VILNIAUS UNIVERSITETAS
MATEMATIKOS IR INFORMATIKOS FAKULTETAS
DUOMENŲ MOKSLO STUDIJŲ PROGRAMA

Duomenų mokslo projektas - kursinis darbas

Socialinių tinklų įtaka kriptovalutų kainoms

Social media impact on cryptocurrencies prices

Darbo atliko: Matas Gaulia,
Jekaterina Sergejeva
Darbo vadovas: Linas Petkevičius

VILNIUS 2022

Turinys

Lentelių sąrašas

Iliustracijų sąrašas

Socialinių tinklų įtaka kriptovaliutų kainoms

Santrauka

Kriptovaliutos daugeliu aspektų skiriasi nuo įprastų finansinių rinkų, jų kainos priklauso nuo visiškai skirtingų dalykų negu įmonių akcijų kainos. Kriptovaliutos yra decentralizuotos, nevaldomos jokių finansinių institucijų, o jų panaudojimas ribotas. Vienas iš pagrindinių dalykų, darančių įtaką kriptovaliutų kainoms yra jų populiarumas ir žmonių nuomonė apie ją, išreiškšta socialiniuose tinkluose. Remiantis šia teorija bus bandoma ištirti ar socialinių tinklų įrašai turi įtakos kriptovaliutos XRP kainai.

Raktiniai žodžiai : Kriptovaliutos; Laiko eilutės; Natūralios kalbos apdorojimas

Social media influence on cryptocurrency prices

Abstract

Cryptocurrencies differ from regular financial markets in many aspects, their prices depend on totally different things than companies' stock prices. Cryptocurrencies are decentralized, not controlled by any financial institutions, their use is very limited. One of the main things that affects cryptocurrency's price is its popularity and people's opinion about it shared on social media. Based on this theory we are going to explore whether social media posts about XRP cryptocurrency impact its price.

Key words : Cryptocurrencies; Time series; Natural language processing

Acronyms

ARIMA Autoregressive Integrated Moving Average. 1

BTC Bitcoin kriptovaliuta. 1

ETH Ethereum kriptovaliuta. 1

GARCH Generalized Auto Regressive Conditional Heteroskedasticity. 1

GRU Gated Recurrent Unit. 1

LSTM Long Short Term Memory. 1

MAE Mean Absolute Eroor. 1

MAPE Mean Absolute Percentage Error. 1

MSLE Mean Squared Logarithmic Error. 1

RMSE Root Mean Squared Error. 1

SPY S&P500 investicinis fondas sekantis didžiausias 500 Jungtinių Amerikos Valstijų kompanijas. 1

SVI Search Volume Index. 1

TCN Temporal Convolutional Networks. 1

VXX Investicinis fondas sekantis kokio stiprio kainų judėjimo tikisi žmonės pamatyti S&P500 per ateinančius 12 mėnsių. 1

XAU Investicinis fondas sekantis aukso kainą. 1

XRP Ripple kriptovaliuta. 1

1 Įvadas

Kriptovaliutos − tai skaitmeniniai, tik virtualioje erdvėje egzistuojantys valiutos tipai, paremti blokų grandinės (angl. *blockchain*) technologija ir leidžiantys anonimiškai atlikti įvairius internetinius mokėjimus be trečiųjų šalių tarpininkavimo. Kriptovaliutos yra decentralizuotos ir neprižiūrimos jokių finansinių institucijų. Kiekvienais metais jų atsiranda vis daugiau, kai kurios greitai išpopuliarėja ir jų kainos sparčiai auga. [AEK+21] straipsnyje rašoma, jog tyrimas, atliktas 2017 metais, parodė, kad tuo metu apie 6 milijonus žmonių visame pasaulyje turėjo kriptovaliutų. Praėjus vos 4 metams, šis skaičius padidėjo iki 300 milijonų. Tačiau nėra aišku kas gali lemti tam tikros kriptovaliutos populiarumą ir atvirkščiai. Stebint akcijų rinką, galima nesunkiai pasakyti kurie faktoriai daro įtaką vienos ar kitos akcijų rūšies kainų kitimams. Mes manome, jog kriptovaliutų populiarumą ir jų kainų augimą gali lemti socialinių tinklų (pvz. *Twitter*) įrašai apie ją, t.y. kuo didesnis žmonių, keliančių įrašus apie tam tikros kriptovaliutos įsigijimą ir giriančių ją, tuo daugiau atsiranda naujų pirkėjų, kriptovaliuta populiarėja ir jos kaina auga. Tuo tarpu, kai žmonės rašo daug negatyvių komentarų apie kriptovaliutą, tai gali sumažinti jos kainą.
Tyrimui mes pasirinkome 1 kriptovaliutą − XRP bei įrašus apie ją iš socialinio tinklo *Twitter*.
Darbe bus panaudojamos dvi strategijos *XRP* kainų prognozavimui. Pirmiausia bus pritaikytas ARIMA modelis laiko eilutėms ir bus bandoma nuspėti *XRP* kainų pokyčius ateityje. Antras būdas, kuris atrodo veiksmingesnis, yra *XRP* kainų prognozavimas remiantis šrašais iš socialinio tinklo Twitter apie šią kriptovaliutą. Pirmiausia kiekvienas duomenų rinkinys atskirai turi būti išanalizuotas. Iš įrašų bus gautas sentimentas (teigiamas, neigiamas, neutralus), šiam tikslui pasiekti bus naudojamas *VADER* modelis. Toliau bus tikrinama hipotezė, kad socialinių tinklų įrašai apie kriptovaliutą daro įtaką jos kainai.

Tikslas

- Patikrinti ar *Twitter* socialinio tinklo įrašai apie XRP kriptovaliutą daro įtaką jos kainų pokyčiams.

Uždaviniai

- Išrinkti geriausią laiko eilučių modelį kriptovaliutų kainoms
- Ištirti socialinių medijų nuomonės reikšimnguma kriptovaliutos kainai
- Atrasti reikšmingus požymius kriptovaliutos kainų progozavimui
- Prognozuoti kriptovaliutos kainas pasitelkiant sentimento informacija

2 Temos aprašymas

Nūdienoje kriptovaliutų kainų prognozavimas ir analizavimas yra gan populiari tema moksliniuose darbuose, juk pavykus sėkmingai prognozuoti ateities kainas, galima gauti didelę finansinę grąžą. Daugelis jau atliktų eksperimentų pagrindinį dėmesį skiria sentimento analizavimui, nes žmonių nuomonė, kuria jie dalinasi socialiniuose tinkluose, gali turėti įtakos kriptovaliutų kainoms.

Toliau pateikta panašių ir reikšmingų darbų, kurie gali būti naudingi ir šiam darbui, apžvalga.

Pirmame darbe [SN18] naudojami LSTM ir ARIMA modeliai siekiant nuspėti tokių finansinių indeksų kaip Nikkei 225 ir NASDAQ ateities kainas, kai vieninteliai duomenys buvo praeities kainos(mėnesiniai duomenys). Pagal RMSE rodiklį, gauta kad LSTM modelis pasirodė daug geriau nei ARIMA, buvo fiksuotas vidutiniškai -87% RMSE sumažėjimas.

