

Porovnanie algoritmov strojového učenia pre tvorbu predikčného modelu ceny bitcoinu

Milan Cibula, Michal Tkáč 

Podnikovohospodárska fakulta so sídlom v Košiciach, Ekonomická univerzita v Bratislave, Slovenská republika

E-mail: milan.cibula@euba.sk; michal.tkac1@euba.sk

Abstract

Comparison of Machine Learning Algorithms for Creation of a Bitcoin Price Prediction Model

With the advancement of machine learning tools, an increasing number of algorithms are being utilized for predicting not only traditional time series data related to financial markets but also those connected to cryptocurrencies. This paper aims to compare various machine learning algorithms used for prediction, in order to identify the one with the greatest practical potential for creating a prediction model of Bitcoin's price as an investment asset. The analysis focuses on supervised learning algorithms, taking into account the nature of the task involving long time series datasets. The paper also describes the exact process of creating and setting up individual models and their parameters, explaining procedures for obtaining and editing datasets, and shows how to evaluate performance of these models. In addition to the analysis of the main subject of research, which is Bitcoin, the paper also uses an analysis of reference cryptocurrencies such as Ethereum, Litecoin and NEO to compare the resulting performances. The processes consisting of editing the analysed datasets, creating individual prediction models, training and testing the performance of models on historical data, and creating, debugging and implementing individual machine learning models were realised through coding in the Python program.

Keywords: Machine learning, Bitcoin, prediction model

JEL Classifiaction: C22, C55, C88

Citation: Tkáč, M., Cibula, M. (2023). Porovnanie algoritmov strojového učenia pre tvorbu predikčného modelu ceny bitcoinu. *Politická ekonomie*, 71, (5), 496–517, <https://doi.org/10.18267/j.polek.1397>

Copyright: © 2023 by the author(s). Licensee Prague University of Economics and Business, Czech Republic. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution License (CC BY NC ND 4.0).

1. Úvod

Trh s kryptomenami, jeho unikátne vlastnosti, popularita a nestálosť podnietila záujem investorov, obchodníkov a výskumníkov o dôkladné štúdium tohto fenoménu, pričom rozmach transakcií Bitcoinu je toho najvýznamnejším príkladom. Na druhej strane v posledných rokoch sa aj algoritmy strojového učenia ukázali ako mocný nástroj na analýzu zložitých dátových súborov a tvorbu relatívne presných predikcií, čo bezpodmienečne viedlo k prepojeniu týchto dvoch rapídne rozvíjajúcich sa FinTech oblastí. Už s príchodom prvých počítačov sa investori a obchodníci snažili vytvoriť modely a obchodné systémy, ktoré by prekonal trh a boli automatizované, bez potreby ľudskej intervencie, čím by sa eliminovali ľudské emócie a chyby. Tradičné prístupy tvorby analýz, investičných modelov a predikcií budúceho vývoja boli poväčšine založené na neustálom manuálnom vylepšovaní pravidiel, algoritmov, vzťahov a výstupov, ktoré boli deterministicky dané užívateľom na základe minulých skúseností. Využitie umelej inteligencie a strojového učenia na druhej strane predstavuje odlišný spôsob vytvárania investičných modelov, predpovedí a investičných stratégií, pretože umožňuje vytvárať modely aj bez presne definovaných pravidiel. Výskum v tomto článku sa opiera o strojové učenie a umelú inteligenciu ako nástroje na vytvorenie predikčných modelov postavených na rozličných algoritmoch strojového učenia podľa ich inherentných vlastností. Jednotlivé modely sú postavené na základe dát historického vývoja ceny podkladového aktíva.

2. Prehľad literatúry

Tradičné metódy predpovedania cien kryptomien zvyčajne spočívali v použití štatistických a ekonometrických modelov. Ekonometrické prístupy kombinujú štatistické a ekonomické teórie s cieľom odhadnúť a predpovedať hodnoty rôznych ekonomických premenných. Pri skúmaní volatility a predpovedaní cien kryptomien pomocou ekonometrie sa výskumníci všeobecne spoliehali na štatistické modely pre časové rady údajov (Khedr *et al.*, 2021). Sovbetov (2018) napr. použil rozšírený Dickey-Fullerov test jednotkových koreňov na preskúmanie faktorov ovplyvňujúcich cenu kryptomien s použitím týždenných údajov za obdobie 8 rokov. Výsledky odhalili, že trhovú beta, objem obchodovania a volatilita ovplyvňujú ceny všetkých piatich kryptomien v krátkodobom aj dlhodobom horizonte. Roy *et al.* (2018) s použitím ročných údajov o bitcoine z rokov 2013 až 2017, použili modely časových radov (autoregresný integrovaný model s kĺzavým priemerom (ARIMA), autoregresný model a model využitím kĺzavých priemerov) na prognózovanie ceny bitcoinu. Zistili, že ARIMA model bol najlepším modelom na predpovedanie ceny bitcoinu. Abu Bakar *et al.* (2019) použili na predpovedanie ceny bitcoinu metódu kĺzavého priemeru o rôznej dĺžke od dvoudňového až po sedmidňový kĺzavý priemer. Výsledky ukázali,

že metóda dvoudňového kľzavého priemeru je lepšou metódou predpovede s najnižším percentom priemernej absolútnej chyby pre všetky obdobia pozorovania. Kjærland *et al.* (2018) použili ekonometrické metódy reprezentované autoregresným modelom s distribuovaným oneskorením a GARCH modelom na štúdium determinantov dynamiky ceny bitcoinu. Výsledky odhalili, že výnosy z indexu S&P 500 majú zásadný význam pri vysvetľovaní dynamiky cien Bitcoinu. Model ARIMA prezentoval Alahmari (2019) na predpovedanie troch hlavných kryptomenových cien (Bitcoin, XRP a ethereum) na základe denných, týždenných a mesačných časových radov. Model ARIMA prekonal ostatné modely z hľadiska strednej kvadratickej chyby (MSE), MAE a RMSE – koreňovej strednej kvadratickej chyby (Khedr *et al.*, 2021).

Chen *et al.* (2020) použili na predpovedanie ceny Bitcoinu lineárny model a algoritmy strojového učenia. Zistili, že lineárne modely boli lepšie ako ostatné modely strojového učenia. Poonodi *et al.* (2020) porovnávali lineárne modely a model SVM (Support Vector Machines) pri predpovedaní ceny Etherea a zistili, že lineárne modely sú lepšie ako model SVM. Akyildirim *et al.* (2021) porovnávali vplyv lineárnych modelov, SVM, neurónových sietí a ansámblových modelov na predikciu kryptomien. Výsledky ich výskumu naznačovali, že algoritmy strojového učenia by mohli robiť krátkodobé predpovede kryptomien. Zistili, že investície založené na algoritmoch strojového učenia dosahovali lepšie výsledky ako náhodný výber (Ren *et al.*, 2022). Na druhú stranu autori Hiransha *et al.* (2018) analyzovali predikčnú výkonnosť hlbokých neurónových sietí (MLP, RNN, LSTM a CNN) na denných zatváracích cenách dvoch akciových trhov, pričom model ARIMA bol použitý ako benchmark. Vo výsledku neurónové siete prekonal spomínaný benchmark. Ďalším typom umelej inteligencie, ktorá sa využívala na predikciu kryptomien, bola oblasť spracovania prirodzeného jazyka (NLP). Ding *et al.* (2015) navrhli hlbokú neurónovú sieť (DNN) riadenú udalosťami. Výsledky modelu naznačovali takmer 6% zlepšenie v porovnaní s najmodernejšími základnými metódami. McNally *et al.* (2018) použili špeciálne typy neurónových sietí RNN a LSTM na predpovedanie ceny finančne inovatívneho aktíva Bitcoin. Výskum sa zaoberal dennými uzatváracími cenami BTC/USD, ktoré boli získané ako priemer hodnôt z piatich hlavných Bitcoinových burz. Okrem dennej uzatváracie ceny zahrnuli vstupné premenné aj informácie o otváracích, najvyšších a najnižších cenách, ako aj špecifické údaje pre Bitcoinovu sieť (mining difficulty a hash rate) a dva jednoduché kľzavé priemery (SMA). LSTM model dosiahol najvyššiu presnosť klasifikácie na úrovni 52 % a RMSE 8 %. Populárny model SARIMA na predpovedanie časových radov bol použitý ako benchmark na porovnanie s modelmi hlbokého učenia. Výsledky ukázali, že nelineárne metódy hlbokého učenia prekonal model SARIMA, ktorý mal slabý výkon s hodnotou RMSE na úrovni 53,74 %.

Pri rozbere literatúry zameranej na predikciu ceny Bitcoinu existuje viacero prác využívajúc textové údaje z platforiem sociálnych médií a iných zdrojov. Autori Georgoula *et al.* (2015) skúmali sentiment pomocou SVM spolu s frekvenciou zobrazení na Wikipédii a hash rate siete.

Matta, Lunesu a Marchesi (2018) analyzovali vzťah medzi cenou Bitcoinu, na sociálnej sieti Twitter a zobrazením pre slovo Bitcoin na Google Trends.

Podľa autorov Li, Zheng a Dai (2020) jedným z obmedzení takýchto štúdií je však často malá veľkosť vzorky a sklon k šíreniu dezinformácií prostredníctvom rôznych (sociálnych) mediálnych kanálov, akým je Twitter alebo na diskusných fórach, ako je Reddit, ktoré sú využívané na nekalé praktiky ovplyvňovania cien kryptoaktív. Nakoľko je likvidita na Bitcoinových burzách značne obmedzená a trh kryptoaktív je neregulovaný, existuje vysoké riziko manipulácie. Z tohto dôvodu autori sentiment zo sociálnych médií ďalej neberú do úvahy.

