

SREDNJA ELEKTRO IN RAČUNALNIŠKA ŠOLA VELENJE
Trg mladosti 3, Velenje
MLADI RAZISKOVALCI ZA RAZVOJ ŠALEŠKE DOLINE

**RAZISKOVALNA NALOGA
NAPOVED SPREMEMBE CENE BITCOINA
NA PODLAGI TWITTERJA**

Tematsko področje: računalništvo

Avtorja:
Bernard Kužnik, 3. letnik
Nik Krnjovšek, 3. letnik

Mentorja:
Rok Urbanc
Islam Mušić

Velenje, 2022

Raziskovalna naloga je bila narejena na ŠC Velenje, Elektro in računalniška šola, 2021/2022.

Mentorja:
Rok Urbanc,
Islam Mušić

Datum predstavitve:
Marec, 2022



By: Avtorji N. Krnjovšek, B. Kužnik, R. Urbanc, I. Mušić

KLJUČNA DOKUMENTACIJSKA INFORMACIJA

ŠD ŠC Velenje, šolsko leto 2021/2022

KG napoved cene kripto valut

AV Krnjovšek Nik, Kužnik Bernard

SA Urbanc Rok, Mušič Islam

KZ 3320 Velenje, SLO, Trg mladosti 3

ZA ŠC Velenje, Elektro in računalniška šola, 2022

LI 2022

IN NAPOVED SPREMEMBE CENE BITCOINA NA PODLAGI TWITTERJA

TD Raziskovalna naloga

IJ SL

JI sl

AI V današnjem svetu se kripto valute povečujejo v popularnosti in načinu služenja denarja. Ker se vedno več ljudi zanaša na pridobivanje denarja na tak način, se je pojavil problem negotovosti in izgube denarja. Sreča je glavni igralec pri vlaganju denarja v kripto valute in ta faktor prepreči veliko ljudem, da bi dobili denar, za katerega so investirali. Nas je zanimalo ali lahko to srečo, ki je prisotna zamenjamo s statistiko, ki nam pove ali bo cena padla ali narasla s pomočjo umetne inteligence.

KEY WORDS DOCUMENTATION

ND ŠC Velenje, 2021/2022

CX crypto price prediction

AU Krnjovšek Nik, Kužnik Bernard

AA Urbanc Rok, Mušić Islam

PP 3320 Velenje, SLO, Trg mladosti 3

PB ŠC Velenje, Elektro in računalniška šola, 2022

PY 2022

BITCOIN PRICE PREDICTION BASED ON TWITTER DATA

LA SL

AL sl/en

AD In today's world cryptocurrencies are increasing in popularity as a way of making a living. As more and more people rely on making money this way, the problem of losing money and your savings is rising. Luck is a major player in investing money in cryptocurrencies and this factor prevents many people from getting their money back and gaining profit. We were interested in whether we could replace the luck factor with statistics that tell us whether the price will fall or rise with the help of artificial intelligence.

SEZNAM OKRAJŠAV IN TUJK

Angl. – polni angleški izraz

Npr. – na primer

Oz. – oziroma

Tj. – to je

% - odstotek

KAZALO VSEBINE

1	Uvod.....	1
1.1	Ideja za izdelavo raziskovalne naloge	1
1.2	Oprelitev problema	1
1.3	Hipoteze.....	1
2	Pregled trenutnih predvidevanj kripto valut z uporabo umetne inteligence	2
2.1	Unicsoft	2
2.2	ScienceDirect.....	2
3	Začetek raziskovalne naloge	3
3.1	Python 3.9.....	3
3.2	Tvit.....	3
3.3	Pycharm	4
3.4	Python knjižnice	4
3.5	Twitter	4
4	Programska oprema in potek.....	5
4.1	Prvi del naloge	9
4.2	Drugi del naloge	9
5	Umetna Inteligenca/UI.....	11
5.1	Računalniški vid	11
5.2	Nevronske mreže	12
5.3	Globoko učenje.....	14
5.4	Popularne nevronske mreže.....	15
6	Praktični del	18
7	Proces delovanja	19
7.1	Pridobivanje podatkov	19
7.2	Filtriranje podatkov	20
7.3	Shranjevanje podatkov	21
7.4	Posodabljanje podatkov	22
7.5	Obdelava podatkov in predvidevanje	22
7.6	Rezultati.....	23
8	Napake in zapleti.....	24
8.1	Pridobivanje podatkov	24
8.2	Filtriranje podatkov	24
8.3	Shranjevanje podatkov	25

8.4	Obdelava podatkov in predvidevanje	25
9	Zaključek.....	26
9.1	Povzetek.....	26
9.2	Zahvala	26
10	Viri in literatura	27
10.1	Viri besedil	27
10.2	Viri slik.....	28
10.3	Viri enačb	29
11	Avtorja raziskovalne naloge.....	30

KAZALO SLIK

Slika 1: python logo, Vir 5	3
Slika 2: PyCharm logo, Vir 6	4
Slika 3: Twitter logo, Vir 7	5
Slika 4: Twint logo, Vir 9.....	5
Slika 5: Pandas logo, vir 10.....	6
Slika 6: Reg Ex logo, vir 11	6
Slika 7: Sentiment chart, vir 12.....	7
Slika 8: CoinGecko logo, vir 13.....	7
Slika 9: Sklearn logo, vir 14.....	8
Slika 10: Primer končnega izgleda Excel datoteke, Vir 8.....	9
Slika 11: Vizualni prikaz stopenj; vir 1.....	11
Slika 12: graf linearne regresije, vir 2	12
Slika 13: Nevronske mreže, Vir 4	13
Slika 14: Globoko učenje, Vir 3.....	14
Slika 15: LSTM celica, vir 15	15
Slika 16: GRU celica, vir 16	16
Slika 17: FFN zgradba, vir 17	17
Slika 18: Uporaba knjižnice twint, lasten vir	19
Slika 19: Filter za smeške, lasten vir.....	20
Slika 20: Filter za linke in #, lasten vir	20
Slika 21: Končni izgled Excel datoteke, lasten vir.....	21
Slika 22: Učenje programa, lasten vir	22
Slika 23: Graf rezultatov, lasten vir.....	23
Slika 24: Smeško unicode primer, lasten vir	25
Slika 25: Slika avtorjev, lasten vir	30

1 UVOD

1.1 Ideja za izdelavo raziskovalne naloge

V današnjem svetu mislim, da ne bi šlo brez kripto valut, ki so med nami sedaj že kar nekaj časa. Bitcoin je ena izmed kripto valut, ki je bila ustvarjena leta 2009 in je sedaj najbolj poznana valuta na svetu. Svet se vedno bolj nagiba na stran vlaganja v kripto valute, ker z malo sreče lahko hitro obogatiš, vendar z malo nesreče pa lahko tudi hitro izgubiš denar. Zaradi te problematike je nastala ideja za to raziskovalno nalogo. Kot že poznano, umetna inteligenca ni nič novega, vendar se vsak dan bolj razvija in je uporabljena pri stvareh, ki so neverjetne. Umetna inteligenca je že v preteklosti bila uporabljena pri kripto valutah za raznorazne stvari, ki se tičejo predvidevanja ali pa beleženja starih podatkov. Pri denarju je pomemben vsak procent, da tega denarja ne izgubiš, tu pride v igro program za predvidevanje cene, ki uporabniku malo olajša odločitev, ali je pametno investirati ali ne.