Kito panašaus darbo [ABR17] autoriai turėdami mėnesinius BTC duomenis nup 2013 iki 2017 metų pritaikė ARIMA modelį ir bandė nuspėti BTC ateities kainas. Buvo nustatyta, kad BTC eilutė nėra stacionari, taip pat, rodiklis MAPE buvo lygus 5.36% . Autoriai pastabose taip pat nurodė kad modelio panaudojimui reikalinga detalesnė paklaidų analizė, nes BTC kainų duomenys yra labai nepastovūs.

Trečiame darbe [SJOL20] naudojami įrašai iš Twitter platformos su žyme, kad įraše yra informacija apie BTC kriptovaliutą. Taip pat minutiniai BTC kainos duomenys. Kadangi socialinių medijų įrašuose nėra jokių reikalavimų, jie dažnai būna nerišlios kalbos, turi akronimų ar sutrumpinimų, emotikonų (angl. emoji). Autoriai sprendė šią problemą pašalindami iš įrašų visus simboius, kurie nebuvo tekstu, suvienodino raidžių didumą, klasifikavo ir pašalino emotikonus. Sentimentui išgauti buvo naudojamas VADER[1] įrankis. Prognozavimui pasirinktas Atsitiktinių miškų regresijos (angl. Random Forest Regression) modelis. Vertinimo rodiklis vidutinė paklaida siekė 37.52%. Autoriai pastebi, kad nors ir gauta per didelė vidutinė paklaida sėkmingai prognozei, tačiau rasta stipri koreliacija tarp BTC kainos procentinio pokyčio ir Twitter įrašų sentimento.

Kito darbo [AEK+21] autorių tikslas buvo nuspėti BTC kriptovaliutos kainos svyravimus (dispersiją). Surinkti 15 minučių dažnumo BTC kainos duomenys (14'000 įrašų) ir įrašai iš Twitter platformos apie BTC kriptovaliutą (30'000'000 įrašų). Kad informacija būtu lengviau apdorojama,iš teksto buvo išgauti tokie požymiai kaip įrašo tipas, jautrios kalbos statu-

[1] Hutto, C.J. and Gilbert, E.E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. Eighth International Conference on Weblogs and Social Media (ICWSM-14). Ann Arbor, MI, June 2014.

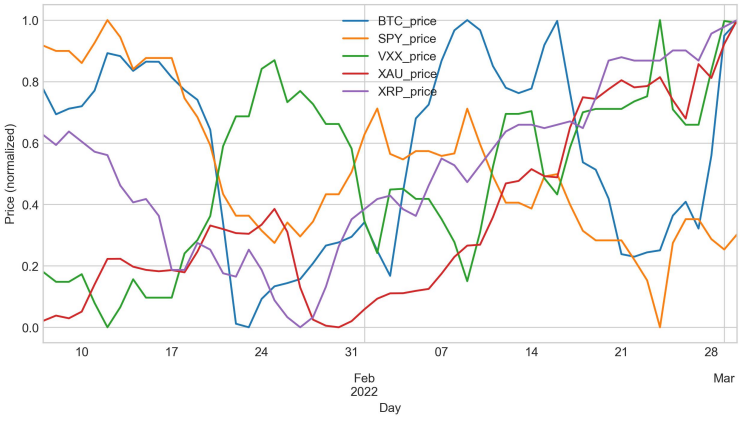
sas, vartotojas, sentimentui gauti naudotas ankstesnėje pastraipoje minėtas VADER[1] modelis. Išbandyti LSTM, GRU, TCN, GARCH prognozavimo modeliai. Įvertinimui ir palyginimui naudoti MAPE, MAE, RMSE, MSLE. Gauta, kad geriausi rezultati (MAPE = 0.2) buvo gauti su TCN modeliu naudojant informaciją apie vartotoją, kuris paskelbė įrašą, kas yra paaiškinama, nes daugelis spekuliuojančių žmonių seka ir kopijuoja tai, ką daro už juos sėkmingesni vartotojai.

Šiame darbe [AHNI18] dėmesys skirtas BTC ir ETH kriptovaliutų kainų prognozei pasitelkiant Twitter įrašais. Surinkti Twitter įrašų duomenys apie BTC ir ETH kriptovaliutas, Google Trends duomenys, dieninis Twitter bendras įrašų skaičius, BTC ir ETH dieninės kainos. Sentimentui gauti panaudotas VADER[1] modelis, sukurti tokie požymiai kaip SVI iš Google Trends duomenų, taip pat paskaičiuoti kainų pokyčiai tarp dienų. Pritaikius tiesinės regresijos modelį nebuvo gauti geri rezultatai, autoriaus teigimu taip gali būti dėl to, kad ryšys tarp kainų ir sukurtų požymių nėra tiesinis. Rezultatuose taip pat pastebėta, kad Google Trends duomenys stipriai koreliuoja su BTC ir ETH kainomis.

3 Pirminė duomenų analizė

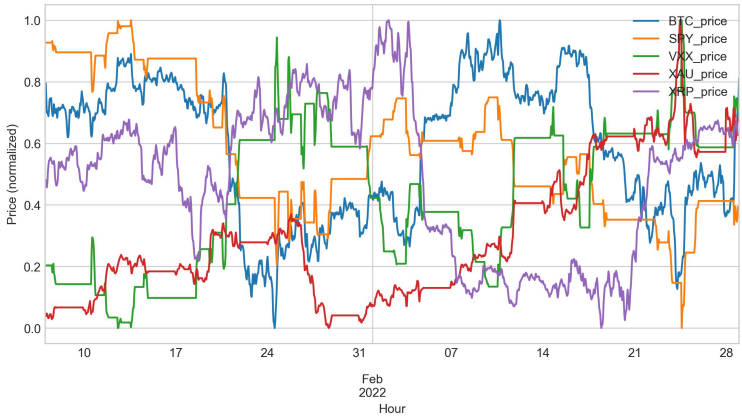
Kriptovaliutų ir kitų aktyvų (angl. assets) kainų duomenis gavome iš švedų internetinio banko svetainės Dukascopy[2]. Gavome tokių aktyvų kaip XRP, BTC, SPY, VXX, XAU kainas dieniniais, valandinias ir minutiniais intervalais, taip pat duomenys buvo prieinami tik arba iš pirkėjo (angl. bid) arba pardavėjo (angl. ask) pusės, tad iš viso pradžioje turėjome 30 failų.
Duomenys buvo nuo 2022-01-07 00:01:00 iki 2022-02-28 23:52:00.
Dieniniai duomenys turėjo 55 eilutes, valandiniai - 1271, minutiniai - 76312.
Po duomenų valymo turėjome 3 duomenų rinkinius kuriame kiekviename buvo 5 aktyvų kainos.
Prieš braižant kainų grafikus, duomenys buvo normalizuoti, nes aktyvų skalės stipriai skyrėsi. Po normalizavimo visos kainos įgyjo reikšmes intervale [0;1].
Dieninių aktyvų kainų grafike (žr. 1 pav.) nesimato tendencijų arba priklausomybių tarp XRP kainų ir kitų finansinių aktyvų kainų. Matome, jog sausio mėnesį XRP kaina greitai mažėjo, kol sausio pabaigoje nepasiekė savo minimumo. Vasario mėnesį šios kritovaliutos kaina, kaip ir XAU, VXX kainos augo.
Valandiniuose aktyvų kainų duomenyse (žr. 2 pav.) ryškiai matosi daug kainų šuolių, sunkiau įžiūrimos ilgalaikės tendencijos. XRP kainos valandiniuose intervaluose keitėsi daug dažniau.
Minutiniuose duomenyse (žr. 3 pav.), kaip ir valandiniuose, kainos kito labai greitai. Nei viename iš trijų - minutinių, valandinių, dieninių duomenų - grafikų nepastebimos stiprios koreliacijos tarp XRP ir kitų finansinių aktyvų kainų, todėl prognozuoti XRP kainą atsižvelgiant į kitų aktyvų kainas nėra prasmės.