Predikciu ceny Bitcoinu môžeme považovať za analogickú s ostatnými finančnými úlohami časových radov, akými sú už spomenuté práce na predikciu akcií, ale aj ostatných investičných aktív obchodovaných na finančných trhoch. Niekoľko autorov ako Persio a Honchar (2016), Gao, Zhang a Yang (2020) a Hiransha *et al.* (2018) využívajú na predikciu cien modely Multi Layer Perceptron (MLP). MLP však analyzuje iba jedno pozorovanie naraz. Na druhej strane výstup z každej vrstvy rekurentnej neurónovej siete (RNN) sa ukladá do kontextovej vrstvy, ktorá je následne prepojená s výstupom z nasledujúcej vrstvy. Vďaka tomu má sieť pamäť, na rozdiel od MLP, a umožňuje LSTM sieti pamätať si aj dlhodobé závislosti. Konštrukciou modelov LSTM alebo konštrukciou zložitejších modelov založených na modeloch LSTM výskumníci zistili, že modely LSTM sú veľmi presné a robustné. Ren *et al.* (2022) Avšak autori poukazujú aj na nevýhodu RNN a LSTM sietí, a to ich výpočtovú náročnosť. Napríklad 50dňová LSTM sieť je porovnateľná s tréningom 50 jednotlivých MLP modelov. Autori Menculini *et al.* (2021) pri predikovaní ceny využívajú model Auto-ARIMA, ktorý porovnávajú s Prophetom, škálovateľným prognostickým nástrojom od Facebooku založeným na zovšeobecnenom aditívnom modeli, s modelmi hlbokého učenia využívajúcimi dlhodobú krátkodobú pamäť (LSTM) a konvolučné neurónové siete (CNN). Modely ARIMA sa často používajú v ekonometrických analýzach a poskytujú dobrý štandard pre skúmaný problém. Výsledky autorov naznačujú, že model Auto-ARIMA a neurónové siete LSTM fungujú podobne pri posudzovanej prognostickej úlohe, zatiaľ čo kombinácia CNN a LSTM dosahuje najlepšiu celkovú presnosť, ale vyžaduje si viac času na výpočet, ladenie. Naopak, Prophet je rýchly a jednoduchý na používanie, no podstatne menej presný.

Na základe charakteru úlohy analýzy dlhých časových radov sú v práci používané algoritmy strojového učenia SARIMA, Support Vector Machines (SVM), Prophet, Long Short-Term Memory (LSTM), DeepAR predstavujúce algoritmy kategórie Supervised learning.

3. Metodológia

Na získanie historických cien nami analyzovaného investičného aktíva Bitcoin, ako aj ostatných referenčných kryptomien využitých za účelom porovnania výsledkov analyzovaných algoritmov

strojového učenia sme využili stránku cryptodatadownload.com. Daný zdroj ponúka po registrácii možnosť stiahnutia datasetov hodinových historických cien rozličných kryptoaktív obchodovaných na viacerých kryptoburzách na základe nimi poskytnutých API (Application Program Interface).

Za účelom dosiahnutia čo najvyššej presnosti výpovednej hodnoty analýz vykonaných prácou, historické údaje o vývoji cien sme čerpali z kryptoburzy Binance, ktorá predstavuje jednu z najväčších kryptobúr z hľadiska množstva obchodovaných mien, tokenov ako aj objemu uzatvorených obchodov.

Pre získanie širšieho obrazu správania sa jednotlivých algoritmov bola, analýza s využitím výhradne historických cien rozdelená na 2 časti, a to na obdobie pred covid-19 a obdobie ovplyvnené pandémiou covid-19. Daným rozdelením dataset „before“ (pred) obsahuje dáta za obdobie od 1. 1. 2018, 0:00 až po 31. 12. 2019, 23:00, čo predstavuje dva súvislé roky. Vzhľadom na fakt, že analyzujeme hodinové dáta a kryptomeny sú obchodované na rozdiel od akcií, či FIAT mien 24/7, daný dataset obsahuje 17 520 údajov historických cien. Dataset „after“ (po covid-19 alebo počas) obsahuje dáta so začiatkom 1. 1. 2020, 0:00 a koncom 30. 12. 2021 0:00 obsahujúci 17 479 záznamov hodinových uzatváracích cien za účelom dosiahnutia porovnateľnosti rozsahu datasetov pred a počas.

Pri získaní spomínaných datasetov bolo potrebné odstrániť nepotrebné časti dát, ktoré by zbytočne ovplyvňovali komputačný čas jednotlivých modelov. Datasets obsahujú špecifické identifikačné označenie definované konkrétnou burzou tzv. „unix“, dané informácie sú pre nás nepodstatné a odstránime ich spolu so stĺpcom „tradedcount“, ktorý v datasetoch pred pandémiou mali hodnotu „NULL“. Rovnako je potrebné prevrátiť datasety na zabezpečenie zostupného poradia dátumov.

Obrázok 1: Úvodná úprava datasetu historických cien

```
In [4]: df = pd.read_csv(dataset) # načítať dáta
df.drop_duplicates(ignore_index=True, inplace=True)

# vymazanie nepotrebných údajov datasetu
del df["unix"]
del df["tradedcount"]

# reverse data - zabezpečiť časovú postupnosť
df = df[::-1].reset_index(drop=True)

# formátovať datetime
df["date"] = pd.to_datetime(df["date"], format="%d/%m/%Y %H:%M")
df.head()
```

Zdroj: vlastné spracovanie v Python 3.8.0

Výsledné dáta, s upraveným poradím a odstránením špecifických údajov zmenárne pre označenie položiek datasetu, obsahujú informácie o hodinových otváracích, najvyšších, najnižších a uzatváracích cenách Bitcoinu (referenčných aktív), informácie o objeme uzatvorených obchodov vyjadrených v množstve obchodovaného podkladového aktíva, ako aj v hodnotovom vyjadrení v USD pre analyzovaný časový rozsah, časové okno (timeframe). Všetky údaje o cenách a objeme obchodovania predstavujú hodnoty pre časové pásmo GMT. Náhľad tvaru našich dát upravených v programe Python je možné vidieť na obrázku 2.

Obrázok 2: Dataset BTC 1. 1. 2020, 0:00–4:00

Out[4]:

	date	symbol	open	high	low	close	Volume BTC	Volume USDT
0	2020-01-01 00:00:00	BTC/USDT	7195.23	7196.25	7175.46	7177.02	480.93	3453949.34
1	2020-01-01 01:00:00	BTC/USDT	7177.02	7230.00	7175.71	7216.27	836.51	6030300.92
2	2020-01-01 02:00:00	BTC/USDT	7216.27	7244.87	7211.41	7242.85	626.60	4530216.13
3	2020-01-01 03:00:00	BTC/USDT	7242.85	7245.00	7220.00	7225.01	748.76	5414508.65
4	2020-01-01 04:00:00	BTC/USDT	7225.01	7230.00	7215.03	7217.27	442.09	3193315.81

Zdroj: vlastné spracovanie v Python 3.8.0

Jednotlivé algoritmy budú pri využívaní a tvorbe modelov vytvárať predikcie pre uzatváracie ceny za nami zvolený časový horizont. Čiže každá predikovaná hodnota predstavuje hodnotu hodinovej uzatváracie ceny. Týmto pre tréovanie a správne fungovanie modelov algoritmy potrebujú mať k dispozícii dátum/čas a k nim prislúchajúcu hodnotu uzatváracie ceny (closing price).

Obrázok 3: Výsledný počet záznamov BTC 1. 1. 2020, 0:00–30. 12. 2021, 0:00

```
In [5]: # ponechať iba uzatváracie ceny a dátum/čas
df = df[["date", "close"]]
df.rename(columns={"date": "ds", "close": "y"}, inplace=True) # premenovať na požadovaný formát
df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 17479 entries, 0 to 17478
Data columns (total 2 columns):
#   Column  Non-Null Count  Dtype
---  -
0    ds      17479 non-null    datetime64[ns]
1    y        17479 non-null    float64
dtypes: datetime64[ns](1), float64(1)
memory usage: 273.2 KB
```

Zdroj: vlastné spracovanie v Python 3.8.0

Ako referenčné meny pre porovnanie so subjektom skúmania Bitcoin sme si vybrali kryptomeny Ethereum, Litecoin a NEO. Vzhľadom na fakt, že Bitcoin predstavuje najväčšieho predstaviteľa trhu kryptomien z hľadiska trhovej kapitalizácie, pri výbere referenčných kryptomien sme sa zamerali na výber takých mien, ktoré by mali čo najvyššiu trhovú kapitalizáciu so zámerom dosiahnutia, čo najlepších hodnôt pre porovnanie správania sa jednotlivých algoritmov.

Pri realizácii výskumu (leto 2022) by sme pri abstrahovaní tzv. stablecoinov, t. j. kryptomien, ktorej hodnoty sú viazané ku forme kolaterálu v pomere 1:1 (Grobys *et al.*, 2021) využili meny, akými sú BNB, Cardano, Solana či TRON, ktoré dnes kumulujú najvyššiu trhovú hodnotu v kryptosfére po Bitcoine, avšak nakoľko je kryptosféra vysoko dynamickým prostredím podiel trhovej kapitalizácie sa relatívne dramaticky menil počas sledovaného obdobia od roku 2018 do súčasnosti. Jednotlivé kryptoburzy začali so zbieraním, resp. ponúkaním záznamov o historických cenách na základe záujmu, trhovej kapitalizácie daných mien v čase. Práve preto sme sa pri výbere referenčných mien rozhodli pre Ethereum, Litecoin a NEO, pre ktoré burzy ponúkajú záznam vývoja historických cien pre nami požadované obdobie, nakoľko všetky dané meny patrili do top 10 kryptomien z hľadiska trhovej kapitalizácie v roku 2018 a v súčasnosti je stále Ethereum na 2. mieste, Litecoin na 19. mieste a NEO na mieste 64.

