4chanLogFile.txt', 'a')
    logFile.write(str(datetime.now().strftime("%H:%M:%S")) + ' - ERROR: ' + str(e) + '\n' + '\tLink: ' + str(URL) + '\n')
    logFile.close()
```

Slika 5: Kod za preuzimanje svih linkova objava

Na Slici 5. prikazana je funkcija `getAllLink()` koja pokušava preuzeti link objave dvadeset puta. Ukoliko ne uspije ni nakon dvadesetog pokušaja objavu sprema u log file kako bi na kraju mogli probati opet preuzeti link objava koje nisu uspješno preuzete radi neke greške.

```
logFile = open('4Chan/4chanLogFile.txt', 'a')
logFile.write(str(datetime.now().strftime("%H:%M:%S")) + ' - Started getting links from page ' + str(i) + '\n')
logFile.close()

while lastPage == False:
    for j in range(CPU):
        processList.append(Process(target=getAllLink, args=(URL + "/" + str(i + j), i+j,headers)))

    for p in processList:
        p.start()

    for p in processList:
        p.join()

    if(q.empty() == False):
        for k in range(q.qsize()):
            pageThread = pageThread + q.get()

    if not pageThread:
        lastPage = True
    else:
        df_all_thread = pd.DataFrame({'Link':pageThread , 'Page': i + CPU - 1})
        df_all_thread.to_csv('4Chan/4chanAllThreadLink.csv',header=None, mode='a', index=False)
        i = i + CPU
        # if(i % 100) == 0:
        #     print("Complete 100 page. Now on page: " + str(i))
        pageThread.clear()
        processList.clear()
```

Slika 6: Paralelno pozivanje `getAllLink()` funkcije

Na Slici 6. prikazan je poziv funkcije `getAllLink()` koja je prethodno objašnjena. Funkcije se izvršavaju paralelno čime se dobiva puno veća brzina. Ovisno o broju jezgri, u našem slučaju 4, toliko puta možemo istovremeno pozvati funkciju. Funkciji proslijedujemo link Business & Finance poddirektorija i povećavamo stranice poddirektorija kako program preuzima linkove objava. Kada program stigne do zadnje stranice u poddirektoriju preuzete linkove sprema u datoteku `4chanAllThreadLink.csv` i time završava preuzimanje linkova objava.

Zatim skripta prolazi kroz sve linkove objava te skida i sprema u tekstualnu datoteku `4chanAllPosts.csv` sve objave i komentare. To radi funkcija `getMessages()` koja je prikazana na Slici 7.

```
def getMessages(thread,headers):
    try:
        tryAgain = 0
        while tryAgain < 20:
            #print("Gettin posts from thread: " + str(thread))
            posts = pd.DataFrame(columns=['Date', 'PostMessage', 'Link'])

            postMessage = []
            postTime = []

            req = requests.get(thread,headers = headers)
            soup = BeautifulSoup (req.content, features="lxml")

            text = soup.find_all("div", class_="text")
            time = soup.find_all("span", class_="time_wrap")

            for _time in time:
                postTime.append(_time.find("time", recursive=False).text)
            for _t in text:
                postMessage.append(_t.text)

            for i in range(len(postMessage)):
                posts = posts.append({'Date': postTime[i], 'PostMessage':postMessage[i], 'Link': thread}, ignore_index=True)

            if posts.empty:
                tryAgain = tryAgain + 1
                print("Trying again get posts from thread No. " + str(tryAgain))
            else:
                posts.to_csv('4chanAllPosts.csv', header=None, mode='a', index=False)
                tryAgain = 100
    except Exception as e:
        print("greska")
        logFile = open('4chanLogFile.txt', 'a')
```

Slika 7: Preuzimanje objava, vrijeme objave i linka objave

Vidimo da skripta pomoću BeautifulSoup-a traži određene klase na stranici te preuzima tekst objave, sve komentare, vrijeme objave i link objave za provjeru točnosti.

| Date                     | PostMessage  | Link                           |
|--------------------------|--|--------------------------------|
| Sun 27 Dec 2020 13:00:03 | How long until crypto gains mass adoption on everyday transactions in the markets? I cannot wait to open my store that runs exclusi  | https://archived.moe/biz/three |
| Sun 27 Dec 2020 14:05:09 | >>25208449jesus retard, fees will be so high people can't buy nothing with it. Only to store as (speculated) value                   | https://archived.moe/biz/three |
| Sun 27 Dec 2020 14:07:37 | >>25208449doubt it will ever happen. the guy with the big stick makes the rules. the guy with the big stick decides what is money. i | https://archived.moe/biz/three |
| Sun 27 Dec 2020 11:43:48 |  | https://archived.moe/biz/three |
| Sun 27 Dec 2020 11:44:35 | >>25207018get up bobo  | https://archived.moe/biz/three |
| Sun 27 Dec 2020 11:44:36 | >>25207018Oh shit were only +12% today on btc, it's over   | https://archived.moe/biz/three |
| Sun 27 Dec 2020 12:39:32 | >>25207018I DIDNT HEAR NO BELL   | https://archived.moe/biz/three |
| Sun 27 Dec 2020 12:58:59 | >>25207018please boboturn NU back into a leveraged BTC short like the good old timesdon't you remember all the fun we had?           | https://archived.moe/biz/three |
| Sun 27 Dec 2020 14:14:09 | >>25207018get fucked loser, missed out on too many gains to be credible  | https://archived.moe/biz/three |

Slika 8: Primjer preuzetih objava u csv datoteci

Dobivena je baza podataka od oko 25 GB svih objava te ćemo dalje te podatke transformirati za učenje modela.

Kako bi na model brže učio prvo trebamo smanjiti podatke tako što ćemo Linux komandom izvući samo objave koje spominju tražene kriptovalute – Bitcoin i Ethereum.

```
grep -E 'bitcoin|BTC' TestData.csv > TestBitcoinData.csv; sed -i '1 i\Date,PostMessage,Link' TestBitcoinData.csv
```

Ovom su linijom grep komandom u zasebnu datoteku TestBitcoinData.csv izvučene sve linije u kojima se spominje riječ bitcoin ili BTC te dodani nazivi stupaca Date, PostMessage i Link. Kako je zadatak predvidjeti da li će cijena rasti ili padati, a ne i točna cijena kriptovalute, u našu tablicu ćemo dodati još indikator rasta ili pada. To je urađeno na sljedeći način.

```
_tmpVar = 0
_indicaotor = 0
_first = True
for index, row in DataFrame.iterrows():
    if(_first):
        _indicaotor=1
        _tmpVar = row['BitcoinPrice']
        _first = False

    elif (_tmpVar <= row['BitcoinPrice']):
        _indicaotor=1
        _tmpVar = row['BitcoinPrice']

    elif (_tmpVar > row['BitcoinPrice']):
        _indicaotor=0
        _tmpVar = row['BitcoinPrice']

    DataFrame.at[index, 'Indicaotor'] = _indicaotor
```

Slika 9: Dodavanje indikator rasta ili pada

Kod je vrlo jednostavan - ukoliko je cijena viša od prethodne cijene upisuje se 1, a ako je cijena niža od prethodne cijene upisuje se 0. Ovime su dobiveni podatci spremni za učenje modela.

## 5.2. Analiza podataka

Nakon uređivanja podataka za učenje sljedeći je korak analiza podataka. Prvo je za analizu izrađen graf koji prikazuje cijenu neke valute i broj spominjanje te valute ili broj objava u kojima se spominje ta valuta. Podatke treba sortirati prema datumu od većeg prema manjem jer su preuzimanjem sortirani obrnuto. Također je potrebno maknuti sate i minute iz stupca „Date“ i ostaviti samo datum. To radimo sa skriptom

*python OrderByDate BitcoinData.csv*

kojoj prosljeđujemo podatke u csv obliku koje želimo testirati.

```
1 import pandas as pd
2 import sys
3
4 DataFrame = pd.read_csv(sys.argv[1])
5
6 DataFrame['Date'] = pd.to_datetime(DataFrame['Date']).dt.date
7
8 DataFrame = DataFrame.sort_values(by=['Date'], ascending = False)
9
10 print(DataFrame.tail(10))
11
12 DataFrame.to_csv(sys.argv[1], mode='w', index=False)
```

Slika 10: Sortiranje po datumu

Kod učitava podatke, pretvara ih u *Date* format te ih zatim sortira po datumu od većeg prema manjem i sprema u istu datoteku. Zatim je potrebno pokrenuti skriptu koja će analizom sentimenta odrediti je li objava bila negativno ili pozitivno nastrojena.

