

**EKONOMICKÁ UNIVERZITA V BRATISLAVE
PODNIKOVĽHOSPODÁRSKA FAKULTA SO SÍDLOM V
KOŠICIACH**

Evidenčné číslo: 107003/D/2022/36069191765156100

**VYUŽITIE FINTECH NÁSTROJOV UMELEJ
INTELIGENCIE NA PODPORU
INVESTIČNEJ STRATÉGIE
KRYPTOMENOVÉHO INVESTORA**

Dizertačná práca

**EKONOMICKÁ UNIVERZITA V BRATISLAVE
PODNIKOVĽHOSPODÁRSKA FAKULTA SO SÍDLOM V
KOŠICIACH**

**VYUŽITIE FINTECH NÁSTROJOV UMELEJ
INTELIGENCIE NA PODPORU
INVESTIČNEJ STRATÉGIE
KRYPTOMENOVÉHO INVESTORA**

Dizertačná práca

Študijný program:	Ekonomika a manažment podniku
Študijný odbor:	3.3.16 Ekonomika a manažment podniku
Školiace pracovisko:	Katedra finančného riadenia podniku
Vedúci záverečnej práce:	doc. Ing. Michal Tkáč, PhD.

Košice 2022

Ing. Milan Cibul'a

Zadanie záverečnej práce (vo vytlačenej verzii nahradit' stranou z AIS-u).

Čestné vyhlásenie

Čestne vyhlasujem, že záverečnú prácu som vypracoval samostatne a že som uviedol všetku použitú literatúru.

Dátum: 22.07.2022

.....

Ing. Milan Cibul'a

Podakovanie

Touto cestou by som rád podakoval vedúcemu mojej dizertačnej práce doc. Ing. Michalovi Tkáčovi, PhD., za cenné rady a odbornú pomoc pri písaní tejto záverečnej práce.

ABSTRAKT

CIBULA, Milan: Využitie FinTech nástrojov umelej inteligencie na podporu investičnej stratégie kryptomenového investora – Ekonomická univerzita v Bratislave. Podnikovohospodárska fakulta so sídlom v Košiciach; Katedra finančného riadenia podniku. – Vedúci záverečnej práce: doc. Ing. Michal Tkáč, PhD.. – Košice: PHF EU, 2022, počet strán 118.

Hlavným cieľom predloženej dizertačnej práce je posúdiť využitie FinTech nástrojov strojového učenia pre tvorbu inovatívneho prístupu investovania a investičnej stratégie kryptomenového investora. Práca sa zaoberá možnosťami využitia FinTech nástrojov na podporu investičných rozhodnutí kryptomenového investora využitím umelej inteligencie na tvorbu investičnej stratégie. Práca sa venuje implementácii strojového učenia v rámci procesov obchodovania s kryptomenou Bitcoin. Na príklade konkrétneho rozhodovacieho procesu aktívneho investora skúma relevantné prístupy tvorby modelov strojového učenia na základe dát historického vývoja ceny podkladového aktíva, využíva umelú inteligenciu pri definovaní sentimentu trhu analýzou textu a poskytuje finálne zhodnotenie výkonnosti, využiteľnosti vytvorených modelov v porovnaní s trhom. Predmetné navrhované stratégie negarantujú dosiahnutie pozitívnych výsledkov na finančných trhoch, ale opisujú možné riziká a prístupy k využitiu inovatívnej umelej inteligencie pri investovaní na finančných trhoch, čím tvoria analytickú podporu pre investičné rozhodovanie investora. Práca je rozdelená do piatich kapitol. Obsahuje päťdesiattri obrázkov a deväť tabuliek. Prvá kapitola je venovaná definovaniu dvoch základných zložiek pojmu FinTech, ktoré využívame pri tvorbe práce, ako aj vymedzeniu pre prácu relevantných prístupov strojového učenia. V ďalšej časti je charakterizovaný cieľ dizertačnej práce, metodika a metódy skúmania. Záverečné kapitoly výsledky práce a diskusia sa zaoberajú podrobnou analýzou tvorby a výsledkov jednotlivých investičných modelov strojového učenia pre dáta historického vývoja ceny ako aj sentimentu na trhu Bitcoin-u, ktoré predstavujú základ investičnej stratégie, ich vlastnosti, vzájomné vzťahy a zdôvodnenie ich výberu.

Kľúčové slová:

Strojové učenie. Bitcoin. FinTech. Investičná stratégia.

ABSTRACT

CIBULA, Milan: The use of FinTech artificial intelligence tools to support the investment strategy of cryptocurrency investor – University of Economics in Bratislava. Faculty of Business Economics with seat in Košice; Department of Corporate Financial Management. – Final thesis supervisor: doc. Ing. Michal Tkáč, PhD.. – Košice: PHF EU, 2022, number of pages 118.

The main goal of the presented dissertation thesis is to assess the use of FinTech machine learning tools for the creation of an innovative investment approach and investment strategy of a cryptocurrency investor. The work deals with the possibilities of using FinTech tools to support the investment decisions of a cryptocurrency investor in the use of artificial intelligence to create an investment strategy. The work deals with the implementation of machine learning in the processes of trading with cryptocurrency Bitcoin. Using the example of a specific decision-making process of an active investor, it examines relevant approaches to creating machine learning models based on historical price data underlying assets, uses artificial intelligence to define market sentiment by text analysis and provides a final evaluation of performance, usability of created models compared to the market. The proposed strategies do not guarantee the achievement of positive results in the financial markets, but describe the possible risks and approaches to the use of innovative artificial intelligence in investing in the financial markets, thus providing analytical support for the investor's investment decisions. The work is divided into five chapters. It contains fifty-three pictures and nine tables. The first chapter is devoted to the definition of the two basic components of the term FinTech, which we use in the creation of work, as well as to the definition of work-relevant approaches of machine learning. The next part characterizes the aim of the dissertation thesis, methodology and research methods. The final chapters results of work and discussion deal with a detailed analysis of the creation and results of individual investment models of machine learning for the data of historical price development as well as sentiment in the Bitcoin market, which represent the basis of investment strategy, their characteristics, interrelationships and rationale for their choice.

Keywords:

Machine learning. Bitcoin. FinTech. Investment strategy.

OBSAH

Úvod	13
1 Súčasný stav riešenej problematiky doma a v zahraničí.....	15
1.1 <i>FinTech</i>	20
1.1.1 Finančne inovatívne produkty.....	21
1.1.2 Finančne inovatívne nástroje	24
1.2 <i>Strojové učenie (Machine learning)</i>	27
1.2.1 ARIMA, SARIMA.....	29
1.2.2 Support Vector Machines (SVM)	30
1.2.3 Prophet	31
1.2.4 Long Short-Term Memory (LSTM).....	32
1.2.5 DeepAR	35
1.2.6 Advantage Actor Critic (A2C)	36
1.2.7 Proximal Policy Optimization (PPO).....	37
1.2.8 Soft Actor Critic (SAC)	38
1.2.9 Twin Delayed DDPG (TD3)	39
1.3 <i>Sentimentálna analýza</i>	39
1.3.1 GloVe.....	42
1.3.2 Valence Aware Dictionary for Sentiment Reasoning (VADER)	43
2 Cieľ práce	45
3 Metodika práce a metódy skúmania	46
3.1 <i>Charakteristika subjektu skúmania</i>	48
3.2 <i>Dataset historických cien</i>	50
3.3 <i>Dataset Tweet-ov</i>	53
3.4 <i>Python</i>	57
3.5 <i>Nastavenie parametrov algoritmov strojového učenia pre tvorbu investičných modelov</i>	58
4 Výsledky práce	65
4.1 <i>Supervised learning – modely založené na historických cenách</i>	69
4.2 <i>Sentimentálna analýza – modely analyzujúce text</i>	80
4.2.1 GloVe.....	80
4.2.2 VADER.....	84
4.2.3 Porovnanie modelov sentimentálnej analýzy s modelmi historických cien	86
4.3 <i>Investičné stratégie</i>	88
4.3.1 Obchodný systém – prahové hodnoty (thresholds)	90
4.3.2 Obchodný systém – škálovacie hodnoty	91

4.3.3	Reinforcement learning.....	93
4.3.4	Porovnanie najlepších systémov a modelov na nových dátach.....	96
5	Diskusia.....	101
	Záver	105
	Zoznam použitej literatúry	106

Zoznam ilustrácií a zoznam tabuliek

Obr. 1 Základné inovatívne obchodné modely FinTech podľa NBS	21
Obr. 2 Základné inovatívne technológie FinTech podľa NBS	24
Obr. 3 RNNs vanishing gradient problem	34
Obr. 4 LSTM.....	35
Obr. 5 A2C štruktúra	36
Obr. 6 SAC štruktúra	38
Obr. 7 Podobné slová sú tesne umiestnené vo vektorovom priestore	42
Obr. 8 Postup spracovania riešenej problematiky	47
Obr. 9 Vývoj ceny BTC/USD október 2013 – júl 2022	49
Obr. 10 Úvodná úprava dataset-u historických cien.....	51
Obr. 11 Dataset BTC 01.01.2020 0:00 - 04:00.....	51
Obr. 12 Výsledný počet záznamov BTC 01.01.2020 0:00 - 30.12.2021 0:00.....	52
Obr. 13 Štruktúra dataset-u pre sentimentálnu analýzu 10.02.2021 - 29.12.2021.....	56
Obr. 14 Výber analyzovaných dát dataset-u pre sentimentálnu analýzu.....	56
Obr. 15 Náhľad na začiatok a koniec dataset-u pre sentimentálnu analýzu	57
Obr. 16 Optuna pre hyperparametre pri LSTM	62
Obr. 17 Výsledok optimalizácie hyperparametrov pre LSTM	64
Obr. 18 Analyzovaný dataset historických cien Bitcoin pred COVID-19	65
Obr. 19 Analyzovaný dataset historických cien Bitcoin počas COVID-19	66
Obr. 20 Analyzovaný dataset historických cien Ethereum pred COVID-19.....	66
Obr. 21 Analyzovaný dataset historických cien Ethereum počas COVID-19.....	67
Obr. 22 Analyzovaný dataset historických cien Litecoin pred COVID-19.....	67
Obr. 23 Analyzovaný dataset historických cien Litecoin počas COVID-19.....	68
Obr. 24 Analyzovaný dataset historických cien NEO pred COVID-19.....	68
Obr. 25 Analyzovaný dataset historických cien NEO počas COVID-19.....	69

Obr. 26 Porovnanie výsledkov ML modelov pre historické ceny Bitcoin-before.....	70
Obr. 27 Porovnanie výsledkov ML modelov pre historické ceny Bitcoin-after.....	71
Obr. 28 Porovnanie výsledkov ML modelov pre historické ceny Ethereum-before	72
Obr. 29 Porovnanie výsledkov ML modelov pre historické ceny Ethereum-after	73
Obr. 30 Porovnanie výsledkov ML modelov pre historické ceny Litecoin-before	73
Obr. 31 Porovnanie výsledkov ML modelov pre historické ceny Litecoin-after	74
Obr. 32 Porovnanie výsledkov ML modelov pre historické ceny NEO-before	75
Obr. 33 Porovnanie výsledkov ML modelov pre historické ceny NEO-after	75
Obr. 34 Zaokrúhlenie jednotlivých tweet-ov nadol na najbližšiu hodinu.....	81
Obr. 35 Zoradený výstupný dataset tweet-ov podľa GloVe	81
Obr. 36 Optimalizácia hyperparametrov GloVe.....	83
Obr. 37 Výsledné predikcie ceny vs. reálne ceny model GloVe	84
Obr. 38 Zoradený výstupný dataset tweet-ov podľa VADER	84
Obr. 39 Výsledné predikcie ceny vs. reálne ceny model VADER	85
Obr. 40 Porovnanie modelov sentimentálnej analýzy s modelmi historických cien.....	86
Obr. 41 Začiatok a koniec testovacieho dataset-u reálne a predikované hodnoty modelov	87
Obr. 42 Nastavenie obchodnej stratégie modelov sentimentu a historických cien	89
Obr. 43 Obchodný systém prahové hodnoty	90
Obr. 44 Obchodný systém prahové hodnoty výsledky	91
Obr. 45 Obchodný systém škálovacie hodnoty	92
Obr. 46 Obchodný systém škálovacie hodnoty výsledky	93
Obr. 47 Obchodný systém RL modely A2C, PPC, SAC, TD3 historické ceny	94
Obr. 48 Obchodný systém RL modely A2C, PPC, SAC, TD3 historické ceny + GloVe ...	95
Obr. 49 Obchodný systém RL modely A2C, PPC, SAC, TD3 historické ceny + VADER	96
Obr. 50 Finálny aktualizovaný dataset historických cien Bitcoin-u 01.01.20 – 11.03.22...	96

Obr. 51 Grafické vyjadrenie ziskovosti finálnych modelov podľa analyzovaných týždňov	98
Obr. 52 % vyjadrenie ziskovosti finálnych modelov podľa analyzovaných týždňov	99
Obr. 53 Porovnanie hodnôt ziskovosti modelov sentimentu a historických cien.....	100
Tab. 1 Features databázy tweet-ov	55
Tab. 2 Výsledky MSE modelov založených na historických cenách	76
Tab. 3 Výsledky RMSE modelov založených na historických cenách	77
Tab. 4 Výsledky MAE modelov založených na historických cenách	78
Tab. 5 Výsledky MAPE modelov založených na historických cenách	79
Tab. 6 Prehľad výsledkov jednotlivých dimenzií model GloVe	82
Tab. 7 Porovnanie hodnôt metrík modelov sentimentu a historických cien.....	87
Tab. 8 Rozdelenie testovacích dát na 10 týždňových testovacích prostredí.....	97
Tab. 9 Porovnanie hodnôt ziskovosti modelov sentimentu a historických cien.....	100

Úvod

Pojem FinTech je jednoducho kombináciou slov finančná a technológia. Popisuje použitie technológie na poskytovanie finančných služieb a produktov spotrebiteľom. To môže byť v oblasti bankovníctva, poisťovníctva, investovania, všetkého čo sa týka financií. Hoci FinTech je relatívne novým slovom, to čo FinTech predstavuje však nie je ničím novým.

Technológia vždy menila finančný sektor. Internet, spolu s rozsiahlym používaním zariadení, ako sú smartfóny a tablety, však zmenu vo finančnom sektore v posledných rokoch výrazne zrýchlil. Dnešný svet je poháňaný bezprecedentným množstvom dostupných údajov, ktoré spolu s neustálym rastom schopností výpočtovej techniky vytvorili v posledných rokoch inovatívne technológie, ktoré umožnili vytváranie nových obchodných modelov po celom svete.

Pojem FinTech teda predstavuje finančné inovácie, ktoré sú dané do pohybu práve technologickým pokrokom. Národná banka Slovenska zaraďuje pod pojem FinTech šesť základných pilierov, nových inovatívnych obchodných modelov, avšak v skutočnosti každá technologická inovácia finančného sektora tvorí FinTech.

Na sektor FinTech sa môžeme pozerat' z viacerých hľadísk, my sa však vzhľadom na cieľ práce spojený s problematikou investícií a aktívneho investovania zameriame na časti, ktoré v súčasnosti stále viac a viac definujú moderný spôsob obchodovania aktívnych traderov vo svete.

Táto práca spája dve základne zložky pojmu FinTech, kde na strane nových inovatívnych obchodných modelov práca analyzuje Bitcoin ako investičné aktívum, ktoré z pohľadu inovatívnych obchodných prístupov predstavuje hlavného predstaviteľa kryptoaktív s najvyššou trhovou kapitalizáciou medzi kryptomenami. Bitcoin rovnako spája viacero FinTech inovatívnych modelov ako technológiu distribuovaných záznamov a Blockchain, ktoré umožňujú jednotlivým investorom využívať Bitcoin ako nový decentralizovaný spôsob platby, presunu majetku, ale aj ako investičné aktívum obchodované v súčasnosti stále do veľkej miery bez regulácie.

Na strane druhej predstaviteľom inovatívnych technológií, ktoré umožňujú tvorbu nových inovatívnych modelov FinTech, práca využíva na inovatívny prístup investovania umelú inteligenciu (AI) a jej reprezentanta strojové učenie (ML), ktoré je často nazývané aj ako srdce umelej inteligencie.

Každý investor či trader by chcel vytvoriť model, obchodný systém, ktorý by prekonal trh, ktorý by bol automatizovaný, fungoval by bez potreby ľudskej intervencie a ktorý by pracoval bez ľudských emócií a chýb. V minulosti tradičné prístupy tvorby analýz, investičných modelov, či predikcií budúceho vývoja boli založené na neustálom manuálnom vylepšovaní výstupov na základe minulých skúsenosti. Umelá inteligencia a strojové učenie by však mohli byť kľúčom k budúcemu vytváraniu investičných modelov, predpovedí a investičných stratégií, pretože nám umožňuje vytvárať modely, aj keď dopredu nevieme, ako by mal výsledok vyzeráť.

Práca využíva strojové učenie a umelú inteligenciu ako FinTech nástroj na vytvorenie investičných modelov postavených na rozličných algoritmoch strojového učenia podľa ich inherentných vlastností. Jednotlivé modely sú postavené na základe dát historického vývoja ceny podkladového aktíva, avšak práca rovnako skúma rozšírenie vstupných dát o údaje sentimentu trhu a algoritmy určené na analýzu prirodzeného textu.

Výstupom jednotlivých investičných modelov postavených na historických údajoch ceny Bitcoin-u, či v kombinácii so sentimentom trhu je predikovaná hodnota ceny v porovnaní s reálnou hodnotou dosiahnutou trhom, kde je výkonnosť vyjadrená rozličnými metrikami vyjadrujúcimi rozdiel, vzdialenosť medzi predikovanou a skutočnou hodnotou.

Využitím rozličných algoritmov strojového učenia, ich rôznych kombinácii spolu s kombinovaním vstupných dát modelov sú stanovené prístupy tvorby s najlepšimi výslednými hodnotami, resp. najväčším potenciálom pre aplikáciu pri tvorbe obchodného systému.

Práca tvorí investičné stratégie na základe predošlej analýzy tvorbou odlišných prístupov pre definovanie parametrov a postupov tvorby signálov na kúpu, držanie alebo predaj kryptoaktíva Bitcoin prostredníctvom obchodných systémov, ktoré sa usilujú o maximalizáciu výkonnosti nie len kombináciou jednotlivých algoritmov, modelov, ale aj rozdielnym prístupom k definovaniu riešeného problému, ktorým je maximalizácia ziskovosti finálnej investičnej stratégie investora použitím konkrétnej investičnej stratégie, kombinácie modelov a dát.

Dané výsledky práca dáva do kontextu so základnými charakteristikami analyzovaného investičného aktíva a trhu, na ktorom pôsobí počas sledovaného obdobia. Práca rovnako berie do úvahy prirodzené obmedzenia, ktorými je definovaný nie inštitucionálny aktívny investor.

1 Súčasný stav riešenej problematiky doma a v zahraničí

Literatúra v predikcii časových radov má z finančného a ekonometrického hľadiska dlhú históriu. V tejto časti bude predstavená literatúra súvisiaca s predmetnou oblasťou práce týkajúcej sa strojového učenia, ako inovatívneho nástroja na predikciu investičných aktív obchodovaných na finančných trhoch.

Xiong et al. (2016) aplikovali neurónovú sieť Long Short-Term Memory na modelovanie volatility akciového indexu S&P 500 s trendmi spoločnosti Google ako indikátormi makroekonomických faktorov, v práci využívali predpovedanie implicitnej volatility, ktoré napomáha pri predpovedaní cien akcií. Okrem OHLC denných historických údajov autori využívajú 25 domácich trendov, ktoré možno považovať za reprezentáciu verejného záujmu (sentiment) ako vstupné údaje. Pokiaľ ide o benchmark, používajú model GARCH ako porovnanie s modelom LSTM, pričom ich výsledky poukazujú na prekonanie výkonnosti modelu GARCH modelom LSTM.

Autori Hiransha et al. (2018) porovnávajú štyri typy hlbokých neurónových sietí (MLP, RNN, LSTM a CNN) na testovanie výkonnosti pri predpovedaní ceny na dvoch akciových trhoch. Súbor údajov denných uzatváracích cien je prevzatý z trhu National Stock Exchange of India (NSE) a New York Stock Exchange (NYSE). V tomto článku bol použitý model ARIMA ako benchmark a predstaviteľ lineárnych modelov. Výsledky práce ukazujú na prekonanie lineárneho modelu ARIMA nelineárnymi modelmi DNN.

Ding et al. (2015) využili prácu z oblasti spracovania prirodzeného jazyka (NLP) a uplatnili ju v predikcii akciového trhu. V tomto článku tím navrhol novú hlbokú neurónovú sieť (DNN) v predikcii akciového trhu riadeného udalosťami. Experimentálne výsledky ukazujú, že navrhovaný model by mohol dosiahnuť takmer 6 % zlepšenie v porovnaní s najmodernejšími základnými metódami.