1.2 Opredelitev problema

Problematika, ki jo skušamo rešiti, je negotovosti pri investiranju denarja v kripto valute. Problematika se povečuje, ker popularnost kripto valut še vedno raste. Rešili bi jo tako, da bi sprogramirali program, ki je zmožen predvideti ceno Bitcoina z vsaj 60-% natančnostjo.

1.3 Hipoteze

- Pridobivanje tvitov z uporabo programskega jezika Python.
- Filtriranje podatkov in shranjevanje v Excel datoteko, ki jo je možno brati.
- Predvidevanje cene na podlagi pridobljene datoteke.
- Preprosta uporaba programa za vse uporabnike.
- 60-% natančnost programa za predvidevanje cene

2 PREGLED TRENUTNIH PREDVIDEVANJ KRIPTO VALUT Z UPORABO UMETNE INTELIGENCE

2.1 Unisoft

Preden sva sploh začela delati sva si pogledala trenutne primere predvidevanj kripto valut s uporabo umetne inteligence. Podjetje Unisoft je naredilo izziv da naredijo takšen program z analiziranjem raznih forumov ki se redno ukvarjajo z trgovanjem kripto valut. Imeli so nekaj izzivov, ampak jim je uspelo ustvariti program v štirih tednih. Glavne zatičke katere so izkusili so bili, najti dobre vire, kateri imajo vpliv na ceno kripto valut ter optimizacijo za čas in virov, ki so jih imeli. Rezultate katere so dosegli so bili 73 % pravilnost glede na to ali bo cena narasla ali padla {Vir 3}.

2.2 ScienceDirect

Osebe *Patrick Jaquart*, *David Dann* in *Christof Weinhardt* so na spletni strani ScienceDirect naredili raziskavo ter uporabili verzije nevronske mreže: GRU(Gated recurrent unit), LSTM(Long short-term memory), FNN(Feedforward Neural Network), LR(Linear Regression), GBC(Gradient Boosting Classifier), RF(Random Forest) da poizkusijo predvideti ceno kripto valut. Uporabili so podatke iz raznih družbenih medijih, kot so Bloomberg, Twitter in Blockchain.com. Rezultati so podobni med vsemi nevronske mreže, kar je bilo okoli 53 %, vendar je LSTM povprečno predvideval najboljše, za okoli 1 %. {Vir 4}.

3 ZAČETEK RAZISKOVALNE NALOGE

Ob začetku raziskovalne naloge smo morali narediti ogromno raziskav, kako bi naš program sploh deloval. Ob tem smo morali poiskati še vse stvari, ki naj bi jih za projekt potrebovali (knjižnice, programski jezik...). Ko smo našli vse, kar smo potrebovali in raziskali vse, kar smo morali vedeti, smo začeli delo.

3.1 Python 3.9

Python je programski jezik in 3.9 nam prikazuje, katero verzijo Pythona porabljam. Python je bil ustvarjen 1991 in je od takrat dobil veliko ime. Lahko ga uporabljamo pri skoraj vseh stvareh, vendar ta jezik najbolj slavi po njegovi enostavnosti in optimizaciji za UI. Zaradi njegovih knjižnic in preprostega učenja jezika ima naziv najboljšega programskega jezika za UI. Kaj je naredilo Python naš jezik izbire je preprosta uporaba in preproste sintakse, ogromno knjižnic, ki so jih naredili posamezniki za javno uporabo, preprosta dostopnost {Vir 5}.



Slika 1: python logo, Vir 5

3.2 Tvit

Tviti (angl. tweets) so podatki iz socialnega omrežja Twitter. Po navadi jih objavijo posamezniki in s svojimi tviti izrazijo neko osebno mnenje ali mišljenje o določeni tematiki. Twitter je ogromna platforma, na kateri se ustvarjajo razni popularni razgovori in eden izmed teh je tudi Bitcoin in ostale kripto valute. Če se cena Bitcoina spremeni, se bo to najprej izvedelo na Twitterju in ravno zato je naša izbira socialnega omrežja Twitter {Vir 6}.

3.3 Pycharm

Pycharm je orodje za urejanje kode napisane v programskem jeziku Python. Pycharm je bil prvič predstavljen leta 2010 iz strani podjetja JetBrains. Še danes je ena izmed najbolj uporabljenih orodij za programiranje v Pythonu. {Vir 7}



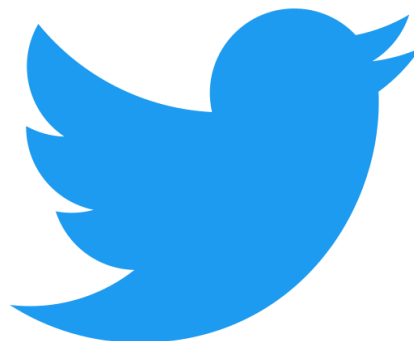
Slika 2: PyCharm logo, Vir 6

3.4 Python knjižnice

Python knjižnice so namenjene temu, da je programiranje lažje. Knjižnico lahko ustvarijo posamezniki in jo pripravijo na javno uporabo, ki je preprosta za inštalacijo (pip3 install [ime knjižnice]). Ko jo namestimo, lahko začnemo uporabljati funkcije, ki so definirane v tej knjižnici. Če jih ne bi imeli pa bi morali funkcije pisati sami, kar bi vzelo ogromno več časa {Vir 8}

3.5 Twitter

Twitter je popularno socialno omrežje, predstavljeno leta 2006. Avtor ali Ustanovitelj tega omrežja je Jack Dorsey. Hitro je postalo glavno omrežje, kjer se razglablja različne tematike glede popularnih tem, med katere je sedaj prišla tudi tema o kriptovalutah. Najpomembnejša med temi je Bitcoin, zato je pogovora o tej kriptovaluti največ {Vir 9}



Slika 3: Twitter logo, Vir 7

4 PROGRAMSKA OPREMA IN POTEK

Pri izdelavi raziskovalne naloge so bili potrebni naslednji pripomoči/oprema:

- Računalnik (Operacijski sistem Windows 10/11)
- Orodje za urejanje kode PyCharm (Professional 2021.2)
- Programski jezik Python (Python 3.9)
- Socialno omrežje Twitter
- Python knjižnice:
 - **Twint** (Uporabno orodje, ki dovoli pridobivanje tvitov iz Twitterja glede na veliko različnih faktorjev, kot so:
 - Različne besede v tvidu
 - Lokacija tvita
 - Jezik tvita
 - Čas objave tvita, itd.)



Slika 4: Twint logo, Vir 9

- **Pandas** (Glavna uporaba te knjižnice je izdelava podatkovnega okvira, ki nam dovoli obdelavo velike količine podatkov. Dovolj tudi obdelovanje CSV (comma separated values), kar smo največ uporabljali pri izdelavi te naloge)