```
from textblob import TextBlob
import pandas as pd
import sys

DataFrame = pd.read_csv(sys.argv[1])
DataFrame['Polarity'] = ''
#DataFrame['Subjectivity'] = ''
print('START')
for index, row in DataFrame.iterrows():
    blob = TextBlob(row['PostMessage'])
    row['Polarity'] = blob.polarity
    #row['Subjectivity'] = blob.subjectivity

DataFrame.to_csv(sys.argv[1], mode='w', index=False)
```

Slika 11: Kod za dodavanje polariteta

Biblioteka TextBlob nam je odradila sav posao, a mi smo trebali samo pozvati funkciju i spremiti novu tablicu u datoteku. Za dodavanje polarnosti teksta skriptu pozivamo komandom

```
python SentimentalAnalysis.py BitcionData.csv
```

Kreiranoj skripti Plot.py šalju se podaci u csv obliku i inicijali valute. Kako bi dobili graf za Bitcoin podatke skriptu trebamo pozvati na sljedeći način:

```
python Plot.py -r B BitcoinData.csv
```

gdje `-r` označava broj redova tj. broj objava po danu, B znači da uspoređujemo cijenu Bitcoina, a BitcionData.csv je datoteka u kojoj se nalazi broj objava.

Na temelju ove skripte moguće je uspoređivati samo valute Bitcoin i Ethereum, dok za uspoređivanje drugih valuta treba preuzeti cijene valuta i dodati u kod liniju koja ih čita.

```
if 'B' in cryptoCurrency:
    BitcoinPrice = pd.read_csv('Price/BitcoinPrice.csv', usecols=['Date', 'Closing Price (USD)'])
    BitcoinPrice.columns = ['Date', 'BitcoinPrice']

if 'E' in cryptoCurrency:
    EthereumPrice = pd.read_csv('Price/EthereumPrice.csv', usecols=['Date', 'Close'])
    EthereumPrice.columns = ['Date', 'EthereumPrice']
```

Slika 12: Čitanje cijene valute u tablice

Za izradu grafova korištena je biblioteka matplotlib.

Skripta može grupirati po broju traženih riječi ili po broju objava na dan. Ovisno o objavi grupiramo je po „dobrim“ ili „lošim“ objavama što je prikazano na slici 13.

```
for i in range(len(DataList)):
    #for _data in DataList:
        if 'B' == cryptoCurrency[i]:
            subData = DataList[i].loc[DataList[i]['Polarity'] >= 0.00]
            DataList[i]['Good Message'] = subData.PostMessage.str.count("BTC|bitcoin")
            subData = DataList[i].loc[DataList[i]['Polarity'] < 0.00]
            DataList[i]['Bad Message'] = subData.PostMessage.str.count("BTC|bitcoin")

        if 'E' in cryptoCurrency[i]:
            subData = DataList[i].loc[DataList[i]['Polarity'] >= 0.00]
            DataList[i]['Good Message'] = subData.PostMessage.str.count("ETH|ethereum|etherum")
            subData = DataList[i].loc[DataList[i]['Polarity'] < 0.00]
            DataList[i]['Bad Message'] = subData.PostMessage.str.count("ETH|ethereum|etherum")

    DataList[i] = DataList[i].drop(columns=['PostMessage', 'Polarity'])
```

Slika 13: Grupiranje po broju traženih riječi

```

#----- po retcima
if(mode == '-r' or mode == '-a'):
    i=0
    for _data in DataListRow:
        subData = DataListRow[i]['Polarity'] >= 0.00
        print(subData)
        DataListRow[i]['Good Message'] = subData.count()
        subData = DataListRow[i].loc[DataListRow[i]['Polarity'] < 0.00]
        DataListRow[i]['Bad Message'] = subData.count()

        DataListRow[i] = __data.groupby(['Date'],as_index=False).count()
    i=i+1

```

Slika 14: Grupiranje po broju objava

Prvo je potrebno zbrojiti objave i grupirati ih po datumu što je napravljeno pomoću Pandas funkcije *groupby()* kojoj se šalje ime stupca koji želimo zbrojiti te matematička operacija koju želimo izvršiti na tom stupcu. Nakon toga potrebno je imenovati novi stupac i dodati naziv valute na početak imena novog stupca.

```

if(mode == '-r' or mode == '-a'):
    for _data in DataListRow:
        AllData = AllData.join(_data.set_index('Date'), on='Date')
    startDate = DataListRow[0].iloc[0]['Date']
    endDate = DataListRow[0].iloc[-1]['Date']

AllData = AllData.fillna(0)

if(mode == '-a'):
    columnsName = columnsName + columnsNameRow
elif(mode == '-r'):
    columnsName = columnsNameRow.copy()

if 'B' in cryptoCurrency:
    columnsName.append('BitcoinPrice')
    BitcoinPrice = BitcoinPrice[BitcoinPrice.Date.between(startDate, endDate)]
    AllData = AllData.join(BitcoinPrice.set_index('Date'), on='Date')

if 'E' in cryptoCurrency:
    columnsName.append('EthereumPrice')
    EthereumPrice = EthereumPrice[EthereumPrice.Date.between(startDate, endDate)]
    AllData = AllData.join(EthereumPrice.set_index('Date'), on='Date')

```

Slika 15: Spajanje cijene i grupiranih podataka po datumu

U kodu prikazanom na slici 15. uzimaju se samo redci cijena za koje imamo podatke. Prvo se uzima prvi i zadnji datum za koje imamo podatke te zatim spajamo cijenu u rasponu između tih dva datuma.

```

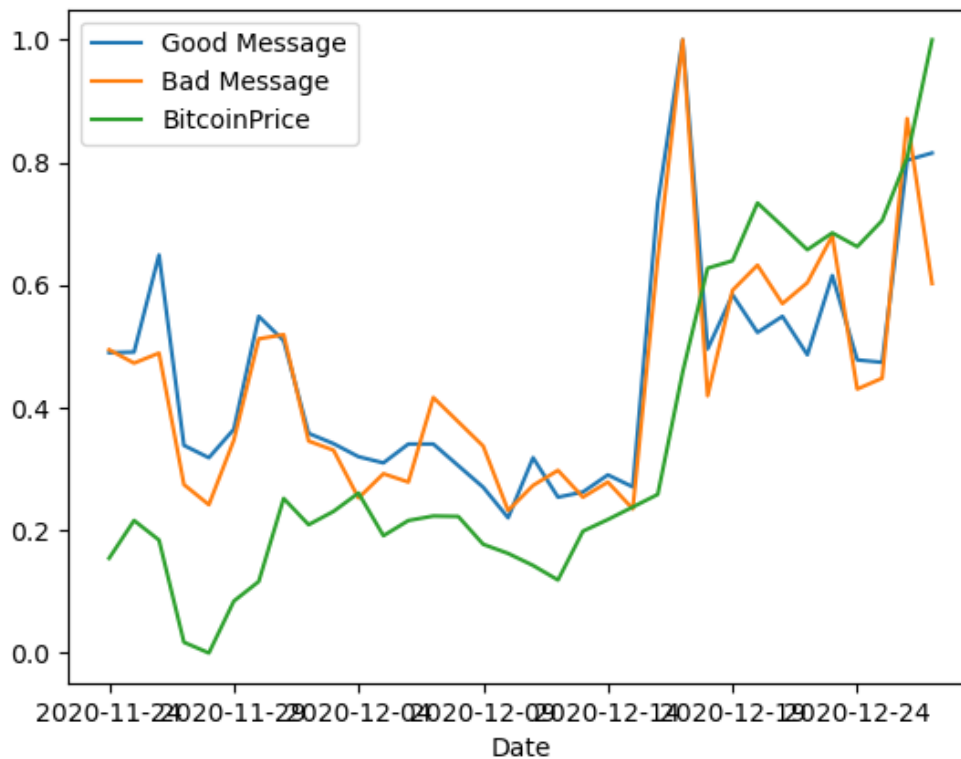
for column in columnsName:
    AllData[str(column)] = (AllData[str(column)] - AllData[str(column)].min()) / (AllData[str(column)].max() - AllData[str(column)].min())

AllData.to_csv('Result/ResultData.csv', mode='w', index=False)
AllData.plot(x='Date', y=columnsName)
plt.savefig('Result/ResultPlot.png')