Pri predpovedaní ceny finančne inovatívneho aktíva Bitcoin autori McNally et al. (2018) využívali špeciálne typy neurónových sietí RNN a LSTM. Analyzovaný dátový súbor predstavoval denné uzatváracie ceny BTC/USD vytvorené ako priemer hodnôt z 5 (v čase písania) hlavných Bitcoin zmenární za obdobie od 19. augusta 2013 do 19. júla 2016. Na hodnotenie výkonnosti modelov autori využívajú Root Mean Squared Error (RMSE) uzatváracej ceny a ďalej kategorizujú predikovanú cenu do kategorickej premennej odrážajúcej: cena hore, dole alebo bez zmeny. Vstupné premenné okrem dennej uzatváracej ceny obsahujú aj informácie o otváracích, najvyšších a najnižších cenách, ako aj špecifické

údaje pre Bitcoin network (mining difficulty a hash rate), a dva jednoduché kľzavé priemery (SMA). Model LSTM v príspevku dosahuje najvyššiu presnosť klasifikácie na úrovni 52 % a RMSE 8 %. Populárny model SARIMA na predpovedanie časových radov je implementovaný ako benchmark pre porovnanie s modelmi hlbokého učenia. Nelineárne metódy hlbokého učenia prekonávajú výsledky modelu SARIMA, ktorá dosiahla slabú výkonnosť s hodnotou RMSE na úrovni 53,74 %.

Pri rozbere literatúry zameranej na predikciu ceny Bitcoin-u 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 Bitcoin-u, Tweet-mi 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 Bitcoin-ových 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 Bitcoin-u 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. Autori ako Persio a Honchar (2016); Gao, Zhang a Yang (2020); či už spomínaní autori Hiransha et al. (2018) využívali vo svojich výskumoch model Multi Layer Perceptron (MLP) na predikciu cien akcií. MLP však analyzuje iba jedno pozorovanie súčasne. Naproti tomu je výstup z každej vrstvy v rekurentnej neurónovej sieti (RNN) uložený v kontextovej vrstve, ktorá má byť naviazaná späť do výstupu s výstupom z nasledujúcej vrstvy. V tomto zmysle sieť získava pamäť na rozdiel od MLP. Sieť LSTM si je tak schopná zapamätať aj dlhodobé závislosti ako už bolo poznamenané vyššie. Autori však poukazujú aj na nevýhodu siete RNN a LSTM, čím je výpočtová náročnosť. Kde napríklad 50-dňová LSTM sieť je porovnateľná s výcvikom 50 jednotlivých modelov MLP.

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.

Saini a Shobana (2020) vo svojom príspevku dávajú do súvislosti dva modely strojového učenia časových radov ARIMA a model Prophet za účelom tvorby predpovedí uzatváracej ceny Bitcoin-u. Metrikou, ktorú autori využívajú je skóre R^2 , ktorého výsledne hodnoty boli takmer rovnaké pre oba modely. Model ARIMA dosiahol výkon s mierne vyšším R^2 skóre na úrovni 94 % a model Prophet na hodnote 93 %. Autori sa pre ďalšie použitie však rozhodli pre predpovedanie pomocou algoritmu Prophet, kvôli jeho inherentnej jednoduchosti, ľahko pochopiteľným metódam a pracovnému postupu.

Vedecký príspevok od Consoli et al. (2022) s názvom „Predpovedanie akciového indexu IBEX-35 pomocou hlbokého učenia a emócií správ“, poskytuje pohľad na spôsob merania informačného obsahu textu v ekonomických a finančných správach za účelom upraviť vnímanie účastníkov trhu a ich očakávania týkajúce sa dynamiky finančných trhov. V ich práci používajú neurónový strojový preklad a prístup hlbokého učenia na extrahovanie emocionálneho obsahu zo španielskych online ekonomických a finančných časopisov. Následne skúmajú úlohu týchto indikátorov emócií založených na správach pri predpovedaní španielskeho akciového indexu IBEX-35 pomocou DeepAR, pokročilej metódy neurónového predpovedania založenej na autoregresívnych rekurentných neurónových sieťach (RNN) fungujúcich v pravdepodobnostnom prostredí. Cieľom je zhodnotiť, či kombinácia bohatšieho súboru informácií vrátane emocionálneho obsahu ekonomických a finančných správ s najmodernejším strojovým učením môže pomôcť pri takejto náročnej predikčnej úlohe.

Model DeepAR je trénovaný prijatím prístupu rolujúceho okna a používa sa na vytváranie predpovedí bodov a hustoty. Celkovo sa DeepAR darí dosahovať dobré obchodné výsledky a prináša lepšie výsledky, keď sú do modelu zahrnuté premenné emócií založené na správach.

Autorka Kilimci 2019 vykonala predikciu smeru založenú na analýze sentimentu Bitcoin-ového trhu pomocou algoritmov hlbokého učenia a modelov vkladania slov (Word

Embedding Models). Cieľom tejto práce je predpovedať smerovanie ceny Bitcoin-u analýzou názorov používateľov v sociálnych médiách, ako je Twitter. Na účely odhadu smerovania Bitcoin-u sa ako architektúry hlbokého učenia používajú rekurentné neurónové siete (RNN), siete s dlhou krátkou pamäťou (LSTM), konvolučné neurónové siete (CNN) a Word2Vec, GloVe a FastText sú aplikované ako modely vkladania slov do experimentov. Aby boli demonštrované prínosy práce, experimenty sa vykonávajú na anglickom súbore údajov Twitter-u. Výsledky experimentu ukazujú, že použitie modelu GloVe ako modelu vloženia slov dosahuje benchmark-ovú výkonnosť s hodnotou presnosti 82,01 % pri odhade smerovania ceny Bitcoin-u.

Autori Pano a Kashef (2020) zasa v ich práci identifikovali optimálnu stratégiu predbežného spracovania Bitcoin-ových tweet-ov zavedených v analýze sentimentu využitím algoritmu lexikónu VADER pre éru COVID-19. Tento dokument použil skóre VADER zo stratégií spracovania textu na vyjadrenie trendu cien Bitcoin-ov, kde autori dospeli k záveru, že skóre VADER z spracovania textu ukazuje významnú krátkodobú koreláciu s cenami Bitcoin-ov.

Príspevok „Predpovedanie trendu akcií pomocou analýzy sentimentu v správach a technických indikátorov v Sparku“, od autorov Kabbani a Usta (2022), v ktorom využívajú pri predikciách rozličné funkcie, ktoré pomáhajú modelu strojového učenia predpovedať, či sa daný deň nachádza v rastúcom alebo klesajúcom trende, tieto funkcie sú technickými indikátormi generovanými z histórie cien akcií. Okrem toho, keďže finančné správy zohrávajú dôležitú úlohu pri zmene správania investora, celkové skóre sentimentu pomocou VADER je vytvárané pomocou všetkých správ zverejnených v daný deň a pridáva sa do modelu ako ďalšia funkcia. Tri rôzne modely strojového učenia sa testujú v platofrme Spark, Logistic Regression, Random Forest a Gradient Boosting Machine. Najvýkonnejší model autori dosiahli pre Random Forest s presnosťou testu 63,58 % a použitím údajov o sentimente analyzovaných algoritmom VADER.

Ďalšími autormi, ktorý využili analýzu sentimentu pre akcie Sensex prostredníctvom algoritmu VADER boli Nitha a Sivakumari (2021), ako ich prístup k investičným odporúčaniam na indickom akciovom trhu. Autori tvrdia, že analýza sentimentu novinových online tituliek a tweet-ov má vplyv na hodnoty akciového trhu. Takže titulky správ sa zhromažďujú spolu s údajmi o investíciách na akciovom trhu za obdobie štúdie. Na základe percentuálnej analýzy sa zistilo, že tento prístup funguje veľmi dobre pre akcie Sensex na indickom akciovom trhu na účely krátkodobej predikcie. Priemerné percento presnosti tohto

modelu je 80 % a uvádza, že presnosť predpovedí je na vysokej úrovni pri predpovedaní krátkodobých a dlhodobých investičných návrhov pre akcie Sensex.

Wang a Zhang (2021) analyzovali aplikáciu hlbokého učenia pre obchodovanie na burze využitím prognostického modelu časových radov, ktorý sa používa na predpovedanie trhovej ceny. Autori aplikujú základnú obchodnú stratégiu založenú na výsledku, pričom vyžívajú päť druhov Reinforcement learning algoritmov, ktorými sú PPO, A2C, DDPG, TD3 a SAC. Ako súbor používajú 20 akcií s najvyššou trhovou kapitalizáciou v indexe S&P 500 od roku 2000 do roku 2020. Autori porovnávali predpovede RL skúmaných algoritmov s algoritmami Supervised learning-u, kde A2C, PPO, TD3 a SAC zaznamenali stabilné zlepšenie ročného výnosu, pričom zaznamenali problémy pri naladení algoritmu DDPG.

Yang, Liu, Zhong a Walid (2020) navrhujú komplexnú stratégiu, ktorá využíva schémy RL algoritmov s cieľom učenia sa pre stratégiu obchodovania s akciami maximalizáciou návratnosti investícií. Školením agenta hlbokého učenia autori získali stratégiu obchodovania v súbore pomocou troch algoritmov založených na kritických aktéroch: Proximal Policy Optimization (PPO), Advantage Actor Critic (A2C) a Deep Deterministic Policy Gradient (DDPG). Súborová stratégia zdedí a integruje najlepšie vlastnosti troch algoritmov, čím sa výrazne prispôsobuje rôznym trhovým situáciám. Aby sa vyhlí veľkej spotrebe pamäte v tréningových sieťach s nepretržitým akčným priestorom, využívajú na spracovanie veľmi veľkých údajov techniku nahraj-keď-potrebuješ (load-on-demand).

Algoritmy sú testované na 30 akciách Dow Jones, ktoré majú primeranú likviditu. Výkonnosť obchodného agenta s rôznymi algoritmami Reinforcement learning-u sa hodnotí a porovnáva s indexom Dow Jones Industrial Average a tradičnou stratégiou alokácie portfólia s minimálnou odchýlkou. Výsledky ukazujú, že kompletná stratégia využitím RL algoritmov prekonáva Dow Jones Industrial Average a metódu alokácie portfólia s minimálnou odchýlkou, pokiaľ ide o Sharpe-ov pomer vyvážením rizika a výnosu pod transakčné náklady. PPO dosiahol najlepšie agregované výsledky s kumulatívnym výnosom v sledovanom období (2016/01/04-2020/05/08) na úrovni 83,0 %.

1.1 FinTech

Význam odvetvia finančných inovácií FinTech možno globálne vidieť na rapídnom raste medzinárodných investícií do FinTech spoločností. Podľa štatistík The Business Research Company (2020) celosvetový trh FinTech dosiahol v roku 2019 hodnotu takmer 111,3 miliárd USD, pričom od roku 2015 rástol zloženým ročným tempom rastu (CAGR¹) na úrovni 7,9 % a do roku 2023 sa očakáva ročný rast 9,2 % CAGR na hodnotu takmer 158 miliárd USD. Daný rastúci podiel FinTechu na celosvetovej ekonomike je predpokladaný aj z dlhodobejšieho hľadiska, kde v roku 2025 podľa predpovedí trh vzrastie na takmer 192 miliárd USD pri CAGR 10,2 % a v roku 2030 na viac ako 325 miliárd USD pri CAGR 11,1%.

V súčasnosti nie je pevne definované čo FinTech presne znamená a čo ho tvorí. Autori Leong a Sung (2018) definujú FinTech ako multidisciplinárny subjekt, ktorý kombinuje financie, technologický a inovačný manažment. Schueffel (2016) vo svojej analýze rozličných definícií pojmu FinTech dospel k záveru, že FinTech je nové finančné odvetvie, ktoré aplikuje technológiu na zlepšenie finančných aktivít.

V tejto práci budeme vychádzať z definície pojmu FinTech, ktorý ponúka Národná banka Slovenska (NBS). Podľa NBS je pojem FinTech skrátená verzia termínu „finančné technológie“, ktorá vo všeobecnosti reprezentuje oblasť technologických inovácií aplikovaných v praxi vo vzťahu k poskytovaniu rôznorodých finančných služieb, ako sú napríklad platobné služby, bankovníctvo, investičné služby alebo poisťovníctvo.

Inými slovami FinTech predstavuje nové inovatívne obchodné modely, prakticky využiteľné procesy alebo produkty, ktoré sú vytvárané na základe inovatívnych nástrojov, technológií. Obchodná spoločnosť, ktorá vo svojej činnosti uplatňuje finančnú inováciu, je FinTech (NBS, 2021).

Z danej skutočnosti plynú dve základné zložky predstavujúce FinTech, ktorými sú:

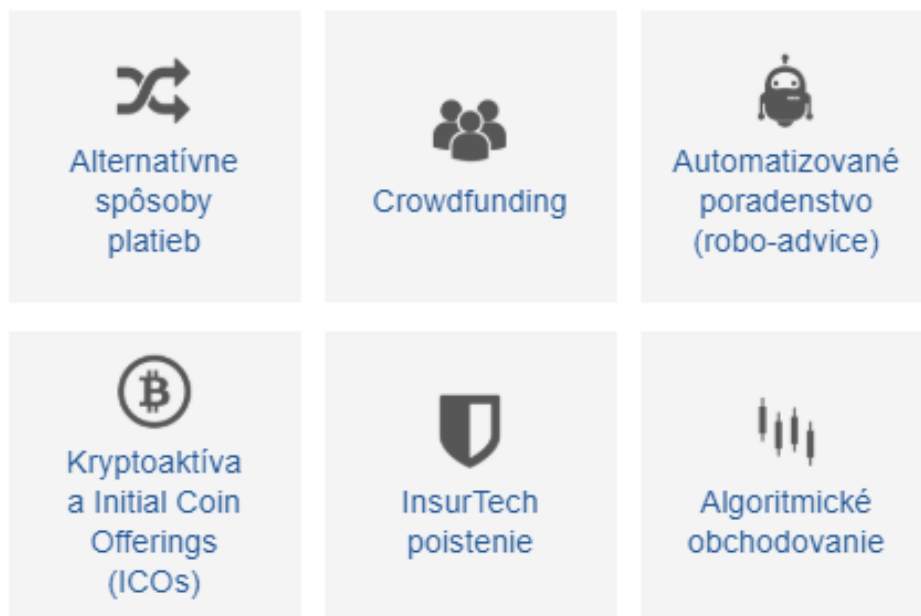
1. finančne inovatívne produkty,
2. finančne inovatívne nástroje.

Táto práca sa zameriava práve na analýzu finančne inovatívneho produktu prostredníctvom nových inovatívnych prístupov FinTech.

¹ **CAGR** - je miera návratnosti, ktorá by sa vyžadovala, aby investícia rástla zo svojej počiatočnej bilancie na konečnú, za predpokladu, že zisky boli reinvestované na konci každého roka životnosti investície.

1.1.1 Finančne inovatívne produkty

FinTech podľa NBS tvorí šesť základných pilierov, ktoré reprezentujú obchodné modely, v ktorých je aplikovaný FinTech v praxi pre vytvorenie nových alebo zlepšenie existujúcich modelov.



Obr. 1 Základné inovatívne obchodné modely FinTech podľa NBS

Zdroj: Národná banka Slovenska, 2021

1. Alternatívne platobné metódy

Medzi alternatívne spôsoby platby patria všetky inovatívne spôsoby platby spotrebiteľa za produkt alebo službu.

Tie predstavujú rozličné typy bezhotovostných platieb od platby kartou, prevodom až po v súčasnosti stále častejšie využívanie bezkontaktných platieb prostredníctvom smartfónu, tabletu, hodínok či náramku.

Dané inovácie v spôsobe platieb predstavujú nie len komfortnejšiu, rýchlejšiu a lacnejšiu platbu pre spotrebiteľa ale rovnako sú zamerané aj na zvýšenie bezpečnosti pri platbách prostredníctvom elektronického bankovníctva a využívania hesiel, QR kódov, či rozličných tokenov na overenie totožnosti a zabezpečenie prevodu peňažných prostriedkov spotrebiteľa (NBS, 2021).

2. Kolektívne financovanie (Crowdfunding)

Crowdfunding je spôsob alternatívneho financovania projektov jednotlivcov a spoločností, vo väčšine ide o financovanie projektov start-upov. Crowdfunding ako kolektívne financovanie nie je definovaný v právnej úprave Slovenskej republiky, resp. Európskej únie, často je využívaný najmä na financovanie v USA (NBS, 2021).

Jedným z najznámejších crowdfundingových platforiem je spoločnosť Kickstarter, kde získavanie vopred stanovenej minimálnej sumy pre financovanie projektu je získavané od veľkej skupiny osôb rozličnými čiastkami z celkovej hodnoty. Podporovatelia najčastejšie dostávajú zľavy na budúci produkt, na ktorý prispeli. Často býva získavanie finančných prostriedkov založené na princípe „všetko alebo nič“, podľa ktorého sa v prípade nevyzbierania vopred určenej cieľovej sumy vrátia finančné prostriedky späť jednotlivým podporovateľom.

3. Automatické poradenstvo

Vo všeobecnosti môžeme do automatizovaného poradenstva zahrnúť všetky kontakty zákazníka, ktorý prostredníctvom webovej stránky využije možnosť zabudovaného chatu, prostredníctvom ktorého hľadá odpoveď na svoje otázky („Chat-advice“).

Automatizované poradenstvo NBS definuje z finančného hľadiska so zameraním sa na investičné poradenstvo ako aj služby riadenia portfólia prostredníctvom viac či menej automatizovaného systému, ktorý sa používa ako nástroj na kontakt s klientom (NBS, 2021).

4. Kryptoaktíva a získavanie kapitálu prostredníctvom ICOs (Initial Coin Offerings)

Kryptoaktíva, sú digitálne aktíva založené na kryptografii (šifrovaní), ktoré sú decentralizované a fungujú na tzv. blockchainovej technológii. Ide zároveň o súhrnné označenie pre elektronické „mince“, či „tokeny“ (NBS, 2021).

Svetová banka (2017) klasifikovala kryptomeny ako podmnožinu virtuálnych či digitálnych mien, ktoré definuje ako digitálne vyjadrenia hodnoty, ktoré sú denominované v ich vlastnej účtovnej jednotke odlišnej od elektronických peňazí, a ktoré využívajú techniky kryptografie na dosiahnutie konsenzu. Predstavujú digitálnu reprezentáciu hodnoty, ktorá má predstavovať peer-to-peer (klient-klient) alternatívu k štátom vydaným zákonným platidlám, sú využívané ako univerzálny prostriedok výmeny, čiže nezávislý od akejkolvek centrálnej banky a môžu byť premenené na zákonné platidlo a naopak.

Ku dňu 26.01.2021 môžeme ako príklad kryptomien uviesť Bitcoin, Ethereum, Tether, Polkadot a Ripple (XRP), ktoré predstavujú päť najväčších kryptomien podľa trhovej kapitalizácie kryptoaktív.

Alternatívna forma financovania označovaná ako „Initial Coin Offerings“ (ICOs) je spôsob zhromažďovania finančných prostriedkov od verejnosti za účelom financovania projektov. Ide o vytvorenie elektronických „mincí“, resp. „tokenov“ a ich následnú ponuku a predaj verejnosti výmenou za fiat meny, alebo častejšie za kryptoaktíva (napríklad bitcoin alebo ether). K takejto ponuke dochádza najčastejšie prostredníctvom internetu a sociálnych médií. ICO je často označovaná ako neregulovaná forma IPO v krypto sfére.

Kryptografické tokeny, predstavujú kryptoaktíva, ktoré na rozdiel od kryptomien ponúkajú aj inú funkciu ako univerzálny prostriedok výmeny hodnoty.

Európsky parlament (2018) rovnako odlišuje od kryptomien kategóriu kryptocenných-papierov, ktoré predstavujú kryptoaktíva umožňujúce spoločnostiam prostredníctvom technológie Blockchain, registráciu, emisiu a prevod bežných akcií a iných podnikových cenných papierov.

5. Nové prístupy poisťovníctva InsurTech

K pojmu „InsurTech“ neexistuje v súčasnosti žiadna právna definícia. Vo všeobecnosti sa týmto pojmom označujú všetky nové technológie využívané v oblasti poisťovníctva, zatiaľ však chýba konzistentnosť medzi poskytovateľmi služieb (NBS, 2021).

PwC (2021) definuje InsurTech ako ekosystém, ktorý spája susedné odvetvia za účelom poskytnutia vyššej hodnoty poisťovateľom a ich zákazníkom. V súčasnosti je InsurTech zameraný na prepájanie susedných odvetví v poľnohospodárstve, zdravotníctve, kybernetickej bezpečnosti, zdieľanej ekonomiky a dopravy.

6. Algoritmické obchodovanie

Podľa autorov Gomber a Zimmermann (2018) je algoritmické obchodovanie považované za nástroj, ktorým profesionálni obchodníci na finančných trhoch pozorujú jednotlivé trhové parametre alebo iné informácie v reálnom čase a automaticky generujú obchodné rozhodnutia bez ľudskej intervencie.