1 pav.: Dieniniai kainų duomenys

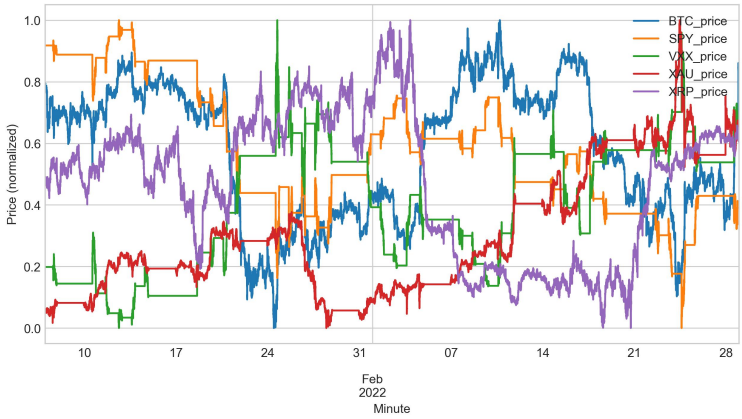


[2] Dukascopy Bank - https://www.dukascopy.com

2 pav.: Valandiniai kainų duomenys

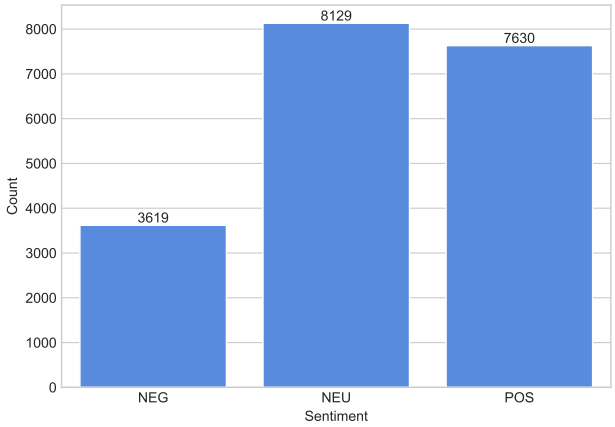


3 pav.: Minutiniai kainų duomenys



Kitas duomenų rinkinys – tai Twitter įrašų duomenys apie XRP. Jame yra 2 stulpeliai: įrašo tekstas ir įrašo paskelbimo laikas minučių tikslumu. Pastebėta, kad įrašų tekste dažnai yra nenaudingos informacijos tokios kaip nuorodos, kitų valiutų ar vartotojų minėjimai, tad ji buvo pašalinta.
Įrašo sentimentui išgauti buvo naudojamas VADER[1] modelis. Iš viso turima 19378 Twitter įrašų apie XRP. Stulpelinė diagrama (žr. 4 pav.) parodo kiek iš jų yra pozityvūs, neigiami ir neutralūs. Diagramoje NEG reiškia neigiamą sentimentą, NEU - neutralų, POS - pozityvų. Iš grafiko aiškiai matyti, kad neutralių įrašų daugiausia (8129), taip gali būti dėl kelių priežasčių, viena iš jų kad modelis nesugeba nustatyti sentimento, tad priskiria neutralų statusą, kita priežastis galėtų būti, kad žmonės tiesiog diskutuoja Twitter erdvėje ir nebando perteikti savo nuomonės apie XRP kriptovaliutą arba prie nereikšmingo įrašo prirašo XRP žymę, kad jų įrašas būtų pamatytas didesnės grupės žmonių. Taip pat pastebėjome kad teigiamų irašų skaičius (7630) stipriai lenkia neigiamų įrašų skaičių (3619), viena iš to priežasčių yra kad teigiamos žinutės skatina kainos kilimą, kas ir yra vienas iš pagrindinių investuotojų tikslų.

4 pav.: Įrašų sentimento histograma



Tuomet buvo nuspręsta pasižiūrėti kaip atrodo bendras dienos įrašų skaičius, nuspalvintas pagal sentimento dominavimą tą dieną.

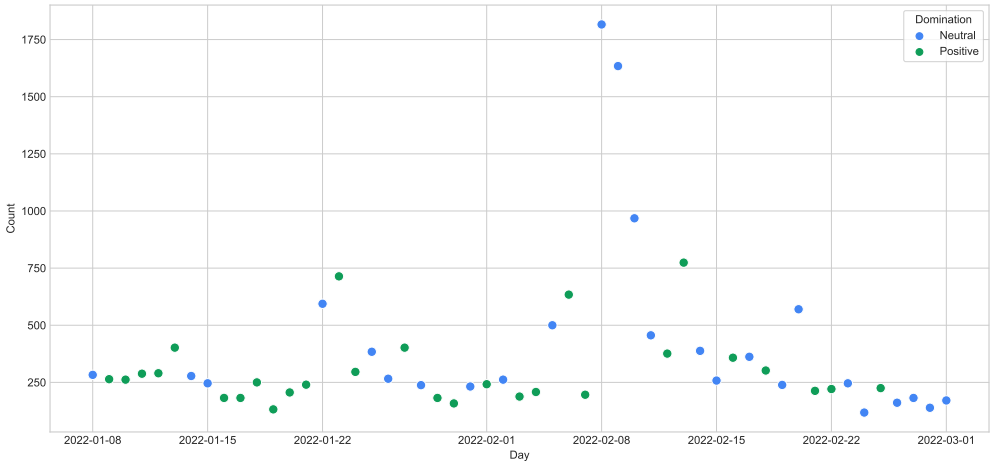
Dieniniuose duomenys (žr. 5 pav.) nebuvo tokios dienos, kad dominuotų neigiami įrašai, manome tai yra dėl to, nes diena yra ilgas laiko tarpas kriptovaliutų kainų judėjimui ir neigiami komentarai yra užgožiami labai dideliu kiekių teigiamų įrašų. Matome kad daugiausia dienų buvo su neutraliu sentimentu. Dvi dienos (vasario 8 ir 9) ryškiai išsiskiria grafike, kadangi tomis dienomis Twitter buvo pasidalinta daugiausia įrašų apie XRP kriptovaliutą ir abi dienas dominavo neutralūs įrašai.

Valandiniuose duomenys (žr. 6 pav.) galime pastebėti, kad buvo labai nedaug valandų su dominuojančiu neigiamu sentimentu, o jeigu dominuojantis sentimentas yra neigiamas, tomis valandomis bendras įrašų skaičius yra mažas. Taip pat padaugėjus įrašų skaičiui, bendru atveju padidėja ir teigiamų irašų skaičiaus dominavimo valandų. Vis dėlto valandą, kai buvo pasidalinta didžiausiu skaičiumi įrašų apie XRP, dominavo neutralūs komentarai.