4. Nastavenie parametrov algoritmov strojového učenia

Použitie údajov bez predošlej normalizácie v algoritmoch využívajúcich akýkoľvek druh opakujúcich sa buniek predstavuje veľký problém. Algoritmy strojového učenia predpokladajú, že číselné dáta sa riadia gaussovským rozdelením pravdepodobnosti, čo však nemusí byť pravda pre všetky dáta, čím sa algoritmy strojového učenia bez predošlého spracovania údajov môžu stať nestabilnými a poskytovať suboptimálne výsledky (Zhang, Qi, 2005).

V dôsledku toho sme pri tvorbe jednotlivých algoritmov normalizovali údaje o vývoji podkladového aktíva v čase pomocou Power Transformer Scikit-Learn Python knižnice strojového učenia. Power Transformer, akým je Box-Cox transformácia, predstavuje automatický spôsob vykonávania týchto normalizácií údajov, čím nám umožní dosiahnuť lepší výkon v širokom rozsahu algoritmov strojového učenia transformáciou údajov na gaussové resp. viac gaussové rozdelenie (Zheng, Casari, 2018). Optimálny parameter na stabilizáciu rozptylu a minimalizáciu šikmosti sa odhaduje pomocou maximálnej pravdepodobnosti. Box-Cox vyžaduje, aby vstupné údaje boli striktné pozitívne. Štandardne sa na transformované údaje použije normalizácia s nulovým priemerom (t. j. priemer a očakávaná hodnota štandardného normálneho rozdelenia je nula) a jednotkovou odchýlkou čo znamená štandardná odchýlka vzorky, ako aj rozptyl budú mať tendenciu smerovať k hodnote 1, keď veľkosť vzorky smeruje k nekonečnu (Scikit-learn, 2022). Autori Makridakis, Spiliotis a Assimakopoulos (2018) dosiahli najlepšie hodnoty normalizácie

v nimi zameranom článku na prístupy normalizácie dát pre algoritmy strojového učenia práve využitím Box-Cox transformácie. My sme v našom prípade rovnako vedeli aplikovať Box-Cox transformáciu, keďže naše dáta sú striktne pozitívne, pretože obsahujú údaje o vývoji ceny v čase, ktorá nikdy nenadobudla záporné alebo nulové hodnoty. Práve využitím Box-Cox transformácie sme dosiahli vyššiu hodnotu presnosti v porovnaní s využitím Yeo-Johnson transformácie, čím v práci pre normalizáciu dát historických cien podkladových aktív využívame transformáciu Box-Cox.

Údaje o historickom vývoji cien analyzovaných aktív sme štrukturovali do okien, pričom dĺžka časového okna nám definuje frekvenciu predikcií algoritmu. Každé okno predstavuje zoznam údajov o dĺžke 32 uzatváracích hodinových cien (okno 32 reprezentuje dĺžku medzi jednotlivými predpoveďami algoritmu), pričom hodnota 33 je ponechaná na predikciu. Hodnota dĺžky okna na úrovni 32 predstavovala výsledok ladenia daného hyperparametra, kde algoritmus hľadal najlepšie výstupné hodnoty v rozmedzí 8 až 128 časových okien, v našom prípade hodín.

Dáta sú rozdelené na tréningové a testovacie dáta, kde na testovanie je využitých posledných 1024 hodín príslušného datasetu. Tréningové údaje sú použité na prispôsobenie, učenie modelu. Testovacie údaje sú použité na poskytnutie nezáujatého hodnotenia konečného modelu prispôbeného využitím tréningových údajov. Model ich pri učení nikdy nepoužíva. To znamená, že model nemá k dispozícii historické údaje vývoja ceny, ktoré sa snaží predpovedať, skôr než danú predikciu vykonať. Daná skutočnosť má za následok, že akékoľvek predpovede predstavujú dobrý odhad skutočného výkonu modelov na akýchkoľvek vopred nedostupných údajoch rovnakej štruktúry.

Okrem toho sú tréningové dáta rozdelené na tréning a validáciu v pomere 0,9 k 0,1. Validačná vzorka predstavuje súbor údajov použitý na poskytnutie nezáujatého hodnotenia modelu prispôbeného množine tréningových údajov pri ladení hyperparametrov modelu. Model niekedy dané dáta používa, avšak nikdy sa z nich neučí, za účelom vyhodnotenia, ktorý model je počas tréningu najlepší, kde model využíva validačné údaje, aby zabránil overfittingu prostredníctvom náhodného výberu údajov z tréningovej množiny. Pojem „overfitting“ predstavuje model, ktorý príliš dobre modeluje tréningové dáta. Tento jav nastáva, keď sa model naučí špecifické charakteristiky a šum v tréningových dátach do takej miery, ktorá negatívne ovplyvňuje výkon modelu pri použití nových (počas tréningu nedostupných) dátach. To v praxi znamená, že šum alebo náhodné fluktuácie typické iba pre tréningové dáta sú identifikované a naučené modelom ako pravidlá. Takéto špecifické pravidlá, ktoré nefungujú v rámci nových údajov znižujú schopnosť modelov zovšeobecňovať, čo predstavuje podstatný nedostatok pri skúmaní nami analyzovaného problému, pretože zle naučené koncepty modelu strojového učenia by mohli predstavovať vysoké straty pri aplikácii modelu na reálne dáta a obchody (Drobetz, Otto, 2021).

Testovacie dáta predstavujú 1 024 hodnôt príslušného datasetu mínus 32 údajov predstavujúcich hodnotu 33 z každého okna využitú na predikcie. Hodnota 996 vzoriek (1024–32) je zaokrúhlená na najbližší násobok hodnoty batch, ktorá je v našom prípade nastavená na hodnotu 256, aby dané vzorky mohol model spracovať v batches (dávках), čím nám ostane 768 testovacích vzoriek dát. Dané platí pre algoritmy, pre ktoré je nevyhnutné poskytnúť modelu hodnotu batch size.

Tréning jednotlivých modelov je postavený tak aby využíval Adam Optimiser pre minimalizovanie MSE loss function. Adam Optimizer je technika, ktorá skracuje čas potrebný na tréningovanie modelu v hlbokom učení. Cesta učenia v mini-batch gradientnom zostupe do lokálneho minima nie je priama. Preto sa nejaký čas pri učení stráca prostredníctvom kľukatého vertikálneho pohybu. Adam Optimizer zväčšuje pri zostupe horizontálny pohyb a redukuje vertikálny, čím sa kľukatá cesta vyrovnáva a tým sa skracuje čas potrebný na tréningovanie modelu.

Jednotlivé modely rovnako obsahujú ohraničenie pre ukončenie všetkých neprimerane dlhých tréningov za predpokladu, že sa výkonnosť modelu s časom už prestala zlepšovať za účelom zabránenia overfittingu, čo nám rovnako umožní vykonať viacero skúšok pri ladení hyperparametrov v rámci časového ohraničenia.

Namiesto testovania všetkých možných kombinácií hyperparametrov, ktoré by exponenciálne rástli s rozsahom, v práci pre jednotlivé modely využívame na ladenie hyperparametrov voľne dostupný rámec optimalizácie Optuna, ktorá umožňuje využívanie najmodernejších algoritmov na vzorkovanie hyperparametrov a orezávanie neperspektívnych pokusov. To pomáha výrazne urýchliť čas a výkon optimalizácie v porovnaní s tradičnými metódami ako napríklad GridSearch. Optuna štandardne implementuje Bayesovský optimalizačný algoritmus (TPE), ktorý využívame pri definovaní optimálnych hyperparametrov aj my (Optuna, 2022).

Obrázok 4: Optuna pre hyperparametre pri LSTM

```
In [10]: def objective(trial):
# generuj hyperparametre použitím optuna
n_batch = trial.suggest_int("log2_batch_size", 4, int(np.log2(max_batch)))
n_epochs = trial.suggest_int("epochs", 2, 50)
n_neurons = trial.suggest_int("neurons", 1, 32)
dropout = trial.suggest_float("dropout", 0, 0.5)
lr = trial.suggest_float("log_learning_rate", -3, -1)

# zhodnot' model na testovacích dátach
return run(log2_batch_size=n_batch, epochs=n_epochs, neurons=n_neurons, dropout=dropout, log_learning_rate=lr)
```

Zdroj: vlastné spracovanie v Python 3.8.0

Optuna má k dispozícii 100 pokusov, ktoré nám poskytnú komplexný pohľad na priestor hľadania a obsahuje časové ohraničenie na úrovni 7 200 sekúnd, čiže 2 hodiny. Model je následne znova spustený s najlepšimi hodnotami, ktoré Optuna našla pre jednotlivé hyperparametre.

