

Socialinių tinklų įtaka kriptovaliutų kainoms

Literatūros apžvalga

DM 3 kursas, 2 grupė, Matas Gaulia, Jekaterina Sergejeva

Žymėjimai

BTC - bitcoin - bitkoinas

ETH - ethereum - etereumas

VADER - Valence Aware Dictionary and sEntiment Reasoner -
sentimento analizei skirtas įrankis, specifiskai pritaikytas socialinių medijų įrašams

BERT - Bidirectional Encoder Representations from Transformers -
transformerių pagrindu veikianti mašininio mokymosi technika skirta natūralios
kalbos apdorojimui

ARIMA - Autoregressive Integrated Moving Average -
autoregresinis integruotas slenkamojo vidurkio modelis

1. FORECASTING ECONOMIC AND FINANCIAL TIME SERIES: ARIMA VS. LSTM

- Duomenys
 - Finansiniai duomenys iš Yahoo Finance svetainės (pvz. Nikkei 225 index, NASDAQ index, t.t.)
 - Nuo 1985 metų sausio iki 2018 metų rugpjūčio (mėnesiniai)
- Metodai
 - Duomenų paruošimas:
 - Duomenų suskirstymas į mokymosi ir testinę aibes
 - Įvertinimas:
 - Root Mean Squared Error (RMSE)
 - ARIMA
 - Long-Short Term Memory (LSTM)
- Rezultatas:
 - Taikant šiuos du modelius, ant visų finansinių duomenų geriau suveikė LSTM modelis, tai rodo RMSE rodiklis.

2. Autoregressive Integrated Moving Average (ARIMA) Model for Forecasting Cryptocurrency Exchange Rate in High Volatility Environment: A New Insight of Bitcoin Transaction

- Duomenys
 - BTC kainų duomenys, nuo 2013 metų sausio mėn. iki 2017 metų spalio (mėnesiniai)
- Metodai
 - ARIMA
 - Parametrų (p , d , q) parinkimas
 - Modelio parametrų vertinimas
 - Liekanų patikrinimas
 - Prognozavimas
 - Mean Absolute Percentage Error (MAPE)
- Rezultatas
 - BTC kainos yra nestacionari laiko eilutė
 - MAPE rodiklis lygus 5.36%
 - Nustatyta, kad prognozavimas naudojant ARIMA modelį yra gan tikslus. Tačiau, reikalinga detalesnė paklaidų analizė, kadangi BTC kainų duomenys yra labai nepastovūs.

3. Forecasting Bitcoin Price Fluctuation by Twitter Sentiment Analysis

- Duomenys:
 - Twitter įrašai apie BTC (60 dienų)
 - BTC kainos duomenys (minutiniai)
- Metodai:
 - Teksto valymas:
 - raidžių sumažinimas
 - minėjimų(\$BTC, @user), ne alfa-numerinių simbolių, pasidalinimų (retweet) pašalinimas
 - emotikonų klasifikavimas
 - Sentimentas:
 - VADER
 - Modelis prognozavimui:
 - Atsitiktinių miškų regresija (angl. Random Forest Regression)
 - Įvertinimas:
 - Maksimali paklaida
 - Minimali paklaida
 - Vidutinė paklaida
- Rezultatas:
 - Egzistuoja stipri koreliacija tarp BTC procentinio pokyčio ir Twitter įrašų sentimentų

	Definition	Value
1.	Number of tweets	92550
2.	Number of prediction data	80491
3.	Maximum value of error (%)	43.83
4.	Minimum value of error (%)	21.84
5.	Average of error (%)	37.52

4. Ask “Who”, Not “What”: Bitcoin Volatility Forecasting with Twitter Data

- Duomenys:
 - Twitter įrašai apie BTC (145 dienos – 30’000’000 įrašų)
 - BTC kainos duomenys (15 minučių dažnis – 14’000 įrašų)
- Metodai:
 - Teksto valymas:
 - Požymių kūrimas pvz.: įrašo tipas, ar yra jautrios kalbos, vartotojas
 - Sentimentas:
 - VADER
 - Prognozavimo modeliai:
 - Autoregresyvus (AR)
 - Long-Short Term Memory (LSTM)
 - Gated Recurrent Unit (GRU)
 - Temporal Convolutional Networks (TCN)
 - Generalized Auto Regressive Conditional Heteroskedasticity (GARCH)
 - Įvertinimas:
 - Mean Absolute Percentage Error (MAPE)
 - Mean Absolute Error (MAE)
 - Root Mean Squared Error (RMSE)
 - Mean Squared Logarithmic Error (MSLE)
- Rezultatas:
 - Semantinė Twitter įrašų informacija nėra pati geriausia priemonė prognozuoti BTC kainos išsibarstymą (dispersiją), labiau tinkama yra informacija apie įrašo kūrėją ir kita įrašą apibūdinanti informacija, taip pat bendras įrašų kiekis per laiko periodą.