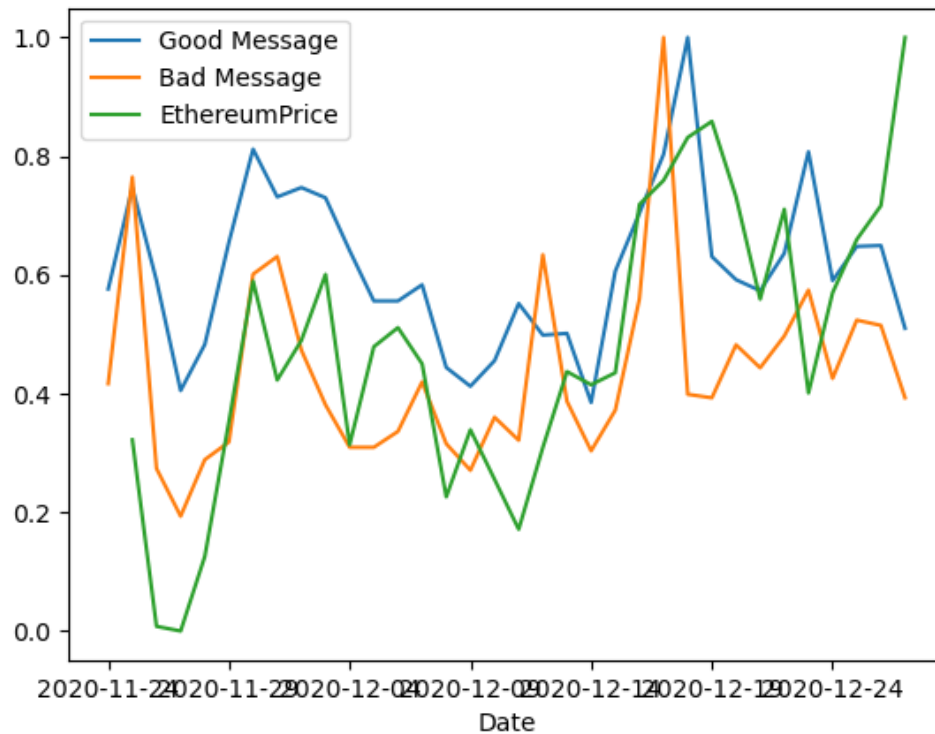
```

Slika 16: Normalizacija podataka

Slika 16. prikazuje normalizaciju svih podataka, te spremanje u datoteku i generiranje grafa.



Slika 17: Prikaz grafa Bitcoin cijene i broj njegovih objava



Slika 18: Prikaz Ethereum cijene i broj njegovih objava

Dobivamo sljedeće grafove sa Slike 17 i Slike 18. Zelenom bojom je prikazan graf koji pokazuje cijenu kriptovalute u nekom vremenskom razdoblju, a grafovi obojani plavom i narančastom bojom su prikazuju broj „dobrih“ odnosno „loših“ objava po danu. Možemo uočiti da broj objava ovisi o cijeni valute neovisno o „dobrim“ ili „lošim“ objavama. Uočavamo da se većinom broj objava reflektira na istom danu, odnosno ako je cijena niža biti će i manji broj objava te ako je cijena valute viša biti će veći broj objava. To nam nije dobro za predviđanje cijene valute unaprijed ali zamjećujemo da se nakon najvećeg broja objava cijena valute naglo povećala.



|    | A          | B                 | C                  | D                  | E          |
|----|------------|-------------------|--------------------|--------------------|------------|
| 1  | Date       | Good Message      | Bad Message        | EthereumPrice      | Indicaotor |
| 2  | 2020-11-22 | 0,514880952380952 | 0,0029761904761904 | 0,190576378440826  | 1          |
| 3  | 2020-11-23 | 0,510057471264368 | 0,419642857142857  | 0,42723710116614   | 1          |
| 4  | 2020-11-24 | 0,576149425287356 | 0,416666666666667  | 0,405836455786808  | 0          |
| 5  | 2020-11-25 | 0,747126436781609 | 0,764880952380952  | 0,249845124760816  | 0          |
| 6  | 2020-11-26 | 0,59051724137931  | 0,273809523809524  | 0,0061409663356323 | 0          |
| 7  | 2020-11-27 | 0,405172413793103 | 0,193452380952381  | 0,0061409663356323 | 1          |
| 8  | 2020-11-28 | 0,482758620689655 | 0,288690476190476  | 0,0973965751057711 | 1          |
| 9  | 2020-11-29 | 0,655172413793103 | 0,318452380952381  | 0,273665343773901  | 1          |
| 10 | 2020-11-30 | 0,811781609195402 | 0,601190476190476  | 0,457243579308509  | 1          |
| 11 | 2020-12-01 | 0,73132183908046  | 0,630952380952381  | 0,327990867567469  | 0          |
| 12 | 2020-12-02 | 0,747126436781609 | 0,473214285714286  | 0,379789882019201  | 1          |
| 13 | 2020-12-03 | 0,729885057471264 | 0,380952380952381  | 0,466009099347139  | 1          |
| 14 | 2020-12-04 | 0,640804597701149 | 0,309523809523809  | 0,243586616979459  | 0          |
| 15 | 2020-12-05 | 0,556034482758621 | 0,309523809523809  | 0,371537805560907  | 1          |
| 16 | 2020-12-06 | 0,556034482758621 | 0,336309523809524  | 0,396495285618104  | 1          |
| 17 | 2020-12-07 | 0,583333333333333 | 0,419642857142857  | 0,349217495243599  | 0          |
| 18 | 2020-12-08 | 0,443965517241379 | 0,31547619047619   | 0,175356553110091  | 0          |
| 19 | 2020-12-09 | 0,41235632183908  | 0,270833333333333  | 0,262961294045786  | 1          |
| 20 | 2020-12-10 | 0,455459770114942 | 0,360119047619048  | 0,198140425283316  | 0          |
| 21 | 2020-12-11 | 0,551724137931034 | 0,321428571428571  | 0,132941136474585  | 0          |
| 22 | 2020-12-12 | 0,498563218390805 | 0,633928571428571  | 0,239890742865976  | 1          |
| 23 | 2020-12-13 | 0,501436781609195 | 0,386904761904762  | 0,338977293886014  | 1          |

Slika 19: Završna tablica

Na slici 19. prikazana je završna tablica i sada su nam podaci spremni za učenje. Tablica sadrži normalizirani broj dobrih poruka, normalizirani broj negativnih poruka, cijenu kriptovalute i indikator rasta. Sljedeći nam je korak napraviti model za predviđanje cijene.

## 6. MODELI PREDVIĐANJA

Da bi nam predviđanje bilo što točnije, moramo probati predvidjeti s više različitih modela i uspoređivati rezultate te odabrati jedan ili više modela koje bi koristili za najtočnije predviđanje. Koristili smo učenje pomoću vektorske autoregresije, neuronskih mreža, algoritam K najbližih susjeda i stablo odluke. Za izradu naših modela koristili smo se sa statsmodels i scikit-learn bibliotekama.

### 6.1. Vektorska autoregresija

Vektorska autoregresija (VAR) statistički je model koji se koristi za praćenje odnosa između više veličina koje se vremenom mijenjaju. Takvi modeli često se koriste u ekonomiji i prirodnim znanostima. VAR model opisuje razvoj skupa od  $k$  varijabli tijekom vremena. Svako vremensko razdoblje je numerirano,  $t = 1, \dots, T$ . Varijable se prikupljaju u vektor,  $y_t$  koji je duljine  $k$ . Vektor je modeliran kao linearna funkcija njegove prethodne vrijednosti. Komponente vektora nazivaju se  $y_{t,i}$  što označava promatranje u vremenu  $t$  i-te varijable. Na primjer, ako prva varijabla u modelu mjeri cijenu kriptovalute tijekom vremena, tada bi  $y_{1,2020}$  označio cijenu kriptovalute u 2020. godini.

VAR modele karakterizira njihov redosljed koji se odnosi na broj ranijih vremenskih razdoblja koje će model koristiti. Nastavljajući gornji primjer, VAR 5. reda mogao bi modelirati cijenu kriptovalute svakog dana kao linearnu kombinaciju posljednjih pet dana cijene te kriptovalute. Zaostatak (eng. lag) je vrijednost varijable u prethodnom vremenskom razdoblju. Dakle, općenito se VAR  $p$ -tog reda odnosi na VAR model koji uključuje zaostajanja za posljednja  $p$  vremenska razdoblja. VAR  $p$ -tog reda označava se “VAR  $p$ ”, a ponekad se naziva i “VAR s  $p$  zaostacima”. Model VAR  $p$ -tog reda zapisan je kao

$$y_t = c + A_1 y_{t-1} + A_2 y_{t-2} + \dots + A_p y_{t-p} + e_t$$

Varijable oblika  $y_{t-i}$  ukazuju na vrijednost te varijable u vremenskim razdobljima i nazivaju se „ $i$ ti zaostatak“ od  $y_t$ . Varijabla  $c$  je  $k$ -vektor konstanti koje služe kao presjek modela.  $A_i$  je vremenski nepromjenjiva ( $k \times k$ ) -matrica i  $e_t$  je  $k$ -vektor pogreške.

Za stvaranje modela korištena je Python biblioteka statsmodels te vektorska autoregresija (eng. Vector Autoregression). Prvo je bilo potrebno učitati podatke i podijeliti ih na skup za učenje i skup za testiranje. Podaci su podijeljeni u omjeru 80% za učenje te preostalih 20% na testiranje.

```

trainData = learningData[:int(0.80*(len(learningData)))]
testData = learningData[int(0.80*(len(learningData))):]

trainData = trainData.drop(columns = ['Date'])
dateList = testData['Date'].tolist()

testData = testData.drop(columns = ['Date'])

```

Slika 20: Podjela na skup za treniranje i testiranje

Nakon podjele na skupove za treniranje i testiranje, treniramo model pomoću `VAR()` funkcije. Podaci za ulaz su cijena valute, datum, polarnost i broj objava. Podaci za treniranje uzeti su od prve objave koju smo preuzeli s foruma pa sve do zadnjih 8 dana. Podaci za testiranje su samo zadnjih 8 dana te smo to koristili i kao rezultate predviđanja. Model nam je najtočniji kad koristimo model s 2 „laga“ ili VAR drugog reda, što znači da modeliramo cijenu kriptovalute svaki dan kao linearnu kombinaciju posljednjih 2 dana cijene kriptovalute. Probali smo testirati do VAR-a petog reda, te smo zaključili da najveću točnost dobivamo s VAR-om drugog reda.