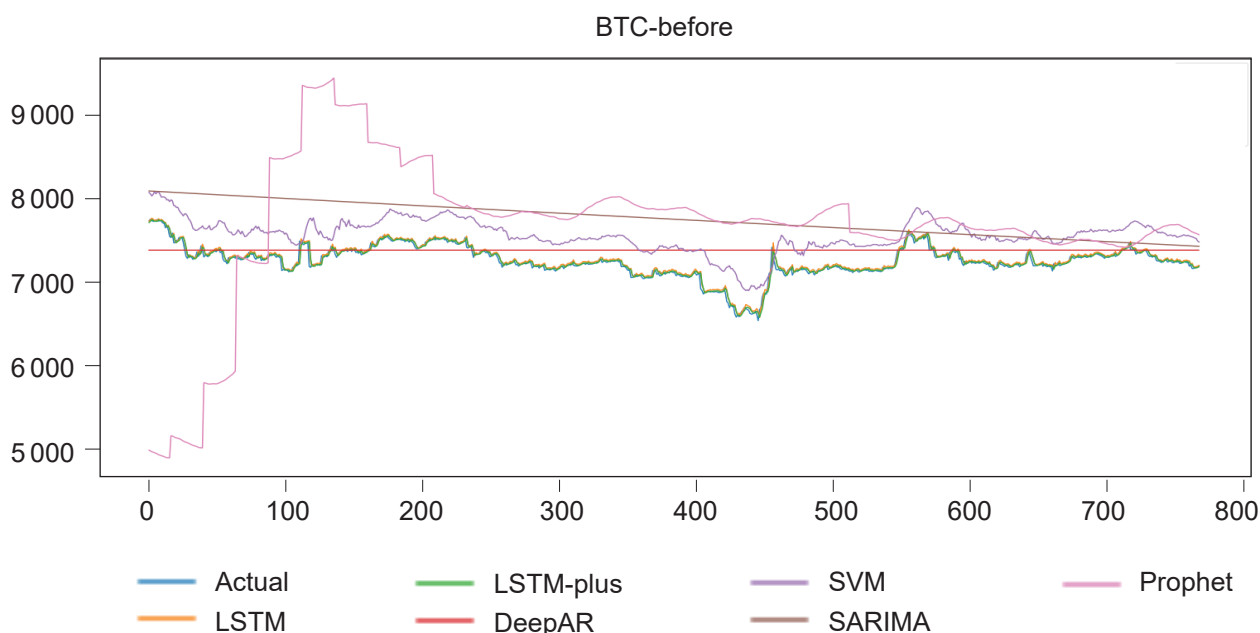
5. Výsledky práce

Na meranie presnosti predikcie (alebo chyby) modelov strojového učenia využívame MSE (Mean Squared Error), RMSE (Root Mean Squared Error), MAE (Mean Absolute Error) a MAPE (Mean Absolute Percentage Error), ktoré nám ako regresné metriky vyjadrujú typ vzdialenosti predikovaných hodnôt od hodnôt skutočných. Každá z metrik má svoje inherentné výhody a nevýhody plynúce z jej konštrukcie. Využitím viacerých metrik a schopnosti porovnania ich výsledných hodnôt dosiahneme ucelený obraz o výkonnosti našich modelov. Keďže ide o metriky chybovosti, ich výsledné hodnoty by mali byť čo najnižšie, pretože nižšia hodnota predstavuje menšiu odchýlku predikovaných hodnôt od reálnych cien trhu a vyššia výsledná hodnota metriky zas predstavuje väčšiu vzdialenosť predikcie od reality.

Obrázok 5 až obrázok 12 predstavujú porovnanie skutočného vývoja ceny analyzovaných podkladových aktív s predikciami o vývoji ceny skúmaných modelov. Os x predstavuje 768 testovacích vzoriek dát ako testovací dataset a os y ponúka informácie o cene daného aktíva.

Už na pohľad prostredníctvom heuristického prístupu môžeme na daných vývojjoch predikcii vidieť viacero problematických modelov. Modely postavené na algoritmoch strojového učenia SARIMA a Prophet zaznamenali naprieč aktívami a datasetmi pre nami stanovený cieľ nepoužiteľné výsledky.

Obrázok 5: Porovnanie výsledkov ML modelov pre historické ceny Bitcoin-before

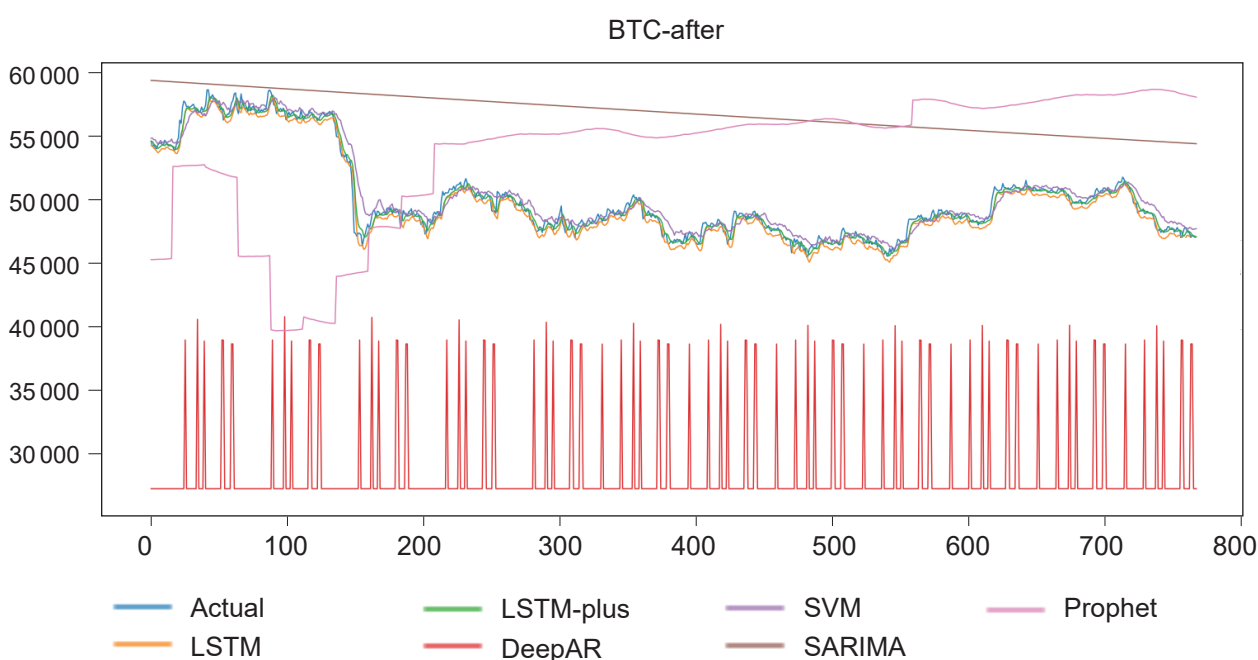


Zdroj: vlastné spracovanie v Python 3.8.0

Pri tvorbe modelu SARIMA máme nízku kontrolu nad tým, ako bude model pracovať, existuje len veľmi málo hyperparametrov s obmedzenou funkčnosťou. Model je jednoduchý na aplikáciu a poskytuje relatívne podrobné štatistiky, ktoré sú však z pohľadu nami analyzovaného problému nepodstatné.

Pohľadom na vývoj modelu SARIMA, výstupom na všetkých skúmaných datasetoch je priamka predpovedajúca všeobecný trend. Z hľadiska presnosti, či využiteľnosti je daný výsledok pre nás nepoužiteľný, nakoľko trading bot (automatizovaný obchodný model) nie je možné na základe takéhoto výstupu vytvoriť.

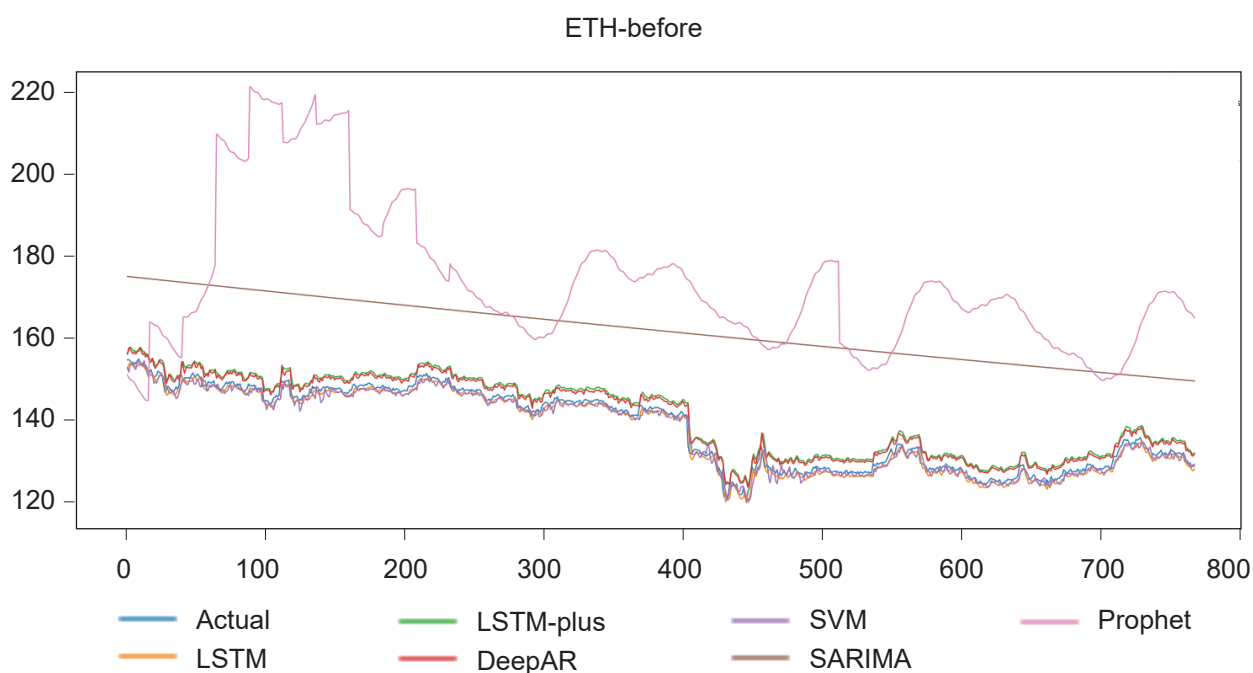
Obrázok 6: Porovnanie výsledkov ML modelov pre historické ceny Bitcoin-after



Zdroj: vlastné spracovanie v Python 3.8.0

Cieľom vývojárov Facebooku pre model Prophet Python API bolo zabezpečiť jeho dostupnosť čo najširšej verejnosti, čím je vytvorený aby bol čo najjednoduchší. Nie je tak zameraný na vývojárov umelej inteligencie a podobne ako pri modeli SARIMA, nemáme prakticky žiadne významné možnosti pri nastavovaní hyperparametrov (facebook.prophet, 2022).