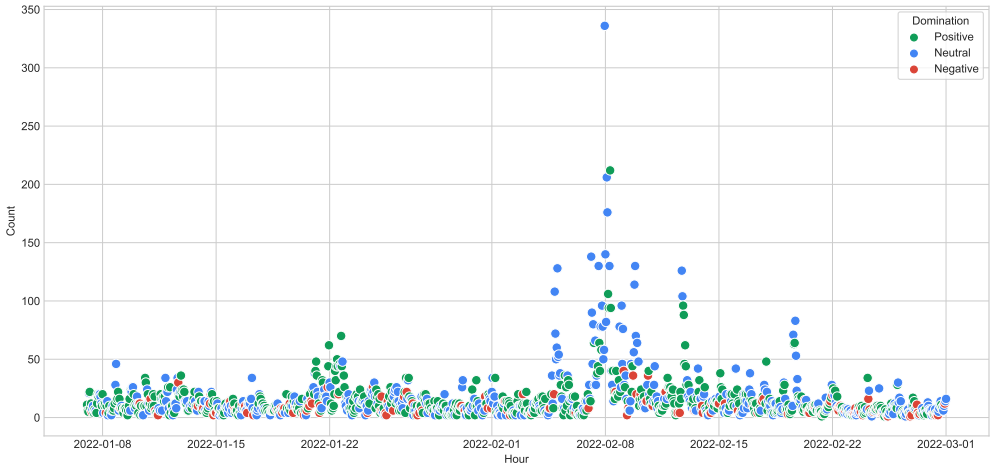
Grafike su minutiniais duomenimis (žr. 7 pav.), daugelis taškų stipriai susigrūda, tačiau paeksperementavus su skirtingais grafikais, buvo padaryta išvada, kad taškų grafikas geriausiai atvaizduoja sentimento dominavimą ir skirtingų minučių reikšmės mažiausiai persidengia. Taip pat matome, kad didėjant įrašų skaičiui, atsiranda vis daugiau minučių, kurių sentimentas yra neigiamas. Tai galima paaiškinti tuo, kad didėjant teigiamiems komentarams, didėja ir vartotojų, norinčių pasisakyti priešingai skaičius, vyksta diskusijos. Minutiniuose duomenyse minutę, kai buvo pasidalinta daugiausia įrašų apie XRP, dominavo teigiamas įrašų sentimentas.

Atrodo, kad minutiniai duomenys turėtų būti naudingiausi nes turi didžiausią sentimento variaciją tarp laiko taškų.

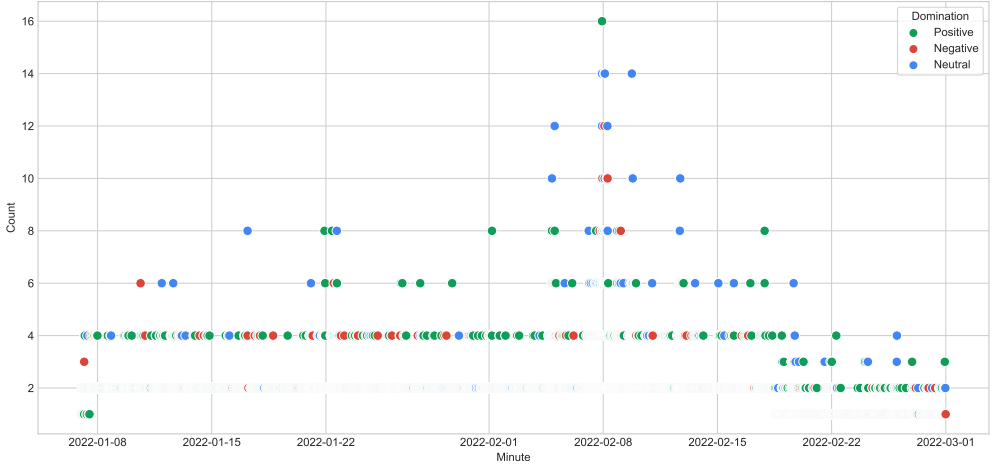
5 pav.: Dieninis irašų skaičius pagal sentimento dominavimą



6 pav.: Valandinis irašų skaičius pagal sentimento dominavimą



7 pav.: Minutinis irašų skaičius pagal sentimento dominavimą

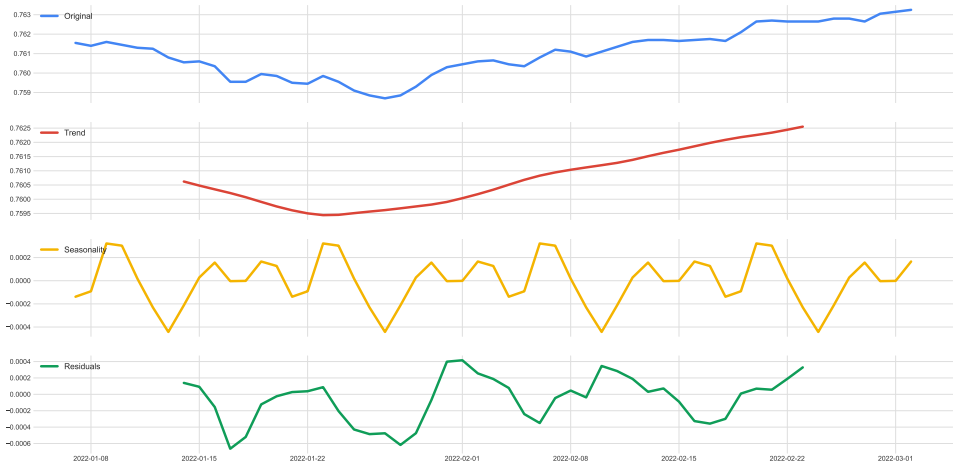


4 Eksperimentai

Pirmiausia buvo patikrinta kaip su turimais duomenimis veikia klasikinis ARIMA modelis su sezonine komponente.

4.1 Dieniniai grafikai

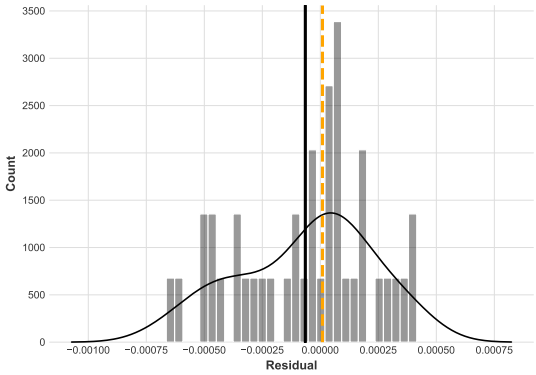
8 pav.: Dieninių duomenų dekompozicija



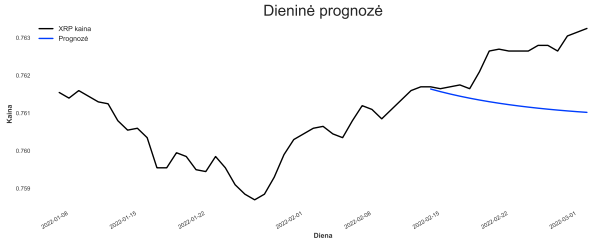
Pradėta nuo dieninių duomenų. Jie buvo išskaidyti į dedamąsias dalis - trendas, sezoniškumas, liekanos (žr. 8 pav.). Dieniniuose duomenyse ryškiai matosi kainos augimo trendas, tačiau dar per anksti priimti hipotezę, kad trendas egzistuoja, reikia patikrinti valandinius ir minutinius duomenis. Taip pat matosi pasikartojanti struktūra sezoniškumo kreivėje. Atskiro dėmesio ir detalesnės analizės reikalauja liekanos. Buvo patikrinta ar dieninių duomenų liekanos normaliai pasiskirsčiusios (žr. 9 pav.). Nors kreivė gana panaši į normaliojo skirstinio kreivę, sunku pasakyti ar liekanos yra normaliai pasiskirsčiusios, kadangi duomenų yra per mažai (tik 55 dienų XRP kainos).