NBS (2021) definuje systém algoritmického obchodovania ako spôsob ako určiť alebo naprogramovať špecifické pravidlá pre komplexné identifikovanie príležitostí na

nákup a predaj investičných pozícií, pričom počítačový systém zabezpečuje určovanie individuálnych parametrov pokynov týkajúcich sa finančných nástrojov, ich generovanie, načasovanie a smerovanie na obchodné miesto alebo vykonanie na základe vopred definovaných pravidiel človekom.

Algoritmické obchodovanie je dôležité pre jednotlivé trhy investičných aktív najmä z hľadiska likvidity, ktorú danému trhu poskytujú, nakoľko v praktickom využití podľa autorov Hasbrouck a Saar (2013), je dosiahnutie vysokých ziskov pre obchodníkov postupne vymazávané vysoko konkurenčným bojom jednotlivých účastníkov finančných trhov, ktorí sa snažia o zabezpečenie čo najrýchlejšieho spojenia s trhom, na ktorom pôsobia, o zabezpečenie najrýchlejšej a najefektívnejšej technológie na tvorbu konkurenčnej výhody prostredníctvom algoritmického obchodovania.

1.1.2 Finančne inovatívne nástroje

Inovatívne obchodné modely sú dané do pohybu prostredníctvom inovatívnych nástrojov, technológií. NBS rovnako rozoznáva šesť základných kategórií inovatívnych nástrojov, ktoré tvoria FinTech.



Obr. 2 Základné inovatívne technológie FinTech podľa NBS

Zdroj: Národná banka Slovenska, 2021

1. Inteligentné zmluvy

Pojem smart contract alebo inteligentná zmluva bol po prvý krát použil autor Nick Szabo už v roku 1996, avšak rozmach inteligentných zmlúv prišiel až s Blockchainovou technológiou. V inteligentných zmluvách sa jednotlivé zmluvné doložky napísané v počítačových programoch automaticky vykonajú, keď sú splnené vopred definované podmienky.

Podľa Zheng et al. (2020) Blockchain umožňuje využitie inteligentných kontraktov, nakoľko ukladá všetky potrebné informácie prostredníctvom distribuovanej siete a tak umožňuje uskutočňovanie transakcií bez potreby dôveryhodnej tretej strany (notár, banka, realitná spoločnosť, a pod.).

2. Biometria

Technológia rozpoznávania špecifických ľudských znakov na zabezpečenie rozličných elektronických finančných aktivít.

Podľa NBS (2021) sem patrí najmä:

1. autentifikácia založená na odtlačkoch prstov zákazníka,
2. autentifikácia pomocou hlasového rozpoznávania,
3. autentifikácia zákazníkov a autorizácia transakcií prostredníctvom overovania biometrických podpisov,
4. autentifikácia prostredníctvom rozpoznávania tváre.

3. Umelá inteligencia (AI) a big data

Autori Khan a Kannan (2014) definujú umelú inteligenciu ako proces demonštrácie inteligencie strojmi, ktorý je odlišný od prirodzenej inteligencie ľudí.

Strojové učenie (ML) je jednou z aplikácií umelej inteligencie, ktorá rozmiestňuje systémy tak, aby vytvorili schopnosť učiť sa automaticky na základe svojho tréningu bez výslovného programovania. ML kladie rovnako dôraz na samostatný vývoj algoritmov, dokáže získavať údaje a umožňuje neustály tréning na základe nových dát bez špecifického naprogramovania človekom (Chiang – Thang, 2016; Pirbhulal et al., 2019).

Pojem big data sa všeobecne používa na opis ukladania a analýzy veľkých a/alebo komplikovaných súborov údajov s využitím rôznych techník vrátane umelej inteligencie (NBS, 2021).

4. Technológia distribuovaných záznamov (DLT)

Technológia distribuovanej účtovnej knihy (záznamov) predstavuje zastrešujúci výraz pre systémy viacerých strán, ktoré fungujú v prostredí bez centrálného operátora, orgánu, bez tretích strán. V súčasnosti najznámejším predstaviteľom DLT je technológia Blockchain, ktorá je považovaná za špecifickú podmnožinu DLT (Pieters et al., 2018).

DLT koncept existoval pred Blockchainom a Bitcoinom, kde Lamport et al. (1982) rozoberal problém byzantských generálov, pričom opísali ako musia počítačové systémy zvládnuť konfliktné informácie. Následný výskum viedol k vzniku prvého algoritmu pre vysoko dostupné systémy, ktoré tolerujú chyby s malým oneskorením (Castro – Liskov, 2002).

Blockchain je mechanizmus, ktorý využíva metódu šifrovania známu ako kryptografia a používa (súbor) špecifických matematických algoritmov na vytvorenie a overenie neustále rastúcej dátovej štruktúry, do ktorej je možné dáta iba pridávať, pričom existujúce dáta (záznamy) nie je možné odstrániť. Blockchain pracuje na princípe reťazca transakčných blokov, ktorý funguje ako distribuovaná účtovná kniha. Blockchain tak napríklad umožňuje existenciu kryptomien, kde v danom prípade predstavuje decentralizovanú účtovnú knihu všetkých transakcií v peer-to-peer sieti. Pomocou tejto technológie môžu účastníci tejto siete potvrdzovať transakcie bez potreby centrálného zúčtovacieho orgánu (World Bank Group, 2017).

5. Mobilná peňaženka

Predstavuje digitálnu peňaženku, ktorá umožňuje spotrebiteľom uskutočňovať prostredníctvom aplikácii finančných inštitúcií na prenosnom zariadení jednu alebo viac platobných služieb. Spotrebiteľ môže napríklad vykonávať bezkontaktné platby platobnou kartou vďaka technológii Near Field Communication (NFC), používať virtuálne karty, ale aj povoľovať inkasá a využívať iné platobné služby prostredníctvom internetu (NBS, 2021).

6. Cloud computing

Cloud computing je poháňaný rozvojom moderných technologických funkcií s cieľom vytvoriť nové výpočtové prostredie, ktoré môže slúžiť rôznym rastúcim obchodným požiadavkám s vyššou flexibilitou a efektívnosťou (Varshney et al., 2020).

Základným konceptom je oddelenie aplikácií, operačného systému a fyzického hardvéru navzájom od seba. V prípade akejkoľvek poruchy alebo napadnutia systému

vírusom, nie je potrebné vypnutie celého systému. Rovnako spoločnosti môžu cloud computing využívať z hľadiska znižovania nákladov (odpadajú vysoké vstupné náklady na serverové vybavenie a výkonné výpočtové zariadenia) ako aj zvýšenia rýchlosti a tým aj celkovej efektivity prác, či konkurencieschopnosti online ponúkaného produktu, služby (Mastelic, 2015).

1.2 Strojové učenie (Machine learning)

Pred príchodom techník strojového učenia využívali investori a vedci rôzne štatistické modely, ktoré často predpokladajú linearitu, stacionárnosť a normálnosť, na popísanie časových radov ceny a objemu obchodovania investičných aktív s cieľom analyzovať a predpovedať ich budúci trend (Chen, 2010; Xie, 2011).

Rapidný vývoj umelej inteligencie, výpočtového počítačového hardvéru (CPU, GPU, RAM, atď.), či prístupu ku bezprecedentnému objemu finančných dát, v posledných rokoch umožnili prudký nárast využitia strojového učenia (ML) v predikcii finančných trhov (Cavalcantea – Brasileirob – Souzab – Nobregab – Oliveir, 2016).

Prehĺbenie výskumu inteligentných algoritmov od lineárnych modelov, cez základné neuronové siete, k modelom hlbokých neurónových sietí (DNN), predstavuje pre investičnú prax neustále nové možnosti využitia. Strojové učenie sa účinne uplatňuje v oblasti rozpoznávania obrázkov a pri analýze textu. Okrem toho sa autori ako Malagrino, Roman, a Monteiro (2018) domnievajú, že pokročilé algoritmy strojového učenia (ML) dokážu zachytiť dynamické zmeny na finančnom trhu, simulovať proces obchodovania s investičnými aktívami a vykonávať automatické investičné rozhodnutia.

Vysokovýkonné počítačové vybavenie, inteligentné algoritmy a finančné big dáta, spolu poskytujú podporu pri rozhodovaní o automatizovanom obchodovaní s investičnými aktívami, čo odborníci v sektore investícií postupne akceptujú. Sila finančne inovatívnych technológií FinTech, týmto pretvára spôsob a možnosti investovania.

Mnohé finančné organizácie postupne prechádzajú od používania tradičnej prediktívnej analýzy k používaniu algoritmov strojového učenia na predpovedanie finančných trendov. S pomocou strojového učenia môžu finanční špecialisti identifikovať zmeny na trhu oveľa skôr ako pri tradičných metódach.

Strojové učenie môžeme definovať, ako schopnosť počítačov učiť sa a konať bez toho, aby boli explicitne naprogramované. Strojové učenie tak hľadá funkciu, vzťah od vstupu X

k výstupu Y, taktiež nám umožňuje riešenie problémov, pri ktorých nevieme dopredu definovať ako by mal výsledok vyzeráť. Rovnako v priebehu rokov tradičné metódy strojového učenia (ML) preukázali silnú schopnosť v predikcii trendov cien investičných aktív (Zhang et al., 2017; Ruta, 2014; Patel et al., 2015).

V súčasnosti existuje veľké množstvo machine learning algoritmov, ktoré sú špecifické pre problém, ktorý sa snažia vyjadriť. V tejto časti sa zameriame na pre prácu relevantné prístupy strojového učenia využívané pri predikcii investičných aktív, čiže na algoritmy s najlepšimi výsledkami pri zameraní na dáta dlhých časových radov, ktoré pri analýze finančných trhov predstavujú chronologickú zbierku pozorovaní vývoja cien a objemu obchodovania (Fu et al., 2005).

Algoritmy strojového učenia môžeme rozdeliť na tri hlavné kategórie:

1. Supervised learning (učenie pod dohľadom, s učiteľom),
2. Unsupervised learning (učenie bez dohľadu, bez učiteľa),
3. Reinforcement learning (učenie formou odmeňovania).

Pri Supervised learning-u vytvára strojové učenie model na základe označených (labelled) údajov, čo algoritmu umožňuje definovať vstupné a výstupné hodnoty. Na týchto dátach sa potom algoritmus trénuje spôsobom, aby vedel pre daný vstup automaticky vyprodukovať požadovaný výstup na základe parametrov vybraného algoritmu. Po natrénovaní je stroj naučený predpovedať očakávaný výstup pre vstupný údajový bod.

Unsupervised learning sa využíva pre neoznačené (unlabelled) údaje, ktoré neboli vopred klasifikované. Cieľom algoritmov učenia bez učiteľa je potom v týchto neklasifikovaných dátach nájsť, modelovať a popísať vzory, koreláciu alebo zhuk v danom súbore údajov. Učenie bez učiteľa môže byť rovnako využité ako nástroj úpravy dát pre úlohy učenia s učiteľom (Bhardwaj et al., 2015).

Reinforcement learning algoritmy predstavujú osobitnú kategóriu strojového učenia, kde pre učenie modelu nie sú využívané označené, či neoznačené údaje. Zahŕňa stratégiu učenia na základe vytvoreného systému (agenta), ktorý je osadený do prostredia, kde sa prostredníctvom interakcie (postupnosť akcií, pozorovaní a odmien) s prostredím učí (Kober et al., 2013).

Podľa Bernal et al. (2012) prístupy strojového učenia, špeciálne práve Supervised learning algoritmy dosahujú pri predikcii investičných aktív sľubné výsledky. Techniky učenia pod dohľadom, ako napríklad Support Vector Machine (SVM) a Recurrent Neural

Networks (RNN), sú schopné naučiť sa predpovedať ceny a trendy na finančných trhoch na základe historických údajov a poskytovať zmysluplnú analýzu historických cien.

Vedci, analytici a obchodníci sa väčšinou zameriavajú na krátkodobé až strednodobé (hodinové, denné, týždenné, mesačné) predpovedanie cien finančných aktív založené na historických údajoch v porovnaní s tvorbou modelov na základe dlhých časových radov.

Autori Shah, Isah, a Zulkernine (2019) poukazujú na existenciu výskumu novších algoritmov strojového učenia, akými sú modely SARIMA a nové prístupy k umelej neurónovej sieti (ANN), konkrétne algoritmus LSTM, ktoré sú porovnávané s existujúcimi prístupmi pri predpovedaní dlhodobých závislostí v dátach a predstavujú pre investorov nové možnosti pri tvorbe modelov.

1.2.1 ARIMA, SARIMA

ARMA kombinuje autoregresné (AR) modely, ktoré sa snažia vysvetliť hybnosť (momentum), stredné efekty reverzie často pozorované na finančných trhoch a modely kľzavého priemeru (MA), ktoré sa zameriavajú na zachytenie šokových efektov pozorovaných v časových radoch. Základným obmedzením modelu ARMA je to, že nezohľadňuje zoskupenie volatility, kľúčový empirický jav v mnohých finančných časových radoch.

ARIMA je prirodzeným rozšírením triedy modelov ARMA a dokáže premeniť nestacionárnu sériu dát na stacionárnu sériu. ARIMA je prispôsobená údajom časových radov, čím odbúrava základné obmedzenia modelu ARMA pri predpovedaní budúcich bodov (Box et al., 2015).

Podľa autorov Šterba a Hil'ovska (2010) sú modely ARIMA známe svojou robustnosťou a efektívnosťou pri predpovedaní finančných časových radov a to najmä pri krátkodobých predikciách často prekonávajú aj populárnejšie ANN techniky.

Rovnako autori Ariyo et al. (2010) podporujú tvrdenie o významnosti modelov ARIMA, ktorí ich radia medzi najvýznamnejšie metódy finančného prognózovania.

Model Auto-Regressive Integrated Moving Averages (ARIMA) je široko používanou technikou pre analýzu finančných trhov. Modely ARIMA pracujú na základe nasledujúcich predpokladov:

1. dátový rad je stacionárny, čo znamená, že priemer a rozptyl by sa nemali meniť s časom. Sériu je možné urobiť stacionárnou pomocou log transformácie alebo diferenciácie série,
2. dáta poskytnuté ako vstup musia byť vo forme jednorozmernej série, pretože ARIMA používa minulé hodnoty na predpovedanie budúcich hodnôt (Hiransha et al., 2018).

ARIMA má tri zložky p , q , d , ktoré predstavujú:

1. **AR** (autoregresná zložka) - predstavuje minulú hodnotu použitú na predpoveď nasledujúcej hodnoty. Výraz AR je v ARIME definovaný parametrom „ p “.
2. **I** (diferenciačná zložka) - predstavuje poradie diferenciácie, ktoré určuje počet opakovaní diferenčnej operácie na sérii, aby sa zabezpečila stacionarita. Výraz I je v ARIME definovaný parametrom „ d “.
3. **MA** (zložka klzavý priemer) - termín sa používa na definovanie počtu minulých chýb predpovede použitých na predpovedanie budúcich hodnôt. Parameter „ q “ v modeli ARIMA predstavuje pojem MA.

Špeciálnym typom ARIMA modelov je model Seasonal ARIMA (SARIMA), ktorý sa využíva za predpokladu, že dataset časového radu obsahuje sezónnosť. Model SARIMA využíva na zohľadnenie jednotlivých období niekoľko parametrov navyše voči modelom ARIMA.

SARIMA môže byť definovaná ako $ARIMA(p,d,q)(P,D,Q)m$, kde zložky p , q , d sú zhodné s modelom ARIMA, pričom novými parametrami sú:

1. m - odkazuje na počet období v každej sezóne,
2. (P, D, Q) - predstavuje (p, d, q) pre sezónnu časť časového radu (Pongdatu – Putra, 2018).

1.2.2 Support Vector Machines (SVM)

Autori Cortes a Vapnik (1995) predstavili Support Vector Machines (SVM), ktoré predstavujú často využívanú alternatívu k využitiu iných modelov kvôli ich schopnosti nájsť globálne maximum a inherentnému talentu odbúrať overfitting pomocou regularizácie a minimalizácie chyby generalizácie.

Základnou myšlienkou SVM je transformovať údaje, ktoré sú nelineárne oddeliteľné v pôvodnom priestore do priestoru vyšších dimenzií, v ktorom môžu byť oddelené jednoduchou nadrovinou. Oddelenie bodov nájdením maximálneho rozpätia medzi podpornými vektormi (support vectors) a nadrovinou umožňuje SVM väčšiu spoľahlivosť v porovnaní s metódami, ako je lineárna regresia, ktoré nachádzajú vhodné, ale nie optimálne oddelenie (Pai et al., 2011).

Autori Meesad et al. (2013) dokázali, že pri správnom definovaní parametrov použitých pre vstupné vektory, modely SVM prekonali modely ako napríklad spätné šírenie umelých neurónových sietí (ANN) z hľadiska minimalizácie chýb a celkovej presnosti modelu.

V dôsledku transformácie údajov do lineárne oddeliteľného priestoru môžu tieto vstupné vektory trpieť tzv. kľatbou rozmerov, ktorá si vyžaduje množstvo pamäte a výpočtového času na tréningovanie (Wang, 2014).

SVM predstavuje vzťah medzi výstupom y a množinou vstupov x vo forme (Cortes – Vapnik, 1995):

$$f(x) = \sum_{i=1}^N w_i \Phi(x) + b \quad (1)$$

kde $\Phi(x)$ predstavuje nelineárne mapovanie x do priestoru vyšších dimenzionálnych znakov, t. j. bázičnú funkciu, a w a b sú parametre naučené z N prípadov tréningových dát.

Hlavnou nevýhodou SVM sú už spomenuté veľmi vysoké požiadavky na výpočtovú silu, čas, SVM má však aj viacero výhod, medzi ktoré patria napríklad:

1. potreba definovať iba dva voľné parametre, hornú hranicu a parameter jadra (kernel),
2. riešenie SVM je jedinečné, optimálne a globálne, pretože učenie SVM sa vykonáva riešením lineárne obmedzeného kvadratického problému,
3. SVM sú založené na princípe minimalizácie štruktúrneho rizika, preto tento typ klasifikátora minimalizuje hornú hranicu skutočného rizika, zatiaľ čo iné klasifikátory minimalizujú empirické riziko (Shin et al., 2005).

1.2.3 Prophet

Prophet (Automatic Forecasting Procedure) alebo Facebook Prophet predstavuje open source softvér vydaný tímom Cora Data Science Facebook-u, ktorý je určený na

vytváranie predpovedí pre jednorozmerné súbory údajov časových radov. Je ľahko použiteľný a navrhnutý spôsobom, aby automaticky našiel dobrú sadu hyperparametrov pre model v snahe vytvoriť použiteľné predpovede údajov s trendmi a sezónnou štruktúrou v predvolenom nastavení (facebook, 2017).

Prophet je odolný voči chýbajúcim údajom a posunom v trende a zvyčajne dobre zvláda odľahlé hodnoty (outliers). Ide o aditívny model pozostávajúci zo štyroch komponentov:

$$y_t = g(t) + s(t) + h(t) + \varepsilon_t \quad (2)$$

Kde jednotlivé komponenty predstavujú:

1. **$g(t)$** - predstavuje trend a cieľom je zachytiť všeobecný trend série. Napríklad počet zobrazení reklám na Facebooku (cena Bitcoin-u) sa pravdepodobne časom zvýši, keď sa k sieti pripojí viac ľudí,
2. **$s(t)$** - je to zložka sezónnosti. Počet zobrazení reklamy (nákupu investičného aktíva) môže závisieť aj od sezóny. Sezónne výkyvy môžu byť veľmi odlišné pre rôzne obchodné časové rady. Druhým komponentom je teda funkcia modelujúca sezónne trendy,
3. **$h(t)$** – je komponent sviatky, či dni pracovného pokoja, ktoré majú jasný vplyv na väčšinu obchodných časových radov (kryptomenový trh je obchodovaný 24/7),
4. Chybová zložka ε_t predstavuje náhodné fluktuácie, ktoré nie je možné vysvetliť modelom (CRAN, 2019).

Aj keď je algoritmus Prophet využívaný hlavne na predpovedanie reklám, či predaja iného typu produktov, jeho jednotlivé komponenty spolu s jednoduchosťou implementácie pre nami skúmaný problém predstavujú zaujímavú alternatívu.

1.2.4 Long Short-Term Memory (LSTM)

Umelé neurónové siete (ANN) sú nelineárne modely, ktoré simulujú procesy medzi neurónmi a synapsiami v mozgu s cieľom modelovať zložité problémy. Ich využitie je populárne pri predpovedaní akciového trhu kvôli svojej povahe čiernej skrinky, ktorá umožňuje používateľom vyskúšať ľubovoľné množstvo premenných (features) na

generovanie použiteľných výsledkov bez toho, aby museli vedieť, akým spôsobom algoritmus funguje a generuje výsledky.

