

Анализа утицаја развоја технологија вештачке интелигенције и криптовалута на цене графичких карти

- Маја Варга P2 5/2024, Никола Симић E2 17/2024 -

У овом документу описан је предлог пројекта из предмета Системи за истраживање и анализу података. Прва два поглавља дефинишу проблем и мотивацију за решавање истог. Нагон тога следи преглед релевантне литературе, скупа података, планиране методологије и евалуације решења.

1. Дефиниција проблема

Потребно је прикупити историју цена различитих модела графичких карти са тржишта Сједињених Америчких Држава, и ујединити је са подацима о потражњи наметнутој применама у сврху копања крипто валута (енг: *cryptocurrency mining*) и у сврху тренирања разних модела вештачке интелигенције. Тако формиран сет података потребно је анализирати, извести закључке и покушати извршити предикцију цене на основу кључних параметара.

2. Мотивација

У претходних десет година, поред већ устаљених примена графичких процесора две значајне области које захтевају снагу графичких процесора су доживеле велике напретке: крипто валуте и вештачка интелигенција. Копање крипто валута и тренирање разних модела из области вештачке интелигенције су почеле да имају значајан утицај на већу потражњу графичких карти, а тиме и на њихове цене. Анализом утицаја ових фактора на цене графичких карти могуће је направити предикцију цена која би пружила крајњим купцима и компанијама значајне информације о томе кад би било најбоље да изврше куповину.

3. Релевантна литература

[\[1\] Risyad Faisal Hadi, Siti Saadah, Didit Aditya \(2023\) „Forecasting of GPU Prices Using Transformer Method“](#)

Циљ рада: У раду је описан начин креирања предикције цене графичких карти са циљем пружања купцима бољи увид у потенцијалну промену цена компоненте да би знали да ли је добар тренутак да је купе.

Методологија: Модели који су коришћени су трансформер неуронских мреже, рекурентне неуронске мреже и LSTM (Long short-term memory) неуронске мреже. Скуп података је подељен на 80% тренинг скуп, 10% валидациони скуп и 10% тестни скуп. Извршена је предикција цене графичких карти за периоде од 8, 16 и 30 дана. Резултати између 3 модела су упоређени коришћењем доле наведених метрика.

Скуп података: У овом раду коришћен је предефинисани скуп податка, преузет са интернета. У њему се налазе цене на дневном нивоу специфично једне графичке карте у периоду од септембра 2020. године до новембра 2022. године (конкретно 783 реда). Скуп података садржи 4 атрибута: датум, цену, прошлу цену и ручно додат атрибут будућа цена.

Евалуација решења: Добијени резултати су евалуирани коришћењем метрика *CC (Coefficient Correlation)*, *RMSE (Root Mean Squared Error)*, *MAPE (Mean Averaged Percentage Errors)*.

Резултати: Трансформер модел је имао већу прецизност у односу на друга два модела. Иако је *CC* метрика била иста на сва три, *RMSE* и *MAPE* метрике су биле значајно мање код трансформера ($RMSE=34.68$, $MAPE=0.82$, посматраном за период од 30 дана) што је указало на већу прецизност у односу на *LTSM* ($RMSE=65.37$, $MAPE=3.00$) и *RNN* ($RMSE=75.01$, $MAPE=3.51$).

Закључак: Иако је скуп података веома сиромашан, у раду је извршено квалитетно поређење модела за обраду *time-series* скупова података које је узето у обзир приликом избора алгоритама коришћених у овом раду.

[2] Linus Wilson (2022) „GPU prices and cryptocurrency returns“

Циљ рада: Рад истражује и анализира корелацију између цене крипто валута и цена графичких карти на секундарном тржишту (енг: *secondary market*).

Методологија: У раду је изведена је хипотеза: „*Only the most productive graphics cards as measured by secondary market price per hashrate will be significantly positively correlated with daily movements in the Ethereum (ETH) price*“ која је тестирана анализом прикупљених података. Коршћењем метода *OLS* регресије (*Ordinary least squares regression*) доказано је да промена цена карти које су погодне и ефикасне за копање *Ethereum* крипто валуте имају зависност од промене цене *Ethereum*-а, што је и описано у хипотези.