```

model = VAR(endog=trainData)
model_fit = model.fit(2)
prediction = model_fit.forecast(model_fit.y, steps=len(testData))

pred = pd.DataFrame(index=range(0,len(prediction)),columns=[cols + 'Pred'])

print(model_fit.summary())

```

Slika 21: Poziv `VAR()` funkcije

## 6.2. Neuronske mreže

Neuronske mreže se sastoje od slojeva koji se dijele na ulazni sloj, jedan ili više skrivenih slojeva i izlazni sloj. Slojevi sadrže čvorove. Svaki čvor ili umjetni neuron povezuje se sa svim čvorovima iz susjednog sloja i ima povezanu težinu i prag. Ako je izlaz bilo kojeg čvora iznad navedenog praga, aktivira se taj čvor i šalje podatke na sljedeći sloj mreže. U suprotnom se podaci ne prenose na sljedeći sloj mreže. Za izradu modela koristili smo model višeslojni perceptron (MLP) za klasifikaciju i regresiju kriptovaluta. Klasifikaciju smo koristili kad smo predviđali indikator rasta ili pada, a regresiju za predviđanje cijena valuta.

Svi modeli osim vektorske autoregresije kodirani su na dosta sličan način te smo samo mijenjali funkciju za treniranje modela te ćemo zato objasniti samo programski kod za izradu MLP modela. Koristili smo zadane parametre što znači da nam model ima 2 skrivena sloja i svaki sloj sadrži 100 neurona.

Aktivacijska funkcija je ispravljena linearna jedinična funkcija (eng. rectified linear unit function) koja glasi  $f(x) = \max(0, x)$ . Koristimo samo jedan „lag“, odnosno trenutna cijena kriptovalute ovisi o jučerašnjoj cijeni kriptovalute. Za optimizaciju težine koristi se „adam“ optimizator, stohastički optimizator temeljen na gradijentu. Za treniranje uzimamo sve dane osim zadnjih 6 dana koje koristimo za testiranje.

```
def Classifier(learningData):  
  
    y = learningData['Indicaotor'].to_numpy()  
  
    learningData.drop('Date',axis='columns', inplace=True)  
    learningData.drop('Indicaotor',axis='columns', inplace=True)  
    X = learningData.to_numpy()  
  
    X_train, X_test, y_train, y_test = train_test_split(X,  
                                                       y,  
                                                       shuffle=False, test_size=0.15)  
  
    clf = MLPClassifier( max_iter=300).fit(X_train, y_train)  
  
    print('\n\nMLP Classifier')  
    print('Test value: ' + str(y_test))  
    print('Predicted value: ' + str(clf.predict(X_test)))  
    print('Score: ' + str(clf.score(X_test, y_test)))
```

Slika 22: Kod za MLP klasifikaciju

Prvo uzimamo stupac u kojim se nalaze indikatori te ga pretvaramo u vektor cijelih brojeva, a zatim to isto radimo za ostale stupce te dobivamo trodimenzionalan vektor u kojem se nalaze broj pozitivnih objava, broj negativnih objava i cijena valute. Zatim vektore dijelimo na skup za testiranje i treniranje pomoću funkcije *train\_test\_split()*, postavljamo da se ne miješaju nego da oстане poredano po datumu što znači da za testiranje uzimamo zadnjih par dana te postavljamo da nam je veličina testa 15% od ukupne veličine. Sljedeće pozivamo funkciju za treniranje modela te postavljamo maksimalni broj iteracija na 300. Za analizu printamo stvarnu vrijednost i predviđenu vrijednost indikatora rasta te točnost testa.

```

def Regressor(learningData):
    y = learningData['BitcoinPrice'].to_numpy()
    Date = learningData['Date']
    learningData.drop('Date',axis='columns', inplace=True)
    learningData.drop('BitcoinPrice',axis='columns', inplace=True)
    X = learningData.to_numpy()

    X_train, X_test, y_train, y_test = train_test_split(X, y,
                                                        shuffle=False, test_size=0.15)

    regr = MLPRegressor(hidden_layer_sizes=(100, 100), max_iter=500).fit(X_train, y_train)

    print('\n\nMLP Regressor')
    print('Test value: ' + str(y_test))
    print('Predicted value: ' + str(regr.predict(X_test)))
    print('coefficient of determination: ' + str(regr.score(X_test, y_test)))

    pred_data = regr.predict(X_test).tolist()
    test_data = y.tolist()
    DatePred = Date[-len(pred_data):]

    df = pd.DataFrame(columns=['Date', 'Price'])
    df1 = pd.DataFrame(columns=['Date', 'Predicted price'])

    df['Date'] = Date
    df['Price'] = test_data
    df1['Date'] = DatePred
    df1['Predicted price'] = pred_data
    df = df.join(df1.set_index('Date'), on='Date')

    df.plot(x='Date', y=['Price', 'Predicted price'])
    plt.savefig('Models/MLPRegressor.png')