	MAPE	MAE	RMSE	MSLE
TCN	0.23 ± 0.01	1.63 ± 0.06	2.58 ± 0.07	0.11 ± 0.01
D-TCN_{User}	0.20 ± 0.01 ***	1.44 ± 0.03 ***	2.36 ± 0.05 ***	0.08 ± 0.01 ***
D-TCN_{Tweet}	0.21 ± 0.01 ***	1.59 ± 0.08 ***	2.45 ± 0.09 **	0.09 ± 0.01 **
D-TCN _{VADER}	0.26 ± 0.02	1.72 ± 0.08	2.70 ± 0.09	0.13 ± 0.01
D-TCN _{Count}	0.32 ± 0.05	1.97 ± 0.19	3.27 ± 0.47	0.16 ± 0.02
D-TCN_{VADER, Tweet, User}	0.21 ± 0.01 ***	1.47 ± 0.05 ***	2.38 ± 0.06 ***	0.09 ± 0.01 ***
D-TCN _{VADER, Tweet}	0.22 ± 0.01	1.56 ± 0.09	2.51 ± 0.09	0.10 ± 0.01
D-TCN _{VADER, User}	0.22 ± 0.01 **	1.51 ± 0.05 ***	2.43 ± 0.09 ***	0.10 ± 0.01 **
D-TCN _{Tweet, User}	0.21 ± 0.01 ***	1.49 ± 0.06 ***	2.41 ± 0.08 ***	0.09 ± 0.01 ***
D-TCN_{Count, Tweet, User}	0.22 ± 0.01 ***	1.46 ± 0.05 ***	2.36 ± 0.10 ***	0.08 ± 0.01 ***
D-TCN _{Count, Tweet}	0.23 ± 0.01	1.59 ± 0.08	2.56 ± 0.11	0.11 ± 0.01
D-TCN _{Count, User}	0.22 ± 0.01	1.53 ± 0.08 *	2.46 ± 0.11 *	0.10 ± 0.01
D-TCN _{Count, VADER, Tweet, User}	0.22 ± 0.01 *	1.48 ± 0.07 ***	2.43 ± 0.10 **	0.10 ± 0.01 **
D-TCN _{Count, VADER, Tweet}	0.24 ± 0.02	1.60 ± 0.07	2.59 ± 0.09	0.11 ± 0.01
D-TCN _{Count, VADER, User}	0.22 ± 0.01 **	1.47 ± 0.07 ***	2.39 ± 0.10 ***	0.10 ± 0.01 ***
D-TCN _{Count, VADER}	0.29 ± 0.02	1.82 ± 0.08	2.95 ± 0.18	0.15 ± 0.01

5. Cryptocurrency Price Prediction Using Tweet Volumes and Sentiment Analysis

- Duomenys:
 - Twitter įrašai apie BTC ir ETH (60 dienų - 30'420'063 įrašų)
 - Google Trends duomenys apie BTC ir ETH (dieniniai)
 - Twitter įrašų kiekis (dieniniai)
 - BTC ir ETH kainos (dieniniai)
- Metodai:
 - Sentimentas:
 - VADER
 - Požymių kūrimas:
 - Search Volume Index (SVI) iš Google Trends duomenų
 - Kainų pokyčiai
 - Įvertinimas:
 - Pearson koreliacija ir atitinkamai jos p reikšmė
- Rezultatas:
 - Sentimentas nėra patikimas indikatorius prognozuoti kriptovaliutų kainų pokyčius, geriau naudoti Google Trends duomenis ir bendrą įrašų kiekį laiko periode

Nuorodos

1. [Forecasting Economic and Financial Time Series: ARIMA vs. LSTM \(arxiv.org\)](#)
2. [Autoregressive Integrated Moving Average \(ARIMA\) Model for Forecasting Cryptocurrency Exchange Rate in High Volatility Environment: A New Insight of Bitcoin Transaction](#)
3. [\(PDF\) Forecasting Bitcoin Price Fluctuation by Twitter Sentiment Analysis \(researchgate.net\)](#)
4. [Ask “Who”, Not “What”: Bitcoin Volatility Forecasting with Twitter Data \(arxiv.org\)](#)
5. [Cryptocurrency Price Prediction Using Tweet Volumes and Sentiment Analysis \(smu.edu\)](#)