Slika 5: Pandas logo, vir 10

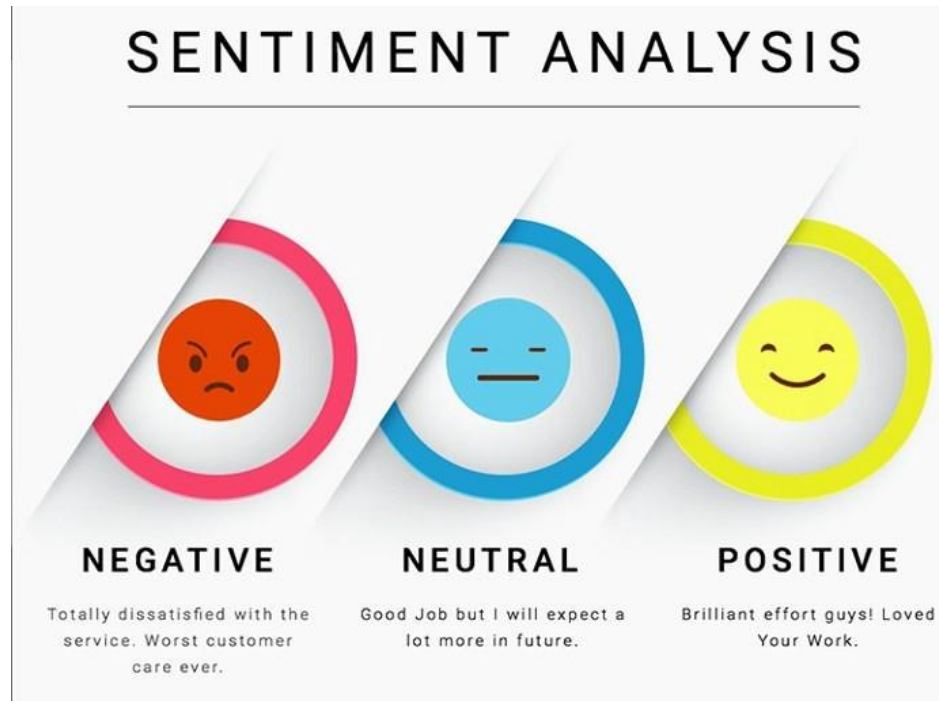
- **Re** (Je knjižnica, ki nam dovoli uporabo Regexa, ta je potreben za lažjo filtracijo podatkov katerih dobimo iz Twitterja)



Slika 6: Reg Ex logo, vir 11

- **SentimentIntensityAnalyzer** (Knjižnica nam dovoli analizirati »čustvenost« tvita.

Iz knjižnice lahko pridobimo ali je tvit pozitiven, negativen, ali nevtralen. Veliko analizatorjev čustva je možnih za uporabo v jeziku Python, ampak smo izbrala tega, ker je posebno namenjen analiziranja socialnih omrežij)



Slika 7: Sentiment chart, vir 12

- **Pycoingecko** (Dovoli nam pridobivanje cen kripto valut v tekočem času. Dovoli tudi več kot samo Bitcoin, tudi druge kripto valute)



Slika 8: CoinGecko logo, vir 13

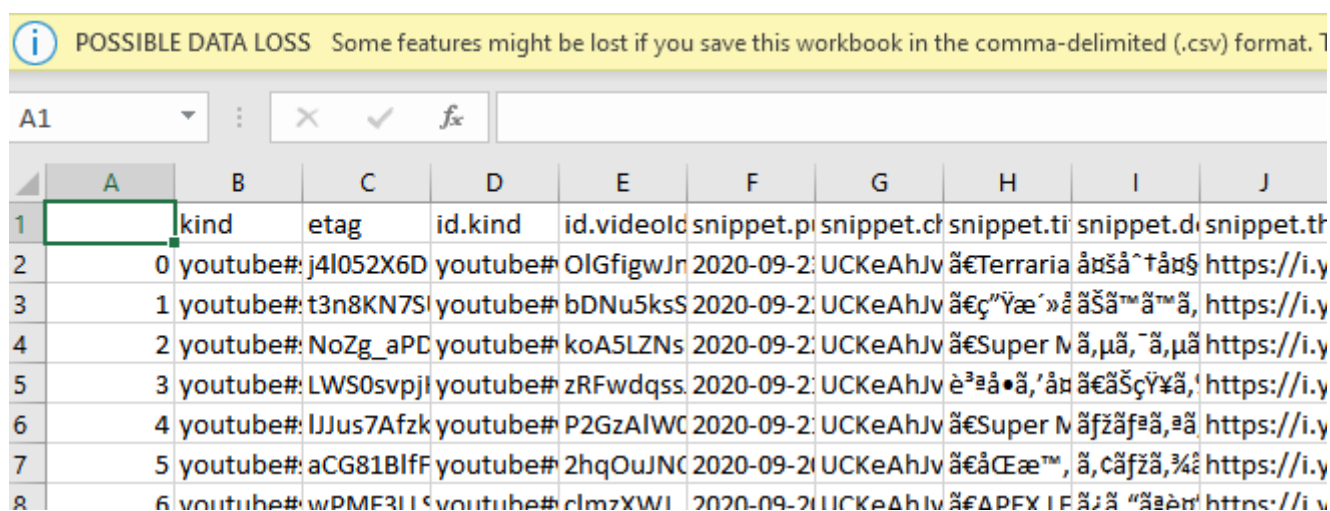
- **Sklearn** (Je glavna knjižnica katero smo uporabljali, pomembna je, ker nam dovoli enostavno uporabo nevronske omrežije in Linearne regresije)



Slika 9: Sklearn logo, vir 14

4.1 Prvi del naloge

Delo smo razdelili na dva dela, ker je potek dela lažji in ne tako kompleksen. Prvi del raziskovalne naloge je bil pridobivanje informacij iz Twitterja, tako da vsakih 5 minut program zakroži in pridobi tvite, ki jih nato filtrira in zabeleži v berljivo Excel datoteko. Za naslednji del programa pa smo morali vedeti podatke, ki so bili zabeleženi v Excel datoteko. (čas objave, pozitivnost/negativnost/nevtralnost tvita v %, število zabeleženih tvitov v enem krogu pridobivanja, trenutno ceno Bitcoina, ceno iz prihodnosti). Zamisel pridobivanja podatkov je bila preprosta saj ne bi bili prvi, ki bi to storili {Vir 1}.



	A	B	C	D	E	F	G	H	I	J
1		kind	etag	id.kind	id.videoid	snippet.p	snippet.cl	snippet.ti	snippet.d	snippet.th
2	0	youtube#:	j4l052X6D	youtube#:	OIGfigwJn	2020-09-2	UCKeAhJv	ã€Terraria	ã€šã^tã€š	https://i.y
3	1	youtube#:	t3n8KN7S	youtube#:	bDNu5ksS	2020-09-2	UCKeAhJv	ã€ç"Yæ'»ã	ãŠã™ã™ã	, https://i.y
4	2	youtube#:	NoZg_aPC	youtube#:	koA5LZNs	2020-09-2	UCKeAhJv	ã€Super M	ã,ã,ã,ã	https://i.y
5	3	youtube#:	LWS0svpjl	youtube#:	zRFwdqss	2020-09-2	UCKeAhJv	è³ãã•ã,	ãããŠçYãã,	https://i.y
6	4	youtube#:	IJJus7Afzk	youtube#:	P2GzAIWC	2020-09-2	UCKeAhJv	ã€Super M	ãfzãfãã,	ãã https://i.y
7	5	youtube#:	aCG81BlfF	youtube#:	2hqOuJNC	2020-09-2	UCKeAhJv	ã€ãCEæ™,	ã,cãfzã,	¾ã https://i.y
8	6	youtube#:	wPME31I	youtube#:	clmzXWI	2020-09-2	UCKeAhJv	ã€ãPEX I	ãããã "ããè"	https://i.y