Ako môžeme vidieť na rozdiel od modelu SARIMA, ktorého výstup ponúka aspoň informáciu o trende z dlhodobého hľadiska, výstup modelu Prophet nám poskytuje predikcie bez akejkoľvek výpovednej hodnoty naprieč analyzovanými datasetmi. Napriek zameraniu algoritmu na modelovanie dát časových radov, potrebuje model dataset zo silnými sezónnymi vplyvmi, ktoré naše datasety vývoja ceny podkladových aktív neobsahujú.

Obrázok 7: Porovnanie výsledkov ML modelov pre historické ceny Ethereum-before

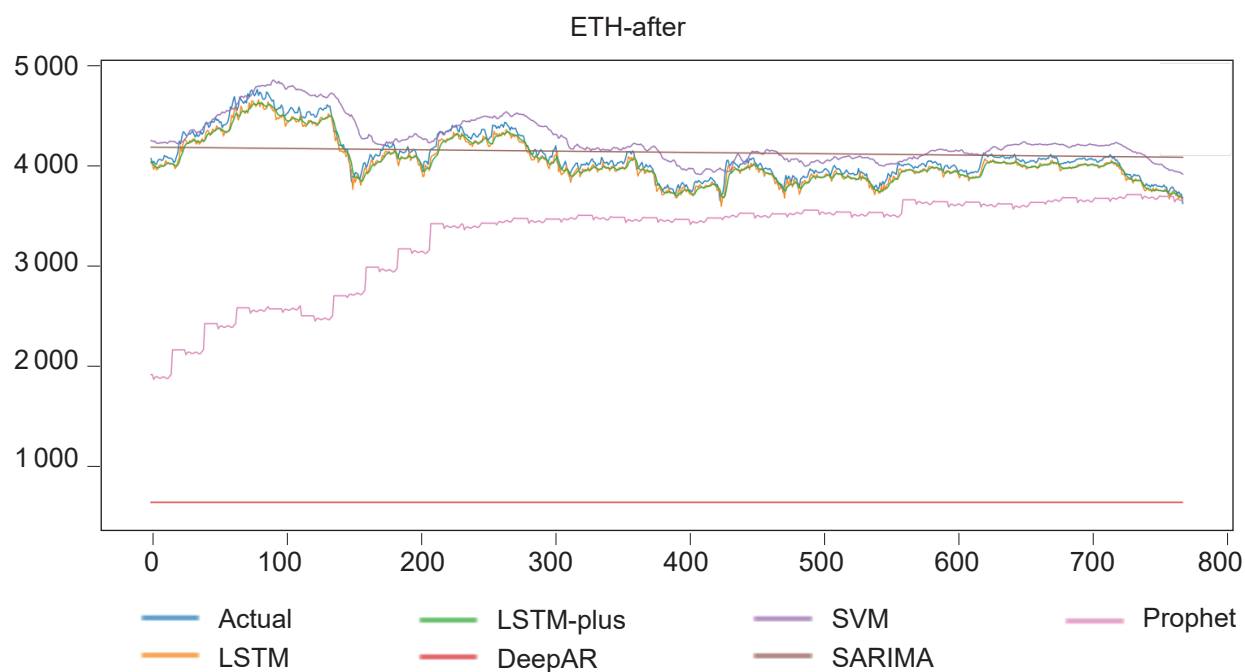
Zdroj: vlastné spracovanie v Python 3.8.0

Ďalším modelom, ktorého výsledky sú veľmi zaujímavé, je algoritmus DeepAR. Daný algoritmus dokázal dosiahnuť veľmi vysoké hodnotenie pri vybraných datasetoch, avšak na druhej strane si nedokázal poradiť s ostatnými analyzovanými súbormi údajov.

Na základe definície algoritmu vývojovým tímom Amazonu by mal DeepAR predstavovať podstatné zlepšenie výkonnosti predikcií časových radov v porovnaní s tradičnými metódami (Amazon, 2019). Vyžaduje viac údajov na tréning ako napríklad SARIMA, či Facebook Prophet. Vychádza z architektúry RNN a využíva bunky LSTM. Aj napriek viacerým pokusom sa neschopnosť daného modelu poskytovať konzistentné výsledky na všetkých datasetoch neodstránila.

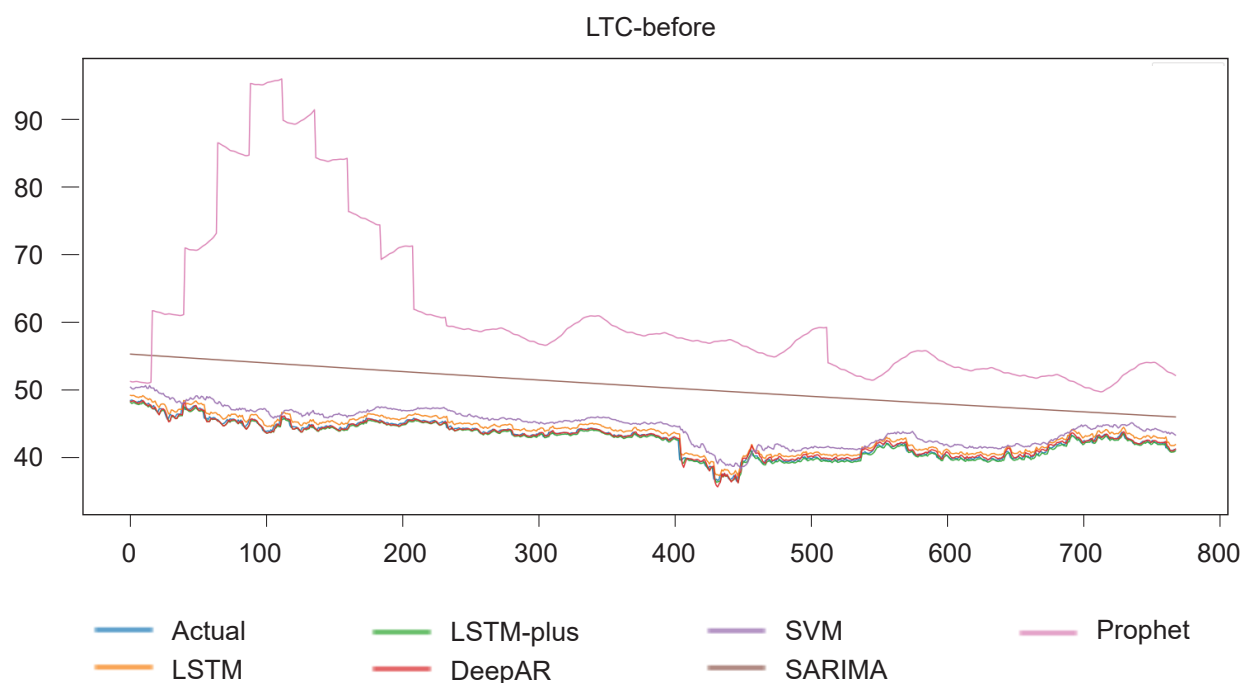
Model postavený na algoritme SVM strojového učenia zaznamenal podstatné zlepšenie hodnôt výsledných metrík v porovnaní s algoritmom SARIMA, či Prophet. Rovnako model SVM zaznamenal výrazné zvýšenie konzistentnosti výsledných hodnôt naprieč datasetmi v porovnaní s algoritmom DeepAR.

Obrázok 8: Porovnanie výsledkov ML modelov pre historické ceny Ethereum-after



Zdroj: vlastné spracovanie v Python 3.8.0

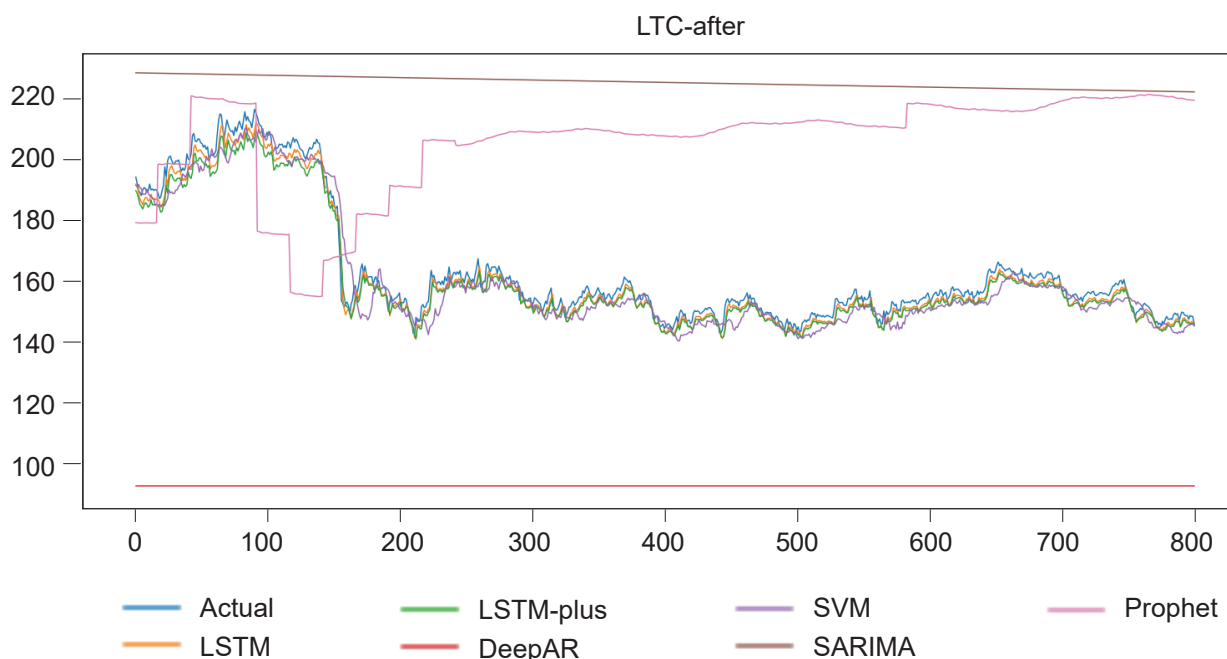
Obrázok 9: Porovnanie výsledkov ML modelov pre historické ceny Litecoin-before