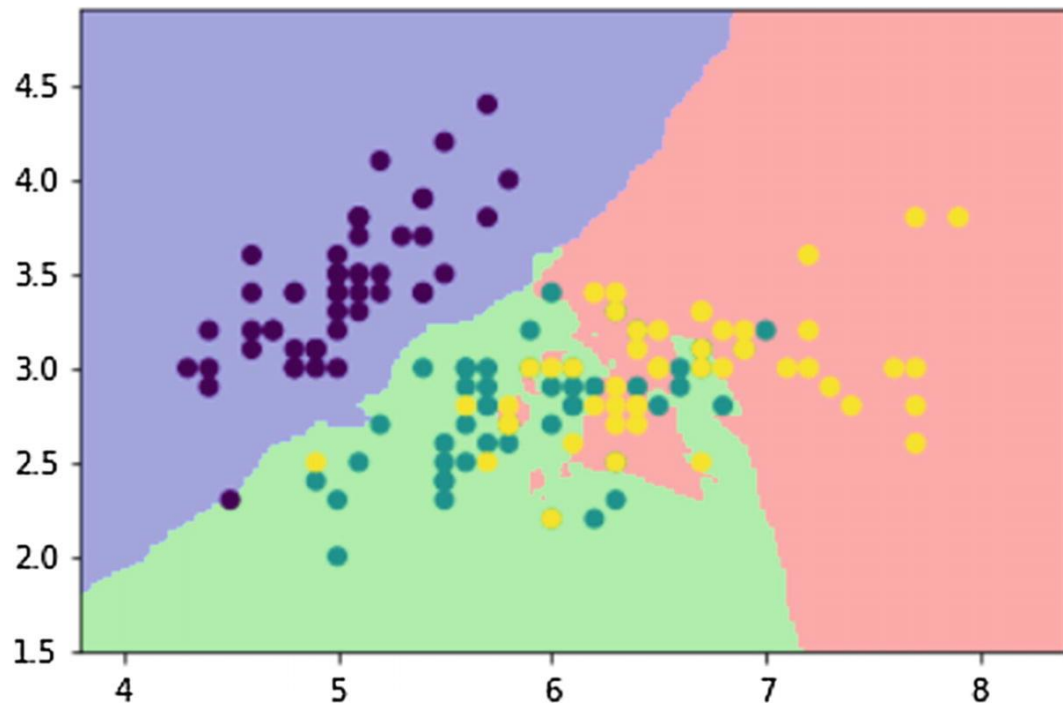
```

Slika 23: Kod za MLP regresiju

Kao i kod klasifikacije uzimamo podatke samo što sada u posebni vektor stavljamo cijenu valute, a u drugi vektor pozitivne poruke, negativne poruke i indikator rasta ili pada. Zatim dijelimo podatke na skup za testiranje i treniranje te pozivamo funkciju za treniranje modela. Koristimo se sa 100 neurona u skrivenim slojevima, a maksimalan broj iteracija je 500. Nakon toga printamo rezultate i koeficijent determinacije, te spremamo podatke za izradu grafa.

### 6.3. Algoritam K najbližih susjeda

Algoritam K najbližih susjeda je jedan od najosnovnijih, ali ipak bitnih klasifikacijskih algoritama u strojnom učenju. Pripada nadziranom učenju i nalazi intenzivnu primjenu u prepoznavanju uzoraka i rudarenju podataka.

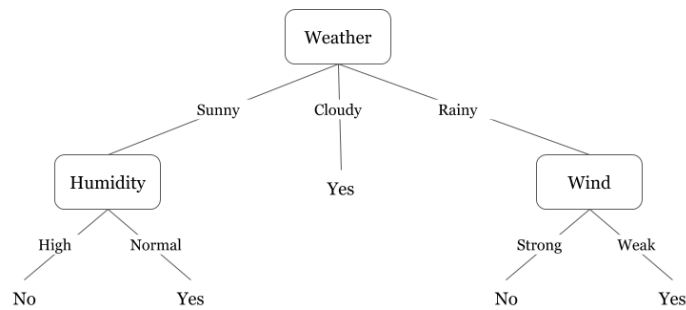


Slika 24: Primjer KNN-a

Primijetite na slici 24. da su većinom slični podaci blizu jedan drugog, odnosno da su podaci grupirani u tri skupine, a svaka skupina ima svoju boju. Vidimo da su slični podaci većinom grupirani na jednom dijelu grafa. Algoritam K najbližih susjeda trenira model računajući udaljenost između točaka na grafu. Za naš model koristili smo Python biblioteke *sklearn.neighbors.KNeighborsClassifier* i *sklearn.neighbors.KNeighborsRegressor*. Koristimo zadane postavke biblioteke što znači da koristimo 5 susjednih čvorova koje model automatski odabere najboljom metodom za te podatke. Funkcija za predikciju je „uniform“, a bodovi se mjere jednako u svakom čvoru. Formula koju koristimo za udaljenost je „minkowski“ sa snagom 2 što nam je ekvivalentno standardnoj Euclideanovoj normi. Kao i za neuronske mreže za treniranje uzimamo sve dane osim zadnjih 6 dana.

#### 6.4. Stablo odlučivanja

Stablo odlučivanja je nadzirana metoda učenja koja se koristi za klasifikaciju i regresiju. Cilj je stvoriti model koji predviđa vrijednost varijable učenjem jednostavnih pravila odlučivanja izvedenih iz značajki podataka. Stablo odlučivanja obično počinje s jednim čvorom koji se grana u moguće ishode. Svaki od tih ishoda dovodi do dodatnih čvorova koji se granaju u druge mogućnosti. To mu daje oblik stabla.



Slika 25: Stablo odluke za izaći van iz kuće ili ne

Za naš rad koristili smo python biblioteku *sklearn.tree.DecisionTreeClassifier* za klasifikaciju i predviđanje indikatora rasta te biblioteku *sklearn.tree.DecisionTreeRegressor* za regresiju i predviđanje cijene valuta. Za učenje modela koristimo zadane značajke, odnosno kriteriji za kvalitetu podijele koristimo gini nečistoću. Gini nečistoća je vjerojatnost pogrešne klasifikacije slučajno odabranog elementa u skupu podataka. Nemamo ograničenje za dubinu stabla te je minimalan broj uzorka za razdvojiti čvor 2, a minimalan broj uzorka na čvoru je 1. Za klasifikaciju ulaza uzimamo broj pozitivnih objava, broj negativnih objava i cijenu, odnosno za regresiju broj pozitivnih objava, broj negativnih objava i indikator rasta ili pada. Za treniranje uzimamo sve dane osim zadnjih 6 dana koje koristimo za testiranje.

## 7. REZULTATI

U ovom poglavlju ćemo prikazati rezultate svih modela onim redom kojim su modeli opisani u prijašnjem poglavlju.

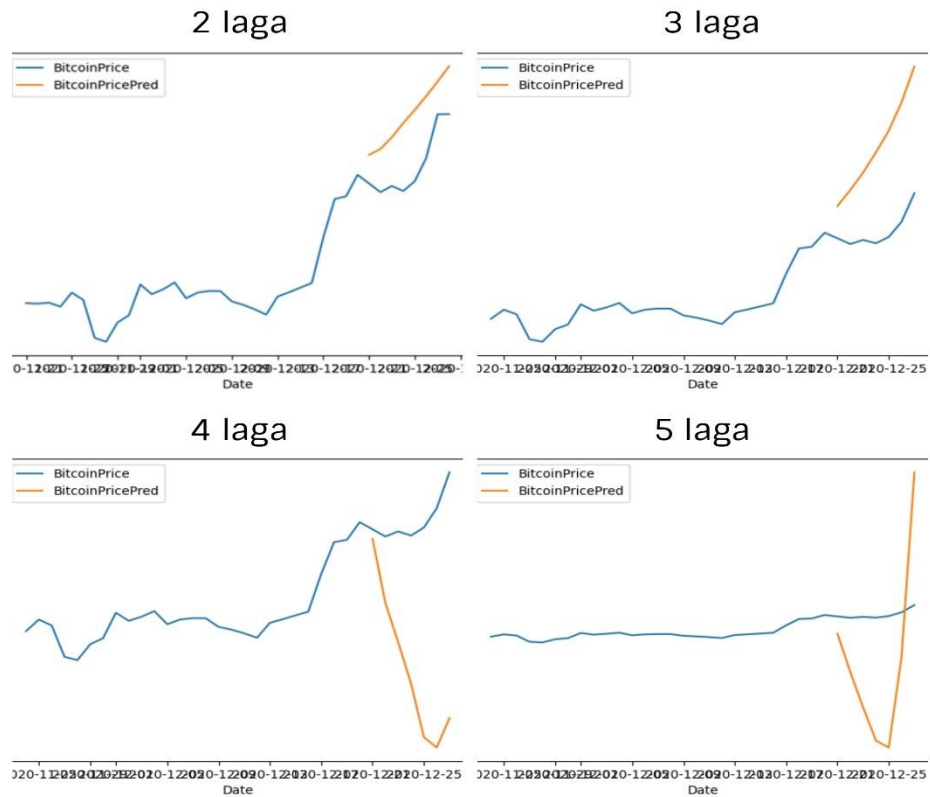
### 7.1. Vektorska autoregresija

Model predviđa cijenu valute i predviđa hoće li cijena rasti ili padati. Imamo dvije vrste rezultata, prve rezultate dobivamo tako da model predviđa cijenu kriptovalute i zatim uspoređuje s jučerašnjom cijenom te kriptovalute. Ako je cijena viša ispisuje ikonu za rast, odnosno ako je cijena niža ispisuje ikonu pada. U drugom slučaju model ne predviđa cijenu valute nego samo indikator rasta ili pada. Iako ne predviđa cijenu kriptovalute cijena je svejedno ulazni parametar u modelu. Rezultate nam ispisuje u terminalu te ispis izgleda kao na slici 26. gdje je prikaz rezultata za Bitcoin valutu. Vidimo da su vrijednosti u jednom i drugom slučaju rast cijene što nam se podudara. U prvom slučaju vidimo da je predviđao cijenu te da ispisuje ikonu rasta jer je cijena viša od jučerašnje.