Скуп података: Како у току писања овог рада графичке карте серије RTX 3000 нису биле на стању, скуп података о ценама графичких карти на секундарном тржишту у периоду од 03.06.2021. до 01.09.2021. је ручно креиран користећи сајт *Stockx*. Подаци о крајњим ценама (енг: *closing price*) крипто валуте *Ethereum* преузети су користећи *Yahoo! Finance*. *Hashrate* графичких карти преузет је са сајта *Minerstat*.

Евалуација решења: Иако у раду није експлицитно наведен начин евалуације *OLS* регресије, претпоставка је да се корисила евалуација методом квадрата грешке која је стандардна евалуација за овај тип регресије.

Резултати: Анализа података указује на тачност хипотезе дефинисане у овом раду. Приказано је да повећање цене *Ethereum* крипто валуте за 1% утиче на пораст цене од 0.19% до 0.22% у зависности од модела графичке карте.

Закључак: Рад представља значајне информације које могу донети доприносе у виду проналажења кључних параметара над скупом података дефинисаним у овом раду. Мана рада је фокус искључиво на графичким картама компаније *Nvidia*.

[3] Oyewola, D. O., Dada, E. G., & Ndunagu, J. N. (2022). “A novel hybrid walk-forward ensemble optimization for time series cryptocurrency prediction”

Циљ рада: Развој и примена хибридног модела за предикцију цена 15 одабраних крипто валути

Методологија: Рад представља хибридни модел заснован на *stacking* ансамблу. Разматрани алгоритми у ансамблу су *ARIMA*, *SARIMA*, *HWES* (статистички модели), *Decision Tree*, *Stochastic*

Gradient Boosting, Random Forest (ML модели), *LSTM, GRU* и *RNN* (*deep-learning* модели). Модел врши предикције за све поменуте алгоритме, након чега бира најбољи из сваке поменуте групе и прави ансамбл од њих, који се користи за финалну предикцију цена крипто валути.

Скуп података: Скуп података је преузет са сајта *Yahoo! Finance* и садржи цене одабраних крипто валути у периоду од 01.01.2018. до 30.06.2021. За сваку валуту је прикупљено 1277 редова, што чини укупно 19155 редова за све крипто валуте. Скуп података садржи датум и коначну цену крипто валуте за тај датум. Скуп података је подељен на нивоу крипто валуте на тренинг (85%, од 01.01.2018. до 31.12.2020.) и тест скуп (15%, од 01.01.2021. – 30.06.2021.).

Евалуација решења: За евалуацију тачности предикција цене за сваку од разматраних крипто валути свих појединачних алгоритама кориштени су *MAE, RMSE* и *MSE*. За евалуацију и оптимизацију предикција коначног модела кориштен је *walk-forward* метод.

Резултати: Рад показује да међу статистичким моделима, *HWES* у 8 од 15 крипто валути показује најбоље резултате, међу *ML* моделима *Stochastic Gradient Boosting* показује најбоље резултате а међу *deep-learning* моделима *GRU*. Такође, рад истиче да је коначни ансамбл креиран на основу најбољих алгоритама постигао боље резултате од сваког појединачног разматраног алгоритма. У раду су табеларно приказане вредности евалуационих метрика за сваки разматрани алгоритам за сваку одабрану крипто валуту.

Закључак: У раду су детаљно описани различити алгоритми погодни за *time-series* скупове података, сличних оним у овом раду. Такође, рад описује примену *walk-forward* методе за евалуацију решења ових алгоритама, која ће бити примењена и у овом раду.