Zdroj: vlastné spracovanie v Python 3.8.0

Výsledné hodnoty modelu SVM sú použiteľné z hľadiska reálnej aplikácie, nie je možné pozorovať žiadne zjavné oneskorenie predikovaných hodnôt od reálnych uzatváracích cien, model neobsahuje známky under- alebo over-fittingu, zvolené hyperparametre pri trénovaní modelu boli správne vzorkované pomocou optuna a rovnako model nezaznamenal problematické výsledky, ktoré by predstavovali predikované hodnoty bez zmeny voči minulému obdobiu.

Obrázok 10: Porovnanie výsledkov ML modelov pre historické ceny Litecoin-after



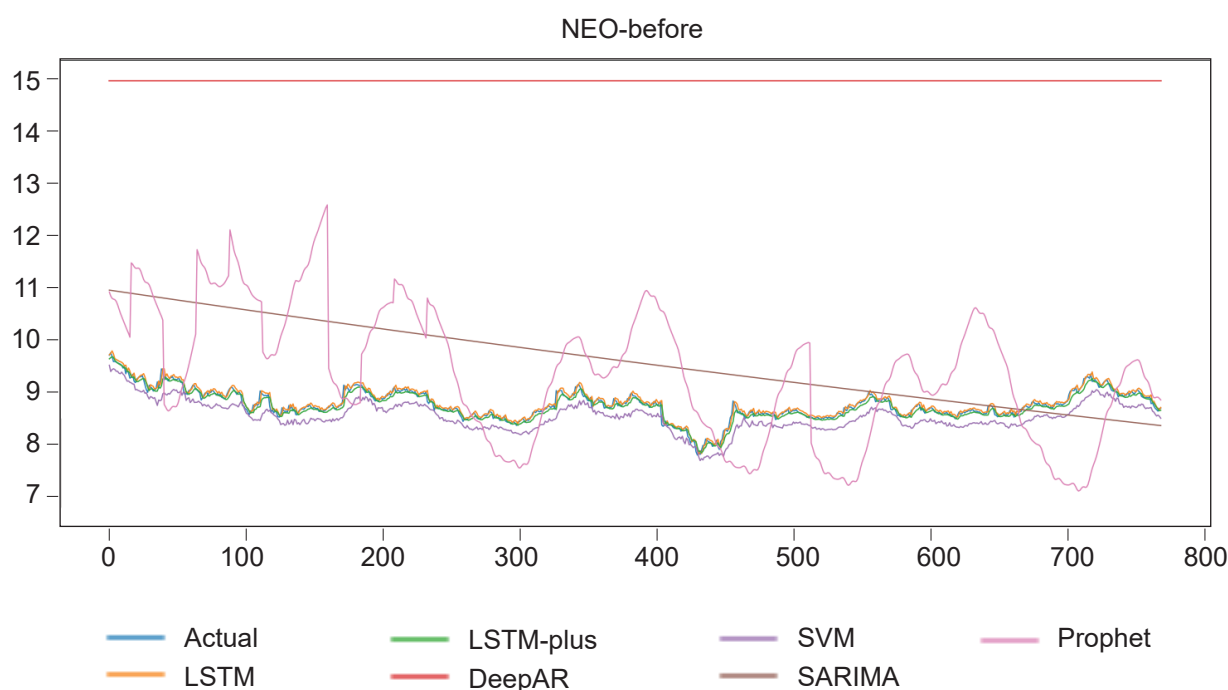
Zdroj: vlastné spracovanie v Python 3.8.0

Posledným testovaným algoritmom strojového učenia je LSTM. Už na prvý pohľad zaznamenal daný model pre všetky datasety vysokú presnosť predpovede. Predikcie sú blízko skutočným hodnotám a model dobre chápe spôsob, akým sa cena mení.

Daný model LSTM predstavuje jednobunkovú sieť nasledovanú jednou vrstvou dense. Vrstva dense predstavuje plne prepojenú vrstvu, ktorá je využívaná v záverečných fázach neurónovej siete. Táto vrstva pomáha pri zmene dimenzionality výstupu z predchádzajúcej vrstvy, čo pomáha modelu jednoducho definovať vzťah medzi hodnotami údajov, s ktorými model pracuje. Medzi optimalizované hyperparametre modelu patrí množstvo epoch a neurónov, batch size, hodnota dropout a learning rate. Vzhľadom na vysoký potenciál modelu LSTM na základe vývoja predikovaných hodnôt modelu, ako aj na problémy s modelom postavenom na algoritme DeepAR, ktorý je postavený práve na architektúre RNN a konkrétne LSTM, sme sa rozhodli vytvoriť ďalší model, ktorý bude predstavovať kombináciu nami testovaných modelov LSTM a DeepAR, ktorý budeme označovať ako LSTM+.

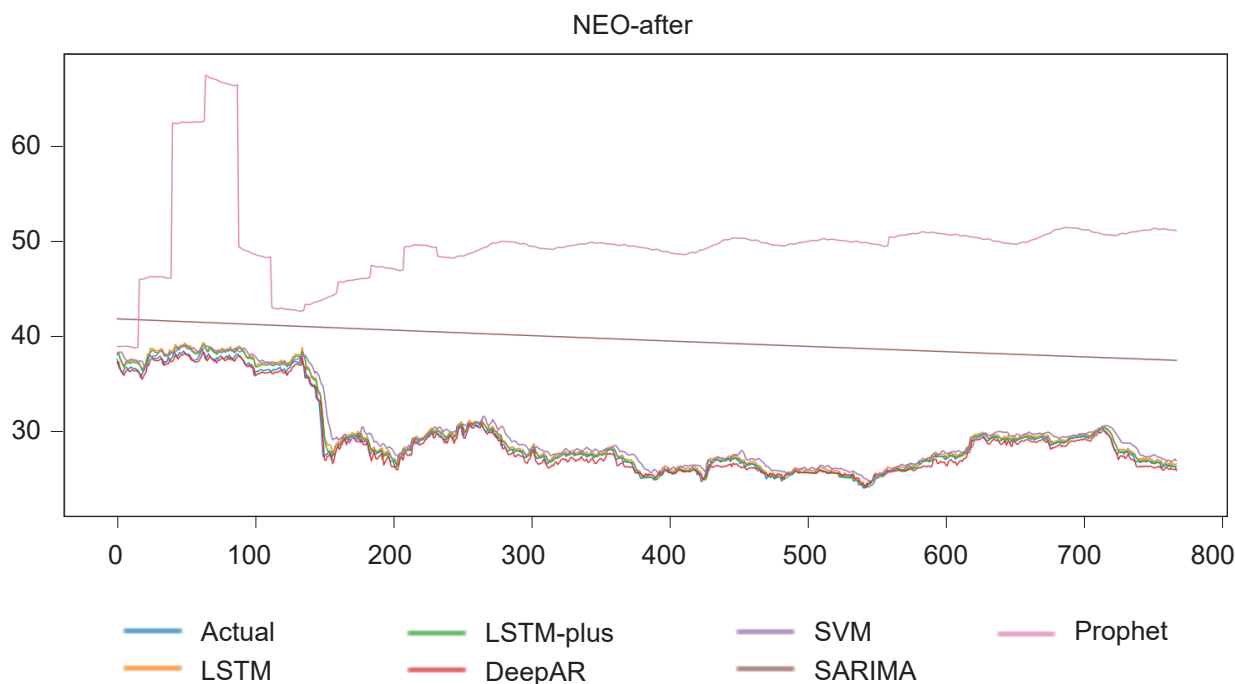
DeepAR je tvorený dvoma LSTM bunkami nasledovanými dvoma vrstvami dense. LSTM+ ako kombinácia modelov využíva jednu LSTM bunku a vrstva dense predstavuje rozširovací hyperparameter voči modelu LSTM a jej veľkosť je optimalizovaná spolu s ostatnými hyperparametrami, ktoré sú totožné s modelom LSTM. Pri testovaní predstavovala v našich modeloch na nami analyzovaných údajoch druhá bunka LSTM, ktorá vracia sekvencie podstatné spomalenie tréningového času a tým aj výpočtovú náročnosť, pričom sme nezaznamenali zvýšenie výkonu výsledného modelu.