Išanalizavus dieninių duomenų dedamąsias, buvo bandoma prognozuoti XRP kainas (žr. 10 pav.). Grafike matosi, kad ARIMA modelis nėra tinkamas dieniniams duomenims, prognozuojamos reikšmės yra toli nuo tikrųjų. Be to, matosi, jog tikroji XRP kaina auga, o ARIMA modelio spėjamos reikšmes eina žemyn.

9 pav.: Dieninių duomenų liekanų pasiskirstymas

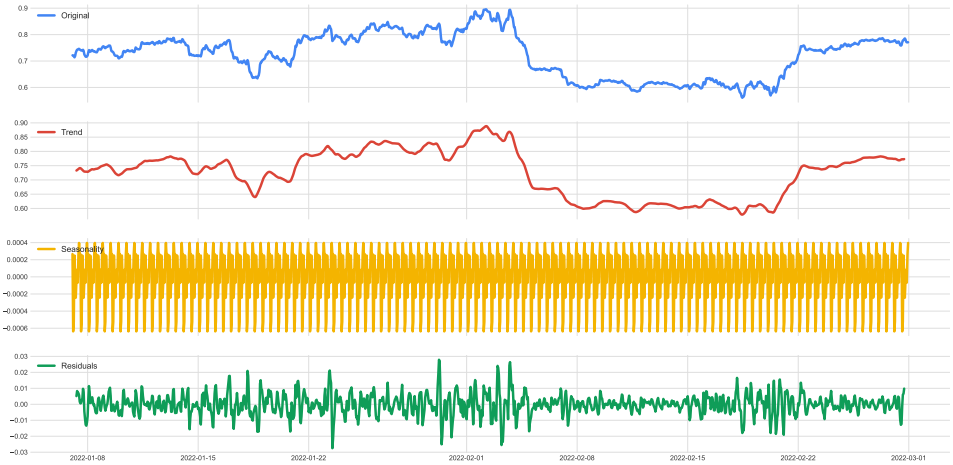


10 pav.: Dieninių duomenų ARIMA rezultatai



4.2 Valandiniai grafikai

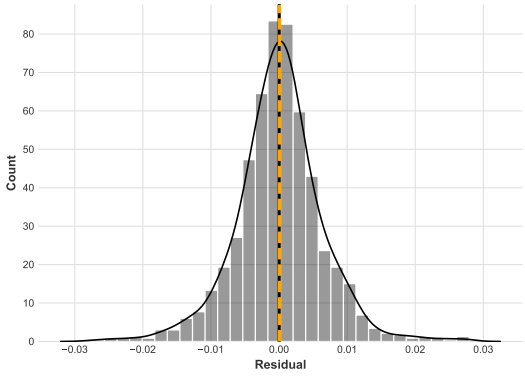
11 pav.: Valandinių duomenų dekompozicija



Tokie pat veiksmai buvo atlikti ir su valandiniais duomenimis. Iš pradžių duomenys buvo išskaidyti į trendą, sezoniškumą, liekanas (žr. 11 pav.).

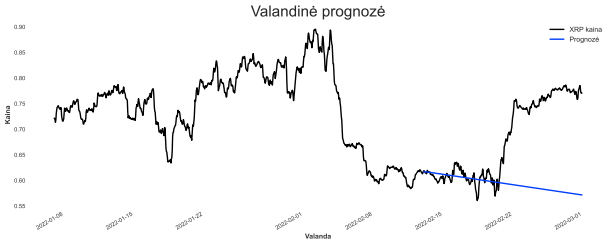
Matosi, kad trendas šiek tiek pašalina triukšmą iš laiko eilutės.

12 pav.: Valandinių duomenų liekanų pasiskirstymas



Taip pat buvo patikrinta prielaida, kad valandinių duomenų liekanos yra normaliai pasiskirsčiusios. Grafike (žr. 12 pav.) matosi, kad liekanos tikrai yra pasiskirsčiusios normaliai, simetriškai apie 0.

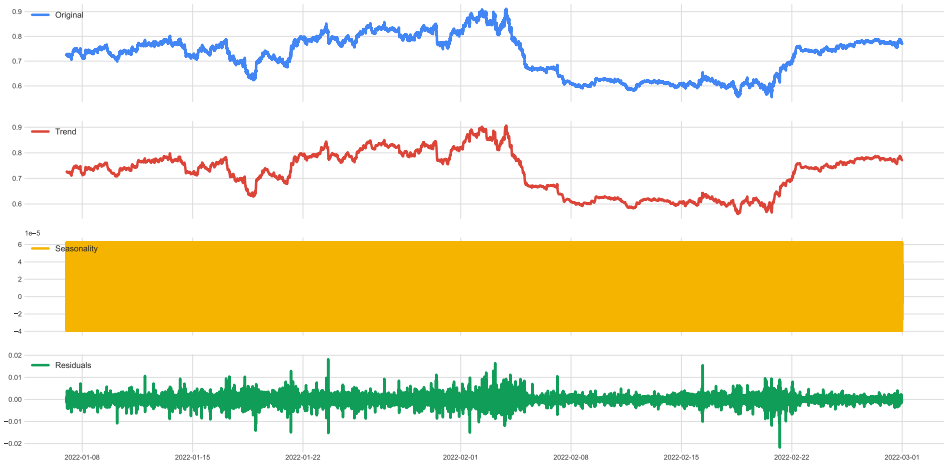
13 pav.: Valandinių duomenų ARIMA rezultatai



Valandiniams duomenims ARIMA modelis (žr. 13 pav.) prognozuoja tiesę einančią žemyn, nors kainos testavimo aibėje didelę laiko dalį kilo.

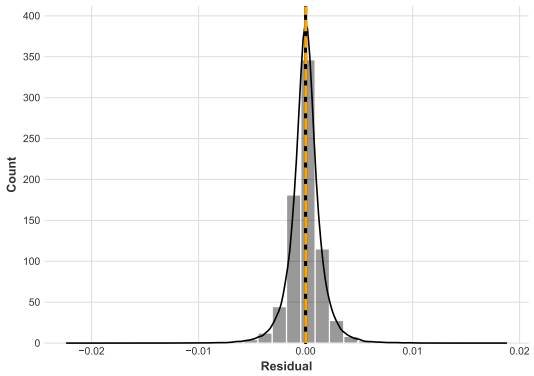
4.3 Minutiniai grafikai

14 pav.: Minutinių duomenų dekompozicija



Visi žingsniai, atlikti su dieniniais ir valandiniais duomenimis, buvo pakartoti su minutiniais duomenimis (žr. 14 pav.). Kadangi minutiniai duomenys yra didelio tankio, jų trendo komponentės atvaizdavimas labai panašus į pradinę eilutę, tiesiog vietomis yra pašalintas triukšmas.