```
2020-12-21 Day before: 0.7332215822712281 Predicted: 0.8205792723445198 Indicator: ▲
2020-12-22 Day before: 0.8205792723445198 Predicted: 0.8469385120488846 Indicator: ▲
2020-12-23 Day before: 0.8469385120488846 Predicted: 0.8980365029430248 Indicator: ▲
2020-12-24 Day before: 0.8980365029430248 Predicted: 0.9606938331645898 Indicator: ▲
2020-12-25 Day before: 0.9606938331645898 Predicted: 1.0180473056510282 Indicator: ▲
2020-12-26 Day before: 1.0180473056510282 Predicted: 1.078159905713521 Indicator: ▲
2020-12-27 Day before: 1.078159905713521 Predicted: 1.142952172183749 Indicator: ▲
2020-12-28 Day before: 1.142952172183749 Predicted: 1.2097128724693893 Indicator: ▲
Only indicator:
2020-12-21 Indicator: ▲
2020-12-22 Indicator: ▲
2020-12-23 Indicator: ▲
2020-12-24 Indicator: ▲
2020-12-25 Indicator: ▲
2020-12-26 Indicator: ▲
2020-12-27 Indicator: ▲
2020-12-28 Indicator: ▲
```

Slika 26: Prikaz rezultata s 2 laga





Slika 27: Prikaz rezultata grafova za Bitcoin valutu

Na kraju dobivamo prikaz grafa koji plavom bojom prikazuje stvarnu cijenu valute dok je narančastom bojom prikazana predviđena cijena valute te na grafu možemo lakše vidjeti koliko nam je model točan. Ako gledamo graf s 2 laga, iako se cijene ne podudaraju model nam je predvidio nagli rast cijene, no nije predvidio pad cijene u naredna tri dana. U ovom slučaju bi nam se isplatilo ulagati u Bitcoin valutu jer je cijena mnogo viša te je model to točno i predvidio. Uočavamo da što više laga imamo to vrijednosti više variraju i točnost je manja.

| Results for equation y3 |             |            |        |       |
|-------------------------|-------------|------------|--------|-------|
|                         | coefficient | std. error | t-stat | prob  |
| const                   | -0.017585   | 0.066429   | -0.265 | 0.791 |
| L1.y1                   | -0.125004   | 0.288576   | -0.433 | 0.665 |
| L1.y2                   | 0.356725    | 0.305385   | 1.168  | 0.243 |
| L1.y3                   | 1.148525    | 0.421523   | 2.725  | 0.006 |
| L1.y4                   | 0.014619    | 0.051551   | 0.284  | 0.777 |
| L2.y1                   | 0.296444    | 0.337988   | 0.877  | 0.380 |
| L2.y2                   | -0.464944   | 0.383413   | -1.213 | 0.225 |
| L2.y3                   | -0.050275   | 0.382117   | -0.132 | 0.895 |
| L2.y4                   | -0.037591   | 0.037359   | -1.006 | 0.314 |

Slika 28: Mjere za regresiju za Bitcoin model s 2 laga

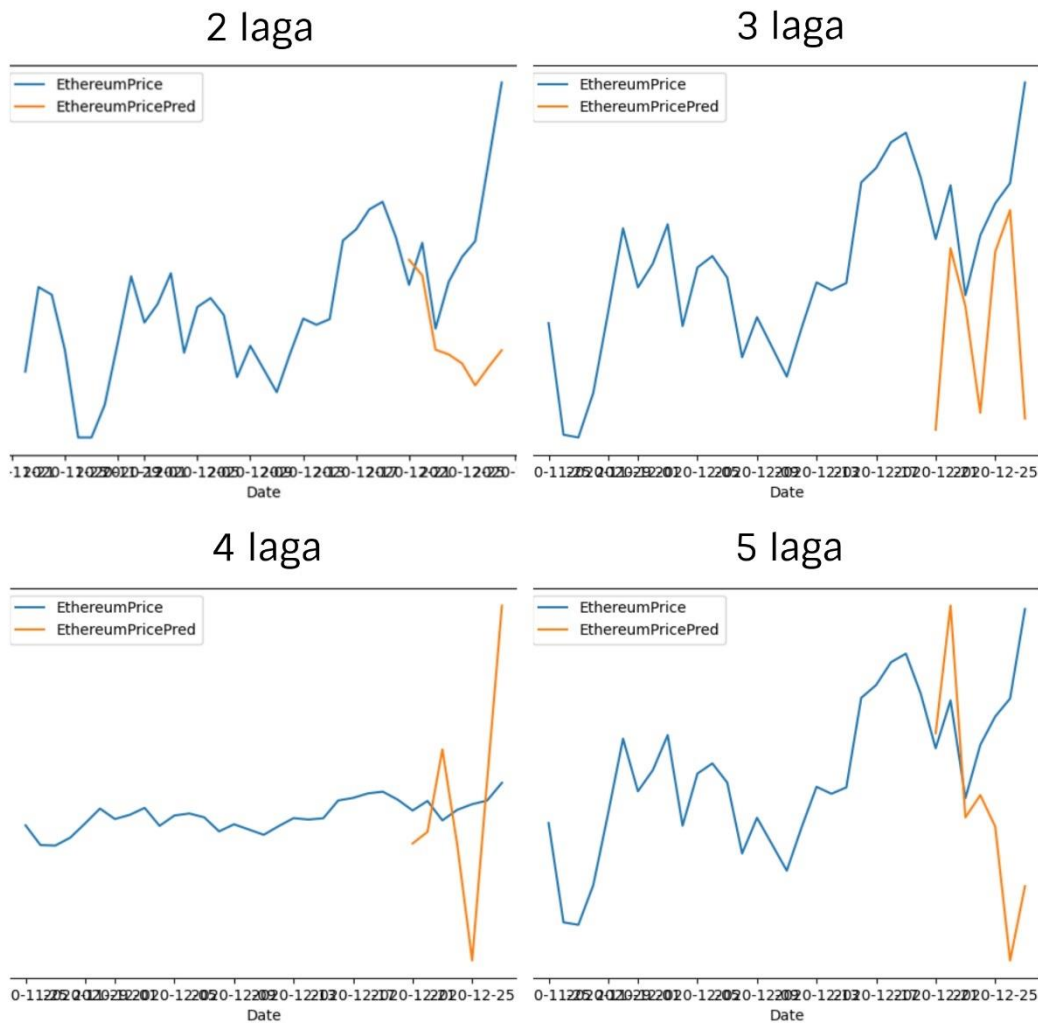
Slika 28. prikazuje odnose za varijablu cijena s drugim varijablama u svakom lagu. Y1 su pozitivne objave, y2 su negativne objave, y3 je cijena valute, a y4 je indikator rasta ili pada. L1 i L2 su lag-ovi. Koeficijent ili relativna standardna devijacija nam pokazuje promjenjivost varijable u odnosu na prosjek vrijednosti. Koeficijent se dobije kao količnik standardne devijacije i srednje vrijednosti.

$$c_v = \frac{\sigma}{\mu}$$

Standardna greška je prosječna vrijednost odstupanja od prosjeka vrijednosti. Ako gledamo standardnu devijaciju za cijenu valute odnosno y3, vidimo da je greška dosta velika te da nam model nije baš najvjerodostojniji. T-statistika je omjer odstupanja procijenjene vrijednosti varijable od njene pretpostavljene vrijednosti do njene standardne greške. T-statistiku računamo formulom

$$t_{\beta} = \frac{\beta - \beta_0}{SE(\beta)}$$

gdje je  $\beta$  procijenjena vrijednost,  $\beta_0$  poznata konstanta i  $SE(\beta)$  standardna greška procijenjene vrijednosti. U ispitivanju značajnosti nulte hipoteze, p-vrijednost je vjerojatnost dobivanja rezultata ispitivanja barem toliko ekstremnih koliko su rezultati stvarno dobiveni, pod pretpostavkom da je nulta hipoteza točna.



Slika 29: Prikaz rezultata grafova za Ethereum valutu

Kao i s Bitcoin valutom i ovdje što više lagova imamo vrijednosti su manje točne te ćemo samo analizirati model s 2 laga. U primjeru u kojem smo probali predvidjeti cijenu Ethereum valute model je predvidio pad te zatim i rast, no cijena valute se značajno razlikuje od predviđene cijene. Iako je prvo predvidio pad te bi u stvarnoj situaciji prvo pričekali i vidjeli danji razvoj cijena, u ovom slučaju model nam nije dao dobar uvid u cijenu valute jer je stvarna cijena mnogo viša nego predviđena te nam se isplati ulagati u ovu kriptovalutu.

Naša prvobitna ideja je bila predvidjeti samo indikator rasta ili pada, a ne točnu cijenu valute. Ovim modelom smo ipak uspjeli dobiti dobru točnost kod predviđanje indikatora.

Tablica 2: Stvarno i predviđeno kretanje cijene za Bitcoin valutu

| Bitcoin             | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---------------------|---|---|---|---|---|---|---|
| Stvarna cijena      | 0 | 0 | 1 | 0 | 1 | 1 | 1 |
| Procijenjena cijena | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

Tablica 3: Stvarno i predviđeno kretanje cijene za Ethereum valutu

| Ethereum            | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---------------------|---|---|---|---|---|---|---|
| Stvarna cijena      | 0 | 1 | 0 | 1 | 1 | 1 | 1 |
| Procijenjena cijena | 1 | 1 | 0 | 0 | 1 | 1 | 0 |

Tablica 2 i tablica 3 nam prikazuju stvarno i predviđeno kretanje cijene za Bitcoin i Ethereum valute. 0 označava pad cijene, a 1 označava rast cijene u odnosu na prethodni dan. Za Bitcoin valutu dobili smo model od 57% točnosti što zadovoljava naša očekivanja. Za Ethereum valutu dobili smo također točnost od 57% s čime smo zadovoljni. Iako ne možemo sa sigurnošću tvrditi hoće li cijena rasti ili padati, model nam može dati mali uvid u cijenu te u koju valutu uložiti a koju prodati.

## 7.2. Neuronske mreže

Slijedi model neuronskih mreža te isto kao i kod VAR-a imamo dva predviđanja. Predviđanje indikatora rasta ili pada i predviđanje cijene valuta. Prvo ćemo analizirati indikator rasta ili pada.

```
MLP Classifier
Test value: [0 1 0 1 1 1]
Predicted value: [1 1 1 1 1 1]
Score: 0.6666666666666666
```

Slika 30: Rezultati MLP modela za Bitcoin valutu

Na slici 30. prvo vidimo listu indikatora s kojima smo testirali, a zatim listu indikatora koje je model predvidio. Vidimo da je od 6 dana točno predvidio 4 dana što nam daje točnost testa od 66%. Isto kao i kod VAR-a, MLP model nam je predvidio rast u svim ostalim danima. Točnost od 66% je dosta visoka te bi u ovom slučaju mogli poslušati model i uložiti u ovu valutu.

```
MLP Classifier
Test value: [0 1 0 1 1 1 1]
Predicted value: [1 1 1 1 1 1 1]
Score: 0.7142857142857143
```

Slika 31: rezultat MLP modela za Ethereum valutu

Slika 31. nam prikazuje rezultate MLP modela za Ethereum valutu. Model je isto predvidio rast u svim ostalim danima, no kako je i u stvarnosti većinom cijena valute rasla iz dana u dan model nam ima točnost od 71% što je veoma visoko.

Sljedeće smo radili MLP model regresije. S ovim modelom ćemo probati predvidjeti cijenu kriptovalute, a ne samo indikator rasta ili pada.