Slika 10 Primer končnega videza Excel datoteke, Vir 8

4.2 Drugi del naloge

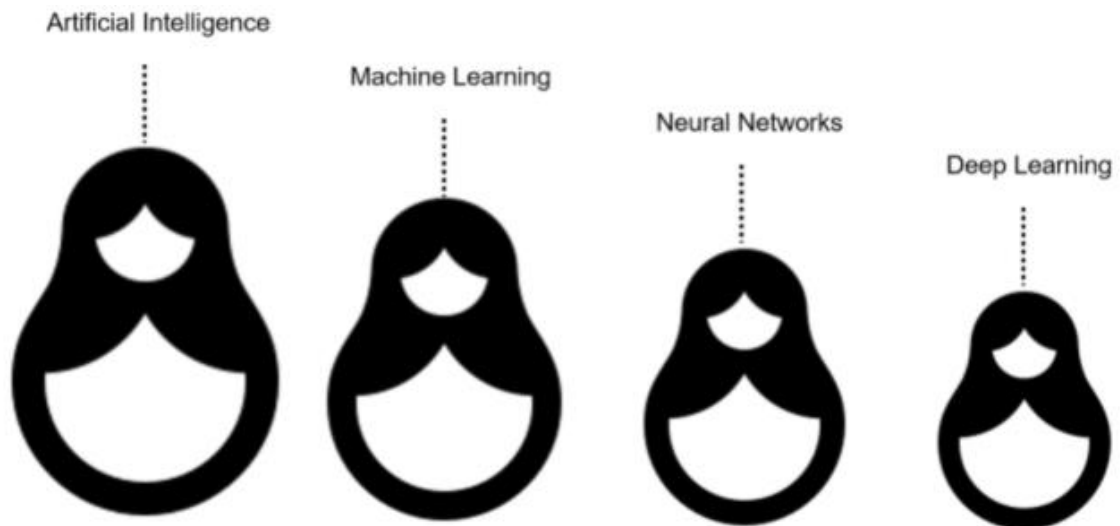
Ko bi uspešno naredili in optimizirali algoritem za pridobivanje le teh informacij bi, lahko prešli na naslednji del raziskovalne naloge, ki je učenje programa s pomočjo UI. Način, kako naučiti program je preprost za razumevanje, in sicer poteka tako, da programu vstaviš Excel datoteko in mu podaš navodila, da za učenje uporabi le 70% podatkov, ostalih 30% pa porabi za testiranje, če se je »naučil pravilno«. Drugi del programa deluje na podlagi nevronske mreže, kamor vstaviš podatke in jih nato program primerja. Ko s primerjanjem konča (se nauči), uporabi ostale podatke za preverjanje natančnosti, ki naj bi bila višja od 60 % po naših hipotezah.

Za svojo raziskovalno nalogo smo uporabili model Linear Regression. To je najbolj standarden model, katerega večino ljudi, ki se prvič soočajo z nevronskimi mrežami. Nima najboljše moči

v predvidevanju podatkov, ampak je relativno ne zakompliciran koncept, ki ga je lažje uporabiti in ga razumeti.

5 UMETNA INTELIGENCA/UI

Umetna inteligenca (angl. Artificial Intelligence) je zmožnost računalnika ali robota, da opravlja naloge, za katere je potrebna človeška inteligenca. UI ima podzvrsti, ki si jih je najlažje predstavljati kot »ruske babuške«, kakor to pove in prikaže Eda Kavlakoglu {Vir 2}.

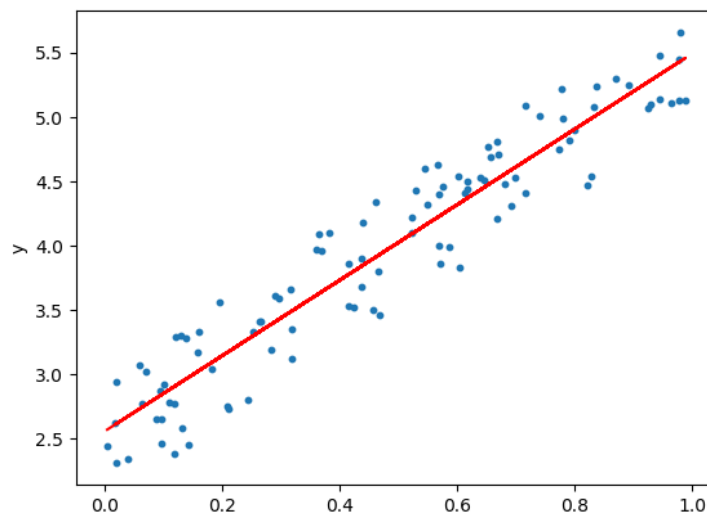


Slika 11: Vizualni prikaz stopenj; vir 1

Kot razvidno na zgornji sliki je UI sestavljen iz treh drugih slojev.

5.1 Računalniški vid

Računalniški vid (angl. Machine learning) je prvi sloj umetne inteligence, ki se uporablja za npr. prepoznavo slike, glasu itd. Narejen je tako, da se lahko izboljšuje in uči iz svojih napak iz preteklosti, ne da bi pri tem pomagal človek. Pri bolj kompleksnih programih pa se uporabljajo zahtevnejše metode, kot so nevronske mreže ali pa globoko učenje (angl. Deep learning). Za ta program smo uporabili algoritem linearne regresije (angl. Linear Regression), ki je manj natančna, ampak je veliko lažja za razumevanje in uporabo {Vir 10}.



Slika 12: graf linearne regresije, vir 2

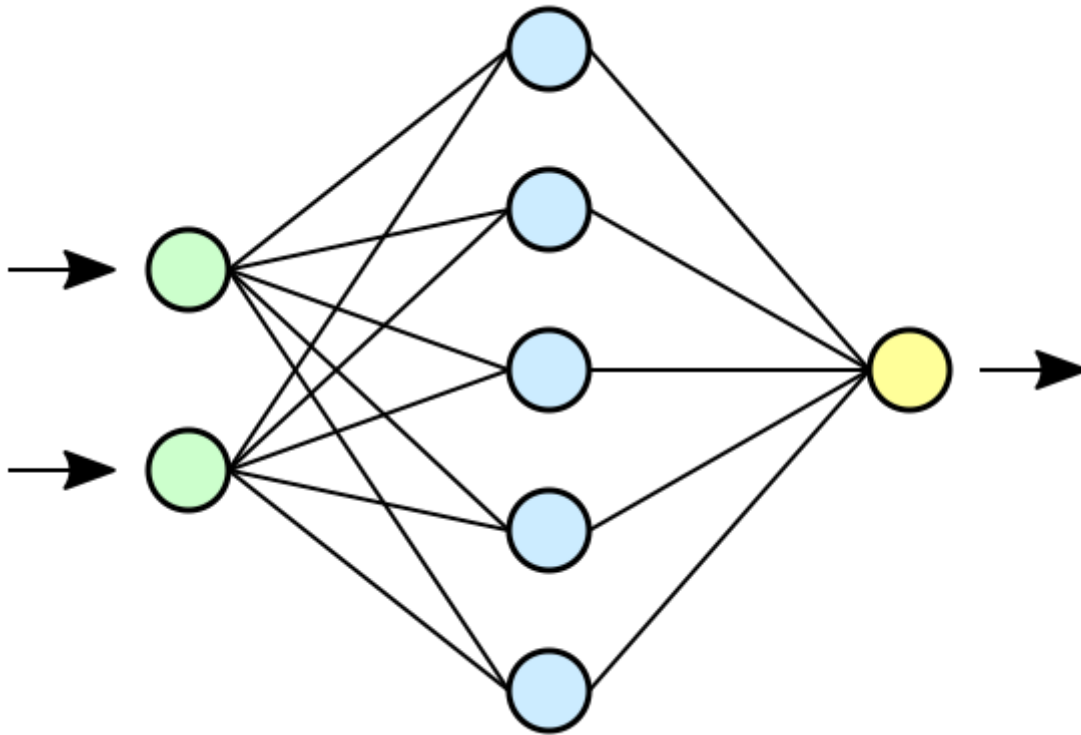
Matematična formula za linearno regresijo je tudi zelo preprosta. Enačba ima obliko $Y = a + b \cdot X$, kjer je Y odvisna spremenljivka (to je spremenljivka, ki gre na os Y), X je neodvisna spremenljivka (tj. narisana je na osi X), b je naklon črta in a je y-presek.