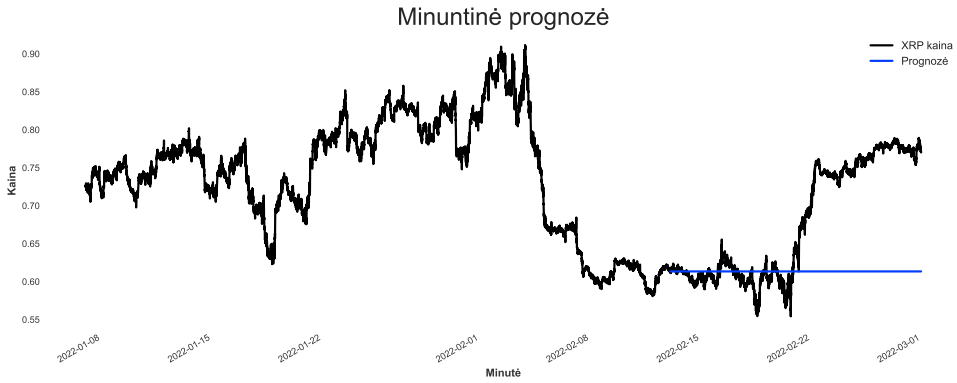
15 pav.: Minutinių duomenų liekanų pasiskirstymas



Tikrinama ar liekanos yra pasiskirsčiusios normaliai, iš grafiko (žr. 15 pav.) matosi, kad jos iš tikrųjų pasiskirsčiusios pagal normalųjį dėsnį.

ARIMA modelio prognozė (žr. 16 pav.), kaip ir su valandiniais duomenimis, prognozuoja tiesę .

16 pav.: Minutinių duomenų ARIMA rezultatai



4.4 Išvados apie ARIMA modelį

ARIMA modelis buvo pratestuotas ant dieninių, valandinių, minutinių grafikų. Su visais duomenimis modelio rezultatai buvo prasti. Taip gali būti dėl to, kad kriptovaliutų kainų duomenys yra labai nepastovūs ir neturi sezoniškumo.

5 Pycaret biblioteka

Nagrinėjant naujausius kriptovaliutų prognozės įrankius buvo rastas Pycaret [Ali20] paketas. Paketas bando automatizuoti modelio parinkimą pritaikydamas visus jame esančius įrankius ir lygina gautas paklaidas, taip gaunant modelį, kurio paklaida yra mažiausia duotiems duomenims. Taip pat pakete yra lengvai prieinami hyperparametrų optimizavimas, modelio diagnozė ir grafikai. Pycaret paketas gali spręsti 4 tipų uždavinius: klasifikavimo, regresijos, klasterizavimo, anomalijų radimo, tačiau šiame darbe jis bus naudojamas tik regresijos uždaviniams.
Kadangi norima pagerinti modelio rezultatus, pycaret paketas testuojamas su agreguotais duomenimis, tai reiškia, kad kiekvienam laiko žingsniui turime kelis kintamuosius: teigiamų įrašų skaičius, neigiamų įrašų skaičius, neutralių irašų skaičius, bendras įrašų skaičius ir suma kiekvienos žinutės sentimento tame laiko tarpe (sentimentas yra perkoduotas taip, kad teigiami yra 1, neutralūs yra 0, o neigiami yra -1). Kaip atsakas buvo pasirinkta prognozuoti procentinį kainų pokytį tarp laiko žingsnių.

5.1 Dieniniai grafikai

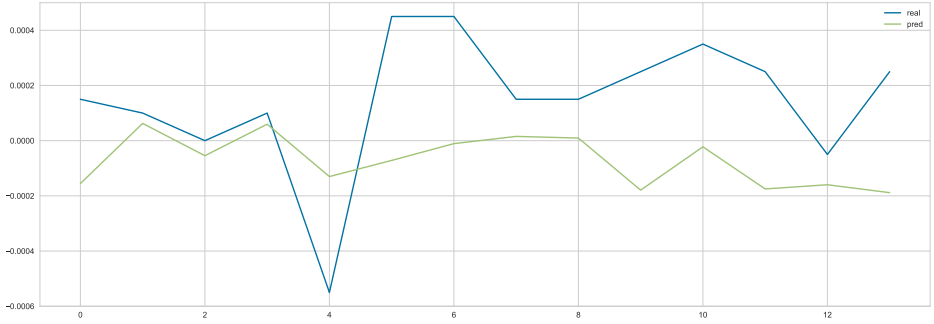
Eksperementuojant su Pycaret biblioteka buvo ieškoma geriausio modelio duomenims. Geriausiu modeliu išrinkta tiesinė regresija (su dieniniais, valandiniais, minutiniais duomenimis), kur XRP kainų pokytis pasirinktas kaip atsakas, o visi kiti stulpeliai − regresoriai.

Valandinių XRP kainų pokyčius pavyko prognozuoti gan neblogai(žr. 17 pav.). Tačiau modelis dažnai prognozuoja, kad pokyčiai bus neigiami (kaina mažės) arba labai artimi 0, netgi jeigu tikroji XRP kaina auga.

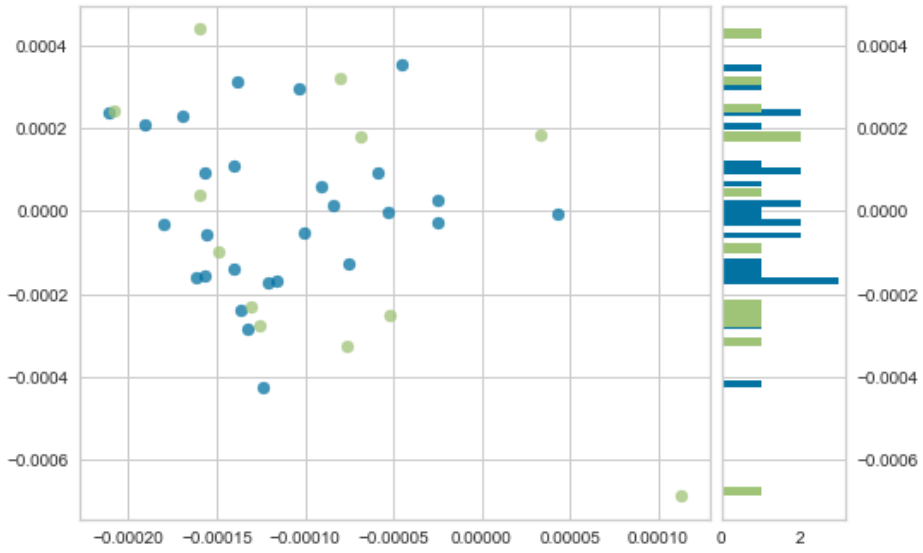
Taip pat pasižiūrėjome kaip pasiskirsčiusios liekanos (žr. 18 pav.).