Existuje mnoho variácií umelých neurónových sietí, ktoré sa dajú použiť na predpovedanie časových radov. Príkladom najjednoduchšieho variantu je Multilayer Perceptrons (MLP), alebo Convolutional Neural Networks (CNN), ktoré kódujú informácie o cene do obrazových formátov. Pre použitie na časové rady finančných trhov sú však vhodnejšie pamäťové siete, ako sú siete Recurrent Neural Networks (RNN) a Long Short-Term Memory (LSTM), ktoré umožňujú zapamätanie a zabudnutie predpovedí, čo môže byť užitočné pri predpovedaní extrémnych zmien v krátkom čase prostredníctvom zapamätania si predchádzajúcich príkladov (Chen – Zhou – Dai, 2015).

V porovnaní s inými modelmi sú RNN a LSTM vhodné na sekvenčné údaje, ako sú ceny akcií, dlhopisov, či Bitcoin-u zatiaľ čo iné modely musia brať nesekvenčnú interpretáciu údajov nezávislú od vzorky, berúc do úvahy iba jednu vstupnú a výstupnú vzorku súčasne.

Autori Di Persio a Honchar (2017) poukazujú v ich analýze na sieť LSTM (Long Short-Term Memory) a jej vysokú efektívnosť pri predikcii časových radov, kde autori použili na hodnotenie ceny akcií Google tri rôzne RNN modely (základnú RNN, LSTM a Gated Recurrent Unit (GRU)) aby vyhodnotili, ktorý RNN algoritmus je najvhodnejší pri práci s finančnými časovými radmi. Z výsledkov bolo zrejmé, že LSTM v päťdňovom horizonte predstihol iné RNN varianty so 72 % presnosťou.

Roondiwala et al. (2017) implementovali sieť LSTM na predpovedanie cien indexu Nifty so vstupnými parametrami denných cien OHLC. Ich výsledky ukazujú, že LSTM dosahuje RMSE 0,00859 pre testovacie dáta, čo sa týka denných percentuálnych zmien.

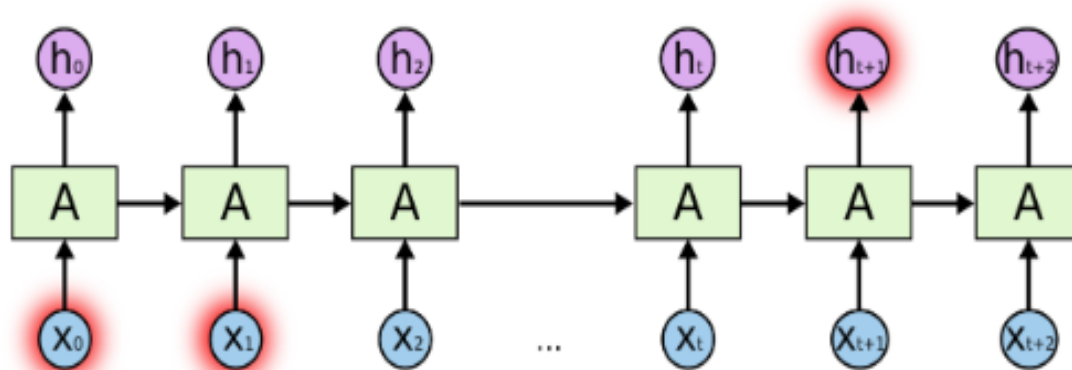
Práca Shah et al. (2018) ukazuje, že model LSTM poskytuje nielen vynikajúce výsledky pre denné predpovede (na jeden deň dopredu), ale poskytuje aj viac než slušné výsledky na predpovedanie v dlhších horizontoch (predpovede na 7 dní) iba s využitím dennej ceny ako funkcie. Autori zámerne využili väčší súbor tréningových údajov (údaje o cenách za 20 rokov), pretože toto obdobie zahŕňa viac býčích a medvedích cyklov trhu. To má za úlohu umožniť modelu LSTM lepšie možnosti učiť sa, a tým zvýšiť nádej na uplatnenie v budúcich časových obdobiach s podobnými pohybmi na trhu.

Siete LSTM sú typom Recurrent Neural Networks (RNNs), ktoré sú schopné naučiť sa závislosť poradia pri problémoch s predikciou sekvencie. Táto vlastnosť sa vyžaduje pri

riešení zložitých problémov, ako je predpoveď časových radov, strojový preklad, či rozpoznávanie reči a jazyka.

Základným problémom používania tradičných neurónových sietí na otázky časových radov je skutočnosť, že štandardné neurónové siete si nie sú schopné zapamätať predchádzajúce informácie a tým pádom ich nemôžu brať do úvahy. Toto je hlavný nedostatok najmä pre dataset-y časových radov z dôvodu autokorelácie (Colah, 2015).

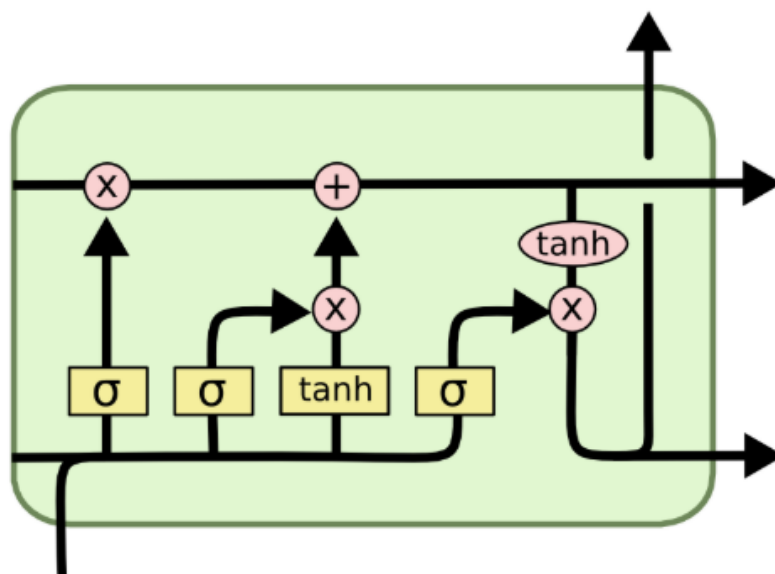
Tento problém riešia práve Recurrent Neural Networks (RNNs). RNN obsahuje vo svojej štruktúre slučky, ktoré im umožňujú zapamätať si a zohľadniť predchádzajúce informácie k súčasnej úlohe. Problémom RNNs pre súbor údajov časových radov je, že keď RNN v určitom okamihu narastajú, začínajú zabúdať na predchádzajúce informácie, pretože nie sú schopné prepojiť (preklenúť) väčšie medzery. Tento jav sa nazýva vanishing gradient problem a keďže nie sú schopné zvládnuť dlhodobé závislosti, nie sú vhodné na predpovedanie dlhých časových radov (Graves, 2009).



Obr. 3 RNNs vanishing gradient problem

Zdroj: Colah, 2015

Vanishing gradient problem vyvoláva pochybnosti o tom, či štandardné RNN môžu skutočne vykazovať významné praktické výhody oproti štandardným neurónovým sieťam bez pamäte pri aplikovaní na časové rady finančných aktív. Tento problém rieši model LSTM ako nadstavba štandardných RNNs.



Obr. 4 LSTM

Zdroj: Colah, 2015

LSTMs sú schopné naučiť sa prekonávať veľké medzery v dataset-e v rámci špeciálnych jednotiek nazývaných bunky (cells). Úspešnosť LSTMs spočíva v tvrdení, že predstavujú nástroj, ktorý rieši technické problémy RNNs.

LSTMs boli explicitne navrhnuté spôsobom, aby boli schopné zabrániť problém dlhodobých závislostí. Pamätanie si informácii za dlhé časové obdobie je ich základnou charakteristikou a nie je to nič čo by sa museli učiť (Gers, Schmidhuber & Cummins, 2000).

Týmto sú LSTMs ako špeciálny druh Recurrent Neural Networks (RNNs) obzvlášť vhodné na riešenie problémov s časovými radmi, a preto sa stali populárnymi pri pokuse o prognózu akciových trhov, čo môžeme využiť aj pri zameraní sa na trh kryptomien a Bitcoin ako investičné aktívum.

1.2.5 DeepAR

Prognostický algoritmus Amazon SageMaker DeepAR je predstavuje Supervised learning algoritmus strojového učenia na predpovedanie jednorozmerných časových radov pomocou rekurentných neurónových sietí (RNN).

Počas tréningu DeepAR akceptuje tréningový súbor údajov a voliteľný testovací súbor údajov. Na vyhodnotenie natrénovaného modelu používa testovací súbor údajov. Vo všeobecnosti množiny údajov nemusia obsahovať rovnakú množinu časových radov. Model

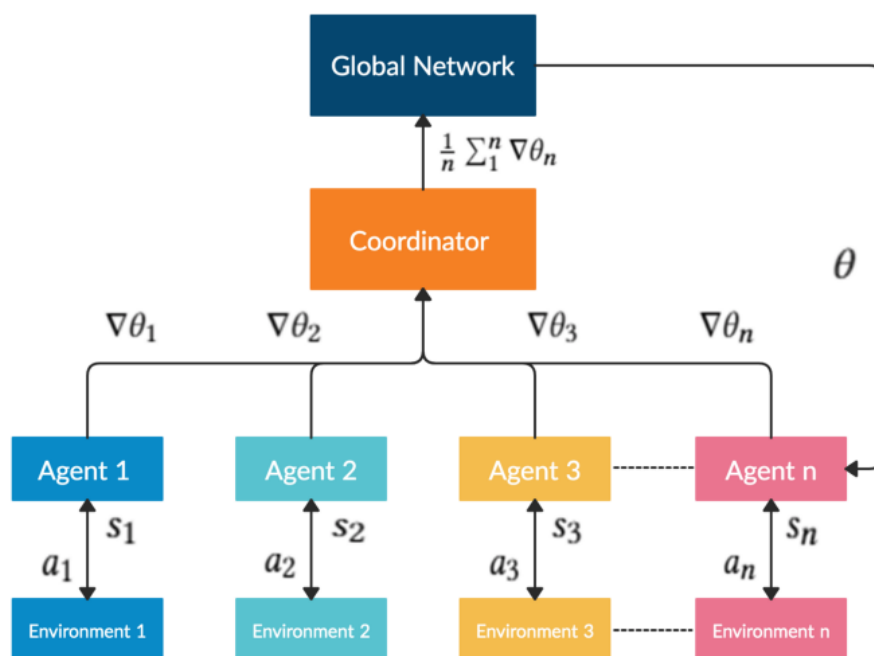
natrénovaný na danej trénovanej množine môžete použiť na generovanie prognóz pre budúcnosť časových radov v trénovanej množine a pre iné časové rady.

Tréningové aj testovacie súbory údajov pozostávajú z jedného alebo, pokiaľ možno, viacerých cieľových časových radov. Každý cieľový časový rad môže byť voliteľne spojený s vektorom časových radov znakov a vektorom kategorických znakov (Amazon, 2019).

Model DeepAR implementuje bunky LSTM v architektúre, ktorá umožňuje simultánne tréningovanie mnohých súvisiacich časových radov údajov. Tento model tak stavia na predchádzajúcich poznatkoch o hlbokom učení pre údaje časových radov a prispôbuje podobnú LSTM architektúru na problém pravdepodobnostného predpovedania (Salinas – Flunkert – Gasthaus – Januschowski, 2020).

1.2.6 Advantage Actor Critic (A2C)

A2C je algoritmus Reinforcement learning-u a ako taký je typickým kritickým algoritmom. A2C je synchronný, deterministický variant algoritmu A3C (Asynchronous Advantage Actor Critic). A2C využíva kópie rovnakého agenta pracujúceho paralelne s rozličnými vzorkami údajov vo viacerých prostrediach.



Obr. 5 A2C štruktúra

Zdroj: Yang, 2020

A2C neodhaduje iba hodnotovú funkciu, ale prostredníctvom kritickej siete odhaduje funkciu výhod. Hodnotenie akcie týmto nezávisí len od toho, aká dobrá, či zlá akcia je, ale algoritmus uvažuje o koľko by mohla byť daná akcia lepšia.

Po ukončení výpočtov všetkých paralelných agentov, A2C použije koordinátora na odovzdanie priemerných gradientov cez všetkých agentov do globálnej siete, čím sieť aktualizuje herca a sieť kritikov.

Prítomnosť globálnej siete zvyšuje rozmanitosť tréningového dataset-u. Aktualizácia synchronizovaného gradientu je výpočtovo efektívnejšia, rýchlejšia a dokazuje vyššiu presnosť pri veľkých dávkach (batch size) v porovnaní s asynchrónnymi implementáciami, kde šum zavedený asynchróniou neposkytuje žiadne výkonnostné výhody voči A2C. A2C predstavuje algoritmus s vysokým potenciálom pre využitie vo finančnom sektore, pre obchodovanie s investičnými aktívami vďaka svojej stabilite, robustnosti a schopnosti pracovať s veľkým množstvom vstupných dát (OpenAI, 2017).

1.2.7 Proximal Policy Optimization (PPO)

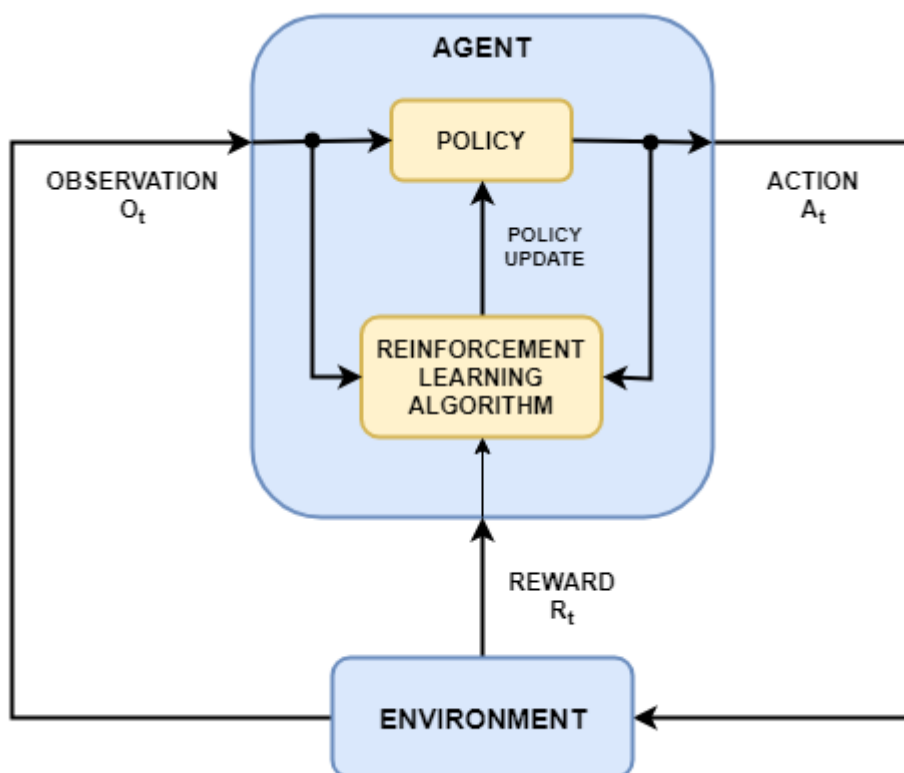
PPO je typ Reinforcement learning algoritmu, ktorý v svojej architektúre kombinuje poznatky z A2C (obsahuje viacero agentov) a TRPO (používa dôveryhodnú oblasť pre zlepšenie aktéra). PPO sa snaží dosiahnuť najväčší možný krok na zlepšenie politiky pomocou údajov, ktoré má model k dispozícii, bez toho aby nastal kolaps výkonu. Hlavnou myšlienkou je, že po aktualizácii by nová politika nemala byť príliš vzdialená od tej starej. Na to používa PPO orezávanie, čím bráni príliš veľkej aktualizácii.

PPO trénuje stochastickú politiku politickým (on-policy) spôsobom. To znamená, že skúma pomocou vzorkovania akcií podľa najnovšej verzie svojej stochastickej politiky. Miera náhodnosti pri výbere akcie závisí od počiatočných podmienok a od tréningového postupu. V priebehu školenia sa politika zvyčajne stáva postupne menej náhodnou, pretože pravidlo aktualizácie ju nabáda k využívaniu odmien, ktoré už našla. To môže spôsobiť, že politika uviazne v lokálnom optime. (Schulman et al., 2017).

PPO predstavuje rovnováhu medzi jednoduchosťou implementácie, zložitou vzorkou a jednoduchosťou ladenia, pričom sa v každom kroku snaží o aktualizáciu, ktorá minimalizuje cost (nákladovú) funkciu. PPO predstavuje algoritmus, ktorý pracuje porovnateľne alebo lepšie ako najmodernejšie prístupy kategórie aktér-kritických algoritmov (OpenAI, 2017).

1.2.8 Soft Actor Critic (SAC)

SAC je hlboký Reinforcement learning rámec pre tréning politík maximálnej entropie v súvislých doménach. SAC je algoritmus, ktorý optimalizuje stochastickú politiku mimopolitickým (off-policy) spôsobom, čím vytvára most medzi stochastickou optimalizáciou politiky a prístupmi v štýle DDPG. SAC je nástupcom Soft Q-Learning SQL a zahŕňa dvojitý Q-learningový trik z TD3 a kvôli prirodzenej stochasticite politiky v SAC, algoritmus tiež ťaží z tzv. vyhladzovania cieľovej politiky (Haarnoja et al., 2018).



Obr. 6 SAC štruktúra

Zdroj: Haarnoja et al., 2018

Ústrednou črtou SAC je regulácia entropie. Politika je trénovaná tak, aby maximalizovala kompromis medzi očakávaným výnosom a entropiou, čo je miera náhodnosti v politike. To má úzku súvislosť s kompromisom medzi prieskumom a využívaním, kde zvýšenie entropie vedie k väčšiemu prieskumu, čo môže neskôr urýchliť učenie. Môže tiež zabrániť predčasnej konvergencii politiky k zlému lokálnemu optimu (OpenAI, 2018).

1.2.9 Twin Delayed DDPG (TD3)

TD3 je priamym nástupcom Deep Deterministic Policy Gradient (DDPG) a je zameraný na riešenie chyby aproximácie funkcie v metódach kritického aktéra. Až donedávna bol DDPG jedným z najpoužívanějších algoritmov pre problémy s nepretržitým riadením, ako je robotika a autonómne riadenie (automatizovaný obchodný systém). Aj keď je DDPG schopný poskytnúť vynikajúce výsledky, má svoje nevýhody. Rovnako ako mnohé RL algoritmy, tréning DDPG môže byť nestabilný a silne závislý na nájdení správnych hyperparametrov pre aktuálnu úlohu (OpenAI, 2018).

Je to spôsobené tým, že algoritmus neustále prehodnocuje Q hodnoty kritickej (hodnotovej) siete. Tieto chyby v odhadoch sa časom nahromadia a môžu viesť k tomu, že agent upadne do lokálneho optima alebo zažije katastrofálne zabudnutie. TD3 rieši tento problém tým, že sa zameriava na zníženie skreslenia nadhodnotenia pozorovaného v predchádzajúcich algoritmoch. To sa deje s pridaním 3 kľúčových funkcií:

1. **využitím skráteného dvojitého Q-Learning** - TD3 sa učí dve funkcie Q namiesto jednej a používa menšiu z dvoch hodnôt Q na vytvorenie cieľov vo funkciách straty chýb Bellman,
2. **oneskorením aktualizácii politiky herca** - TD3 aktualizuje politiku (a cieľové siete) menej často ako funkcia Q,
3. **vyhladením šumu cieľovej politiky** - vyhladenie Q pozdĺž zmien v činnosti (Fujimoto – Hoof – Meger, 2018).

1.3 Sentimentálna analýza

Investičnú stratégiu je možné chápať ako proces rozhodovania, kedy držať aktíva alebo kedy ich predat, aby sme dosiahli výhodnejšie hodnoty rizika a výnosu. S vývojom algoritmickej obchodných prístupov nastáva postupný presun od tradičných metód analýzy k metódam založených na počítačových programoch.

Výber vhodného typu investičnej stratégie závisí od preferencií investora, jeho rizikového profilu, časového horizontu investovania, či od skúseností investora s alokovaním investičných aktív. Vo všeobecnosti je tak investičná stratégia odvodená od základných obmedzení tvorených špecifickými podmienkami investora a od charakteristík

aktíva, na ktoré sa investor zameriava, čím investičná stratégia predstavuje spôsob akým investor zhodnocuje majetok, ktorý má k dispozícii.

Táto práca pri tvorbe analyzovaných modelov, investičných stratégií vychádza najmä z obmedzení plynúcich s nami skúmaného FinTech investičného aktíva Bitcoin, ako aj zo základných obmedzení investora jednotlivca.

