

Upoštevanje indeksa strahu in pohlepa v predvidevanju cene Bitcoin skozi dolgi kratkoročni spomin

Nataša Ošep Ferš, Aleš Zamuda, SMIEE

Univerza v Mariboru, Fakulteta za elektrotehniko, računalništvo in informatiko,
Koroška cesta 46, 2000 Maribor, Slovenija

E-pošta: natasa.osep@um.si, ales.zamuda@um.si

Considering a Fear and Greed Index in Bitcoin Price Prediction Through Long Short-Term Memory

This paper represents a methodology for cryptocurrency Bitcoin market price prediction, composed of technical analysis and fundamental analysis using sentimental analysis. Specifically, Crypto Fear & Greed Index (CFGFI) is considered in the methodology, which is made up of indicators of all analyses. Neural network with Long Short-Term Memory (LSTM) model is trained on these indicators. Prediction on data used is reported when using index CFGFI (improved version), compared against the prediction without this index (basic version).

1 Uvod

Kriptovalute so dandanes splošno znane in privlačne naložbe za marsikaterega vlagatelja [1]. Pred napovedovanjem njihovih vrednosti je treba premisliti, ali so te klasificirane kot valute, sredstva ali naložbe [2]. V tem prispevku je kriptovaluta Bitcoin [3] obravnavana kot naložba v dolar (BTCUSD) in zato se da potegniti vzporednico med njo in vrednostnimi papirji (VP). Ideja izvedbe tega prispevka temelji na izhodiščih [2] in pristopu [4] ter spada v operacijske raziskave [5]. Podobno kot v [4], je tudi v tem prispevku napoved prihodne cene BTCUSD osnovana na tehnični (računalniški) in temeljni analizi. Poleg že uporabljenih indeksov je v tem prispevku na novo vključen še indeks CFGFI. Napoved temelji na Hipotezi prilagodljivega trga AMH (ang. *Adaptive Market Hypothesis*), ki predstavlja alternativo Hipotezi učinkovitega trga EMH (ang. *Efficient Market Hypothesis*). Osnove EMH je v svojem delu [6] leta 1970 postavil E. F. Fama. Hipoteza trdi, da so vse tržne informacije vkalkulirane v cene VP. Tako cene odražajo poštene (notranje, prave) vrednosti VP. Z aktivnim upravljanjem premoženja, trga torej ni mogoče premagati, saj le-ta nikoli ni podcenjen, niti precenjen. V njegovem poznejšem delu [7] Fama poudarja, da je treba preučiti vsak scenarij posebej in se odločiti, kdaj je učinkovitost trga dovolj dobra aproksimacija za trg in kdaj je boljše uporabiti kakšni drugi model poenostavitve.

Leta 2004 je Lo v [8] predstavil AMH. Predlagal je novi okvir, ki na podlagi evolucije - prilagoditve in naravne selekcije dokazuje interakcijo med tržno učinkovitostjo in vedenjskimi financami. Po AMH trenutne cene

odražajo toliko informacij, kot to narekuje spreminjajoče se okolje.

V nadaljevanju tega prispevka so v razdelku 2 predstavljena sorodna dela. V razdelku 3 je podana tehnična metodologija napovedovanja, osnovana na psevdokodi, ki je implementirana na modelu nevronske mreže. V razdelku 4 je podan eksperiment, kjer je v fazi predprocesiranja opisana močna pozitivna korelacija med količino tvitov in obsegom trgovanja, kar je potrjeno tudi z eksperimentom. Več o modelu in podatkih je predstavljenega v podrazdelku 4.1. Rezultati eksperimenta so podrobneje predstavljeni v podrazdelku 4.2. V razdelku 5 je podan zaključek, nato sledi še literatura.

2 Sorodna dela

V [4] avtorji združijo tehnično in temeljno analizo skozi aplikacije podatkovnih znanosti in tehnike stojnega učenja. Za razliko od naše raziskave, napovejo trend gibanja cen delnic (20 največjih podjetij po tržni kapitalizaciji iz indeksa NASDAQ100). Problem napovedi delniškega trga pretvorijo v nalogo klasifikacije časovnih vrst. Na podlagi informacij, ki so pridobljene iz časovnih vrst cen in sentimenta, z nevronske mreže z veriženjem naprej ustvarijo robusten model, ki je sposoben napovedati prihodni tržni trend. S simulacijo dosežejo več kot 80% letnega donosa.