Regresijos uždaviniuose labai svarbus kovariančių reikšmingumas. Grafike pavaizdavus 7 skirtingų kovariančių rekšmingumus, matome, jog didžiausią įtaką XRP kainų pokyčiams daro VXX kainų pokytis.

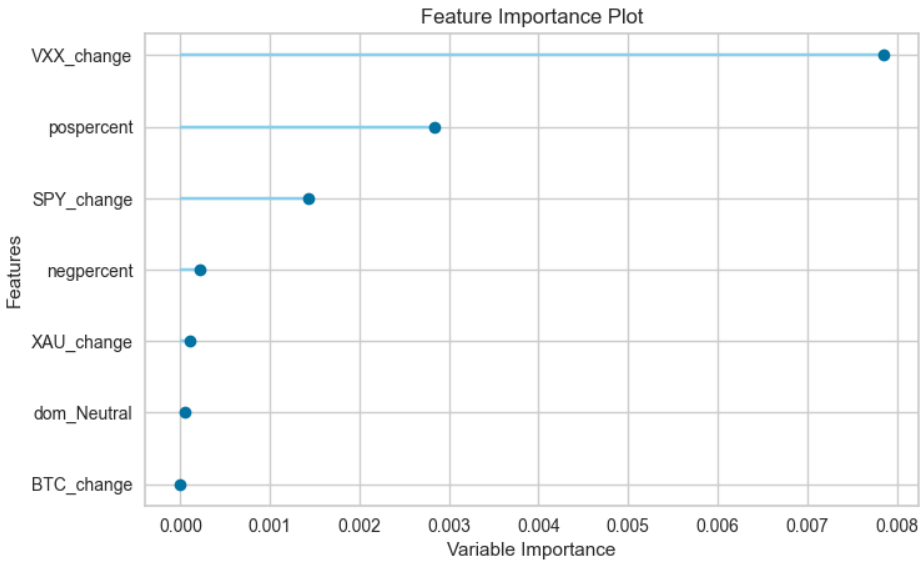
17 pav.: Dieninių XRP kainų pokyčių prognozavimas



18 pav.: Dieninių XRP kainų pokyčių liekanos



19 pav.: Dieniniai duomenys, kovariančių reikšmingumas



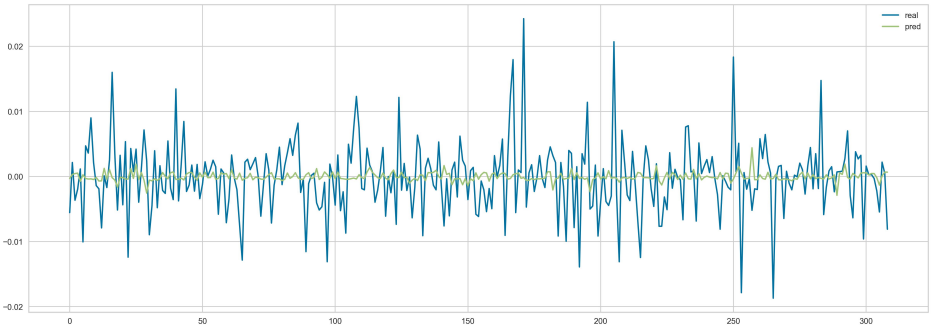
5.2 Valandiniai grafikai

Su valandiniais duomenimis, kainų pokyčiai pagal grafiką buvo nuspėjami daug prasčiau negu su dieniniais ir svyravo apie 0 (žr. 20 pav.). Nors tikrosios XRP kainų pokyčių reikšmės buvo didesnės.

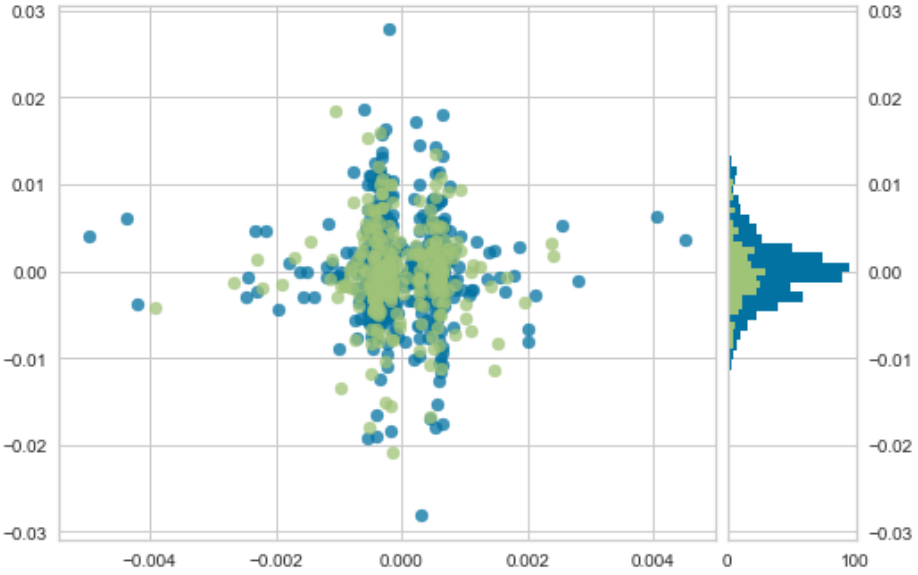
Taip pat valandinių XRP kainų pokyčių liekanos atitinka normalaus pasiskirstymo grafiką. (žr. 21 pav.).

Tarp visų kovariančių reikšmingiausia ir vėl buvo išrinkta VXX kainų pokyčių kovariantė. Antras pagal reikšmingumą buvo SPY kainų pokytis (žr. 22 pav.).

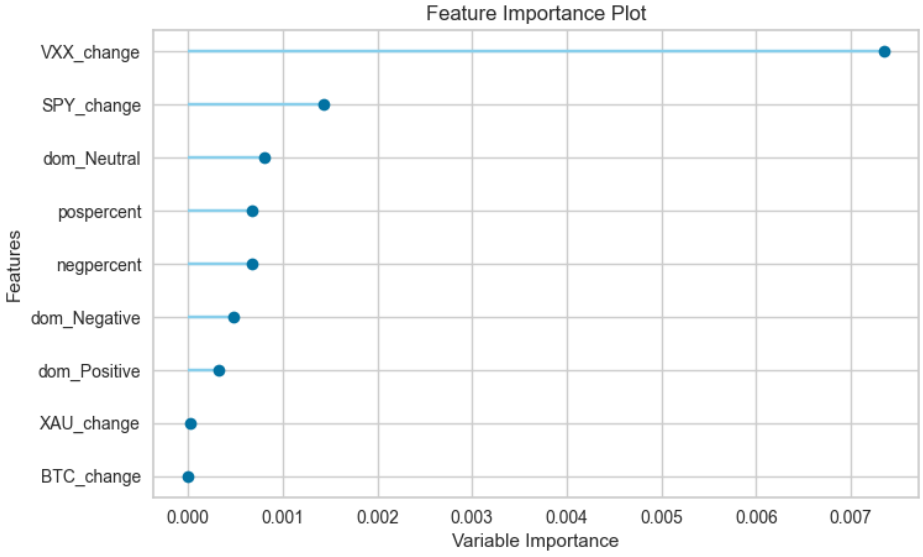
20 pav.: Valandinių XRP kainų pokyčių prognozavimas