$$a = \frac{(\sum y)(\sum x^2) - (\sum x)(\sum xy)}{n(\sum x^2) - (\sum x)^2}$$
$$b = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2}$$

Enačba 1: Linearna regresija, Vir 1

5.2 Nevronske mreže

Nevronska mreža (angl. Neural network) je način obdelave informacij, ki deluje na podoben princip kot človeški možgani, ko razmišljamo. Delujejo tako, da najprej dobijo podatke s katerimi lahko delajo in nato jih preverijo ter izpišejo svoje ugotovitve. Iz tega izhaja izraz umetna inteligenca. Prvi pomisleki glede nevronskih mrež so nastali že leta 1943, vendar so se v svetu tehnologije začele razvijati šele po letu 1958, ko so predstavili prvi računalnik imenovan »Perceptron«, ki naj bi predstavljal, kako se človeški možgani učijo prepoznavati objekte, ki jih vidijo. Današnje dni se uporabljajo nevrnske mreže skoraj v vsaki aplikaciji (YouTube, Tiktok...) {Vir 11}.



Slika 13: Nevronske mreže, Vir 4

$$f \left(b + \sum_{i=1}^n x_i w_i \right)$$

Enačba 2: nevronske mreže, Vir 2

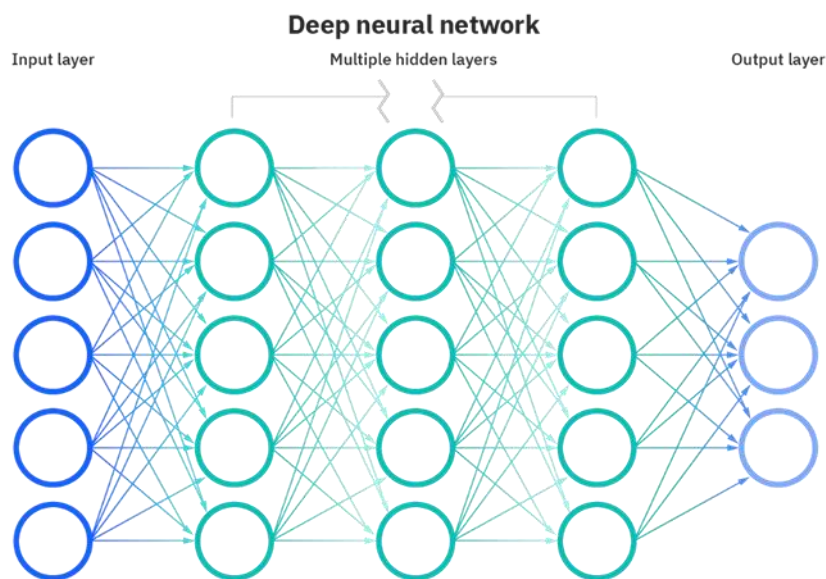
Za navedeno enačbo velja, da je:

- b = pristranskost (angl. bias)
- x = vnos (angl. input)

- w = uteži (angl. weights)
- n = število vnosov
- i = števec, ki se povečuje od 1 do n

5.3 Globoko učenje

Globoko učenje (angl. Deep learning) je sloj UI, ki se uporablja za najzahtevnejše programe z umetno inteligenco, kot npr. avtomobili, ki ne potrebujejo voznika, da se varno vozijo po cestah z ostalimi vozniki na cesti. Pri takšnih primerih je zelo pomembna natančnost izhodne plasti, in globoko učenje ima zelo veliko natančnost. Sestavljeno je iz vhodne plasti, ogromno skritih plasti, kjer so podatki obdelani večkrat, da je rezultat kar se da natančen. Ko so podatki obdelani, pa je na koncu še izhodna plast, ki izpiše rezultat obdelanih podatkov. Kar mnogo ljudi ne opazi je, da je globoko učenje navadna nevronska mreža, vendar je sestavljena iz mnogo več skritih plasti in je zato natančnejša {Vir 12}.



Slika 14: Globoko učenje, Vir 3

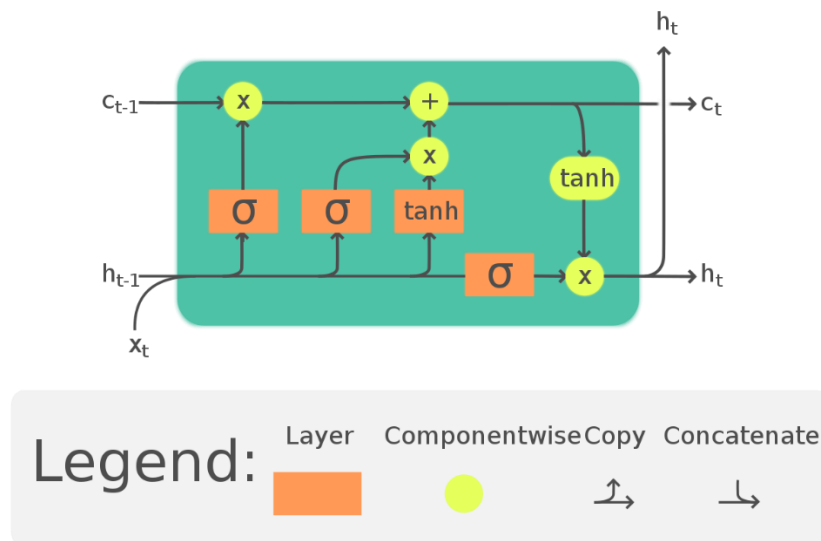
5.4 Popularne nevronske mreže

Dandanes je veliko različnih nevronskih mrež, ki so odprte za javno uporabo. Predstavljamo nekaj nevronskih mrež, ki bi lahko bile uporabljene v tej raziskovalni nalogi in bi imele enake ali boljše rezultate kot trenutna naloga .

Long short-term memory(LSTM)

Je arhitektura umetne ponavljajoče se nevronske mreže (RNN). Uporablja se na področju globokega učenja. Obdeluje lahko ne samo posamezne podatkovne točke (kot so slike), temveč tudi celotne zaporedje podatkov (kot je govor ali video). LSTM je na primer uporaben za naloge, kot so ne segmentirano, povezano prepoznavanje rokopisa, prepoznavanje govora in odkrivanje anomalij v omrežnem prometu.

Zgradba: Skupna enota LSTM je sestavljena iz celice, vhodnih vrat, izhodnih vrat in vrat za pozabo. Celica si zapomni vrednosti v poljubnih časovnih intervalih, tri vrata pa uravnava pretok informacij v celico in iz nje {Vir 13}.