Obrázok 11: Porovnanie výsledkov ML modelov pre historické ceny NEO-before



Zdroj: vlastné spracovanie v Python 3.8.0

Model LSTM+, ako znázorňujú zobrazené predikované hodnoty, dosahuje vysokú presnosť. Medzi modely s najlepšimi výsledkami teda patria SVM, LSTM a LSTM+. Na definovanie benchmarku pre predikcie postavené na historických cenách je nutné porovnať presné hodnoty jednotlivých modelov pre analyzované aktíva a časové obdobia. Hodnoty pre jednotlivé metriky sú zobrazené v tabuľke 1 až 4.

Obrázok 12: Porovnanie výsledkov ML modelov pre historické ceny NEO-after

Zdroj: vlastné spracovanie v Python 3.8.0

Metrika MSE zisťuje priemer druhej mocniny rozdielu medzi reálnou hodnotou uzatvárajcej ceny a hodnotou predikovanou modelom. Vzhľadom na faktor kvadratury je náchylnejší na outlierov (odľahlé hodnoty) ako iné metriky, čo môže mať za následok precenenie toho, ako zlý model v skutočnosti je.

Pohlľadom na výsledky metriky MSE pre modely, ktoré tvoria predikciu na základe predošlého vývoja ceny, môžeme vidieť skutočnosti, ktoré plynú z opisu grafických výstupov jednotlivých modelov. Modely SARIMA a Prophet nadobúdajú hodnoty prevyšujúce niekoľkonásobne odchýlky najlepších modelov prakticky pre všetky aktíva a časové obdobia datasetov.

Model SVM predstavuje výrazné zlepšenie voči SARIMA a Prophet, avšak nezaznamenal najlepšie hodnoty ani pre jedno aktívum v žiadnom období. Benchmarkom s najlepšími hodnotami je model LSTM, ktorého predikcie nadobudli najnižšie hodnoty chybovosti pre štyri datasety z celkových ôsmich skúmaných, a to konkrétne pre aktíva Ethereum pred aj počas pandémie, Litecoin počas pandémie a NEO pre predpandemické obdobie. Model LSTM je nasledovaný modelom LSTM+ s najlepšími výsledkami pre tri skúmané datasety pre obe obdobia aktíva Bitcoin a Litecoin pred pandemiou. Model postavený na algoritme DeepAR dosiahol najlepšiu hodnotu pre dataset NEO počas pandémie.

Tabuľka 1: Výsledky MSE modelov založených na historických cenách

MSE	SARIMA	SVM	Prophet	LSTM	DeepAR	LSTM+
BTC_before	291,7612	132,7216	1479,606	2,3451028	70,31276	1,913486
BTC_after	351,7156	808,2249	1196,508	4,6211723	6203,033	2,790392
ETH_before	4,124122	0,002292	14,62428	0,0017118	0,009384	0,013189
ETH_after	52,14384	43,72808	1396,267	6,2219308	15547,57	7,197815
LTC_before	4,003699	0,004804	4,169022	0,0008192	0,000163	0,000163
LTC_after	3,910639	0,037089	3,64331	0,0105977	6,98969	0,019436
NEO_before	0,075330	7,53E-05	0,016679	6,159E-06	5,002656	7,02E-06
NEO_after	0,090528	0,001163	0,493257	0,0003961	0,000268	0,000292

Zdroj: vlastné spracovanie

Model DeepAR dosiahol správnu funkčnosť na základe nami požadovaných vlastností okrem benchmarkovej hodnoty NEO počas pandémie aj pre datasety Ethereum a Litecoin pred pandemiou, kde práve pri Litecoine dosiahol takmer totožnú hodnotu s benchmarkom pre daný dataset, ktorým je pri MSE LSTM+ model. Pre ostatné datasety model DeepAR nedosiahol použiteľné výstupy.

Metrika RMSE zodpovedá druhej odmocnine priemeru druhej mocniny rozdielu medzi cieľovou hodnotou a hodnotou predpovede uzatváracej ceny modelu. Zaoberá sa penalizáciou menších chýb vykonávaných MSE odmocnením hodnôt a je menej náchylný na problémy plynúce z odľahlých hodnôt. V porovnaní s MSE je však náročnejší na komputácie.

Pri porovnaní výsledkov metriky RMSE s MSE pre vývoj najlepších hodnôt chybovosti, resp. rozdielu predikovaných údajov od tých reálne dosiahnutých trhom, rozloženie najlepších výsledných hodnôt pre jednotlivé datasety je totožné s vývojom, ktorý nastal pri metrike MSE, s LSTM modelom najlepším v štyroch prípadoch, modelom LSTM+ v troch a DeepAR v jednom prípade pre dataset NEO počas, kde rovnako ako pri metrike MSE však daná hodnota bola takmer totožná s druhým najlepším modelom pre daný dataset, ktorým je LSTM+.

Tabuľka 2: Výsledky RMSE modelov založených na historických cenách

RMSE	SARIMA	SVM	Prophet	LSTM	DeepAR	LSTM+
BTC_before	0,616358	0,41571	1,388009	0,0552587	0,302577	0,049915
BTC_after	6,767296	1,025854	12,4818	0,775703	28,41983	0,602771
ETH_before	0,231733	0,001728	0,436378	0,0014929	0,003496	0,004144
ETH_after	0,260568	0,238616	1,348353	0,0900082	4,499359	0,09681
LTC_before	0,00725	0,002501	0,023299	0,0010328	0,000461	0,00046
LTC_after	0,071358	0,006949	0,068876	0,0037147	0,0954	0,005031
NEO_before	0,009911	0,000313	0,001473	8,955E-05	0,00809	9,56E-05
NEO_after	0,010857	0,001231	0,025343	0,0007181	0,000591	0,000616

Zdroj: vlastné spracovanie

Metrika MAE predstavuje priemernú hodnotu rozdielu medzi hodnotou reálne dosiahnutou trhom a predpovedanou hodnotou modelu. Je robustnejší voči odľahlým hodnotám ako MSE. Keďže inherentne MAE poskytuje absolútnu hodnotu zvyšku (rozdielu), nedáva nám predstavu o smere chyby modelu, čiže či sú predikované hodnoty podhodnotené alebo nadhodnotené (Zheng, 2011).

Výsledky chybovosti podľa MAE potvrdzujú výkonnosť jednotlivých modelov a ich rozloženie podľa MSE a RMSE. Z hľadiska benchmarku nastali zmeny pre dataset Litecoin pred pandémiou a NEO počas pandémie, pre ktoré si navzájom vymenili prvé a druhé miesto modely DeepAR a LSTM+.

Tabuľka 3: Výsledky MAE modelov založených na historických cenách

MAE	SARIMA	SVM	Prophet	LSTM	DeepAR	LSTM+
BTC_before	0,501599	0,405112	0,944013	0,039256	0,233748	0,031284
BTC_after	5,624252	0,708065	10,97338	0,6314443	27,63318	0,403306
ETH_before	0,021668	0,001341	0,038965	0,0012466	0,00336	0,004019
ETH_after	0,200502	0,206434	1,052136	0,0796736	4,488971	0,082824
LTC_before	0,006713	0,002392	0,019315	0,0009779	0,000319	0,00035
LTC_after	0,062838	0,00544	0,060802	0,0031418	0,091911	0,004308
NEO_before	0,008488	0,000291	0,001257	6,631E-05	0,008081	7,1E-05
NEO_after	0,00919	0,000908	0,02356	0,0005351	0,000478	0,000415

Zdroj: vlastné spracovanie

MAPE je jedným z najpoužívanějších meradiel presnosti predpovedí vďaka svojim výhodám nezávislosti na mierke a jednoduchosti interpretácie výsledkov. MAPE má však aj značnú nevýhodu, keďže vytvára nekonečné alebo nedefinovateľné hodnoty pre nulové alebo takmer nulové skutočné hodnoty. Daná vlastnosť tvorí požiadavku, aby meraný dataset obsahoval striktné pozitívne údaje. MAPE je vyjadrený ako priemer absolútnych percentuálnych chýb prognóz. Chyba je definovaná ako skutočná mínus prognózovaná hodnota (Kim, Kim, 2016).

Výsledné hodnoty metriky MAPE kopírujú zistenia z predošlých metrík. Rozdelenie najlepších hodnôt je totožné s rozdelením pri metrike MAE. Všetky výsledné hodnoty pre najlepšie modely LSTM a LSTM+ dosiahli podľa modelu výkonnosť, ktorá naznačuje správne fungovanie predpovedí.

Tabuľka 4: Výsledky MAPE modelov založených na historických cenách

MAPE	SARIMA	SVM	Prophet	LSTM	DeepAR	LSTM+
BTC_before	5,363 %	4,295 %	9,809 %	0,417 %	2,519 %	0,332 %
BTC_after	8,781 %	1,090 %	16,356 %	0,968 %	41,988 %	0,617 %
ETH_before	12,031 %	0,749 %	21,408 %	0,700 %	1,886 %	2,254 %
ETH_after	3,689 %	3,919 %	19,083 %	1,476 %	83,987 %	1,533 %
LTC_before	11,855 %	4,296 %	33,360 %	1,759 %	0,582 %	0,628 %
LTC_after	30,273 %	2,546 %	28,760 %	1,454 %	42,412 %	1,960 %
NEO_before	7,206 %	2,548 %	10,780 %	0,582 %	71,050 %	0,621 %
NEO_after	25,028 %	2,375 %	61,297 %	1,338 %	1,264 %	1,034 %

Zdroj: vlastné spracovanie

Výstupom analýzy modelov a jednotlivých algoritmov strojového učenia pre potreby predikcie budúcej hodnoty investičného aktíva na základe informácií o vývoji jeho ceny v minulosti sú modely postavené na algoritmoch LSTM a LSTM+, ktoré dosiahli najlepšie hodnoty presnosti, či chybovosti.