V članku [9] avtorji zgradijo in implementirajo več modelov nadzorovanega učenja, da bi raziskali različne vidike, ki vplivajo na stabilnost in dnevno gibanje cen kriptovalut. Uporabijo celostni pristop, ki vključuje vse razpoložljive informacije: finančne informacije, informacije na podlagi trgovnega obsega (aktivnosti) v verigi blokov in metrike razvoja programske opreme na osnovi GitHub aktivnosti. Njihovi cilji so klasifikacija določene kriptovalute upoštevanje stabilnost in ceno, napoved dnevne tendence s pomočjo regresije ter opredelitev najustreznejših kriterijev takšne analize [9]. Primerjajo različne nevronske mreže za različne kriptovalute: preproste mreže z veriženjem naprej, povratne mreže RNN (ang. *Recurrent Neural Networks*) skupaj z njihovimi izboljšavami, kot so povratne mreže z dolgim kratkoročnim spominom LSTM (ang. *Neural network with LongShort-Term Memory*) in povratne enote z vrati. LSTM je učinkovita, na gradientu osnovana metoda, ki je bila razvita kot rešitev problema časovne zahtevnosti učenja RNN [10]. Pri

merjava [9] pokaže tudi, da največ obetajo RNN.

V članku [11] je predstavljena metoda za predvidevanje sprememb dolarških cen Bitcoin (BTCUSD) in Ethereum (ETHUSD). Avtorji kot vhode v linearni regresijski model vključijo analizo sentimenta, Google Trends in količino tvitov. Z analizo tvitov ugotavljajo, da smer cen bolj kot tviter sentiment, narekuje količina tvitov. Nadalje ugotavljajo, da z vključitvijo Google Trends lahko natančno predvidijo smer spremembe cen. Analiza sentimenta tvitov je narejena s Twitter API in Pythonovo knjižnico Tweepy. Za ekstrakcijo in meritve subjektivnih čustev oziroma mnenja je uporabljen VADER (ang. *Valence Aware Dictionary and sEntiment Reasoner*). Na podlagi analize sentimenta je opredeljeno, ali so tviti v splošnem pozitivni ali negativni in ustvarjena je dnevna lestvica sentimenta, ki jo avtorji potem primerjajo s spremembo dnevne cene. Ugotovljajo, da analiza sentimenta ni konsistentna s spremembo cen pri padajočem trendu in zato je ne vključijo v model. Analizo sentimenta uporabijo še za ločevanje tvitov, ki izražajo neko mnenje (subjektivnost) in tvitov, ki podajajo samo informacije brez opredelitve mnenja (objektivnost). Google Trends prikazuje priljubljenost iskanja določene teme relativno glede na iskanje drugih tem. Da se ga uporabiti kot posredno mero, ki kaže povezavo med zanimanjem za kriptovalute in gibanjem njihovih cen. Avtorji pokažejo močno pozitivno korelacijo med omenjenima spremenljivkama ter med količino tvitov in ceno.

V članku [12] avtorji raziskujejo povezavo med donosnostjo in volatiliteto ter tviti na podlagi minutnih cen in tvitov 30 delnic iz Dow Jones Industrial Average indeksa za 13 dni. V študiji ugotavljajo, da je skoraj 90% donosa delnic ustvarjenega na podlagi tvitov, da pa ne velja obratno in manj kot 10% tvitov izhaja iz donosa delnic.

Alternative.me je platforma za optimizacijo povezav med različicami programske opreme in alternativnimi izdelki (<https://alternative.me>). Na <https://alternative.me/crypto/fear-and-greed-index> se vsak dan analizirajo čustva in mnenja ljudi v zvezi z Bitcoinom iz različnih virov, ki se potem akumulirajo v CFGI. To je preprosta številka med 0 in 100, pri čemer 0 pomeni ekstremni strah, 100 pa ekstremni pohlep. Ekstremni strah je nakupna priložnost, saj je le-ta lahko znak, da so vlagatelji preveč zaskrbljeni. Nasprotno je izjemni pohlep lahko prodajna priložnost, saj je pohlep znak za popravek na trgu. Indeks CFGI sestavljajo naslednji dejavniki: volatiliteta (25%), tržni zagon / obseg (25%), družabna omrežja (15%), tržna prevlada (10%), trendi (10%) in ankete (15%).