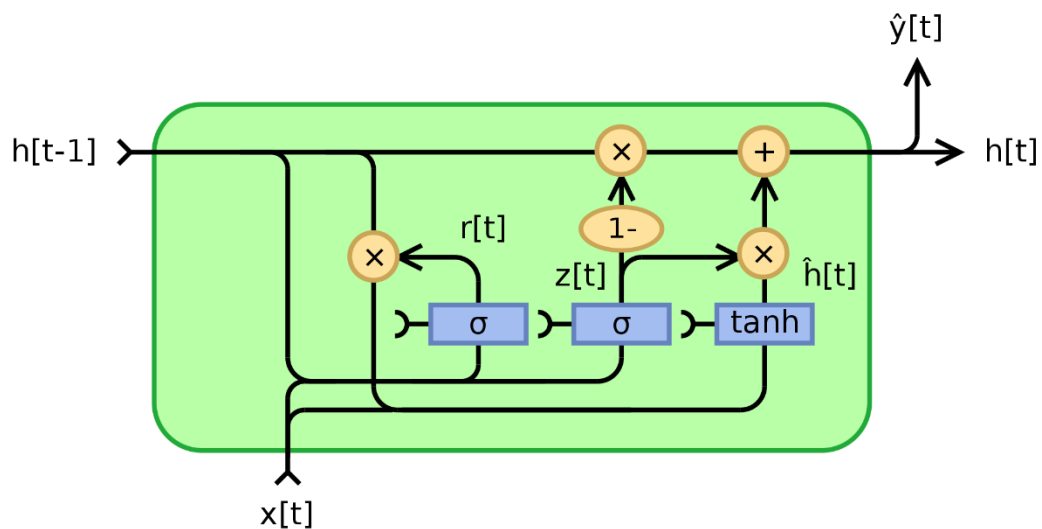


Slika 15: celica LSTM, vir 15

Gated recurrent unit(GRU)

GRU je kot dolgotrajni kratkoročni pomnilnik (LSTM) z vrati za izhod. Ima manj parametrov kot LSTM, saj nima izhodnih vrat. Ugotovljeno je bilo, da je delovanje GRU pri določenih nalogah modeliranja večglasne glasbe, modeliranja govornih signalov in obdelave naravnega jezika podobno kot pri LSTM. Pokazalo se je, da GRU-ji kažejo boljšo učinkovitost pri nekaterih manjših in manj pogostih nizih podatkov {Vir 14}

Zgradba:



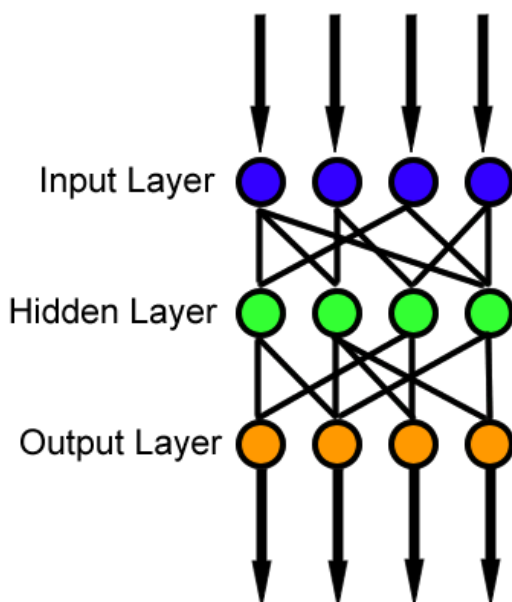
Slika 16: celica GRU, vir 16

Feedforward neural network(FNN)

V tej nevronske mreži povezave med nevroni ne tvorijo zanke. Kot taka se razlikuje od svojih potomcev, ponavljajočih se nevronske mreže.

Bila je prva in najpreprostejša vrsta umetne nevronske mreže. V tem omrežju se informacije premikajo samo v eni smeri – naprej – od vhodnih nevronov, preko skritih nevronov (če obstajajo) in do izhodnih vozlišč. V omrežju ni ciklov ali zank. Njena moč večinoma izhaja iz tega, koliko skritih nevronov je v mreži {Vir 15}.

Zgradba:



Slika 17: zgradba FFN, vir 17

6 PRAKTIČNI DEL

Po teoretičnem delu se končno lahko prestavimo na praktični del, kako naj bi program deloval in kako nam je uspelo, da deluje. Med izdelavo smo šli čez dva različna načina, kako bi program deloval in drugi način, ki ga bomo opisali, se je izkazal za boljšega in bolj natančnega. Za izdelavo celotne raziskovalne naloge smo porabili okoli 5 mesecev, za praktični del pa 4 mesce od tega.

Za pomoč smo prebrskali splet, da bi našli nam potrebne podatke za izdelavo raziskovalne naloge, vendar nismo našli nič nam koristnega, ker se je izkazalo, da je bila naša ideja dokaj originalna.

7 PROCES DELOVANJA

7.1 Pridobivanje podatkov

Kot že omenjeno smo podatke pridobili s knjižnico twint. Twint je knjižnica narejena za programski jezik Python in je namenjena pridobivanju podatkov iz socialnega omrežja Twitter. Pridobi lahko tvite o neki temi s samo 5 vrsticami kode. Najprej navedemo ključno besedo/besede ali temo, ki jo mora twint vsebovati, nato izberemo, koliko tvitov zabeleži in pa za koliko časa nazaj. Ko zaženemo program, nam ta v terminal izpiše vse podatke o tvitih, ki smo jih želeli. Ko dobimo podatke iz Twitterja dobimo se podatke za trenutno Bitcoin ceno s pomočjo knjižnice Pycoingecko, ki nam to omogoča.

```
# twint config
c = tw.Config()
c.Store_csv = True
c.Custom_csv = ["date", "time", "username"]
c.Output = "twitter.csv"
c.Search = "bitcoin"
c.Lang = "en"
c.Limit = 150
```

Slika 18: Uporaba knjižnice twint, lasten vir

7.2 Filtriranje podatkov

Preden podatke shranimo v datoteko jih je treba prefiltrirati, da shranimo samo uporabne informacije. Za filtriranje podatkov smo uporabili knjižnico re (regular expressions). Re je koristna knjižnica, ker ti omogoča pridobivanje samo željene informacije iz daljšega niza znakov (angl. string). Iz dobljenega niza znakov odstranimo vse #, smeškote in ostale nepotrebne stvari, kot so avtor tvita ipd.

```
emoji_pattern = re.compile("[  
    U" \U0001F600-\U0001F64F"  
    U" \U0001F300-\U0001F5FF"  
    U" \U0001F680-\U0001F6FF"  
    U" \U0001F1E0-\U0001F1FF"  
    U" \U00002500-\U00002BEF"  
    U" \U00002702-\U000027B0"  
    U" \U00002702-\U000027B0"  
    U" \U000024C2-\U0001F251"  
    U" \U0001f926-\U0001f937"  
    U" \U00010000-\U0010ffff"  
    U" \u2640-\u2642"  
    U" \u2600-\u2B55"  
    U" \u200d"  
    U" \u23cf"  
    U" \u23e9"  
    U" \u231a"  
    U" \ufe0f" # dingbats  
    U" \u3030"  
"]+", flags=re.UNICODE)
```

Slika 19: Filter za smeške, lasten vir

```
def cleaner(tweet):  
    tweet = re.sub("@[A-Za-z0-9]+", "", tweet)  
    tweet = re.sub("(?:@|http?:://|https?:://|www)\S+", "", tweet)  
    tweet = " ".join(tweet.split())  
    tweet = emoji_pattern.sub('', tweet)  
    tweet = tweet.replace("#", "").replace("_", " ")  
    return tweet
```

Slika 20: Filter za povezave in #, lasten vir

7.3 Shranjevanje podatkov

Za shranjevanje vseh sedaj urejenih tvitov smo uporabili knjižnico pandas, katere namen je urejanje in shranjevanje Excel datotek s pomočjo Pythona. Da bi podatke shranili v Excel datoteko smo morali najprej ustvariti prazno mrežo in nato vanjo dodati podatke. Ko imamo virtualno mrežo s podatki jo lahko shranimo v Excel datoteko.