Modely SARIMA, Prophet a DeepAR pre účely takejto predpovede nedosiahli použiteľné výstupy. Model SVM dosahoval sľubné a konzistentné výsledky naprieč analyzovanými datasetmi, avšak bol prekonaný v každom období a pre každé aktívum algoritmi LSTM a LSTM+.

Model LSTM dosiahol pri všetkých metrikách najväčší počet najnižších odchýlok. Pri abstrahovaní od benchmarkovej hodnoty modelu DeepAR, ktorá pri každej metrike predstavovala najlepšiu hodnotu pre jeden dataset, predstavuje model LSTM+ z hľadiska početnosti rovnako úspešný model, akým je model LSTM.

Najlepším modelom z hľadiska hlavného analyzovaného objektu skúmania, ktorým je investičné aktívum Bitcoin, je model LSTM+, keďže práve tento model dosiahol pre dané aktívum najlepšie hodnoty, schopnosť porozumenia vývoju ceny v čase pre oba analyzované časové horizonty pred a počas pandémie covid-19.

6. Záver

Výstupom analýzy modelov a jednotlivých algoritmov strojového učenia pre potreby predikcie budúcej hodnoty investičného aktíva na základe informácií o vývoji jeho ceny v minulosti sú modely postavené na algoritmoch LSTM a LSTM+, ktoré dosiahli najlepšie hodnoty presnosti či chybovosti.

Modely SARIMA, Prophet a DeepAR pre účely takejto predpovede nedosiahli použiteľné výstupy. Model SVM dosahoval sľubné a konzistentné výsledky naprieč analyzovanými datasetmi, avšak bol prekonaný v každom období a pre každé aktívum algoritmi LSTM a LSTM+.

Model LSTM dosiahol pri všetkých metrikách najväčší počet najnižších odchýlok. Pri abstrahovaní od benchmarkovej hodnoty modelu DeepAR, ktorá pri každej metrike predstavovala najlepšiu hodnotu pre jeden dataset, predstavuje model LSTM+ z hľadiska početnosti rovnako úspešný model, akým je model LSTM.

Najlepším modelom z hľadiska hlavného analyzovaného objektu skúmania, ktorým je investičné aktívum Bitcoin, je model LSTM+, keďže práve tento model dosiahol pre dané aktívum najlepšie hodnoty, schopnosť porozumenia vývoju ceny v čase pre oba analyzované časové horizonty pred a počas pandémie covid-19.

Literatúra

- Abu Bakar, N., Rosbi, S., Uzaki, K. (2019). Forecasting cryptocurrency price movement using moving average method: A case study of bitcoin cash. *International Journal of Advanced Research*, 7(12), 609–614, <https://doi.org/10.21474/ijar01/10188>
- Akyildirim, E., Goncu, A., Sensoy, A. (2021). Prediction of cryptocurrency returns using machine learning. *Annals of Operations Research*, 297, 3–36, <https://doi.org/10.1007/s10479-020-03575-y>

- Alahmari, S. A. (2019). Using Machine Learning ARIMA to Predict the Price of Cryptocurrencies. *The ISC International Journal of Information Security*, 11(3), 139–144.
- Chen, Z., Li, C., Sun, W. (2020). Bitcoin price prediction using machine learning: An approach to sample dimension engineering. *Journal of Computational and Applied Mathematics*, 365, 112395, <https://doi.org/10.1016/j.cam.2019.112395>
- CryptoDataDownload (2022). BINANCE EXCHANGE DATA. *CryptoDataDownload* [online]. [cit. 2022-07-12]. Dostupné z: <https://www.cryptodatadownload.com/data/binance/>
- Di Persio, L., Honchar, O. (2016). Artificial Neural Networks architectures for stock price prediction: Comparisons and applications. *International Journal of Circuits, Systems and Signal Processing*, 10, 403–413.
- Ding, X., Zhang, Y., Liu, T., et al. (2015). Deep learning for event-driven stock prediction. *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence (IJCAI 2015)*, 2327–2333. Dostupné z: <https://www.ijcai.org/Proceedings/15/Papers/329.pdf>
- Drobetz, W., Otto, T. (2021). Empirical asset pricing via machine learning: evidence from the European stock market. *Journal of Asset Management*, 22, 507–538, <https://doi.org/10.2139/ssrn.3640631>
- Facebook (2017). PROPHET Forecasting at scale. *Facebook Open Source* [online]. [cit. 2022-07-12]. Dostupné z: <https://facebook.github.io/prophet/>
- Gao, P., Zhang, R., Yang, X. (2020). The application of stock index price prediction with neural network. *Mathematical and Computational Applications*, 25(3), 53, <https://doi.org/10.3390/mca25030053>
- Georgoula, I., Pournarakis, D., Bilanakos, C., et al. (2015). Using time-series and sentiment analysis to detect the determinants of bitcoin prices. *Proceedings of the Mediterranean Conference on Information Systems (MCIS)* [cit. 2022-07-12]. Dostupné z: <https://aisel.aisnet.org/cgi/viewcontent.cgi?referer=&httpsredir=1&article=1017&context=mcis2015>
- Hiransha, M., Gopalakrishnan, E. A., Menon, V. K., et al. (2018). NSE stock market prediction using deep-learning models. *Procedia Computer Science*, 132, 1351–1362, <https://doi.org/10.1016/j.procs.2018.05.050>
- Khedr, A. M., Arif, I., El-Bannany, M., et al. (2021). Cryptocurrency price prediction using traditional statistical and machine-learning techniques: A survey. *Intelligent Systems in Accounting, Finance and Management*, 28(1), 3–34, <https://doi.org/10.1002/isaf.1488>
- Kim, S., Kim, H. (2016). A new metric of absolute percentage error for intermittent demand forecasts. *International Journal of Forecasting*, 32(3), 669–679, <https://doi.org/10.1016/j.ijforecast.2015.12.003>
- Kjærland, F., Khazal, A., Krogstad, E. A., et al. (2018). An analysis of bitcoin's price dynamics. *Journal of Risk and Financial Management*, 11(4), 63, <https://doi.org/10.3390/jrfm11040063>

- Li, Y., Zheng, Z., Dai, H. N. (2020). Enhancing bitcoin price fluctuation prediction using attentive LSTM and embedding network. *Applied Sciences*, 10(14), 4872, <https://doi.org/10.3390/app10144872>
- Makridakis, S., Spiliotis, E., Assimakopoulos, V. (2018). Statistical and Machine Learning forecasting methods: Concerns and ways forward. *PloS ONE*, 13(3), e0194889, <https://doi.org/10.1371/journal.pone.0194889>
- Matta, M., Lunesu, I., Marchesi, M. (2015). The Predictor Impact of Web Search Media on Bitcoin Trading Volumes. *Proceedings of the 7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K)*, 1(1), 620–626, <https://doi.org/10.5220/0005618606200626>
- McNally, S., Roche, J., Caton, S. (2018). Predicting the price of bitcoin using machine learning. *Proceedings of the 26th Euromicro International Conference on Parallel, Distributed and Network-Based Processing (PDP)*, 339–343, <https://doi.org/10.1109/PDP2018.2018.00060>
- Menculini, L., Marini, A., Proietti, M., et al. (2021). Comparing prophet and deep learning to ARIMA in forecasting wholesale food prices. *Forecasting*, 3(3), 644–662, <https://doi.org/10.3390/forecast3030040>
- Poongodi, M., Sharma, A., Vijayakumar, V., et al. (2020). Prediction of the price of Ethereum blockchain cryptocurrency in an industrial finance system. *Computers & Electrical Engineering*, 81, 106527, <https://doi.org/10.1016/j.compeleceng.2019.106527>
- Ren, Y.-S., Ma, C.-Q., Kong, X.-L., et al. (2022). Past, present, and future of the application of machine learning in cryptocurrency research (2022). *Research in International Business and Finance*, 63, 101799, <https://doi.org/10.1016/j.ribaf.2022.101799>
- Roy, S., Nanjiba, S., Chakrabarty, A. (2018). Bitcoin price forecasting using time series analysis. *Proceedings of the 21st International Conference of Computer and Information Technology (ICCIT)*, 1–5, <https://doi.org/10.1109/iccitechn.2018.8631923>
- Scikit-learn (2022). Sklearn.preprocessing.PowerTransformer. *Scikit-learn* [online]. [cit. 2022-07-12]. Dostupné z: <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.PowerTransformer.html>
- Sovbetov, Y. (2018). Factors influencing cryptocurrency prices: Evidence from bitcoin, ethereum, dash, bitcoin, and monero. *Journal of Economics and Financial Analysis*, 2(2), 1–27.
- Xiong, R., Nichols, E. P., Shen, Y. (2016). Deep Learning Stock Volatility with Google Domestic Trends. *Computational Finance*, <https://doi.org/10.48550/arXiv.1512.04916>
- Zhang, G. P., Qi, M. (2005). Neural network forecasting for seasonal and trend time series. *European Journal of Operational Research*, 160(2), 501–514, <https://doi.org/10.1016/j.ejor.2003.08.037>
- Zheng, A., Casari, A. (2018). *Feature Engineering for Machine Learning: Principles and Techniques for Data Scientists*. Sebastopol: O'Reilly Media. ISBN 978-1491953242.
- Zheng, S. (2011). Gradient descent algorithms for quantile regression with smooth approximation. *International Journal of Machine Learning and Cybernetics*, 2, 191–207, <https://doi.org/10.1007/s13042-011-0031-2>