2.1 LSTM

LSTM je ena izmed oblik RNN. Uporabljena je zato, ker se dobro izkaže pri različnih učnih problemih s sekvencijskimi podatki, prav tako pa tudi pri dolgoročnih časovnih odvisnostih [13] kot izboljšava za osnovno RNN. RNN imajo verige ponavljajočih se modulov, ki so sestavljeni iz vhodne in izhodne plasti ter ene ali več vmesnih plasti [14]. V običajni RNN ima ponavljajoči se modul preprosto strukturo, kot je ena tangens plast. Ponavljajoči se modul pri LSTM je spominska celica [15], zgrajena iz 4 skritih (pod)plasti: tangens plast in 3 sigmoidne pla-

sti. Na vsaki sigmoidni plasti se nahajajo vrata, ki celici omogočajo optimalno prepuščanje oziroma odstranjevanje podatkov [16]. V spominski celici plasti LSTM se uporabijo operacije, ki se jih da predstaviti z naslednjimi enačbami [17]:

$$\mathbf{f}_t = \sigma(\mathbf{W}_f[\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_f), \quad (1)$$

$$\mathbf{i}_t = \sigma(\mathbf{W}_i[\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_i), \quad (2)$$

$$\tilde{\mathbf{C}}_t = \tanh(\mathbf{W}_C[\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_C), \quad (3)$$

$$\mathbf{C}_t = \mathbf{f}_t \mathbf{C}_{t-1} + \mathbf{i}_t \tilde{\mathbf{C}}_t, \quad (4)$$

$$\mathbf{o}_t = \sigma(\mathbf{W}_o[\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_o), \quad (5)$$

$$\mathbf{h}_t = \mathbf{o}_t \tanh(\mathbf{C}_t), \quad (6)$$

kjer so \mathbf{f}_t vrata, ki pozabljajo, \mathbf{i}_t so vhodna in \mathbf{o}_t izhodna vrata. Vrata vzdržujejo stanje spominske celice \mathbf{C}_t . $\mathbf{W}_f, \mathbf{W}_i, \mathbf{W}_C, \mathbf{W}_o$ so matrike uteži, $\mathbf{b}_f, \mathbf{b}_i, \mathbf{b}_C, \mathbf{b}_o$ so vektorji pristranskosti in σ ter \tanh aktivacijski funkciji. \mathbf{x}_t je vhodni in \mathbf{h}_t izhodni vektor. $\tilde{\mathbf{C}}_t$ označuje kandidate stanja celice v času (epoha) t . Primer analize vpliva parametrov modela LSTM na finančno napoved, kot je število učenj, je podan v [16].

3 Metoda

V tem razdelku je predstavljena psevdokoda za eksperiment napovedi cene BTCUSD z izboljšanim pristopom, ki vključuje indeks CFGI. Iz platforme alternative.me so zajeti podatki o indeksu CFGI in shranjeni kot datoteka CFGI.csv, ki je v ta postopek podana kot vhodni parameter. V eksperiment smo podali še začetno stanje za ključna števila in tako dosegli ponovljivost eksperimentov za sicer neodvisne zagone (za RUNNR od 1 do 50).

Require: technical.csv, sentimental.csv, CFGI.csv

Ensure: napoved cen in ovrednotenje eksperimenta

- 1: preberi vhodne datoteke csv, jih stolpično združi in določi njihov vrstni red
- 2: skaliraj podatke z uporabo predmeta za skaliranje v Python (MinMaxScaler)
- 3: vhodno podatkovno množico loči na prvih 200 zapisov kot učno in ostalih 91 zapisov kot testno množico
- 4: podatke preoblikuj v 3D obliko za nadzorovano učenje (vzorci, časovni koraki, značilke).
- 5: nastavi konfiguracijske parametre: sekvenčni model, plast LSTM s 6 celicami, ena izhodna plast z 1 izhodom, funkcija izgube MAE, delitev podatkov za validacijo 8:2
- 6: **for** $epocs = 50$ and $batchsize = 6$ **do**
- 7: $uči_{ADAM}$, uporabi konfiguracijske parametre
- 8: **end for**
- 9: testiraj naučeni model na testnih podatkih
- 10: inverzno skaliraj napovedane podatke in jih vrni kot podatkovni okvir ter ovrednoti eksperiment

Po postopku v vrsticah od 1 do 4 se podatki predprocesirajo. V vrstici 2 so ti skalirani v obseg [0,1] s pomočjo predmeta MinMaxScaler iz razreda scikit-learn [18]. V vrstici 5 nastavimo konfiguracijske parametre modela. V vrsticah 6–8 se zgradi model, ki se uči skozi 50 ponovitev, v serijah velikosti 6. Pri učenju sta uporabljena

funkcija izgube povprečne absolutne napake MAE (ang. *Mean absolute error*) [19] in optimizator parametrov prilagodljive trenutne ocene ADAM (ang. *Adaptive Moment Estimation*) [20]. Za LSTM se uporabijo enačbe od (1) do (6) iz podrazdelka 2.1. V vrstici 9 se izračuna uspešnost učenja, izvede se testiranje z izvedbo napovedi na testni množici, vrstica 10 pa vrne rezultate.