	A	B	C	D	E	F	G	H
1	key	datetime	pos	neg	neu	tweet_nur	currentprice	newprice
2	1	2022/3/2 -	17.897	9.736	100.366	128	44123	44147
3	2	2022/3/2 -	16.507	8.559	84.933	110	44216	44154
4	3	2022/3/2 -	14.523	9.088	89.395	113	44151	44129
5	4	2022/3/2 -	18.617	10.277	79.1	108	44129	44096
6	5	2022/3/2 -	16.543	5.443	84.015	106	44096	44204
7	6	2022/3/2 -	14.124	7.264	76.609	98	44204	44214
8	7	2022/3/2 -	16.127	9.767	85.105	111	44214	44164
9	8	2022/3/2 -	14.537	6.718	64.746	86	44164	44159
10	9	2022/3/2 -	24.651	12.443	126.911	164	44159	44215
11	10	2022/3/2 -	19.139	6.883	75.979	102	44215	44238
12	11	2022/3/2 -	16.628	8.611	76.755	102	44238	44282
13	12	2022/3/2 -	16.85	10.871	105.278	133	44282	44074
14	13	2022/3/2 -	14.741	8.842	91.418	115	44074	44000
15	14	2022/3/2 -	22.666	9.765	94.566	127	44000	43952
16	15	2022/3/2 -	18.298	7.882	98.819	126	43952	43930
17	16	2022/3/2 -	18.703	7.821	91.475	119	43930	43984
18	17	2022/3/2 -	25.004	12.501	116.492	155	43984	43906
19	18	2022/3/2 -	21.972	8.462	100.565	131	43906	44115
20	19	2022/3/2 -	19.103	8.726	88.169	117	44071	44060
21	20	2022/3/2 -	25.321	15.76	137.917	180	44060	44039

Slika 21: Končni videz Excel datoteke, lasten vir

7.4 Posodabljanje podatkov

Za večjo ali kakšno koli natančnost programa moramo seveda podatke sproti posodabljati, da program ostane na tekočem ter je zmožen pravi čas sporočiti, ali bo cena padla ali narasla. Kako program na novo naučimo je, da se naprej prvi del (pridobivanje podatkov) zakroži in pridobi nove podatke, ki jih nato vstavi v drugi del (predvidevanje cene). Novo pridobljene podatke zapiše v novo vrsto in tako se večja datoteka. Dobra stvar tega je, da večja, kot je datoteka, lažje je programu predvideti ceno kriptovalute

7.5 Obdelava podatkov in predvidevanje

Obdelava podatkov in predvidevanje cene teče v drugem delu programa. V nalogi smo uporabili Python knjižnico sklearn. S pomočjo te knjižnice lahko uvozimo Linearno Regresijo ter njihov način selekcioniranja podatkov, da jih lahko nevronskega modela obdeluje. Prikazano v spodnji sliki vidimo, da naša koda prebere datoteko, v kateri so vsi podatki katere smo zbrali ter jih postavi v vsake v svojo tabelo. Nato uporabi uvoženo komando `train_test_split`, da selekcionira podatke na način ki je berljiv za model Linearno Regresijo. Kličemo komando `regr.fit()`, ki nauči naš model. Za tem pa še kličemo `regr.predict()` in `regr.score()`, da dobimo rezultate katere izvozimo v csv datoteko za nadaljnjo obdelavo.

```
data = pd.read_csv("final.csv")

train = data.filter(items=['pos', 'neg', 'neu', 'tweet_num', 'currentprice'], axis=1)
test = data['newprice']

X_train, X_test, y_train, y_test = train_test_split(train, test, test_size=0.3, random_state=2)

regr = LinearRegression()

regr.fit(X_train, y_train)

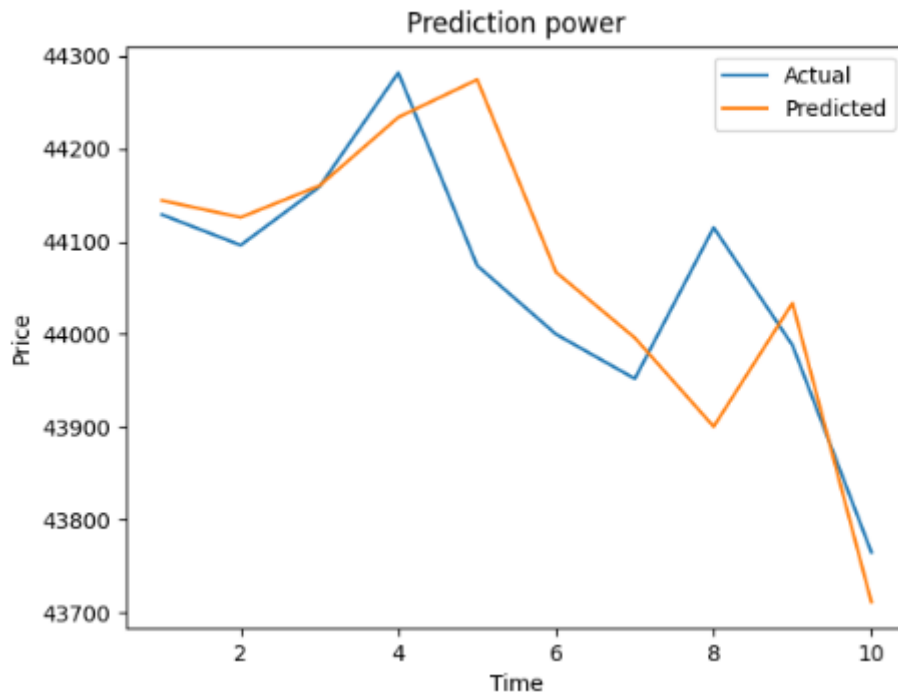
pred = regr.predict(X_test)

print(regr.score(X_test, y_test))
```

Slika 22: Učenje programa, lasten vir

7.6 Rezultati

Končna moč našega modela je 54,6 %. Z daljšim časom pridobivanja podatkov in s eksperimentacijo različnih modelov nevronske mreže bi lahko zvišali moč naše mreže. Kot različne faktorje bi lahko tudi testirali različne knjižnice za vrednotenje tvitov, razna druga socialna omrežja, ter tudi načine filtriranja podatkov. Spodnji graf je reprezentacija rezultatov v vizualni obliki.



Slika 23: Graf rezultatov, lasten vir

8 NAPAKE IN ZAPLETI

Seveda raziskovalna naloga ni potekala brez zapletov ali napak programa. Za vsako napako smo si vzeli čas in jo podrobno preverili.

8.1 Pridobivanje podatkov

Najdaljši del raziskovalne naloge je bil pridobivanje podatkov, ker je knjižnica twint malo zastarela in ne deluje najbolje z novimi verzijami Pythona. Problem je tudi v tem, da se Twitterjev API nenehno posodablja in knjižnica ni posodobljena zraven.

- Could not find the Guest token in HTML:

Ta napaka se je pojavila, ko smo prvič začeli pridobivati tvite s knjižnico twint. Napaka je nastala, ker je twint pošiljal Twitterju napačen simbol (angl. token). Ko smo z raziskovalno nalogo začeli, na spletu še ni bilo veliko znano, kako problem rešiti, zato je izdelava programa morala počakati kakšen teden, ko je nekdo na spletu objavil rešitev. Problem smo rešili tako, da smo inštalirali twint iz kopije kode za to knjižnico, ki je delovala.