4. Скуп података

Скуп података ће бити ручно креиран. Подаци о моделима и ценама графичких карти биће *scrape*-овани са сајта [PCPartPicker](#) за период од 01.12.2022. до 01.12.2024. Број разматраних модела графичких карти је приближно 1200. Како су цене на овом сајту представљене на интерактивном графику који није могуће преузети, потребно је направити скрипту која *screenshot*-ује екран и провлачи слику кроз алгоритам за препознавање текста са слике употребом библиотеке [PyTesseract](#). Детаљне карактеристике за сваки модел графичке карте биће *scrape*-оване са сајта [TechPowerUp](#). Основни подаци који ће бити преузети су: *Base Clock, Boost Clock, Memory Clock, Memory Size, Memory Type, Memory Bus, Memory Bandwidth, TDP, L1 Cache, L2 Cache, Core Number*. Из разлога што новији модели графичких карти поседују додатне карактеристике битне за разматрање (*Cuda, Tensor cores, FLOPS*) и они ће бити преузети. Подаци о историјским ценама крипто валута биће преузети са сајта [CoinGecko](#) за период од 01.12.2022. до 01.12.2024. Подаци о профитабилности копања крипто валута за одређене карте ће бити преузети са сајта [WhatToMine](#). Перформансе (*Speedup*) графичких карти за потребе тренирања модела вештачке интелигенције биће преузете са сајта [LambdaLabs](#). Ови подаци ће бити прикупљени сличном методом као подаци о моделима и ценама графичких карти са сајта [PCPartPicker](#). Апроксимација тренда популарности (број од 1-100 који представља степен популарности појма претраге, о овом случају „*ai training*“.) тренирања вештачке интелигенције за период од 01.12.2022. до 01.12.2024 ће бити преузета са сајта [Google Trends](#).

5. Методологија

Рад се састоји од две повезане целине: 1. Анализа утицаја цена крипто валути на цену графичких карти и 2. Анализа утицаја тренда вештачке интелигенције на цену графичких карти. Да би био покривен шири спектар карактеристика које утичу на перформансе у копању крипто валута, одабране су две валуте: *Monero* (представник *core-intensive* алгоритама) и *Ethereum Classic* (представник *memory-intensive* алгоритама).

У оквиру прве целине, први корак је издвајање кључних карактеристика графичких карти које их чине погодним за копање крипто валути. Кључна мера перформансе графичких карти у овом случају је профитабилност, која је забележена за мањи број карти на сајту [WhatToMine](#). Стога, први циљ у оквиру овог корака је процена профитабилности свих осталих разматраних карти на основу постојећих података. Други корак јесте предикција цене графичке карте на основу историјских цена те карте и крипто валута за претходно наведени период. Крајњи циљ јесте поређење и анализа добијених резултата како би били изведени закључци о њиховим потенцијалним сличностима и међусобним утицајима.

У оквиру друге целине, први корак је издвајање кључних карактеристика графичких карти које их чине погодним за тренирање модела вештачке интелигенције. Кључна мера перформансе графичких карти у овом случају је убрзање (*Speedup*), која је забележена за мањи број карти на сајту [LambdaLabs](#). Стога, први циљ у оквиру овог корака је процена убрзања свих осталих разматраних карти на основу постојећих података. Други корак јесте предикција цене графичке карте на основу историјских цена те карте и апроксимације тренда тренирања вештачке интелигенције за претходно наведени период. Крајњи циљ ове целине је идентичан крајњем циљу претходне целине.

За потребе процене недостајућих вредности у првом кораку обе целине биће испробани регресиони алгоритми *Decision Tree* и *Random Forest*. За потребе предикције цене на основу историјских података о ценама и трендовима (*time-series*) биће коришћени алгоритми *LSTM* и *ARIMA*. За потребе анализе потенцијалних сличности и међусобних утицаја свих добијених резултата користиће се статистички тестови и алгоритми рачунања коефицијента корелације међу обележјима.

6. Метод евалуације

Резултате предикције профитабилности (у оквиру прве целине) и убрзања (у оквиру друге целине) биће евалуирани коришћењем метрика *RMSE* и *MAE*. На основи њих ће бити одабран коначан алгоритам који ће се користити за предикцију профитабилности и убрзања. За евалуацију резултата предикције цена графичких карти биће коришћен *walk-forward* метод коришћењем *RMSE* и *MAE* метрике за рачунање грешке између стварне и предиктоване цене. Скуп података ће бити подељен на тренинг, валидациони и тестни скуп у односу 80:10:10.