Teória rozlišuje tri základné prístupy k predikcii finančných trhov predstavujúcich investičné stratégie, ktoré využívajú fundamentálnu analýzu, technickú analýzu (chartovanie) a sentimentálnu analýzu (Nassirtoussi et al., 2014; Prem Sankar et al., 2015; Ahmadi et al., 2018).

Primárnym obmedzením, ktoré nám Bitcoin definuje je práve pri rozdelení stratégií na fundamentálne, technické a sentimentálne. Bitcoin predstavuje hlavného reprezentanta kryptoaktív, ktoré predstavujú decentralizovaný trh (trh bez oficiálnych riadiacich entít). Ak investor nechce využívať výhradne technickú analýzu používajúcu vývoj ceny aktíva v čase a nemá k dispozícii oficiálne informácie použiteľné pre fundamentálnu analýzu, rozšírenie vstupných informácií pre model, investičnú stratégiu je možné práve prostredníctvom sentimentálnej analýzy.

Analýza sentimentu, často nazývaná aj dolovanie názorov (opinion mining), je aktívna oblasť štúdia v oblasti spracovania prirodzeného jazyka, ktorá je zameraná na analýzu názorov, pocitov, hodnotení, postojov a emócií ľudí prostredníctvom výpočtového spracovania subjektivity v texte. Ide o metódu analýzy textu, ktorá zisťuje polaritu (pozitívnu, neutrálnu alebo negatívnu) v celom texte, čo sa dá urobiť v celom dokumente, odseku, vete alebo dokonca len v tweet-e na sociálnych médiách (Pedrycz – Chen, 2016).

Analýza sentimentu je užitočná aj pre odborníkov z praxe, vedcov a výskumníkov, najmä v oblastiach ako ekonómia, financie, sociológia, marketing, reklama, psychológia a politológia, ktoré sa vo veľkej miere spoliehajú na údaje o interakcii medzi človekom a počítačom.

Stručne povedané, sentimentálnu analýzu možno použiť na:

- analyzovanie príspevkov na sociálnych médiách počas určitého časového obdobia s cieľom zistiť sentiment konkrétneho publika,
- sledovať zmienky o značke na sociálnych sieťach a automaticky ich kategorizovať podľa naliehavosti zmeny,

- automaticky smerovať zmienky zo sociálnych médií k členom tímu služby, ktorí sú na to najvhodnejší,
- automatizovať niektorý alebo všetky tieto procesy,
- získať hlboký prehľad o tom, čo sa deje na kanáloch sociálnych médií (Hutto – Gilbert, 2014).

Existujú dva prístupy pre analýzu sentimentu, ktoré sa bežne používajú, označované ako prístup strojového učenia a prístup založený na lexikóne.

Pri využívaní algoritmov strojového učenia pre vykonanie sentimentálnej analýzy je nevyhnutné zabezpečiť algoritmom klasifikované (často rozsiahle) trénovacie/testovacie dáta. Následne je algoritmus trénovaný na používanie týchto údajov, aby bol schopný predikovať klasifikáciu sentimentu pre vopred nedostupné dáta.

Sentimentálna analýza vykonaná prostredníctvom algoritmu strojového učenia tak závisí od tréningovej sady dát, ktorá by mala obsahovať aspoň všetky „features“, ktoré budú obsiahnuté v dátach, ktoré nemal algoritmus k dispozícii na učenie.

Veľké množstvo údajov potrebných pre algoritmy strojového učenia je výpočtovo nákladné, pokiaľ ide o spracovanie CPU, požiadavky na pamäť a čas na školenie/klasifikáciu, čo predstavuje problém pri investovaní v reálnom živote.

Rovnako podľa Hutto a Gilbert (2014) algoritmy strojového učenia často odvodzujú vlastnosti z tzv. čiernej skrinky, ktoré ľudia nedokážu vysvetliť, a preto je ťažké ich modifikovať, čím je dôležité použiť pri tvorbe sentimentálnej analýzy ideálne algoritmy, ktoré boli vytvorené primárne za účelom analýzy textu.

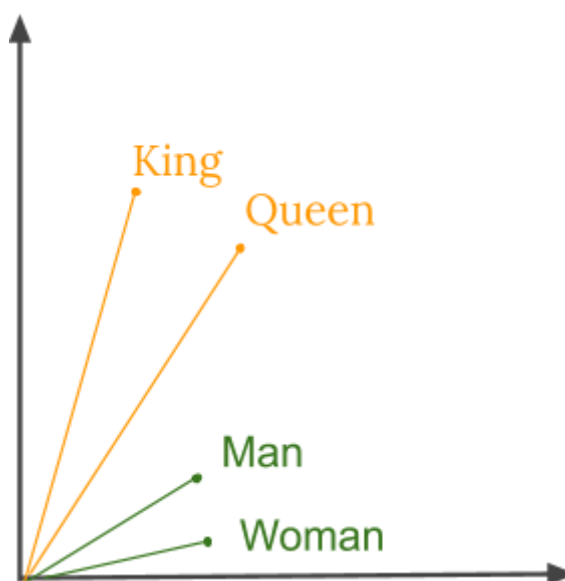
V našej práci preto budeme analyzovať oba prístupy pre analýzu sentimentu, a to prostredníctvom Unsupervised learning algoritmu strojového učenia GloVe a prístupu založeného na lexike VADER, ktorý je založený na slovníku slov spolu s informáciou o ich polarite.

Výhodou použitia prístupu založeného na lexike je, že zoznam sentimentových slov a ich polaritu možno vyhľadávať veľmi rýchlo, údaje nie je potrebné označovať, nie sú potrebné trénovacie údaje, čím je celý proces analýzy výpočtovo prijateľnejší (Annett – Kondrak, 2008).

1.3.1 GloVe

Ľudia vždy vynikali v chápaní jazykov. Pre ľudí je ľahké pochopiť vzťah medzi slovami, ale pre počítače nemusí byť táto úloha jednoduchá.

Word embeddings (vkladanie slov) je v podstate formou reprezentácie slov, ktorá spája ľudské chápanie jazyka s chápaním stroja. Predstavuje reprezentácie textu v n-rozmernom priestore, kde slová, ktoré majú rovnaký význam, majú podobné zastúpenie. To znamená, že dve podobné slová sú reprezentované takmer podobnými vektormi, ktoré sú veľmi blízko umiestnené vo vektorovom priestore.



Obr. 7 Podobné slová sú tesne umiestnené vo vektorovom priestore

Zdroj: GreatLearning, 2020

GloVe (Global Vectors for Word Representation) predstavuje Unsupervised learning algoritmus zameraný na definovanie vektorových reprezentácií slov. Učenie sa vykonáva na agregovaných globálnych word-word (slova a slova) štatistikách spoločného výskytu a výsledné reprezentácie predstavujú zaujímavé lineárne subštruktúry vektorového priestoru slova.

GloVe je v podstate log-bilineárny model s váženým cieľom najmenších štvorcov. Hlavnou intuíciou, ktorá je základom modelu, je jednoduché pozorovanie, že pomery pravdepodobnosti spoločného výskytu word-word majú potenciál na zakódovanie určitej formy významu.

Tréningovým cieľom GloVe je naučiť sa slovné vektory tak, aby sa ich bodový súčin rovnal logaritmu pravdepodobnosti spoločného výskytu slov. Vzhľadom na skutočnosť, že logaritmus pomeru sa rovná rozdielu logaritmov, tento cieľ spája (logaritmus) pomerov pravdepodobností spoločného výskytu s vektorovými rozdielmi vo vektorovom priestore slova. Pretože tieto pomery môžu zakódovať určitú formu významu, tieto informácie sa zakódujú aj ako vektorové rozdiely. Z tohto dôvodu výsledné slovné vektory fungujú veľmi dobre v úlohách slovnej analógie, ako sú tie, ktoré sa skúmali v balíku word2vec od spoločnosti Google (Pennington – Socher – Manning, 2014).

1.3.2 *Valence Aware Dictionary for Sentiment Reasoning (VADER)*

Valence Aware Dictionary and Sentiment Reasoner open-source knižnica analyzátor sentimentu je založená na pravidlách lexikónu chránená pod licenciou MIT. Model analýzy sentimentu využívajúci VADER je citlivý na polaritu (pozitívnu, negatívnu a neutrálnu) aj intenzitu emócií textu. Je k dispozícii v balíku NLTK pre Python od roku 2014 a možno ho použiť priamo na neoznačené textové údaje, čo znamená, že na rozdiel od iných metód analýzy sentimentu, ktoré si pred použitím vyžadujú tréning o súvisiacom texte, je VADER pripravený na analýzu bez akéhokoľvek špeciálneho nastavenia.

Sentimentálna analýza VADER sa opiera o slovník, ktorý hľadá lexikálne črty intenzity emócií známe ako skóre sentimentu a na základe zloženého skóre (nastaveného v rámci analýzy) VADER určí, či je text negatívny, neutrálny alebo pozitívny. Príkladom je skóre pre slovo „komfort“, ktoré je nižšie (menej pozitívne) ako pre slovo „eufória“. Výsledné skóre sentimentu textu sa potom získa ako súčet intenzity každého slova v analyzovanom texte.

VADER je schopný pochopiť základný kontext analyzovaných slov, fráz, a keďže slovo „love“ je s vysokým pozitívnym skóre, fráza „did not love“ sa považuje za negatívny výrok s vysokým negatívnym skóre. Pokúša sa tiež zachytiť a ohodnotiť textové prvky bežné v neformálnom online texte, ako sú veľké písmená, výkričníky a emotikony, čo je ideálne na použitie pre analýzu textu na sociálnych médiách.

VADER má mnoho výhod oproti tradičným metódam analýzy sentimentu, ktoré by sa dali zhrnúť takto:

- poskytuje najlepšie výsledky v triede pri analýze sentimentu textu typu sociálnych médií,

- nepotrebuje tréningové údaje, ale je skonštruovaný zo zovšeobecniteľného, valenčného, ľuďmi spravovaného štandardného lexikónu sentimentu,
- je dostatočne rýchly na to, aby sa dal používať online so streamovanými dátami,
- netrpí výrazným kompromisom medzi rýchlosťou a výkonom (Hutto – Gilbert, 2014).

Ako každý nástroj analýzy sentimentu, aj VADER by mal byť hodnotený kriticky v kontexte predpokladov, ktoré vytvára ohľadom komunikácie. VADER bol vyvinutý v polovici roka 2010 predovšetkým na analýzu anglického jazyka na stránkach sociálnych médií (Twitter). Kontext príspevkov na sociálnych médiách je väčšinou neformálny a obsahuje vzory používania jazyka a funkcií, ktoré sa líšia od formálnych vyhlásení, článkov alebo inej formálnej komunikácie. Vzhľadom na účel nami analyzovaného problému je dané zameranie sentimentálnej analýzy VADER ideálne.

VADER bol však od jeho uvedenia vyvinutý ako všeobecný analyzátor sentimentu a Hutto a Gilbert už v roku 2014 dosiahli jeho využitím lepšie výsledky analýzy sentimentu v porovnaní so strojovým učením, či dokonca v porovnaní s ľuďmi.

2 Cieľ práce

Práca sa zaoberá možnosťami využitia FinTech nástrojov na podporu investičných rozhodnutí. Hlavným cieľom predloženej dizertačnej práce je posúdiť využitie FinTech nástrojov strojového učenia pre tvorbu inovatívneho prístupu investovania a investičnej stratégie kryptomenového investora.

Práca sa zaoberá možnosťami využitia FinTech nástrojov na podporu investičných rozhodnutí kryptomenového investora využitím umelej inteligencie na tvorbu investičnej stratégie. Práca sa venuje implementácii strojového učenia v rámci procesov obchodovania s kryptomenou Bitcoin. Na príklade konkrétneho rozhodovacieho procesu aktívneho investora skúma relevantné prístupy tvorby modelov strojového učenia na základe dát historického vývoja ceny podkladového aktíva, využíva umelú inteligenciu pri definovaní sentimentu trhu analýzou textu a poskytuje finálne zhodnotenie výkonnosti, využiteľnosti vytvorených modelov v porovnaní s trhom. Predmetné navrhované stratégie negarantujú dosiahnutie pozitívnych výsledkov na finančných trhoch, ale opisujú možné riziká a prístupy k využitiu inovatívnej umelej inteligencie pri investovaní na finančných trhoch, čím tvoria analytickú podporu pre investičné rozhodovanie investora.

Na splnenie tohto hlavného cieľa je potrebné splniť nasledujúce čiastkové ciele:

1. vymedziť teoretické východiska riešenia danej problematiky doma i v zahraničí,
2. identifikovať algoritmy strojového učenia vhodné na predikciu kryptoaktív,
3. vytvoriť súbor finančne inovatívnych obchodných modelov postavených na inovatívnom strojovom učení so zameraním na predikciu vývoja ceny Bitcoin-u využitím historických cien,
4. rozšíriť obchodné modely o dáta získane analýzou textu predstavujúce informácie o sentimente na trhu, ktoré sú analyzované algoritmami na analýzu prirodzeného textu,
5. vytvoriť a zhodnotiť výkonnosť investičných stratégií prostredníctvom obchodných systémov na reálnych údajoch,
6. prezentovať prínos dosiahnutých výsledkov pre vedu a prax a odhadnúť možnosti ďalšieho výskumu.

3 Metodika práce a metódy skúmania

Riešenia dizertačnej práce vychádzajú z metód skúmania, prostredníctvom ktorých sme boli schopní popísať v teoretickej časti práce prepojenie medzi FinTech inovatívnymi modelmi a FinTech inovatívnymi nástrojmi, kde práca poskytuje ucelený obraz problematiky využitia umelej inteligencie, strojového učenia, ako inovatívneho investičného nástroja pri tvorbe investičnej stratégie.

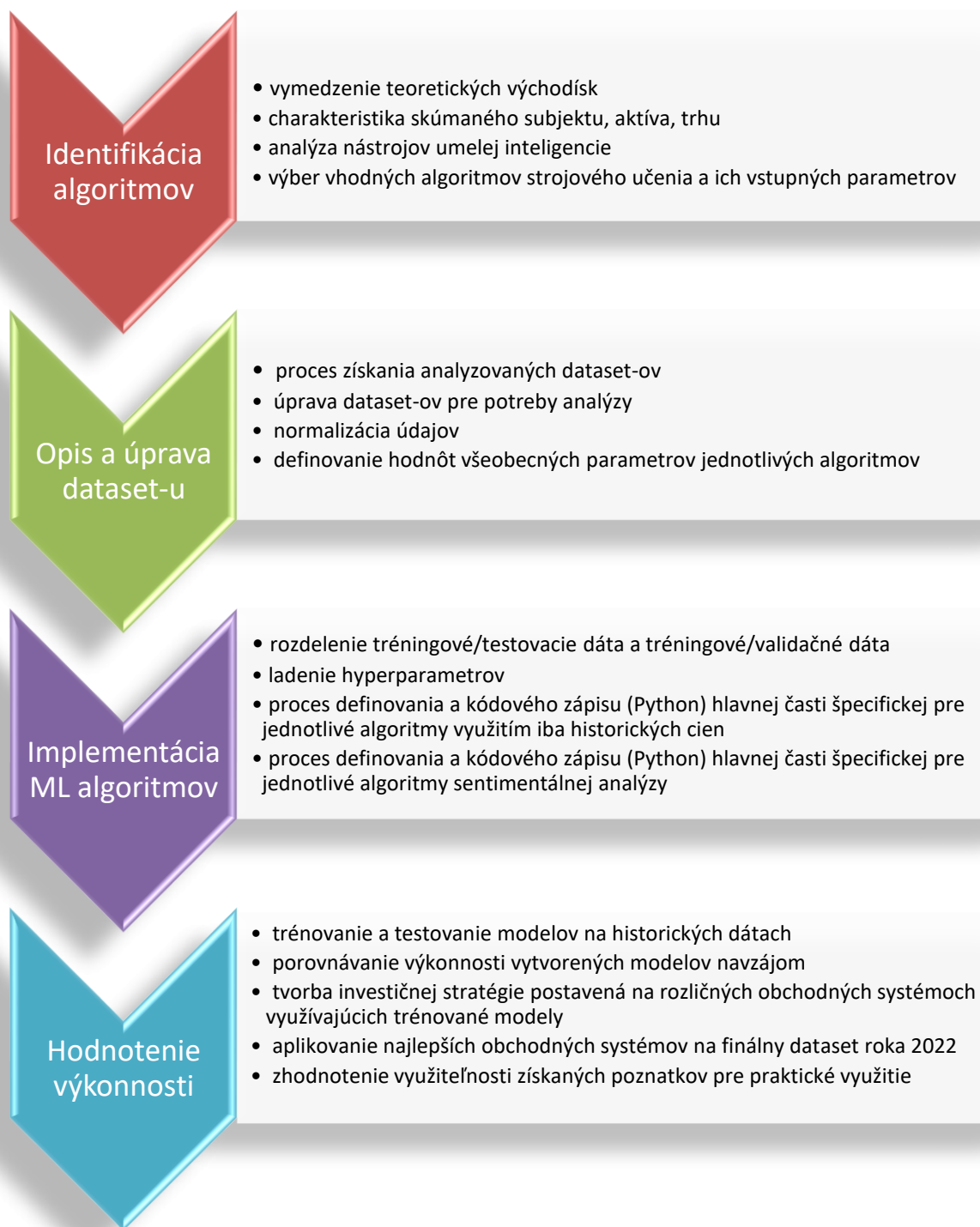
Analyzované finančné aktívum Bitcoin zabezpečuje prepojenie dvoch základných zložiek FinTech-u, kde Bitcoin ako inovatívny finančný model, produkt, existuje na základe ďalších FinTech zložiek, akými sú Blockchain alebo technológia distribuovaných záznamov. Nadobudnuté poznatky sú využívané pri zostavovaní analytickej časti záverečnej práce.

Práca porovnáva správanie sa, čiastkovú i celkovú výkonnosť investičných modelov postavených na rozličných algoritmoch strojového učenia zameraných na analýzu dát časových radov, ktoré porovnáva s benchmarkom predstavujúcim trh.

Proces úpravy analyzovaných dataset-ov, tvorba jednotlivých investičných modelov, tréovanie a testovanie výkonnosti predikčných modelov na historických dátach, ako aj v kombinácii s analýzou textu pre definovanie sentimentu trhu, meranie výkonnosti na reálnych dátach, či tvorba, ladenie a implementácia jednotlivých modelov strojového učenia prebiehajú prostredníctvom kódovania v programe Python.

V neposlednom rade je práca zameraná na swing aktívneho investora, čo znamená, že jeho každodenná práca je spojená s finančnými trhmi, má finančné vzdelanie alebo/a investičné vedomosti, skúsenosti, prax. Nie je však investičnou inštitúciou (hedge fond, investičná banka, a pod.), čo má za následok základné obmedzenia v jednotlivých zdrojoch, ktoré má skúmaný investor k dispozícii, akými sú napríklad výpočtová sila, či analytické možnosti, ktoré v práci berieme do úvahy.

Ako zobrazuje Obr. 8, postup spracovania práce vychádza z identifikácie vhodných algoritmov strojového učenia pre analýzu.



Obr. 8 Postup spracovania riešenej problematiky

Zdroj: Vlastné spracovanie

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

prístupy pre tvorbu modelov postavených na analýze textu a tvorby sentimentálnej analýzy patria VADER ako prístup založený na lexikóne a Unsupervised learning algoritmus strojového učenia GloVe. Predstaviť algoritmov strojového učenia pre Reinforcement learning využitých prácou tvoria Advantage Actor Critic (A2C), Proximal Policy Optimization (PPO), Soft Actor Critic (SAC) a Twin Delayed DDPG (TD3).

3.1 Charakteristika subjektu skúmania

Kryptomena predstavuje v najjednoduchšej podobe peer-to-peer (užívateľ – užívateľ) verziu elektronickej hotovosti, ktorá umožňuje vymieňať finančný majetok medzi jednotlivými stranami (vlastníkmi) priamo prostredníctvom online platby bez nutnosti využitia tretej strany (finančnej inštitúcie) (Brito – Castillo, 2013).

Bitcoin ako základný reprezentant sveta kryptomien bol po prvý krát spomenutý autorom pod pseudonymom Satoshi Nakamoto (2008) v príspevku s názvom „Bitcoin: A Peer-to-Peer Electronic Cash System“.