```
MLP Regressor
Test value: [0.65759211 0.68502618 0.66263448 0.70532093 0.80760654 1. ]
Predicted value: [0.23914051 0.38505958 0.22852306 0.34066321 0.43021142 0.4147343 ]
coefficient of determination: -11.154667005399222
```

Slika 32: Rezultat MLP regresije za Bitcoin valutu

Na slici 32. vidimo da je predviđena cijena ipak puno niža od stvarne cijene. Koeficijent determinacije nam pokazuje koliko vrijednost jedne varijable ovisi o odnosima drugih varijabla gdje najbolja moguća vrijednost iznosi 1. Mi smo dobili koeficijent determinacije od -11.15 što je jako loše i ne možemo puno zaključiti gledajući u ovaj model, osim da broj pozitivnih i negativnih objava ne utječe na cijenu valute.

```
MLP Regressor
Test value: [0.55903681 0.71063551 0.40097754 0.56987133 0.65950895 0.71657941
1. ]
Predicted value: [0.320323 0.54306544 0.3835567 0.51993438 0.54957449 0.54908472
0.48621389]
coefficient of determination: -0.8956880551159159
```

Slika 33: Rezultat MLP regresije za Ethereum valutu

MLP regresija za Ethereum valutu nam je dala puno bolje rezultate i donekle prati stvarnu cijenu valute. Koeficijent determinacije je -0.86 što je još uvijek nisko za točno predviđanje cijene, ali možda nam može dati uvid u daljnji rast ili pad cijene.

### 7.3. Algoritam k najbližih susjeda

Sljedeći nam je model algoritma k najbližih susjeda te ćemo prvo analizirati klasifikaciju i regresiju za Bitcoin valutu, a zatim za Ethereum valutu.

```
KNeighbors Classifier
Test value: [0 1 0 1 1 1]
Predicted value: [1 1 1 1 1 1]
Score: 0.6666666666666666

KNeighbors Regressor
Test value: [0.65759211 0.68502618 0.66263448 0.70532093 0.80760654 1.          ]
Predicted value: [0.44010298 0.6867167  0.45260882 0.42179691 0.35833974 0.44907711]
coefficient of determination: -6.681601692305905
```

Slika 34: Rezultat KNN klasifikacije i regresije za Bitcoin valutu

Na slici 34. vidimo rezultate KNN modela za Bitcoin valutu. Kod klasifikacije smo samo predviđali indikator rasta te smo dobili točnost od 66% što zadovoljava naša očekivanja. Model je također kao i kod Neuronskih mreža predvidio samo rast valute. Kod regresije se baš i ne možemo pouzdati u naš model jer se cijene uveliko razlikuju od stvarnih cijena te nas koeficijent determinacije od -6.68 također navodi da ne vjerujemo puno ovom predviđanju.

```
KNeighbors Classifier
Test value: [0 1 0 1 1 1 1]
Predicted value: [1 1 1 1 1 1 1]
Score: 0.7142857142857143

KNeighbors Regressor
Test value: [0.55903681 0.71063551 0.40097754 0.56987133 0.65950895 0.71657941
1.          ]
Predicted value: [0.59082174 0.78866027 0.57713944 0.64689029 0.60414296 0.60414296
0.47407232]
coefficient of determination: -0.6267602161120363
```

Slika 35: Rezultat KNN klasifikacije i regresije za Ethereum valutu

KNN model za Ethereum valutu nam je dao malo bolje rezultate. Klasifikacija od 71% točnosti nam je vrlo visoka te je regresija s koeficijentom determinacije od -0.63 također puno bolja. Klasifikaciji možemo vjerovati, ali s predviđenom cijenom valute moramo biti oprezni jer se cijene i dalje dosta razlikuju.

#### 7.4. Stablo odlučivanja

Zadnji model nam je stablo odluke te ćemo njega analizirati redom isto kao i algoritam k najbližih susjeda, prvo Bitcoin te Ethereum valutu.

```
Decision Tree Classifier
Test value: [0 1 0 1 1 1]
Predicted value: [1 1 1 0 1 1]
Score: 0.5

Decision Tree Regressor
Test value: [0.65759211 0.68502618 0.66263448 0.70532093 0.80760654 1.          ]
Predicted value: [0.73384798 0.63958541 0.22267434 0.15444986 0.45811066 0.25856881]
coefficient of determination: -12.349162025280691
```

Slika 36: Rezultat stabla odluke klasifikacije i regresije za Bitcoin valutu

Slika 36. nam prikazuje rezultat stabla odlučivanja klasifikacije i regresije za Bitcoin valutu te možemo odmah zaključiti da smo dobili najgore rezultate od svih modela do sad. Model za klasifikaciju ima točnost od 50% što nije baš obećavajuće uzeći u obzir da može predvidjeti samo rast ili pad. Prvi put do sad da nam je model predvidio pad cijene, ali je nažalost na krivi dan predvidio pad. Kod regresije cijena kriptovalute se značajno razlikuje te nam je koeficijent determinacije od -12.35 dovoljni pokazatelj da ne možemo puno vjerovati ovom modelu.

```
Decision Tree Classifier
Test value: [0 1 0 1 1 1]
Predicted value: [0 1 1 0 1 1]
Score: 0.7142857142857143

