

SIAP – predlog projekta

Predrag Glavaš E2 38/2021 i Nikola Mijonić E2 45/2021

- **Definicija problema**

Cilj projekta je predviđanje cene visoko špekulativnih kriptovaluta kao što su Dogecoin, Shiba Inu i Idena kako bi se time smanjio potencijalni investicioni rizik za nove investitore. Analiza će se sprovesti na osnovu istorijskih podataka, indikatora popularnosti u okviru GoogleTrend-a, podataka o razvojnem timu i zajednici. Testiraće se Dollar Cost Average (DCA) investiciona strategija sa različitim parametrima.

- **Motivacija**

Od početka pandemije Korona virusa tržište kriptovaluta je doživelo vrtoglavi skok, većinom zbog do sada neviđenog priliva novih kupaca koji prvi put kupuju kriptovalute sa ciljem ostvarivanja profita. U takvoj klimi se veliki broj njih opredeljuje da investira u određene valute na osnovu njihove trenutne popularnosti na društvenim mrežama. Kriptovalute koje su u ovom periodu postale popularne su izuzetno volatilne i špekulativne valute bez tehnološke i upotrebne vrednosti. S tim u vezi manipulacije koje se vrše nad njima dovode do ogromnih gubitaka za investitore te je potrebno efikasno predvideti njihovu vrednost i odrediti adekvatnu investicionu strategiju za smanjenje rizika.

- **Relevantna literatura**

[1] Wang, Yu, and Runyu Chen. "Cryptocurrency price prediction based on multiple market sentiment." *Proceedings of the 53rd Hawaii International Conference on System Sciences*. 2020.

Tema rada je predviđanje cene kriptovaluta na osnovu podataka sa 3 različita marketa i analize sadržaja i broja komentara na različitim forumima.

Metodologija se sastoji od dobavljanja podataka, preprocesiranja, a zatim korišćenja različitih algoritama mašinskog učenja i dubokog učenja: LSTM, CNN, BPNN, RBF, SVR.

Skup podataka je kreiran preuzimanjem podataka preko API-a relevantnih menjačnica za prvi kvartal 2019. godine. Prikupljani su podaci za valute Bitcoin (BTC), Ethereum (ETH), Tether USD (USDT), EOS, Ripple (XRP), Litecoin (LTC). Atributi koji su korišćeni su: vreme, najveća cena, najniža cena, zapremina, cena na otvaranju, cena na zatvaranju. Komentari i diskusije sa društvenih mreža su prikupljani dnevno za period od trećeg kvartala 2018. do prvog kvartala 2019.

Evaluacija modela je vršena po 2 indikatora, prvi je koren proseka kvadrata greške (RMSE) a drugi je srednje apsolutno odstupanje (MAE).

Uočeno je da korišćenje komentara o određenoj kriptovaluti od pre 1-7 dana doprinosi boljoj tačnosti modela što se tumači težnjom kupaca da kupuju bazirano na informacijama koje su pročitali

par dana pre same kupovine. SVR i metode dubokog učenja su se pokazale kao bolje od RBF i BPNN, ali znatno sporije u predviđanju rezultata.

Zaključak je da komentari i diskusije nastale par dana pre transakcije najviše utiču na kupce/prodavce da izvrše transakciju. Upotreba SVR algoritma sa Sigmoid i linearnom kernel funkcijom bi mogla biti korisna u predviđanju cene špekulativnih kriptovaluta ako se uzme u obzir sentiment komentara i diskusija korisnika.

[2] Wołk, Krzysztof. "Advanced social media sentiment analysis for short-term cryptocurrency price prediction." *Expert Systems* 37.2 (2020): e12493.

Tema rada je kratkoročno predviđanje cene pet kriptovaluta na osnovu istorijskih podataka, sentiment analize podataka sa tvitera i podataka sa Gugl trenda.

Metode koje su korišćene su linearna regresija metodom najmanjih kvadrata, SVR, SGD, Gradient boosting, Ada boost, stabla odlučivanja, ElasticNet, višeslojni perceptron i Bajesova regresija.

Skup podataka je kreiran preuzimanjem podataka sa internet bazirane platforme Crypto-compare. Sentiment analiza tvitova je vršena pomoću Sentiment Reasoner-a i Valence Aware Dictionary-a. Isključivane su promjenljive sa manjom korelacijom od umerene.