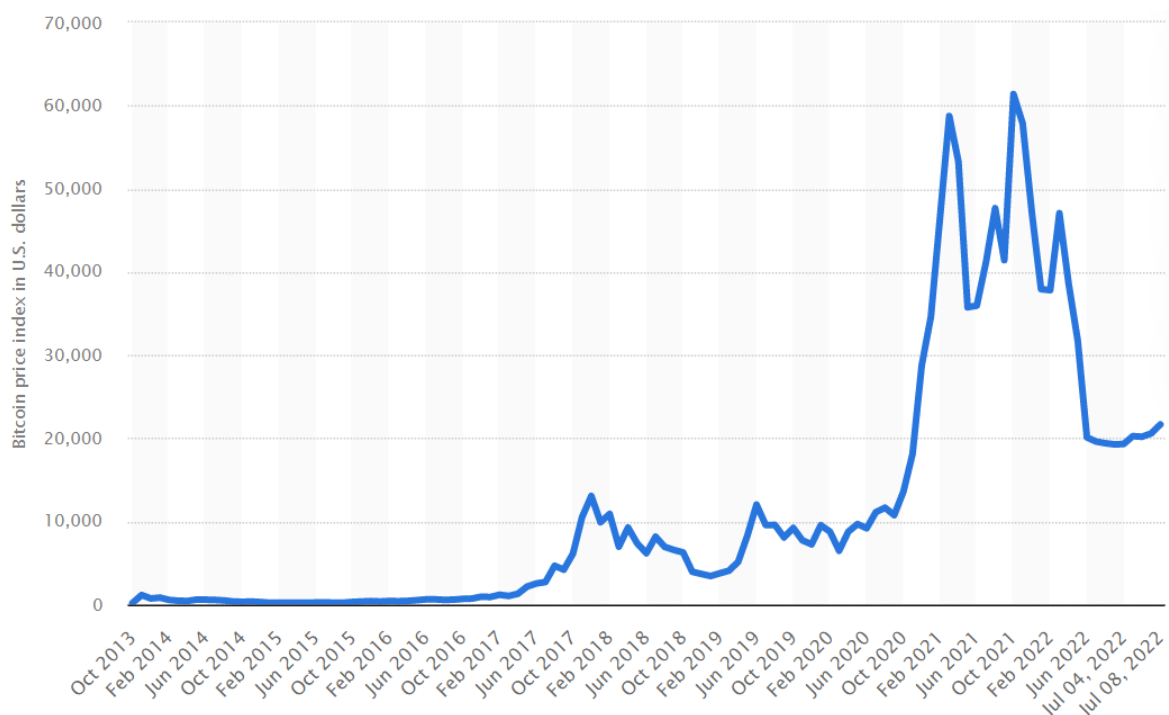
Bitcoin funguje na verejne dostupnom softvéri (Bitcoin Core), ktorý si môže stiahnuť ktokoľvek a ktorý predstavuje decentralizovanú a plne distribuovanú peer-to-peer sieť (bitcoin.org, 2021).

Bitcoin sa skladá z dvoch virtuálnych komponentov. Prvým je verejná elektronická účtovná kniha (ledger), ktorá sa nazýva Blockchain. Táto účtovná kniha je distribuovaná a každý počítač pripojený priamo k Bitcoin-ovej sieti (node) disponuje jej úplnou kópiou, v ktorej sa nachádzajú všetky transakcie medzi Bitcoin-ovými peňaženkami zoskupené v blokoch. Bloky sú zoradené postupne a sekvencia blokov tvorí účtovnú knihu. Z danej skutočnosti pochádza pojem Blockchain, čiže sekvenčné poradie alebo reťaz blokov. Správna postupnosť, ako aj konzistencia účtovnej knihy je podporená kryptografiou.

Páry kľúčov predstavujú druhý virtuálny komponent tvoriaci Bitcoin. Dvojica kľúčov sa skladá z dvoch veľkých čísel (kľúčov), ktoré sú matematicky príbuzné. Tento vzťah umožňuje osobe, ktorá pozná jedno z týchto čísel, vykonať činnosť, ktorú môže znalec druhého čísla overiť, ale nemôže danú činnosť znova vytvoriť (pretože by to vyžadovalo výpočet druhého kľúča, ktorý je neúmerne zložitý).

Ak je jeden z kľúčov v páre utajený, zatiaľ čo druhý je vyzradený, umožňuje to držiteľovi tajného kľúča preukázať sa verejnosti ako jeho vlastník. V takom prípade sa tajný kľúč nazýva súkromný kľúč (private key) a prezradený kľúč sa nazýva verejný kľúč (public

key). Verejný kľúč predstavuje Bitcoin-ovú adresu (podobná číslu účtu v banke), zatiaľ čo súkromný kľúč umožňuje vytvárať transakcie patriace k tejto adrese (Deloitte, 2016).



Obr. 9 Vývoj ceny BTC/USD október 2013 – júl 2022

Zdroj: statista, 2022

Ak by Bitcoin umožňoval iba zúčtovacie transakcie, nemohol by fungovať, pretože by tu neboli žiadne zostatky na prevod. Preto je raz za blok povolený špeciálny typ transakcie, ktorý vytvára nové Bitcoin-y (mining). Povolené množstvo nových Bitcoin-ov v každom bloku je vopred určené a degresívne v priebehu času. Každý Bitcoin je rozdeliteľný na 100 000 000 jednotiek a maximálne (finálne) množstvo Bitcoin-ov, ktoré bude možné vytážiť predstavuje hodnotu 21 000 000 (Taylor, 2018).

Bitcoin ako hlavný predstaviteľ kryptomien predstavuje investičné aktívum obchodované prostredníctvom Bitcoin-ového trhu a viac či menej regulovaných zmenárni kryptoaktív. Pri pohľade na hodnotové vyjadrenie ceny Bitcoin-u v čase ako poukazuje Obr. 9, prudký nárast hodnoty robí z Bitcoin-u stále častejšie špekulatívne aktívum pre investorov. Jeho neregulovaný charakter predstavuje aj napriek najvyššej kapitalizácii spomedzi kryptoaktív, náchylnosť na nekalé praktiky, ktoré sú už v súčasnosti na tradičných trhoch akcií, dlhopisov, či fiat mien minulosťou. Daná skutočnosť robí z Bitcoin-u stále vysoko rizikové aktívum z hľadiska volatility a vysokých percentuálnych korekcií ceny, kde môžeme bežne pozorovať dennú smerodajnú odchýlku na úrovni niekoľko desiatok percent.

Rovnako nesmieme zabudnúť na problémy s bezpečnosťou držania akéhokoľvek kryptoaktíva, kde ak vlastník stratí svoj súkromný kľúč, všetky prostriedky uložené na danej adrese (peňaženke) sú nenávratne stratené, rovnaká skutočnosť platí v prípade, že kryptoaktíva zašle na nesprávnu adresu.

3.2 Dataset historických cien

Na získanie historických cien nami analyzovaného FinTech inovatívneho investičného aktíva Bitcoin, ako aj ostatných referenčných kryptomien využitých za účelom porovnania výsledkov pre prvotné zúženie množstva analyzovaných algoritmov strojového učenia sme využili stránku cryptodatadownload.com. Daný zdroj ponúka po registrácii možnosť stiahnutia dataset-ov 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, analýza s využitím výhradne historických cien bola rozdelená na 2 časti a to na obdobie pred COVID-19 a obdobie ovplyvnené pandémiou. Daným rozdelením dataset „before“ (pred) obsahuje dáta za obdobie od 01.01.2018 0:00 až po 31.12.2019 23:00 čo predstavuje 2 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 v 01.01.2020 0:00 a koncom 30.12.21 0:00 obsahujúci 17 479 záznamov hodinových uzatváracích cien za účelom dosiahnutia porovnateľnosti rozsahu dataset-ov pred a počas.

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

```
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["tradeount"]

# 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()
```

Obr. 10 Úvodná úprava dataset-u historických cien

Zdroj: Vlastné spracovanie v Python 3.8.0.

Výsledné dáta s upraveným poradím a odstránením špecifických údajov zmenárne pre označenie položiek dataset-u obsahujú informácie o hodinových otváracích, najvyšších, najnižších a uzatváracích cenách Bitcoin-u (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 Obr. 11.

```
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

Obr. 11 Dataset BTC 01.01.2020 0:00 - 04:00

Zdroj: Vlastné spracovanie v Python 3.8.0.

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

```
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
```

Obr. 12 Výsledný počet záznamov BTC 01.01.2020 0:00 - 30.12.2021 0:00

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.

V čase písania (máj 2022) by sme pri abstrahovaní tzv. stablecoin-ov² využili meny akými sú BNB, Cardano, Solana či TRON, ktoré dnes kumulujú najvyššiu trhovú hodnotu v kryptosfére po Bitcoin-e, 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 ma mieste 64.

Posledným využívaným dataset-om historických cien sú údaje pre Bitcoin, na ktorých sú testované finálne modely, algoritmy, investičné stratégie práce rozšírené o

² **Stablecoin** – token, kryptomena, ktorej hodnota je viazaná ku forme kolaterálu v pomere 1:1. Najčastejšou formou kolaterálu je FIAT mena používajúca rezervy v USD (Grobys et al., 2021).

obdobie od 29.12.2021 23:00 do 11.03.2022 0:00 predstavujúcich 1 706 záznamov hodinovej uzatváracej ceny.

3.3 Dataset Tweet-ov

Vo finančnom sektore a investovaní sa každý deň generuje veľké množstvo údajov. Dáta, ktoré prichádzajú vo forme nových správ, oficiálnych údajov riadiacich inštitúcií, alebo dokonca neoficiálnych aj nepravdivých správ, ktoré ovplyvňujú verejnú (ale aj inštitucionálnu) mienku o danom finančnom sektore, aktíve, a rozhodnutie kúpiť, predat' alebo držať, čím ovplyvňujú samotnú cenu.

Preto je pre každého obchodníka veľmi dôležité, aby mal prehľad o najnovších správach, aktualizáciách, aby mal všetky možné informácie na zabezpečenie čo najlepšieho rozhodnutia pri zadávaní objednávok, a tak spracovanie týchto informácií zaberie obchodníkovi veľkú časť dňa.

Analýza sentimentu sa dá využiť nielen ako spôsob zlepšenia služieb zákazníkom, pri marketingu produktov a pod., ale v posledných rokoch je to aj spôsob, ako môže obchodník automatizovať proces informovanosti vo svete finančných správ. Zistilo sa, že dokonca aj sentiment na sociálnych médiách má moc ovplyvniť ceny finančných aktív (Wang – Su – Duxbury, 2022).

Sociálne médiá sú v prípade tejto práce však oveľa dôležitejšie, keďže sa v našej analýze sentimentu zameriavame na Bitcoin ako finančné aktívum a predstaviteľa finančného inovatívneho nástroja (FinTech), ku ktorému chýbajú oficiálne vyhlásenia, správy, keďže infraštruktúra, na ktorej je Bitcoin postavený je decentralizovaná.

Najväčšia komunita investorov, traderov kryptoaktív na svete je práve na sociálnej sieti Twitter, čo predstavuje dôvod, prečo na definovanie sentimentu trhu v čase práca využíva tweet-y jednotlivých používateľov.

Pri získavaní dataset-u o tweet-och užívateľov sme boli schopní nami požadované údaje získať zo stránky kaggle.com, ktorá ako taká predstavuje komunitu pre užívateľov strojového učenia a vedy o údajoch, kde stránka jej registrovaným užívateľom ponúka rôznorodé dataset-y pre rozličnú problematiku, ako aj diskusiu pre užívateľov ohľadom problematiky strojového učenia.

Twitter rovnako ponúka možnosť využiť ich API na získanie požadovaných údajov. Pri danej možnosti však existujú určité obmedzenia, ktoré je potrebné vziať do úvahy pred

vytvorením analýzy sentimentu. Je dôležité vedieť, že API Twitter-u pre užívateľa, ktorý nie je súčasťou ich developerskej platformy je zameraný na relevantnosť a nie na úplnosť. To znamená, že niektoré tweet-y a používatelia môžu vo výsledkoch vyhľadávania chýbať. Nie všetky tweet-y budú indexované alebo prístupné cez vyhľadávanie API pre daného užívateľa.

S cieľom spravovať objem žiadostí od vývojárov na Twitter developerskú platformu (a tiež z dôvodu bezpečnostných problémov) sú stanovené limity na počet žiadostí, ktoré je možné vykonať. Daný limit predstavuje 180 žiadostí na 15-minútové okno (nie vývojárske konto). Na overenie na používateľa a na žiadosť je možné vyžiadať maximálne 100 tweet-ov, čo dáva celkovo 18 000 tweet-ov za 15-minútové okno (Twitter, 2020).

Tweet-y sú extrahované prostredníctvom Python knižnice Tweepy, ktorá slúži na prístupenie Twitter API a tým sú dolované na základe obmedzení rýchlosti, od začiatku spustenia kódu denne až do jeho ukončenia.

Tým, že sme boli schopní získať dataset tweet-ov zo spomínaného zdroja, jednoducho overiť správnosť jeho údajov na základe kontroly vybraných príspevkov so zverejnenými užívateľmi v čase, zabezpečili sme si uvoľnenie počítačného výkonu, ktorý by bol inak blokovaný kódom na dolovanie tweet-ov.

Dataset obsahuje tweet-y užívateľov, ktorí vo svojom príspevku použili hashtag #Bitcoin a, alebo #btc. Obsahuje vyhovujúce príspevky so začiatkom dňa 10.02.2021 a koncom dataset-u dňa 29.12.2021. Celkovo dataset obsahuje 1 999 534 vyhovujúcich tweet-ov.

Analyzovaný dataset obsahuje dáta rozdelené do 13 kategórií ako znázorňuje Tab. 1. Pre našu analýzu však dáta všetkých kategórií nie sú potrebné, práca pri tvorbe modelov analýzy textu využíva dátum/čas kedy daný príspevok bol zverejnený, text konkrétneho príspevku ako aj množstvo follower-ov užívateľa, ktorý daný príspevok pridal.

Informácia o počte follower-ov je pre analýzu veľmi dôležitá, nakoľko podstata celej sentimentálnej analýzy je založená v našom prípade nie len na sentimente konkrétneho užívateľa, ktorý ho vyjadruje prostredníctvom svojho príspevku na sociálnej sieti, ale záleží aj na rozsahu akým dokáže ovplyvniť ďalších používateľov sociálnej siete, prostredníctvom zhliadnutia daného príspevku.

Tab. 1 Features databázy tweet-ov

Č.	Stĺpce	Popis
1	user_name	meno používateľa na platforme Twitter
2	user_location	používateľom definované umiestnenie profilu tohto účtu
3	user_description	používateľom definovaný reťazec UTF-8 popisujúci ich účet
4	user_created	čas a dátum, kedy bol účet vytvorený
5	user_followers	počet followerov, ktorý má účet momentálne
6	user_friends	počet priateľov, ktorý má účet momentálne
7	user_favourites	počet obľúbených, ktoré má účet momentálne
8	user_verified	hodnota true znamená, že používateľ má overený účet
9	date	čas a dátum UTC, kedy bol tweet vytvorený
10	text	skutočný text Tweet-u v UTF-8
11	hashtags	všetky ostatné hashtagy uverejnené v tweet-e spolu s #Bitcoin & #btc
12	source	pomôcka používaná na uverejnenie tweet.u, tweet.y z webovej stránky Twitter majú zdrojovú hodnotu - web
13	is_retweet	označuje, či tento tweet odoslal overujúci používateľ

Zdroj: Vlastné spracovanie

Počet follower-ov týmto pre nás predstavuje váhu dôležitosti sentimentu daného príspevku, kde príspevok používateľa s vyšším počtom odberateľov má vyššiu váhu ako používateľa s nižším počtom follower-ov.

```
In [7]: # nahratie tweet-ov
tweets_df = pd.read_csv(tweets_file)
tweets_df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1999534 entries, 0 to 1999533
Data columns (total 13 columns):
#   Column                Dtype
---  -
0   user_name             object
1   user_location         object
2   user_description      object
3   user_created          object
4   user_followers        float64
5   user_friends          object
6   user_favourites       object
7   user_verified         object
8   date                  object
9   text                  object
10  hashtags              object
11  source                object
12  is_retweet            object
dtypes: float64(1), object(12)
memory usage: 198.3+ MB
```

Obr. 13 Štruktúra dataset-u pre sentimentálnu analýzu 10.02.2021 - 29.12.2021

Zdroj: Vlastné spracovanie v Python 3.8.0.

Práca rovnako ako pri historických cenách využíva na testovanie finálnych modelov, algoritmov, investičné stratégie práce aktualizovaný dataset tweet-ov rozšírený o obdobie od 30.12.2021 do 10.03.2022 predstavujúci dodatočných 434 391 záznamov príspevkov voči východiskovému dataset-u.