Decision Tree Regressor
Test value: [0.55903681 0.71063551 0.40097754 0.56987133 0.65950895 0.71657941
1.          ]
Predicted value: [0.45019856 0.71867455 0.58946188 0.85864598 0.71867455 0.71867455
0.43699726]
coefficient of determination: -1.182607119494504
```

Slika 37: Rezultat stabla odluke klasifikacije i regresije za Ethereum valutu

Stablo odlučivanja za Etheruem valutu nam također daje puno točnije rezultate te zaključujemo da nam je kod svih modela nam Ethereum valuta puno točnija. Moguće da cijena Ethereum valute više ovisi o objavama na društvenim mrežama jer je ipak manje poznata od Bitcoin valute te svaka objava može imati veći utjecaj na cijenu. Također je moguće da je za Ethereum valutu napravljena bolja analiza sentimenta jer se o Ethereum valuti ne priča općenito kao i za Bitcoin. Model za klasifikaciju nam je dao točnost od 71% je veoma visoko, te je čak i točno predvidio pad cijene. Kod regresije i dalje imamo jako niski koeficijent determinacije od -1.18 ali puno je više i vjerodostojne nego kod Bitcoin valute.

## 7.5. Analiza rezultata

I prije nego smo počeli raditi na projektu znali smo da je teško ili gotovo nemoguće predvidjeti točno kretanje valute te da bi za točnije rezultate trebali koristiti više parametra od broja pozitivnih i negativnih objava na 4Chanu. Ipak, dobili smo neke rezultate i možemo ih usporediti te vidjeti što nam daje najbolje rezultate za pojedinu valutu.

Tablica 4: Tablica rezultata

|                    | Bitcoin |             | Ethereum |             |
|--------------------|---------|-------------|----------|-------------|
|                    | Točnost | Koeficijent | Točnost  | Koeficijent |
| VAR                | 57%     | -0.02       | 57%      | -0.01       |
| NN                 | 67%     | -11.15      | 71%      | -0.90       |
| KNN                | 67%     | -6.69       | 71%      | -0.63       |
| Stablo odlučivanja | 50%     | -12.35      | 71%      | -1.18       |

Kad bi gledali samo indikatore rasta ili pada Neuronske mreže i algoritam k najbližih susjeda nam imaju najveću točnost. Na prvu ispada da nam je model dosta točan, ali nažalost ti rezultati ne moraju biti točni jer nismo imali puno podataka za testiranje i točnost od 67% gdje može predvidjeti samo dvije opcije nije toliko visoka točnost. Kod koeficijenta najbolje rezultate nam daje VAR model, ali koeficijent od -0.02 i -0.01 nam je i dalje dosta nizak. Na kraju možemo zaključiti da nam model nije uspješan, a razlog tome može biti da nam analiza sentimenta nije dobra ili jednostavno broj objava i analiza sentimenta nisu dovoljni za predviđanje cijena. Visoku točnost kod predviđanja cijena je veoma teško ili gotovo nemoguće dobiti jer ovisi i o mnogo drugih faktora, no možemo nadograđivati model sve dok ne dobijemo veću točnost modela.



## 8. ZAKLJUČAK

Umijeće predviđanja cijena kriptovaluta težak je zadatak mnogih istraživača i analitičara. Investitori su zapravo vrlo zainteresirani za istraživanje područje predviđanja cijena kriptovaluta jer je za dobro i uspješno ulaganje nužno znanje buduće situacije kretanja tržišta. Dobri i učinkoviti sustavi predviđanja pomažu trgovcima, investitorima i analitičarima pružajući dodatne informacije budućeg usmjerenja. Bitcoin i Ethereum, dvije najveće kriptovalute u smislu tržišne kapitalizacije, predstavljaju preko 160 milijardi dolara u kombiniranoj vrijednosti. Obje kriptovalute doživjele su značajne promjene cijena kako na dnevnoj tako i na dugoročnoj procjeni vrijednosti. Twitter se sve više koristi kao izvor vijesti i utječe na odluke o kupnji valute informiranjem korisnika o vrijednosti i popularnosti. Kao rezultat toga, brzo razumijevanje utjecaja tweetova na smjer cijena može pružiti prednost pri kupovini i prodaji korisniku ili trgovcu kriptovaluta.

Zbog nemogućnosti analize Twittera uzrokovane restrikcijama koje imaju, umjesto Twittera analiziran je 4Chan i njegov poddirektoriji *Bussines & Finance*. Korištenjem vektorskog modela koji uzima za ulazne podatke broj objava ili broj spominjanja neke valute, uspješno je predviđen smjer promjene cijena. Iako točnost modela nije dovoljno visoka za efikasno korištenje modela, daljnjim unaprjeđenjem i korištenjem ovog modela korisnik može donijeti bolje i informiranije odluke o kupnji i prodaji neke kriptovalute.

Iako je ideja projekta dosta jednostavna, projekt ima još puno prostora za daljnji razvoji i novo unaprjeđenje. Glavna ideja za daljnji razvoj bila bi veća točnost modela koja bi se dobila korištenjem drugih modela i tehnologija za učenje te uspoređivanjem rezultata. Boljom semantičkom analizom objava također bi dobili veću točnost modela kao i dodavanjem novih značajki u podatke poput broja komentara ili "lajkova" na objavama ali bi se za to trebalo uzimati podatke s neke druge društvene mreže jer na 4Chan-u nema „lajkova“. Zanimljivo bi bilo napraviti i usporedbu dobivenih rezultata s cijenama drugih kriptovaluta ili promjenama burzovnih indeksa kako bi istražili postoji li korelacija u kretanjima vrijednosti.

Analizom tržišnog raspoloženja i broja objava u vezi s određenom kriptovalutom dobili bi brzo i jednostavno očitavanje trenutnog stanja tržišta u stvarnom vremenu i dodatnu podršku u odluci o ulaganju. Za uspješno korištenje programa trebalo bi napraviti automatsko skidanje objava na dnevnoj bazi te automatski predviđati rast ili pad cijene za sljedeći dan ili tjedan dana unaprijed.

## 9. POPIS SLIKA

|   |    |
|---|----|
| Slika 1: centralizirani i blockchain sistem .....   | 6  |
| Slika 2: Distribuirani sustav(Izvor: Usman M. Sheikh, Virtual currency: current regulatory and civil litigation trends, 2018) ..... | 11 |
| Slika 3: Centraliziran i decentraliziran sustav (Izvor : Stewart Martins, Understanding Blockchain: A Simple Approach) .....        | 11 |
| Slika 4: 10 najzastupljenijih kriptovaluta (Izvor: Coin Market) .....   | 13 |
| Slika 5: Kod za preuzimanje svih linkova objava .....   | 20 |
| Slika 6: Paralelno pozivanje getAllLink() funkcije .....  | 20 |
| Slika 7: Preuzimanje objava, vrijeme objave i linka objave .....  | 21 |
| Slika 8: Primjer preuzetih objava u csv datoteci .....  | 21 |
| Slika 9: Dodavanje indikator rasta ili pada .....   | 22 |
| Slika 10: Sortiranje po datumu .....  | 23 |
| Slika 11: Kod za dodavanje polariteta .....   | 23 |
| Slika 12: Čitanje cijene valute u tablice .....   | 24 |
| Slika 13: Grupiranje po broju traženih riječi .....   | 24 |
| Slika 14: Grupiranje po broju objava .....  | 25 |
| Slika 15: Spajanje cijene i grupiranih podataka po datumu .....   | 25 |
| Slika 16: Normalizacija podataka .....  | 26 |
| Slika 17: Prikaz grafa Bitcoin cijene i broj njegovih objava .....  | 26 |
| Slika 18: Prikaz Ethereum cijene i broj njegovih objava .....   | 27 |
| Slika 19: Završna tablica .....   | 28 |
| Slika 20: Podjela na skup za treniranje i testiranje .....  | 30 |
| Slika 21: Poziv VAR() funkcije .....  | 30 |
| Slika 22: Kod za MLP klasifikaciju .....  | 31 |
| Slika 23: Kod za MLP regresiju .....  | 32 |
| Slika 24: Primjer KNN-a .....   | 33 |
| Slika 25: Stablo odluke za izaći van iz kuće ili ne .....   | 34 |
| Slika 26: Prikaz rezultata s 2 laga .....   | 35 |
| Slika 27: Prikaz rezultata grafova za Bitcoin valutu .....  | 36 |
| Slika 28: Mjere za regresiju za Bitcoin model s 2 laga .....  | 37 |

|   |    |
|---|----|
| Slika 29: Prikaz rezultata grafova za Ethereum valutu .....                         | 38 |
| Slika 30: Rezultati MLP modlea za Bitcoin valutu.....                               | 39 |
| Slika 31: rezultat MLP modela za Ethereum valutu .....                              | 40 |
| Slika 32: Rezultat MLP regresije za Bitcoin valutu.....                             | 40 |
| Slika 33: Rezultat MLP regresije za Ethereum valutu .....                           | 40 |
| Slika 34: Rezultat KNN klasifikacije i regresije za Bitcoin valutu.....             | 41 |
| Slika 35: Rezultat KNN klasifikacije i regresije za Ethereum valutu .....           | 41 |
| Slika 36: Rezultat stabla odluke klasifikacije i regresije za Bitcoin valutu .....  | 42 |
| Slika 37: Rezultat stabla odluke klasifikacije i regresije za Ethereum valutu ..... | 42 |

## **10. POPIS TABLICA**

|  |    |
|--|----|
| Tablica 1 Usporedba centraliziranih i decentraliziranih kriptovaluta ..... | 10 |
| Tablica 2: Stvarno i predviđeno kretanje cijene za Bitcoin valutu .....    | 39 |
| Tablica 3: Stvarno i predviđeno kretanje cijene za Ethereum valutu.....    | 39 |
| Tablica 4: Tablica rezultata.....  | 43 |

## **11. POPIS PRILOGA**

Uz ovaj rad priložen je CD na kojem se nalazi rad

## 12. LITERATURA

1. Antony Lewis (2015), A gentle introduction to blockchain technology – Bits on blocks, <https://bitsonblocks.net/2015/09/09/a-gentle-introduction-to-blockchain-technology/> [pristupano 25.05.2021.]
2. Branka Vuleta (2021), How Much Data is Created on the Internet Each Day? – Seed Scientific, <https://seedscientific.com/how-much-data-is-created-every-day/> [pristupano 29.07.2021.]
3. Today's Cryptocurrency Prices by Market Cap - Coin Market, <https://coinmarketcap.com/> [pristupano 29.07.2021.]
4. Domina Hozjan, (2017), Blockchain, PMF, Zagreb
5. Henry He (2018), Blockchain Basics –Fork
6. Holtz-Eakin, D., Newey, W., and Rosen, H. S., (1988), Estimating Vector Autoregressions with Panel Dana
7. Ipsos. (2018). Cracking the code on cryptocurrency. ING.
8. Što je Bitcoin – Hrvatski Bitcoin portal, <https://crobotcoin.com/bitcoin/sto-je-bitcoin/> [pristupano 25.05.2021.]
9. Understanding blockchain – Blockchain explained <https://blockchainexplained.org/understanding-blockchain/> [pristupano 26.05.2021.]
10. Nick Darlington (2021), Blockchain For Beginners: What Is Blockchain Technology? A Step-by-Step Guide – Blcok geek, <https://blockgeeks.com/guides/what-is-blockchain-technology/> [pristupano 20.06.2021.]
11. Term frequency – RyteWiki, [https://en.ryte.com/wiki/Term\\_Frequency](https://en.ryte.com/wiki/Term_Frequency) [pristupano 24.06.2021.]
12. Rachel Wolff (2020), Semantic analysis, explained – Monkey learn, <HTTPS://MONKEYLEARN.COM/BLOG/SEMANTIC-ANALYSIS/>
13. Kit Smith (2019), 126 amazing social media statistic and fact – Brandwatch, <https://www.brandwatch.com/blog/amazing-social-media-statistics-and-facts/>
14. Henry He (2018), Blockchain Basics – Fork – C sharp corner, <https://www.c-sharpcorner.com/article/blockchain-basic-fork/>
15. Hyndman, Rob J; Athanasopoulos, George, (2018), Vector Autoregressions
16. Imran Bashir, (2017), Mastering Blockchain
17. Kit Smith, Micro Focus Blog 126 Amazing Social Media Statistics and Facts, (2019)
18. Lansky, Jan. (2018), Possible State Approaches to Cryptocurrencies. Journal of Systems Integration. 8. 10.20470/jsi.v9i1.335.

19. Nakamoto, Satoshi. (2009), Bitcoin: A Peer-to-Peer Electronic Cash System. Cryptography Mailing list at <https://metzdowd.com>.
20. Rachel Wolff, Semantic Analysis In Linguistics: An Introduction, (2020)
21. Sasha Ivanović, Jacob Saur, Lisa Thomson (2018), What is Blockchain Technology? A Step-by-Step Guide For Beginners
22. Stewart Martins (2018), Understanding Blockchain: A Simple Approach
23. Usman M. Sheikh, (2018), Virtual currency: current regulatory and civil litigation trends
24. Scikit-learn, <https://scikit-learn.org/stable/index.html> [pristupano 02.09.2021.]