- CRITICAL:root:twint.run:Twint:Feed:noDataExpecting value:

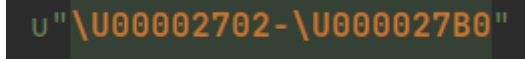
Napaka se je pojavila, ko je imel Twitter posodobitev za svoj API iz katerega se dobijo tviti. Kot že omenjeno je twint malo zastarela knjižnica in zato se je ta napaka pojavila. Rešitev je bila preprosta, in sicer, da knjižnice nisi smel naložiti preko »PyCharm import library« ampak si ga moral naložiti s komando v konzoli.

8.2 Filtriranje podatkov

Pri samem filtriranju podatkov sicer ni bilo napak v kodi, je pa nastal zaplet, ki nam je vzel kar nekaj časa, da smo poiskali rešitev.

- Brisanje smeškov iz tvitov:

Kako smo brisali # iz Tvitov je tako, da smo uporabili knjižnico re in vnesli znake, ki jih ne želimo. S smeški pa nastane problem, ker nimajo svojega posebnega znaka in smo morali uporabiti unicode za vse smeške, ki so prisotni v tvitu.



Slika 24: Smeško unicode primer, lasten vir

8.3 Shranjevanje podatkov

Shranjevanje podatkov nam je pri raziskovalni nalogi povzročalo najmanj problemov. Nastal je le en zaplet in tudi ta zaplet smo hitro rešili.

- Brisanje obstoječih podatkov:

Problem je nastal na začetku shranjevanja podatkov v Excel datoteko. Shranjevanje je delovalo normalno, ampak je program vedno prvo vrstico prepisal z novimi podatki in nikoli shranil. Problem smo rešili tako, da smo definicijo nove datoteke zapisala izven zanke.

8.4 Obdelava podatkov in predvidevanje

Največja težava je nastala pri predvidevanju cene in natančnosti programa, ki je najprej bila 99 %. To se sicer sliši zelo dobro, vendar problem je, da je cena Bitcoina zelo velika in zato če je program predvideval ceno za 50 € več ali manj, se je vseeno štelo kot dobro predvidevanje. To smo popravili, tako da smo z uporabo knjižnice pandas izračunali razliko med trenutno ceno in predvideno ceno ter nato naučiti program, kakšna razlika je prevelika in kakšna je sprejemljiva.

9 ZAKLJUČEK

9.1 Povzetek

V današnjem svetu je umetna inteligenca uporabljena povsod. Eden izmed večjih primerov so razni algoritmi, ki določajo, kaj nam je predlagano na socialnih omrežjih. Z novimi napredki v umetni inteligenci smo se odločili da tudi mi naredimo raziskave na tem in da pomagamo napredku, tudi če samo malo na področju umetne inteligence. V raziskovalni nalogi je bil velik poudarek na pridobivanje podatkov za nevronske mreže, saj je to najpomembnejši del. To je tudi razlog, zakaj vsako podjetje zbira naše podatke in jih trži, za napredek lastnih algoritmov. Ugotovili smo veliko, še posebej ko se zgodi kakšna napaka. Opazili smo, da smo se največ naučili iz ovir na naši poti do končnega rezultata. Uspelo nam je pridobiti obetavne rezultate, ki bi se lahko še izboljšali.

9.2 Zahvala

Posebna zahvala gre najinima mentorjema Roku Urbanu in Islamu Mušiću za odlično idejo, pomoč pri raziskovalni nalogi ter za jeklene živce, ki verjamem da so bili potrebni, ko se je kje zataknilo. Brez njiju ta raziskovalna ne bi bila mogoča in tudi ne bi nastala. Od te raziskovalne naloge sva odnesla veliko znanja in novih izkušenj, zato še enkrat hvala.

10 VIRI IN LITERATURA

Viri, ki niso navedeni, so ustvarjeni iz naše strani.

10.1 Viri besedil

Vir 1: <https://towardsdatascience.com/how-to-scrape-tweets-from-twitter-59287e20f0f1>

Vir2:<https://www.ibm.com/cloud/blog/ai-vs-machine-learning-vs-deep-learning-vs-neural-networks>

Vir 3: <https://unicsoft.com/portfolio/artificial-intelligence-for-cryptocurrency-price-prediction/>

Vir 4: <https://www.sciencedirect.com/science/article/pii/S2405918821000027>

Vir 5: <https://www.python.org/doc/essays/blurb/>

Vir 6: <https://computer.howstuffworks.com/internet/social-networking/networks/twitter1.htm>

Vir 7: <https://en.wikipedia.org/wiki/PyCharm>

Vir 8: <https://docs.python.org/3/reference/import.html>

Vir 9: <https://www.lifewire.com/what-exactly-is-twitter-2483331>

Vir 10: https://en.wikipedia.org/wiki/Machine_learning

Vir 11: https://en.wikipedia.org/wiki/Artificial_neural_network

Vir 12: https://en.wikipedia.org/wiki/Deep_learning

Vir 13: https://en.wikipedia.org/wiki/Long_short-term_memory

Vir 14: https://en.wikipedia.org/wiki/Gated_recurrent_unit

Vir 15: https://en.wikipedia.org/wiki/Feedforward_neural_network

10.2 Viri slik

Vir 1: <https://www.ibm.com/cloud/blog/ai-vs-machine-learning-vs-deep-learning-vs-neural-networks>

Vir 2: <https://medium.com/@shafaqshaikh10/linear-regression-8628af6861cc>

Vir 3: <https://www.ibm.com/cloud/learn/neural-networks>

Vir 4: https://commons.wikimedia.org/wiki/File:Neural_network.svg

Vir 5: [https://en.wikipedia.org/wiki/Python_\(programming_language\)](https://en.wikipedia.org/wiki/Python_(programming_language))

Vir 6: <https://www.jetbrains.com/pycharm>

Vir 7: <https://commons.wikimedia.org/wiki/File:Twitter-logo.svg>

Vir 8: <https://twitter.com/domossu/status/1308706533711073281/photo/1>

Vir 9: <https://github.com/twintproject>

Vir 10: <https://pandas.pydata.org/about/citing.html>

Vir 11: <https://www.javatpoint.com/regex>

Vir12: <https://medium.com/analytics-vidhya/sentiment-analysis-with-vader-label-the-unlabeled-data-8dd785225166>

Vir 13: <https://www.coingecko.com/en/branding>

Vir 14: https://commons.wikimedia.org/wiki/File:Scikit_learn_logo_small.svg

Vir 15: https://en.wikipedia.org/wiki/Long_short-term_memory

Vir 16: https://en.wikipedia.org/wiki/Gated_recurrent_unit

Vir 17: https://en.wikipedia.org/wiki/Feedforward_neural_network

10.3 Viri enačb

Vir 1: https://www.statisticshowto.com/probability-and-statistics/regression-analysis/find-a-linear-regression-equation/#google_vignette

Vir 2: <https://medium.com/coinmonks/the-mathematics-of-neural-network-60a112dd3e05>

11 AVTORJA RAZISKOVALNE NALOGE

Nik Krnjovšek in Bernard Kužnik sta dijaka 3. letnika Elektro in računalniške šole Velenje. Za raziskovalno nalogo sta se odločila, ker oba zanima področje umetne inteligence. Imata zanimanje tudi za izdelavo aplikacij in upravljanje s podatkovnimi bazami. Veselita se, da sta imela priložnost izdelovanja raziskovalne naloge s pomočjo čudovitih mentorjev. V prihodnosti se želita ukvarjati s programiranjem in umetno inteligenco. Spodnja slika prikazuje avtorja raziskovalne naloge.



Slika 25: Slika avtorjev, lastni vir