```
In [8]: # ponechať iba dané stĺpce dát
tweets_df = tweets_df[["user_followers", "date", "text"]]
tweets_df["date"] = pd.to_datetime(tweets_df["date"], errors="coerce")
tweets_df["date"].isna().sum()
```

Obr. 14 Výber analyzovaných dát dataset-u pre sentimentálnu analýzu

Zdroj: Vlastné spracovanie v Python 3.8.0.

Finálne dáta pre tvorbu sentimentálnej analýzy pred špecifickými úpravami potrebnými pre rozdielne algoritmy znázorňuje Obr. 15.

Out[9]:

	user_followers	date	text
0	8534.0	2021-02-10 23:59:04	Blue Ridge Bank shares halted by NYSE after #b...
1	6769.0	2021-02-10 23:58:48	🤖 Today, that's this #Thursday, we will do a "...
2	128.0	2021-02-10 23:54:48	Guys evening, I have read this article about B...
3	625.0	2021-02-10 23:54:33	\$BTC A big chance in a billion! Price: \487264...
4	1249.0	2021-02-10 23:54:06	This network is secured by 9 508 nodes as of t...
...	...	...	...
1999529	529.0	2021-12-29 19:53:40	OlegBarca's Stream №547! Live in few mins!\n\n...
1999530	1117.0	2021-12-29 19:53:36	Crypto: MistToken\nSymbol: #MistSwap \nMistSwa...
1999531	26.0	2021-12-29 19:53:33	Just 🗑 for 3x , what about you ? \n\n@AccelDef...
1999532	67.0	2021-12-29 19:53:31	@Bitboy_Crypto A \$Inari Trailer Video has been...
1999533	11162.0	2021-12-29 19:53:31	CNBC has a crypto special tonight... you know ...

Obr. 15 Náhľad na začiatok a koniec dataset-u pre sentimentálnu analýzu

Zdroj: Vlastné spracovanie v Python 3.8.0.

3.4 Python

Pre implementáciu jednotlivých modelov postavených na strojovom učení potrebných pre našu analýzu v práci využívame Python verziu 3.8.0.. Využitie Python-u predstavuje jednu z najefektívnejších možností tvorby modelov strojového učenia, hlavne vďaka špeciálne vyvinutým knižniciam Python-u, ktoré boli tvorené s hlavným zameraním na čo najrýchlejšie a najjednoduchšie vytváranie a implementáciu algoritmov strojového učenia. Hlavnými knižnicami Python-u využitých pri tvorbe a analýze modelov a obchodných stratégií, sú:

1. **Tensorflow** - používa sa pri písaní algoritmov a vykonávaní náročných komutácií, ktoré zahŕňajú neurónové siete,
2. **Keras** - táto knižnica zjednodušuje implementáciu neurónových sietí a zameriava sa na hlboké učenie (deep learning),
3. **Scikit-Learn** - jedna z najlepších knižníc na prácu s komplexnými dátami, sada Python-ových modulov pre strojové učenie a dolovanie dát,
4. **NumPy** - výpočty vedeckých / matematických údajov, spracovanie polí pre čísla, reťazce, záznamy a objekty,

5. **Pandas** - vysoko výkonné a ľahko použiteľné dátové štruktúry a analytické nástroje,
6. **NLTK** – alebo Natural Language Toolkit predstavuje poprednú platformu, rozšírenie na vytváranie programov Python pre prácu s údajmi ľudského jazyka.

Toto sú hlavné knižnice Python-u, ktoré využívame pri procese kódovania algoritmov strojového učenia, avšak nie sú to všetky. Použili sme Anaconda Distribution ako open-source platformu pre prácu s vedeckými dátami, ktorá obsahuje nielen viacero vedeckých balíkov a predinštalovaných knižníc, ale rovnako umožňuje jednoduché vyhľadávanie a inštaláciu ďalších balíkov potrebných pri tvorbe modelov strojového učenia prostredníctvom Python-u, priamo cez prostredie anakondy.

V neposlednom rade využívame Spyder ako Python IDE vývojové prostredie, ktoré umožňuje pokročilé úpravy, interaktívne testovanie, debugging a ďalšie. Výber vývojového prostredia pre Python je iba subjektívna záležitosť, preto autor môže pri tvorbe využívať akékoľvek vývojové prostredie, ktoré vyhovuje jeho preferenciám.

3.5 Nastavenie parametrov algoritmov strojového učenia pre tvorbu investičných modelov

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 a Yeo-Johnson transformácia predstavuje automatický spôsob vykonávania týchto normalizácií údajov, čím nám umožní dosiahnuť lepší výkon v širokom rozsahu algoritmov strojového učenia transformáciou údajov na gaussovske resp. viac gaussovske rozdelenie (Zheng – Casari, 2018).

V súčasnosti Power Transformer podporuje Box-Coxovu transformáciu a Yeo-Johnsonovu transformáciu. Optimálny parameter na stabilizáciu rozptylu a minimalizáciu šikmosti sa odhaduje pomocou maximálnej pravdepodobnosti. Box-Cox vyžaduje, aby vstupné údaje boli striktne pozitívne, zatiaľ čo Yeo-Johnson podporuje pozitívne, nulové aj negatívne údaje. Štandardne sa na transformované údaje použije normalizácia s nulovým priemerom³ a jednotkovou odchýlkou⁴ (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 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 okien, v našom prípade hodín.

Predlžovanie časového okna predstavuje potencionálny problém z hľadiska zvýšenia výpočtového výkonu (a najmä času) potrebného na vyjadrenie novej hodnoty predikcie, signálu, ktorá musí nastať v reálnom čase. Algoritmus na základe novej hodnoty uzatváracej ceny, ktorú získa z trhu musí vypočítať novú predikovanú hodnotu pre modely využívajúce iba historické ceny, pri modeloch rozšírených o dáta sentiment-u musí vyjadriť agregovaný sentiment za poslednú hodinu (timeframe) a na základe architektúry investičného modelu vypočítať novú predpoveď, signál predstavujúci interakciu ceny a sentimentu. Predpoveď

³ **Nulový priemer** - priemer a očakávaná hodnota štandardného normálneho rozdelenia je nula

⁴ **Jednotkový rozptyl** - znamená, že štandardná odchýlka vzorky, ako aj rozptyl budú mať tendenciu smerovať k hodnote 1, keďže veľkosť vzorky smeruje k nekonečnu

k nasledujúcej hodinovej uzatváracej cene je tak kombináciou za 31 predchádzajúcich hodinových údajov rozšírených o novo vyjadrenú hodnotu 32. Východným nastavením pre všetky modely je časové obmedzenie na výpočet jednotlivých procesov a doručenie modelu novej relevantnej hodnoty predikcie, signálu, pre obdobie 33 na úrovni, ktorá je menšia ako dĺžka časového okna (timeframe), čiže jednej hodiny. Toto obmedzenie zabezpečí, že žiaden model a za žiadnych okolností nebude mať k dispozícii nové hodnoty reálne dosiahnuté trhom, ktoré by využil pre vyjadrenie predikcie. So zvyšujúcim sa dĺžkou okna sa však predlžuje rozptyl medzi uzatváracou cenou, ktorú algoritmus pri výpočte berie do úvahy a reálnou hodnotou, za ktorú bude môcť po výpočte obchodný systém aktívum nakúpiť. Za predpokladu, že by v rámci času menšieho ako jedna hodina model nedokázal vyjadriť nový signál, predikciu, výstupom za dané obdobie bude 0, čiže chýbajúca hodnota.

Hyperparameter modelu je konfigurácia, ktorá je externá voči modelu a ktorej hodnotu nemožno odhadnúť z údajov. Hyperparametre sú tak premenné, ktoré riadia proces učenia modelu strojového učenia. V konečnom dôsledku určujú, ako sa model naučí špecifický vzťah medzi vstupom a predpoveďami. Často sa používajú v procesoch na pomoc pri odhadovaní parametrov modelu, sú špecifikované odborníkom, môžu byť nastavené pomocou heuristiky. Autori Kuhn a Johnson (2013) definujú možnosti ak nemôžeme poznať najlepšiu hodnotu pre modelový hyperparameter pri riešení špecifickej úlohy, máme sa zamerať na hodnoty použité pri iných modeloch alebo hľadať najlepšiu hodnotu metódou pokus-omyl.

Dáta sú rozdelené na tréningové⁵ a testovacie⁶ dáta, kde na testovanie je využitých posledných 1024 hodín príslušného dataset-u. 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 validáciu⁷ v pomere 0,9 k 0,1, za účelom vyhodnotenia, ktorý model je počas tréningu najlepší, kde model využíva

⁵ **Tréningový dataset** - vzorka dát použitých na prispôsobenie, učenie modelu.

⁶ **Testovací dataset** - dáta použité na poskytnutie nezaujatého hodnotenia konečného modelu prispôbeného množine tréningových údajov. Model ju pri učení nikdy nevidí.

⁷ **Validačný dataset** - vzorka údajov použitá na poskytnutie nezaujatého hodnotenia modelu prispôbeného množine tréningových údajov pri ladení hyperparametrov modelu. Model niekedy dané dáta vidí avšak nikdy sa z nich neučí.

validačné údaje aby zabránil overfitting-u 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 nuansy a šum v tréningových dátach do takej miery, ktorá negatívne ovplyvňuje výkon modelu na nových (počas tréningu nedostupných) dátach. To znamená, že šum alebo náhodné fluktuácie v tréningových dátach sú zachytené a naučené ako koncepty modelom. Problémom je, že tieto koncepty sa nevzťahujú na nové údaje a negatívne ovplyvňujú 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 1024 hodnôt príslušného dataset-u 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 (dávkach), 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 trénovanie modelu v hlbokom učení. Cesta učenia v mini-batch gradient-nom 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 trénovanie modelu.

Adam Optimizer je tvorený kombináciou dvoch optimalizátorov hlbokého učenia, ktorými sú Momentum Optimizer a RMSprop Optimizer. Adam Optimizer je prezentovaný autormi Kingma a Ba (2017) prostredníctvom jeho hlavných výhod, medzi ktoré patria hlavne:

- jednoduchosť na implementáciu,
- výpočtová efektivita,
- nízke nároky na pamäť,

- dobre sa hodí pre problémy, ktoré sú veľké z hľadiska údajov a/alebo parametrov,
- vhodné pre nestacionárne ciele,
- hyperparametre majú intuitívnu interpretáciu a zvyčajne vyžadujú malé ladenie.

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 overfitting-u, 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.

```
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)
```

Obr. 16 Optuna pre hyperparametre pri LSTM

Zdroj: Vlastné spracovanie v Python 3.8.0.

Bayesovský optimalizačný algoritmus postupuje pri hľadaní optimálnych hyperparametrov nasledovným spôsobom:

1. najskôr náhodne vyberie podmnožinu hyperparametrov a zoradí ich na základe ich skóre,
2. hyperparametre sú ďalej rozdelené do dvoch skupín na základe preddefinovaného kvantilu,

3. tieto dve skupiny sa potom modelujú do odhadovaných hustôt $l(x_1)$ a $g(x_2)$ pomocou Parzenových odhadov (odhadov hustoty jadra),
4. hľadá hyperparametre s najvyšším očakávaným zlepšením [najnižšie $l(x_1)/g(x_2)$],
5. hyperparameter s najvyšším očakávaným zlepšením sa vyhodnotí, triedi a opäť rozdelí. Tento proces sa opakuje, kým sa nedokončí rozpočet a vráti sa najlepšie hyperparametre (Optuna, 2022).

Jednotlivé hyperparametre využívané v práci sa menia podľa modelov, avšak medzi hlavné patria:

- 1) **batch size** - je termín používaný v strojovom učení a označuje počet tréningových príkladov použitých v jednej iterácii predtým ako je model aktualizovaný. Veľkosť batch môže byť jedna z troch možností:
 - a) batch mode kde je batch size rovná celkovému súboru údajov, čím sú hodnoty iterácie a epochy ekvivalentné,
 - b) mini-batch mode kde je veľkosť batch väčšia ako jedna, ale menšia ako celková veľkosť súboru údajov. Zvyčajne ide o číslo, ktorým možno rozdeliť na celkovú veľkosť množiny údajov (tréningové dáta),
 - c) stochastic mode kde sa veľkosť batch rovná jednej. Preto sa gradient a parametre neurónovej siete aktualizujú po každej vzorke.

V praxi mini-batch mode predstavuje najrýchlejší/najefektívnejší typ z hľadiska iteračného času resp. počítačnej náročnosti. Vo všeobecnosti sa využívajú hodnoty pre mini-batch (batch size) na úrovni 64, 128, 256 alebo 512 (Bengio, 2012).
- 2) **epochs** - je hyperparameter, ktorý definuje koľkokrát učiaci algoritmus prejde cez celý tréningový súbor údajov. Jedna epoch-a znamená, že každá vzorka v tréningovom dataset-e mala možnosť aktualizovať parametre interného modelu. Jedna epoch-a pozostáva z jednej alebo viacerých batch,
- 3) **neurons** - v modeloch hlbokého učenia predstavujú uzly, cez ktoré prechádzajú údaje a výpočty. Neuróny prijímajú jeden alebo viac vstupných signálov, ktoré môžu pochádzať zo súboru nespracovaných údajov, alebo z neurónov umiestnených na predchádzajúcej vrstve neurónovej siete. Po prijatí vstupu/ov vykonávajú výpočty a posielajú cez výstupnú synapsiu výstupné údaje,

- 4) **learning rate** – rýchlosť učenia definuje, ako rýchlo sieť aktualizuje svoje parametre. Nízka rýchlosť spomaľuje proces učenia, ale plynule konverguje. Vyššia rýchlosť zasa urýchľuje učenie, ale nemusí konvergovať,
- 5) **dropout** - je technika regularizácie, pri ktorej sú vyhodnené náhodné neuróny za účelom zvýšenia presnosti validácie, čím sa zvyšuje zovšeobecňujúca sila modelu (Cruz – Pyrcz, 2021).

Optuna má k dispozícii 100 pokusov, ktoré nám poskytnú komplexný pohľad na priestor hľadania a obsahuje časové ohraničenie na úrovni 7200 sekúnd, čiže 2 hodiny. Výnimku z časového ohraničenia je nutné poskytnúť modelom analyzujúcim text (slová), keďže majú na spracovania väčšie množstvo vstupných dát.

```
In [11]: # cieľ minimalizovať loss function
study = optuna.create_study(direction="minimize")
study.optimize(objective,n_trials=100,timeout=7200) # stop po 2 hodinách
# ponechaj najlepší pokus
best_trial = study.best_trial|
best_trial

e: 8.121892460621893e-05.
[I 2022-01-09 00:40:49,890] Trial 93 finished with value: 0.00018833221110986
s': 27, 'neurons': 4, 'dropout': 0.012120793345202073, 'log_learning_rate': -
8.121892460621893e-05.
[I 2022-01-09 00:42:29,155] Trial 94 finished with value: 0.00123078131582587
s': 29, 'neurons': 2, 'dropout': 0.029662974319012332, 'log_learning_rate': -
8.121892460621893e-05.
[I 2022-01-09 00:43:30,262] Trial 95 finished with value: 0.00524437054991722
s': 17, 'neurons': 3, 'dropout': 0.05919468677733711, 'log_learning_rate': -1
8.121892460621893e-05.
[I 2022-01-09 00:44:50,938] Trial 96 finished with value: 0.00045725808013236
s': 24, 'neurons': 5, 'dropout': 0.011021066426083453, 'log_learning_rate': -
8.121892460621893e-05.