Evaluacija modela je vršena po 2 indikatora: srednja greška (ME), koeficijent determinacije R^2 .

Uočeno je da pozitivan sentiment tvitova ne mora nužno indikovati pozitivan rast cene kriptovaluta jer korisnici nastavljaju sa pozitivnim komentarima i u fazama silaznog trenda. Tvitovi sa negativnim sentimentom u kombinaciji podacima sa Gugl trenda daju najbolje prediktivne rezultate.

Zaključak je da komentari i diskusije nastale par dana pre transakcije najviše utiču na kupce/prodavce da izvrše transakciju. Upotreba SVR algoritma bi mogla biti korisna u predviđanju cene špekulativnih kriptovaluta ako se uzme u obzir sentiment komentara i diskusija korisnika.

[3] Sridhar, Sashank, and Sowmya Sanagavarapu. "Multi-Head Self-Attention Transformer for Dogecoin Price Prediction." *2021 14th International Conference on Human System Interaction (HSI)*. IEEE, 2021.

Tema rada je predviđanje cene Bitkoina i analiza uticaja sentimenta na cenu DOGE kriptovalute na osnovu istorijskih informacija, sentiment analize tvitova i analize tvitova Elona Maska .

Metoda koja je korišćena je ARIMA.

Skup podataka je kreiran preuzimanjem podataka sa Tvitera Elona Maska od 2009. do 2021. Podaci o Bitkoinu su preuzeti za interval od 2012. do 2021. a za DOGE od 2014. do 2021. Atributi koji su korišćeni su: vreme, najveća cena, najniža cena, zapremina, cena na otvaranju, cena na zatvaranju.

Evaluacija modela je vršena uz pomoć Akaike's Information Criterion (AIC), Bayesian Information Criterion (BIC).

Uočeno je da ovako napravljen ARIMA model ima prediktivnu tačnost od 96% u proseku pri predviđanju cene Bitkoina. Takođe je uočeno da ne postoji značajna korelacija između tvitova Elona Maska i promene cene DOGE kriptovalute.

Zaključak je da ARIMA model može biti izuzetno koristan za predviđanje cene Bitkoina, a takođe da tvitovi Elona Maska kao ni generalno sentiment tvitova ne mora nužno biti dobar pokazatelj za predviđanje cene kriptovaluta.

• Skup podataka

Podaci o kriptovalutama će se dobavljati upotrebom CoinGecko API-ja (<https://www.coingecko.com/en/api/documentation>), Yahoo Finance (<https://finance.yahoo.com/>) i Coincodex-a (<https://coincodex.com/>). Relevantni podaci za ovaj rad su oni koji se menjaju na dnevnom nivou.

Atributi koji se uzimaju u razmatranje:

- Vreme
- Open price
- Close price
- Low price
- High price
- Oznaka valute
- Market cap
- Trading volume
- Podaci o zajednici
- Podaci o razvojnom timu

Ciljno obeležje je close price čija vrednost predstavlja informaciju o ceni kriptovalute na kraju dana. Podaci dobavljeni sa YahooFinance će se proširiti informacijama o zajednici i razvojnom timu dobavljenom sa CoinGecko API-ja kao i sa indikatorom popularnosti koji nudi GoogleTrend.

• Metodologija

Za predviđanje cena biće korišćen ARIMA model i Support vector regression (SVR) algoritam. Parametri ARIMA modela (p, q, d) biće eksperimentalno utvrđeni i po potrebi menjani nakon evaluacije kako bi se poboljšala tačnost modela. Za kernel funkciju SVR-a će se koristiti Sigmoid i Linearna funkcija. Na osnovu predviđenih cena kriptovaluta biće testirana DCA investiciona strategija na različitim vremenskim intervalima i količinama uloženog novca.

• Evaluacija rešenja

Skup podataka će biti podeljen na trening i test skupove u odnosu 75:25. ARIMA model će biti evaluiran korišćenjem Akaike's Information Criterion (AIC) i Bayesian Information Criterion (BIC). SVR model će biti evaluiran prema koeficijentu determinacije (R^2), srednjoj prosečnoj grešci (MAE) i korenu proseka kvadrata grešaka (RMSE).