4 Eksperiment

V naslednjem podrazdelku so najprej predstavljeni podatki in model, sledijo rezultati eksperimenta.

4.1 Podatki in model

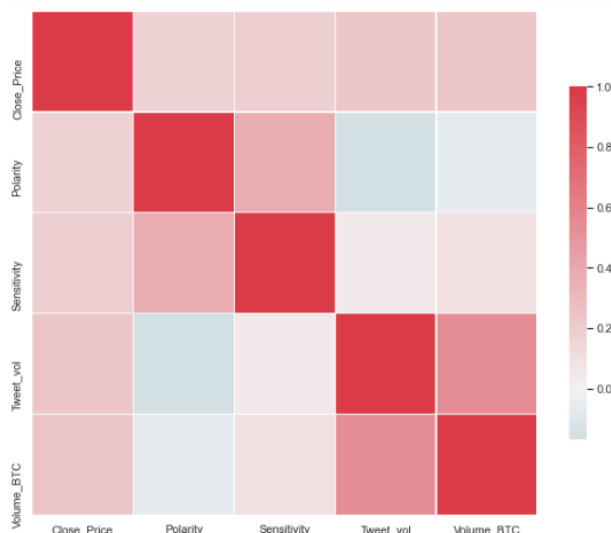
Podatki o ceni in tvitih BTCUSD so pridobljeni s spletne strani Kaggle (<https://www.kaggle.com/paul92s/bitcoin-lstm-model-with-tweet-volume-and-sentiment/data>). Kaggle je podružnica Google LLC in spada med največje skupnosti znanstvenikov iz podatkovnih znanosti ter izvajalcev strojnega učenja. Tam so na voljo številna orodja in viri za uporabo v podatkovni znanosti. V časovno vrsto so zajete minutne/urne serije, v obliki z vejico ločenih podatkov (ang. *Comma Separated Value*), za obdobje 14 dni, od 11.7.2018 do 24.7.2018. Vrednosti indeksa CFGI so pridobljene s pomočjo programskega vmesnika Fear & Greed Index API, v skladu s pogoji uporabe alternative.me. Gre za dnevne podatke, ki se v našem eksperimentu zaradi časovne poravnave z drugimi vhodnimi podatki, 24 ur ne spreminjajo.

Podatki zaključna cena (*Close_Price*), polarnost in senzitivnost indeksa sentimenta (*Polarity*, *Sensitivity*), obseg tvitov (*Tweet_vol*), obseg trgovanja z BTCUSD (*Volume_BTC*) in indeks CFGI (*CFGI*) so uvoženi, pretvorjeni v ure, združeni v eno tabelo in urejeni, kot kaže slika 1. Zaključno ceno BTCUSD za trenutno uro predvidevamo na podlagi preteklih 3 zapisov iz te tabele. Izboljšana različica pristopa vsebuje stolpec z indeksom CFGI, osnovna pa tega stolpca ne vsebuje. CFGI in cena BTCUSD sta pričakovano namreč pozitivno korelirana, dokler indeks ne doseže skrajnih vrednosti in takrat se pričakuje obrat trenda. Programska koda za osnovno različico pristopa je napisana v jeziku Python z uporabo vmesnika Keras in knjižnice TensorFlow in objavljena na spletni strani Kaggle (<https://www.kaggle.com/paul92s/bitcoin-lstm-model-with-tweet-volume-and-sentiment>).

| | Close_Price | Polarity | Sensitivity | Tweet_vol | Volume_BTC | CFGI |
|---------------------|-------------|----------|-------------|-----------|------------|------|
| 2018-07-11 20:00:00 | 6350.00 | 0.102657 | 0.216148 | 4354.0 | 986.73 | 29.0 |
| 2018-07-11 21:00:00 | 6356.48 | 0.098004 | 0.218612 | 4432.0 | 126.46 | 29.0 |
| 2018-07-11 22:00:00 | 6361.93 | 0.096688 | 0.231342 | 3980.0 | 259.10 | 29.0 |
| 2018-07-11 23:00:00 | 6368.78 | 0.103997 | 0.217739 | 3830.0 | 81.54 | 29.0 |
| 2018-07-12 00:00:00 | 6380.00 | 0.094383 | 0.195256 | 3998.0 | 124.55 | 33.0 |

Slika 1: Združeni in urejeni podatki.