21 pav.: Valandinių XRP kainų pokyčių liekanos



22 pav.: Valandiniai duomenys, kovariančių reikšmingumas



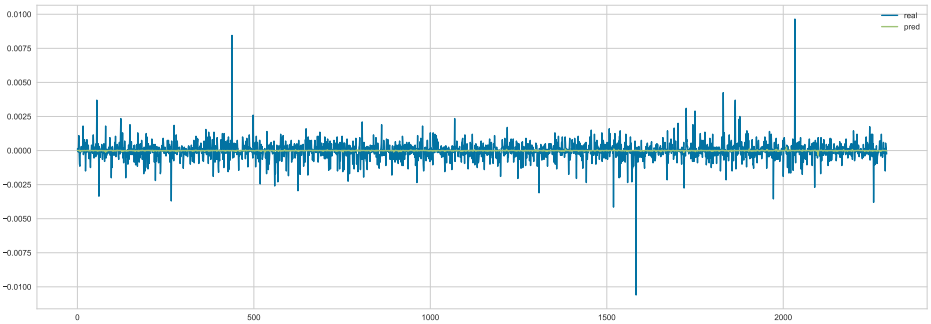
5.3 Minutiniai grafikai

Nors minutinių kainų duomenų rinkinys buvo didžiausias iš visų ir turėjo daug įrašų apie finansinių aktyvų kainas, netgi Pycaret geriausiam išrinktam modeliui nepavyko gerai prognozuoti XRP kainų pokyčius (žr. 23 pav.). Grafike matosi, jog prognozuojamos reikšmės visur labai arti nulio, tai reiškia, kad modelis prognozuoja, kad kainos judesiai bus nežymus arba išvis nebus.

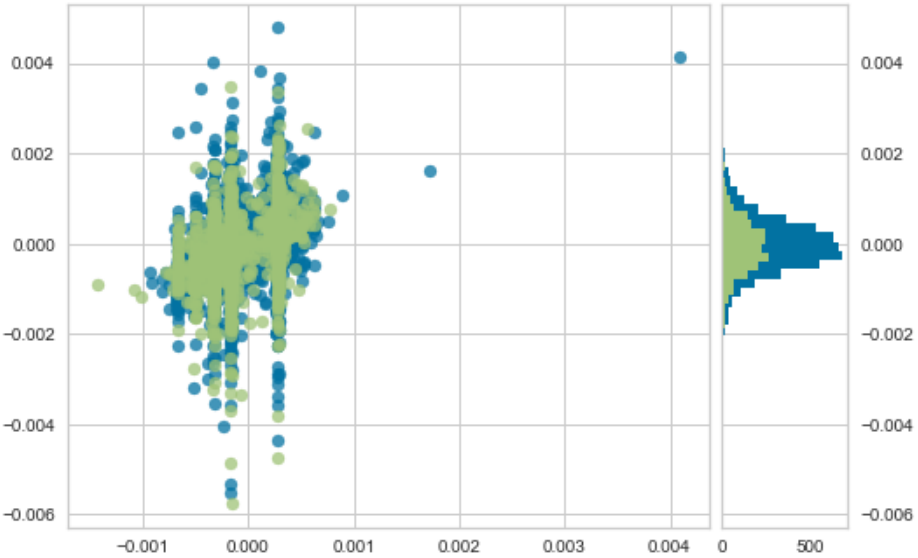
Kaip ir su valandiniais duomenimis, gavome, kad minutinių duomenų liekanos atitinka normaliai pasiskirsčiusio kintamojo grafiką (žr. 24 pav.).

Su minutiniais duomenimis (kaip ir su valandiniais, ir su dieniniais) VXX kainų pokytis turi didžiausią įtaką XRP kainų svyravimams (žr. 25 pav.).

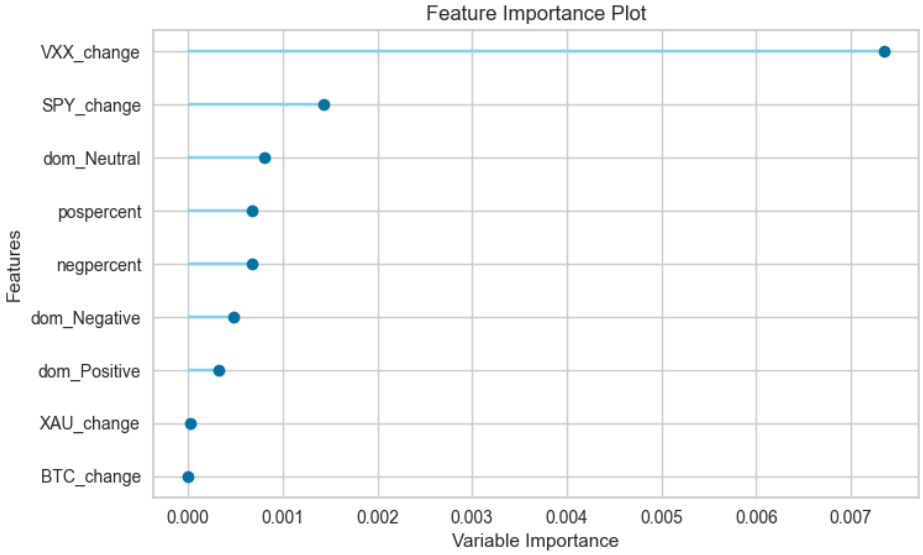
23 pav.: Minutinių XRP kainų pokyčių prognozavimas



24 pav.: Minutinių XRP kainų pokyčių liekanos



25 pav.: Minutiniai duomenys, kovariančių reikšmingumas



6 Modelio sudarymas ir tyrimas

7 Išvados bei rekomendacijos

8 Priedai

A Pirmas priedas

B Antras priedas

Literatūra

[ABR17] Nashirah Abu Bakar and Sofian Rosbi. Autoregressive integrated moving average (arima) model for forecasting cryptocurrency exchange rate in high volatility environment: A new insight of bitcoin transaction. *International Journal of Advanced Engineering Research and Scinece*, 4, 11 2017.

[AEK+21] M. Eren Akbiyik, Mert Erkul, Killian Kaempf, Vaiva Vasiliauskaite, and Nino Antulov-Fantulin. Ask "who", not "what": Bitcoin volatility forecasting with twitter data, 2021.

[AHNI18] Jethin Abraham, Danny W. Higdon, Johnny Nelson, and Juan Ibarra. Cryptocurrency price prediction using tweet volumes and sentiment analysis. 2018.

[Ali20] Moez Ali. *PyCaret: An open source, low-code machine learning library in Python*, April 2020. PyCaret version 1.0.

[SJOL20] Otabek Sattarov, Heung Jeon, Ryumduck Oh, and Jun Lee. Forecasting bitcoin price fluctuation by twitter sentiment analysis. pages 1–4, 11 2020.

[SN18] Sima Siami-Namini and Akbar Siami Namin. Forecasting economics and financial time series: ARIMA vs. LSTM. *CoRR*, abs/1803.06386, 2018.