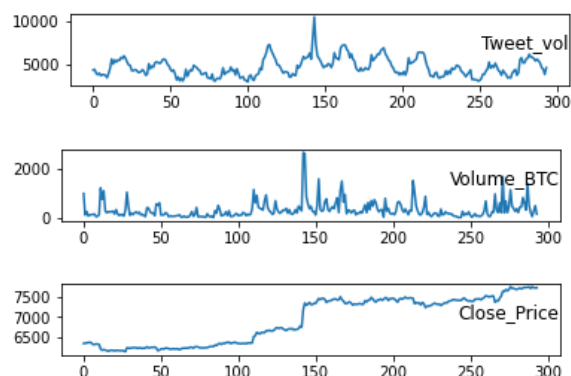
Na sliki 2 so prikazane korelacije med vhodnimi značilkami. Razvidna je močna pozitivna korelacija med količino tvitov in obsegom trgovanja ter zmerne negativna korelacija med količino tvitov in polarnostjo.



Slika 2: Korelacije med značilkami.

4.2 Rezultati eksperimenta

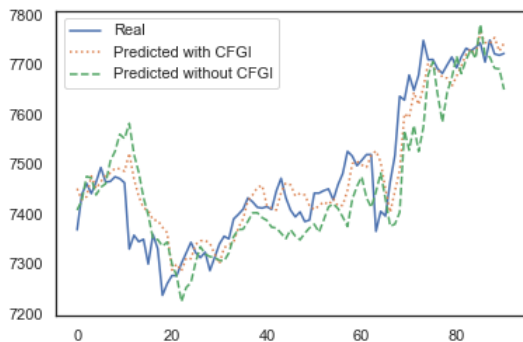
Na sliki 3 je prikaz vizualizacije časovne vrste s količino tvitov, obsegom trgovanja in zaključno ceno BTCUSD.



Slika 3: Vizualizacija časovne vrste.

Za vrednotenje napovedi modela se izračuna kriterijska funkcija napake povprečnega kvadrata RMSE (ang. *Root Mean Square Error*). RMSE predstavlja korenjeno vrednost povprečja kvadratov razlik med napovedano in resnično vrednostjo, pri čemer popolno ujemanje predstavlja vrednost 0.

Iz slike 4 je razvidno, kako uspešno je naš pristop (izboljšana različica) z LSTM napovedal zaključne cene BTCUSD za testno množico, v primerjavi z osnovno različico pristopa in z resničnimi zaključnimi cenami BTCUSD za isto obdobje. Izmerjena vrednost RMSE po RMSE medialnega zagona (izmed 50 ponovitev) osnovne različice za to napovedovanje z LSTM pokaže vrednost 75,59 USD in z indeksom CFGI razširjene različice 74,85 USD. Torej, razlika med medialnima vrednostima RMSE našega izboljšane pristopa in predhodno obstoječega osnovnega pristopa znaša 0,74 USD, v prid predlaganega novega pristopa s CFGI.



Slika 4: Primerjava resničnih (*Real*) in napovedanih cen BTCUSD na podatkih s CFGI (*with*) oz. brez CFGI (*without*).

5 Zaključek

V prispevku je implementirana metoda za predvidevanje zaljučnih cen BTCUSD, ki temelji na hipotezi prilagodljivega trga AMH. Uporabljeni računalniški model v tej metodi je nevronska mreža LSTM z dvema plastema. Vhodni podatki so indeksi tehnične analize in analize sentimenta. S prispevkom je kot vhodni podatek na novo vključen indeks CFGI, ki ga sestavljajo indeksi obeh analiz. V fazi predprocesiranja je ugotovljena močna pozitivna korelacija med količino tвитov in obsegom trgovanja. Izmerili smo vpliv vključitve indeksa CFGI na točnost rezultatov napovedi tako, da smo izračunali napako RMSE za osnovno in izboljšano različico pristopa. Primerjava modelov primerov pokaže, da več obeta izboljšana različica, saj je njena medialna vrednost RMSE boljše (manjša) v primerjavi z osnovno različico.

V nadaljevanju raziskav bi lahko nadgradili predstavljen pristop in metode ter tako pripravili še naprednejše tehnike za predvidevanje vrednosti kriptovalut.

Zahvala

Ta prispevek je nastal s podporo ARRS raziskovalnega programa P2-0041, COST akcij IC1406 in CA15140 ter EU Obzorja 2020 projekta DAPHNE (št. 957407).

Literatura

- [1] M. S. Brown and B. Douglass, "An event study of the effects of cryptocurrency thefts on cryptocurrency prices," in *2020 Spring Simulation Conference (SpringSim)*, 2020, pp. 1–12.
- [2] A. Zamuda, V. Crescimanna, J. C. Burguillo, J. Matos Dias, K. Wegrzyn-Wolska, I. Rached, H. González-Vélez, R. Senkerik, C. Pop, T. Cioara, I. Salomie, and A. Bracciali, *Forecasting Cryptocurrency Value by Sentiment Analysis: An HPC-Oriented Survey of the State-of-the-Art in the Cloud Era*. Cham: Springer International Publishing, 2019, pp. 325–349.
- [3] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," *Decentralized Business Review*, p. 21260, 2008.
- [4] A. P. Ratto, S. Merello, Y. Ma, L. Oneto, and E. Cambria, "Technical analysis and sentiment embeddings for market

trend prediction," *Expert Syst. Appl.*, vol. 135, pp. 60–70, 2019.

- [5] A. Zamuda, *Operacijske raziskave logističnih, transportnih in ekonomskih sistemov: zbrano gradivo*. Univerza v Mariboru, Fakulteta za elektrotehniko, računalništvo in informatiko, 2020.
- [6] E. F. Fama, "Efficient capital markets: A review of theory and empirical work," *The Journal of Finance*, vol. 25, no. 2, pp. 383–417, 1970.
- [7] —, "Efficient capital markets: II," *The Journal of Finance*, vol. 46, no. 5, pp. 1575–1617, 1991.
- [8] A. W. Lo, "The adaptive markets hypothesis," *The Journal of Portfolio Management*, vol. 30, no. 5, pp. 15–29, 2004.
- [9] I. Z. Emiris, E. Christoforou, and A. Florakis, "Neural networks for cryptocurrency evaluation and price fluctuation forecasting," in *MARBLE 2019 - The 1st International Conference on Mathematical Research for Blockchain Economy*, Santorini, Greece, May 2019.
- [10] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 11 1997.
- [11] J. Abraham, D. Higdon, J. Nelson, and J. Ibarra, "Cryptocurrency Price Prediction Using Tweet Volumes and Sentiment Analysis," *SMU Data Science Review*, vol. 1, 2018.
- [12] P. J. de Jong, S. Elfayoumy, and O. Schnusenberg, "From returns to tweets and back: An investigation of the stocks in the dow jones industrial average," *Journal of Behavioral Finance*, vol. 18, pp. 54–64, 2017.
- [13] K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink, and J. Schmidhuber, "Lstm: A search space odyssey," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 10, pp. 2222–2232, 2017.
- [14] T. Fischer and C. Krauss, "Deep learning with long short-term memory networks for financial market predictions," *European Journal of Operational Research*, vol. 270, no. 2, pp. 654–669, 2018.
- [15] X. Pang, Y. Zhou, P. Wang, W. Lin, and V. I. C. Chang, "An innovative neural network approach for stock market prediction," *The Journal of Supercomputing*, vol. 76, pp. 2098–2118, 2018.
- [16] S. Siami-Namini and A. S. Namin, "Forecasting economics and financial time series: ARIMA vs. LSTM," *CoRR*, vol. abs/1803.06386, 2018.
- [17] A. Thakkar and K. Chaudhari, "Crest: Cross-reference to exchange-based stock trend prediction using long short-term memory," *Procedia Computer Science*, vol. 167, pp. 616–625, 2020, international Conference on Computational Intelligence and Data Science.
- [18] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, no. 85, pp. 2825–2830, 2011.
- [19] C. Sammut and G. I. Webb, Eds., *Mean Absolute Error*. Boston, MA: Springer US, 2010, pp. 652–652.
- [20] R. N. Singarimbun, E. B. Nababan, and O. S. Sitompul, "Adaptive moment estimation to minimize square error in backpropagation algorithm," in *2019 International Conference of Computer Science and Information Technology (ICoSNIKOM)*, 2019, pp. 1–7.