Out[11]: FrozenTrial(number=41, values=[8.121892460621893e-05], datetime_start=datetime
ime_complete=datetime.datetime(2022, 1, 8, 23, 24, 33, 548421), params={'log2
pout': 0.0007204497165631254, 'log_learning_rate': -1.4999106420438988}, dist
ion(high=8, low=4, step=1), 'epochs': IntUniformDistribution(high=50, low=2,
2, low=1, step=1), 'dropout': UniformDistribution(high=0.5, low=0.0), 'log_le
-3.0)), user_attrs={}, system_attrs={}, intermediate_values={}, trial_id=41,
```

Obr. 17 Výsledok optimalizácie hyperparametrov pre LSTM

Zdroj: Vlastné spracovanie v Python 3.8.0.

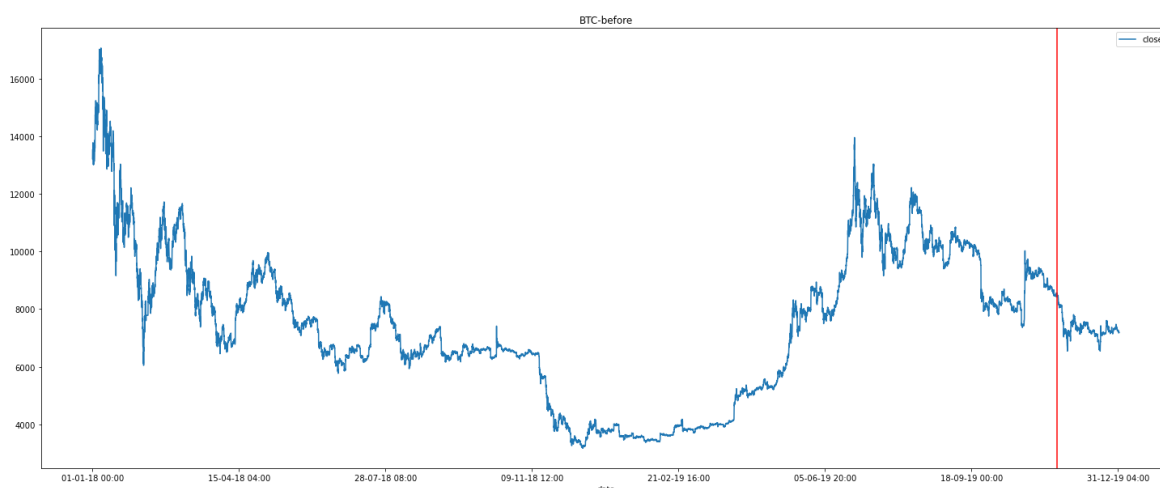
Model je následne znova spustený s najlepšimi hodnotami, ktoré Optuna našla pre jednotlivé hyperparametre.

4 Výsledky práce

Subjektom skúmania práce je kryptoaktívum Bitcoin, avšak v práci pre porovnanie výsledkov výkonnosti základných modelov postavených na vývoji historických cien používame aj referenčné meny pre subjekt skúmania predstavujúce tzv. Altcoins, čiže alternatívne kryptomeny voči hlavnému predstaviteľovi kryptoaktív Bitcoin-u.

Pre porozumenie dosiahnutým výsledkom je potrebné mať informácie ohľadom analyzovaného dataset-u, čiže ohľadom vývoja ceny subjektu skúmania a referenčných mien v nami analyzovaných obdobiach predstavujúcich dataset-y pred pandémiou COVID-19 a počas pandémie.

Jednotlivé obdobia predstavujú zásadný rozdiel vo vývoji na daných trhoch, kde obdobie pred pandémiou predstavuje medvedí trh⁸ (bear market) spolu s obdobím kumulácie aktív. Časový horizont ovplyvnený pandémiou COVID-19 naopak predstavuje fázu kumulácie nasledovanú býčím trendom (bull market).



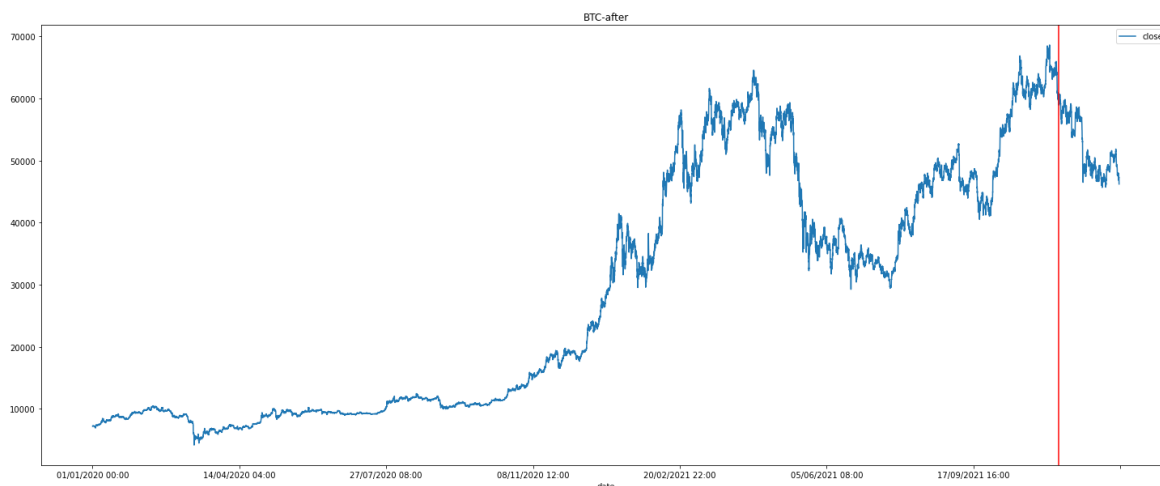
Obr. 18 Analyzovaný dataset historických cien Bitcoin pred COVID-19

Zdroj: Vlastné spracovanie v Python 3.8.0.

Konkrétne vývoj ceny BTC/USD za obdobie od 01.01.2018 0:00 až po 31.12.2019 23:00 predstavuje pokles z dovtedy maximálnej hodnoty 19776.99 BTC/USD dosiahnutej

⁸Guggenheim investments (2019) pri definovaní jednotlivých cyklov historického vývoja na býčie, medvedie alebo obdobia kumulácie aktív, definoval koniec býčieho trendu, ako prepad hodnoty investičného aktíva pod jeho vrchol, pričom jeho hodnota bola nižšia ako vrchol počas signifikantného časového obdobia. Naopak začiatok býčieho trendu je v čase, keď je na trhu vytvorený nový vrchol, ktorý prekonáva vrchol posledného známeho býčieho trendu. Medziobdobia predstavujú fázu kumulácie aktív.

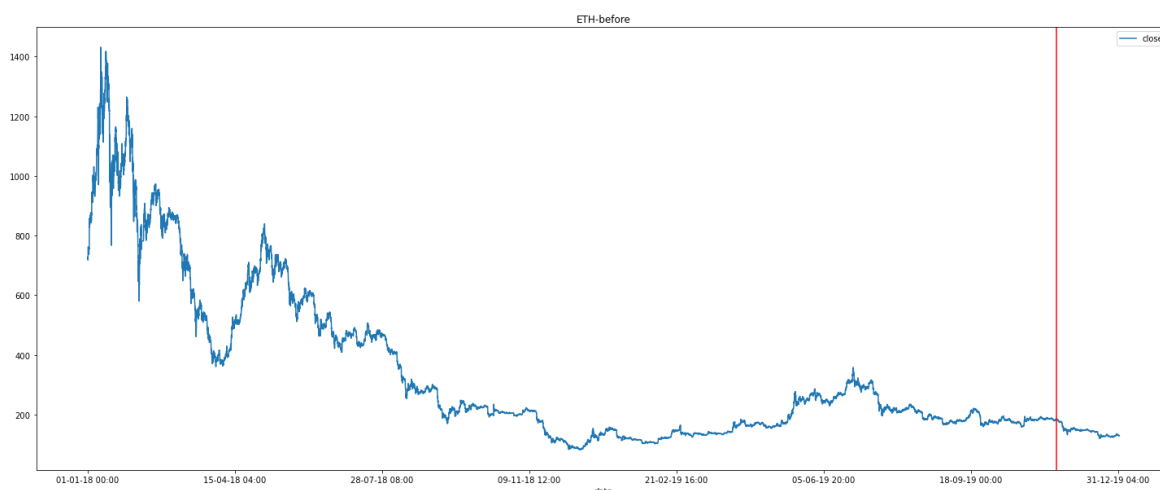
v decembri 2017, odkedy bol Bitcoin v klesajúcom trende nasledovanom lokálnym minimom na úrovni 3126.36 BTC/USD v decembri 2018. Táto hodnota súčasne predstavovala minimum medvedieho trhu. Nárast hodnoty BTC/USD na úroveň 13854.63 v júni 2019 nepredstavoval finálnu zmenu medvedieho trendu ale skôr pokračovanie kumulácie jednotlivých investorov.



Obr. 19 Analyzovaný dataset historických cien Bitcoin počas COVID-19

Zdroj: Vlastné spracovanie v Python 3.8.0.

Kumulácia pokračovala aj do dataset-u so začiatkom 01.01.2020 0:00 a koncom 30.12.21 0:00. Táto fáza trhu bola ukončená minimom v marci 2020 na úrovni 3892.74 BTC/USD predstavujúceho minimálne ocenenie Bitcoin-u v danom dataset-e.

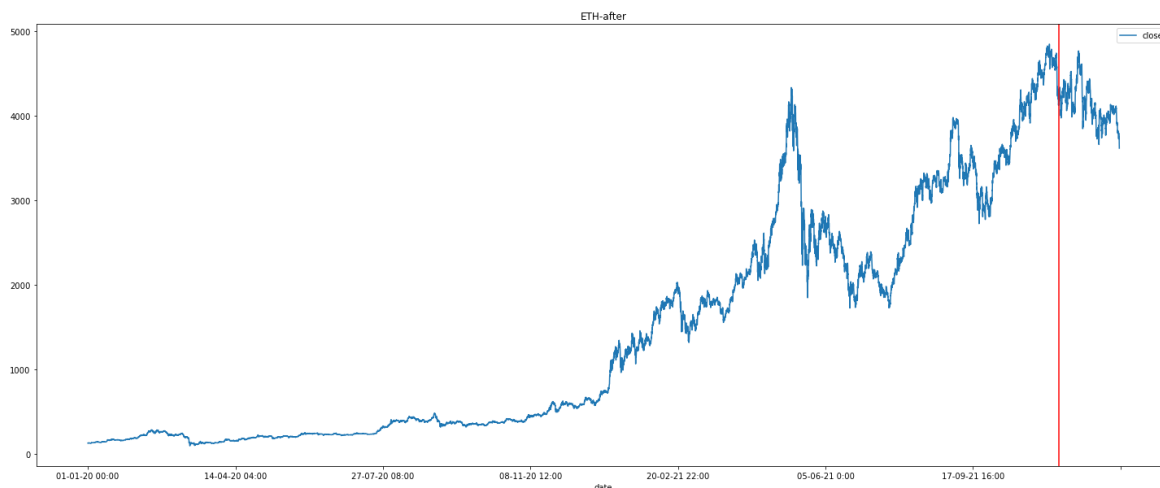


Obr. 20 Analyzovaný dataset historických cien Ethereum pred COVID-19

Zdroj: Vlastné spracovanie v Python 3.8.0.

Nasleduje postupný rast hodnoty BTC voči USD s doterajším ATH (All Time High) na úrovni 69 000 BTC/USD dosiahnutého v novembri 2021.

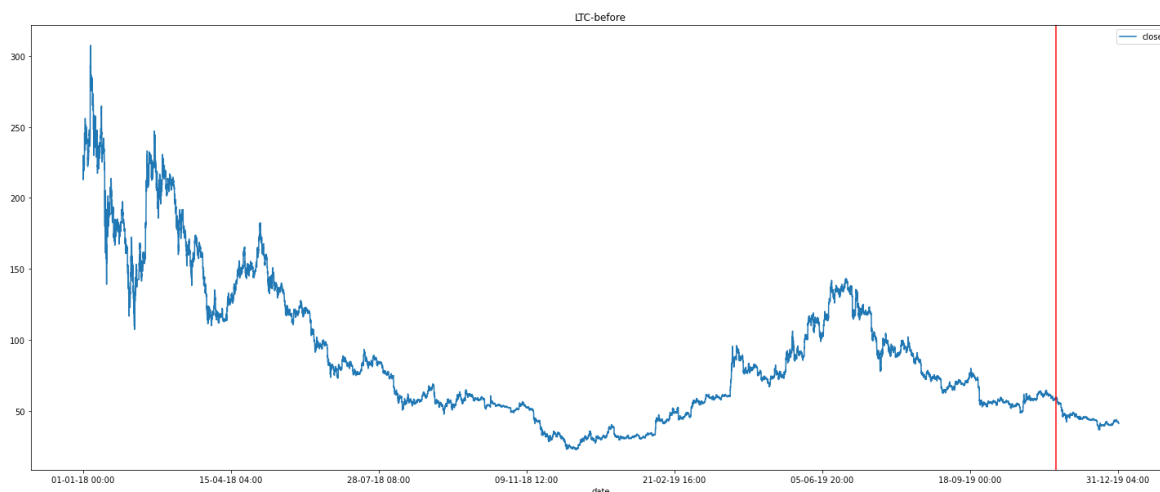
Vývoj analyzovaných referenčných altcoins je totožný z hľadiska jednotlivých fáz trhu s vývojom ceny Bitcoin-u.



Obr. 21 Analyzovaný dataset historických cien Ethereum počas COVID-19

Zdroj: Vlastné spracovanie v Python 3.8.0.

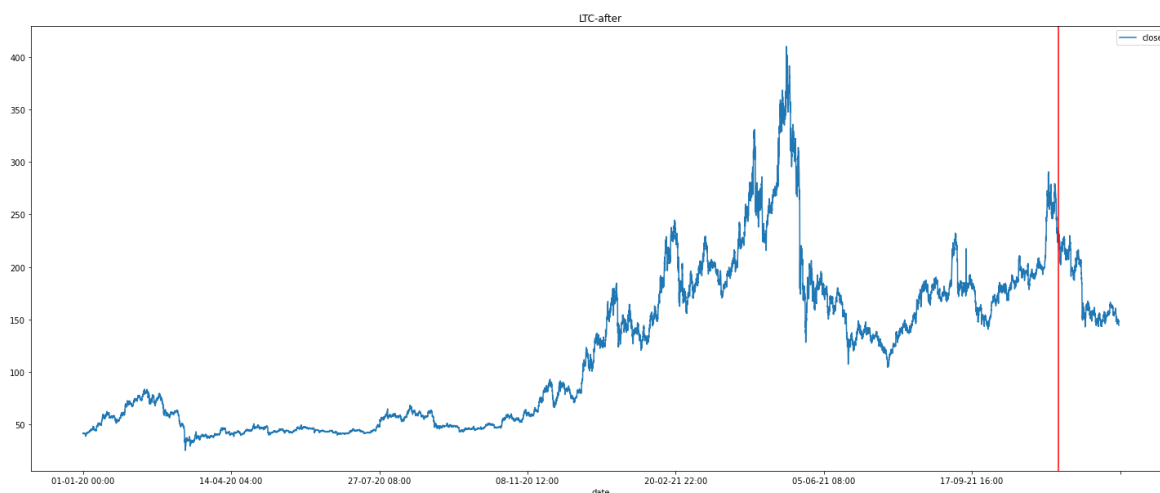
Jednotlivé najvýznamnejšie prepady a prírastky cien nastávali v obdobiach, ktoré sa zhodujú so zmenami v cenovom vývoji Bitcoin-u. Počas analyzovaných období tak nenastala tzv. Altseason, ktorá je definovaná zvyšovaním podielu altcoins na celkovej kapitalizácii kryptomenového trhu na úkor kapitalizácie Bitcoin-u.



Obr. 22 Analyzovaný dataset historických cien Litecoin pred COVID-19

Zdroj: Vlastné spracovanie v Python 3.8.0.

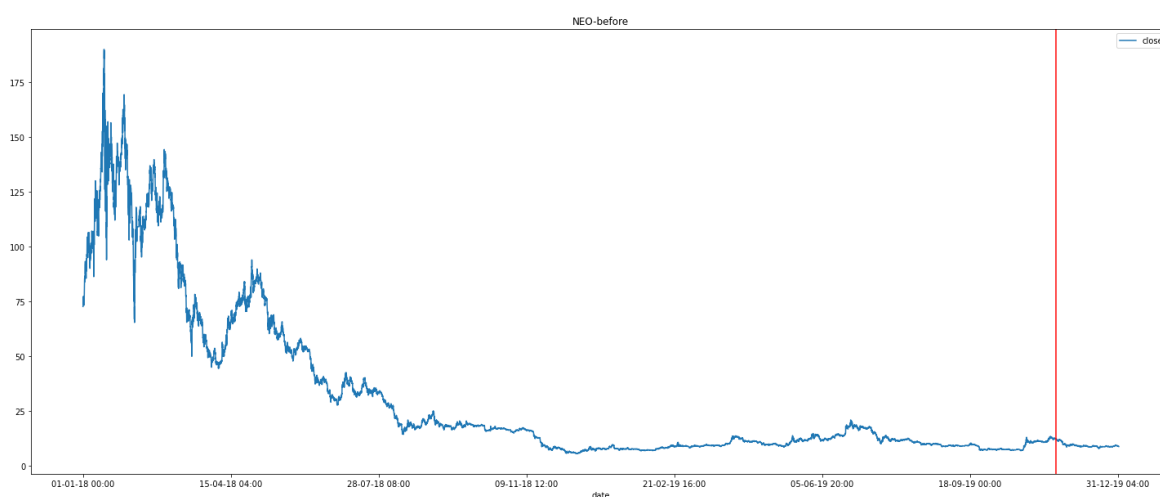
Počas skúmaných období je tak vývoj ceny všetkých kryptomien ovplyvňovaný aktuálnym sentimentom na trhu Bitcoin-u. Tým sú pre nás dataset-y historických cien porovnateľné, resp. potenciálne rozdiely vo výkonnosti modelov strojového učenia nebudú ovplyvnené úplne odlišným vývojom ceny podkladových aktív.



Obr. 23 Analyzovaný dataset historických cien Litecoin počas COVID-19

Zdroj: Vlastné spracovanie v Python 3.8.0.

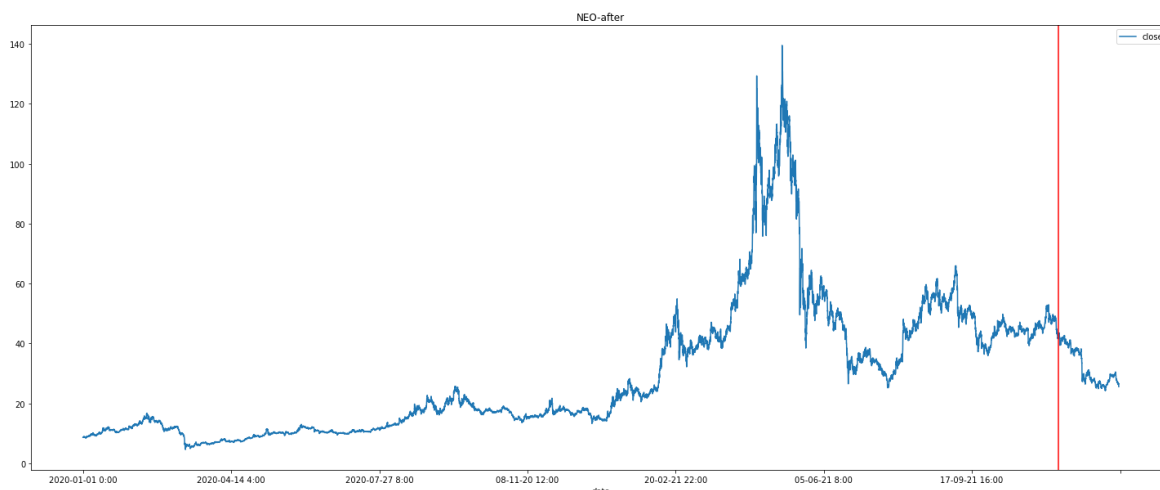
Ako môžeme vidieť vo vývoji ceny jednotlivých altcoins, so zvyšujúcou kapitalizáciou daného aktíva rovnako rastie aj reakcia na zmeny, ktoré nastali na Bitcoin-ovom trhu.



Obr. 24 Analyzovaný dataset historických cien NEO pred COVID-19

Zdroj: Vlastné spracovanie v Python 3.8.0.

Daná skutočnosť plynie z faktu, že v prípade fázy trhu, ktorá nie je ovplyvnená altseason, jednotliví investori primárne investujú do Bitcoin-u, čiže kryptomeny, ktorá má najvyššiu kapitalizáciu a tým investori predpokladajú nižšiu smerodajnú odchýlku, riziko vysokého poklesu. Až následne investujú, či diverzifikujú svoje investície do kryptoaktív medzi ďalšie kryptomeny.



Obr. 25 Analyzovaný dataset historických cien NEO počas COVID-19

Zdroj: Vlastné spracovanie v Python 3.8.0.

4.1 Supervised learning – modely založené na historických cenách

Pre definovanie počiatočnej výkonnosti jednotlivých algoritmov strojového učenia pri predikcii založenej iba na historickom vývoji uzatváracích cien v čase, ako aj so zameraním sa na zúženie množstva skúmaných algoritmov pri nami analyzovanom probléme predikovania dlhých časových radov na modely s najlepšimi výsledkami, v práci vytvárame modely postavené na algoritmoch:

- SARIMA,
- Support Vector Machines (SVM),
- Prophet,
- Long Short-Term Memory (LSTM),
- DeepAR.

Nastavenie potrebných parametrov pre modely algoritmov strojového učenia sú popísané v časti metodika práce.

Na meranie presnosti predikcie (alebo chyby) modelov strojového učenia využívame MSE⁹, RMSE¹⁰, MAE¹¹ a MAPE¹², ktoré nám ako regresné metriky vyjadrujú typ vzdialenosti predikovaných hodnôt od hodnôt skutočných. Každá z metrík má svoje inherentné výhody a nevýhody plynúce z jej konštrukcie. Využitím viacerých metrík 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.

Obr. 26 až Obr. 33 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 dataset-mi pre nami stanovený cieľ nepoužiteľné výsledky.



Obr. 26 Porovnanie výsledkov ML modelov pre historické ceny Bitcoin-before

Zdroj: Vlastné spracovanie v Python 3.8.0.

⁹ Mean squared error

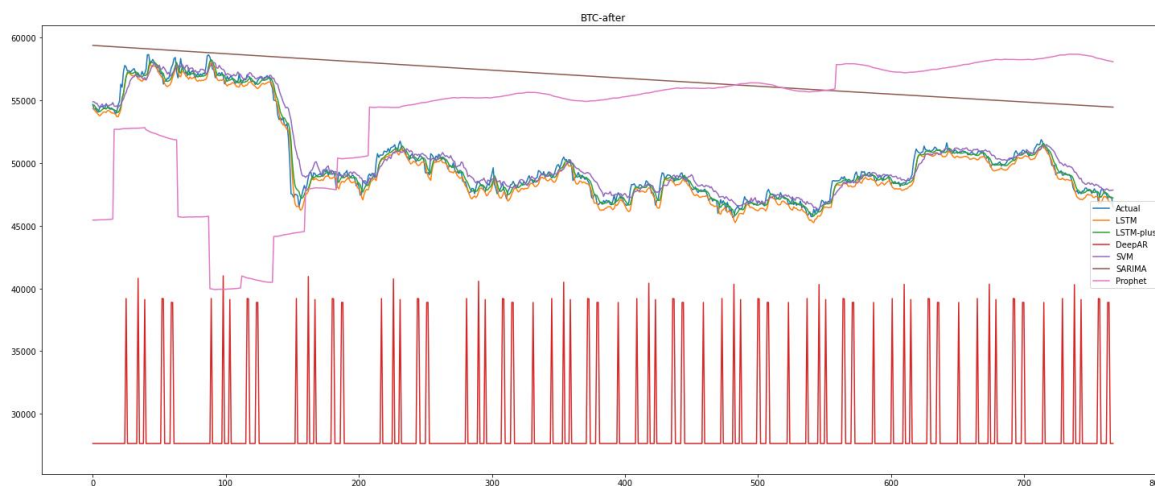
¹⁰ Root mean squared error

¹¹ Mean absolute error

¹² Mean absolute percentage error

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.

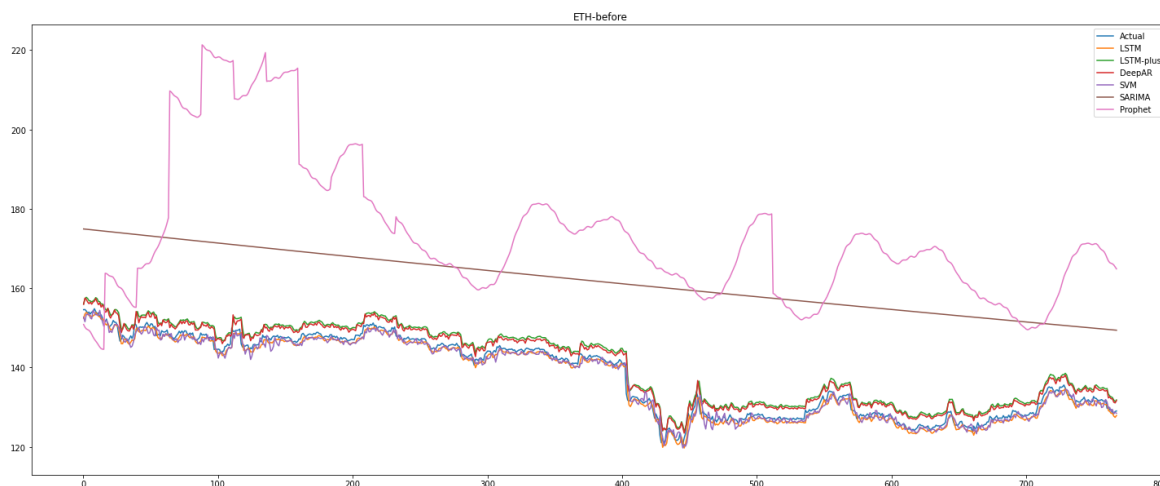


Obr. 27 Porovnanie výsledkov ML modelov pre historické ceny Bitcoin-after

Zdroj: Vlastné spracovanie v Python 3.8.0.

Cieľom vývojárov Facebook-u 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 dataset-mi. Napriek zameraniu algoritmu na modelovanie dát časových radov, potrebuje model dataset zo silnými sezónnymi vplyvmi, ktoré naše dataset-y vývoja ceny podkladových aktív neobsahujú.



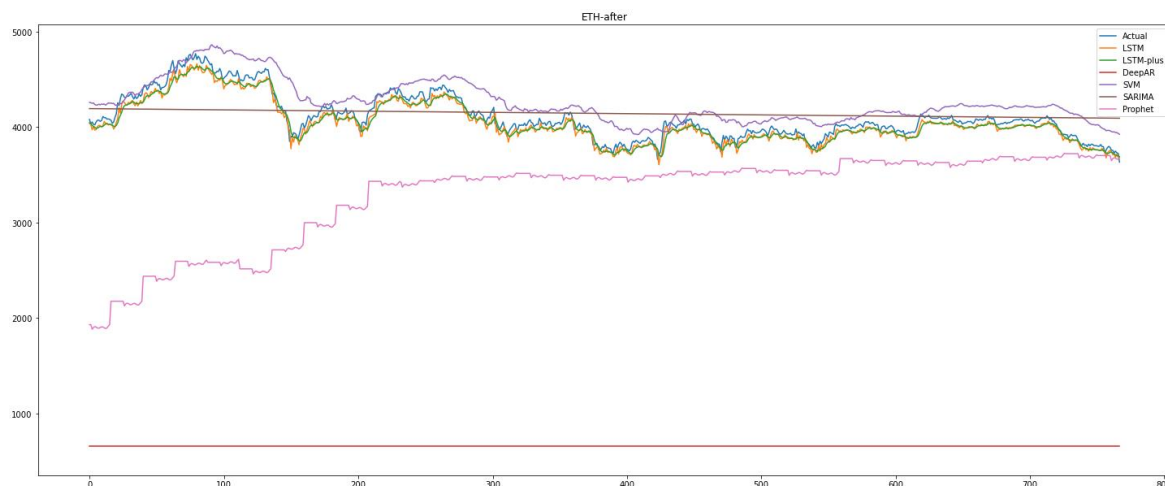
Obr. 28 Porovnanie výsledkov ML modelov pre historické ceny Ethereum-before

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 dataset-och, 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 trénovanie 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 dataset-och neodstránila.

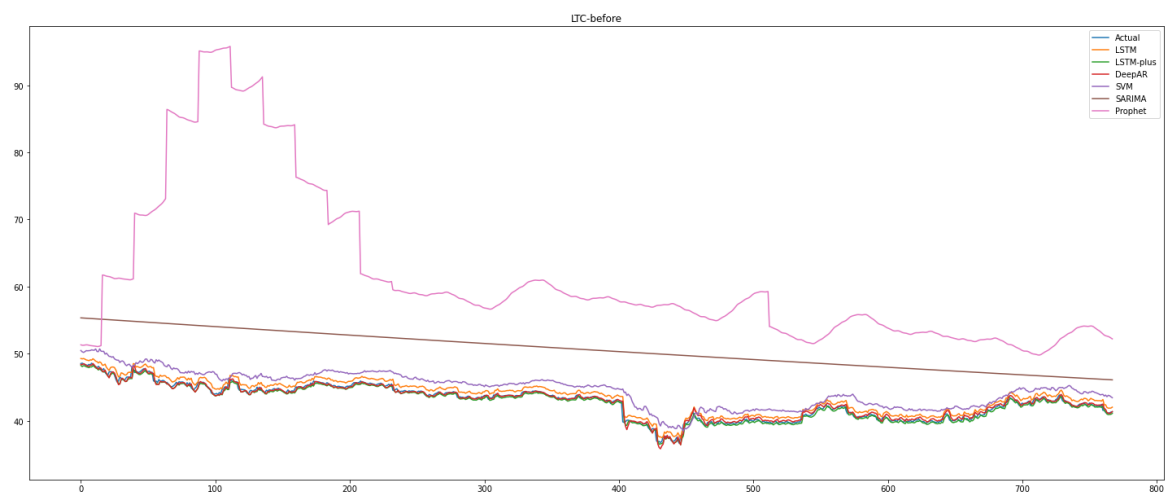
Nakoľko je konzistentnosť pri tvorbe automatizovaného obchodného systému veľmi podstatná algoritmus DeepAR nebudeme ďalej v práci využívať ako benchmark.



Obr. 29 Porovnanie výsledkov ML modelov pre historické ceny Ethereum-after

Zdroj: Vlastné spracovanie v Python 3.8.0.

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 dataset-mi v porovnaní s algoritmom DeepAR, čím pre podmienky analýzy predstavuje potenciálny algoritmus ako benchmark.



Obr. 30 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-fitting-u, 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.



Obr. 31 Porovnanie výsledkov ML modelov pre historické ceny Litecoin-after

Zdroj: Vlastné spracovanie v Python 3.8.0.

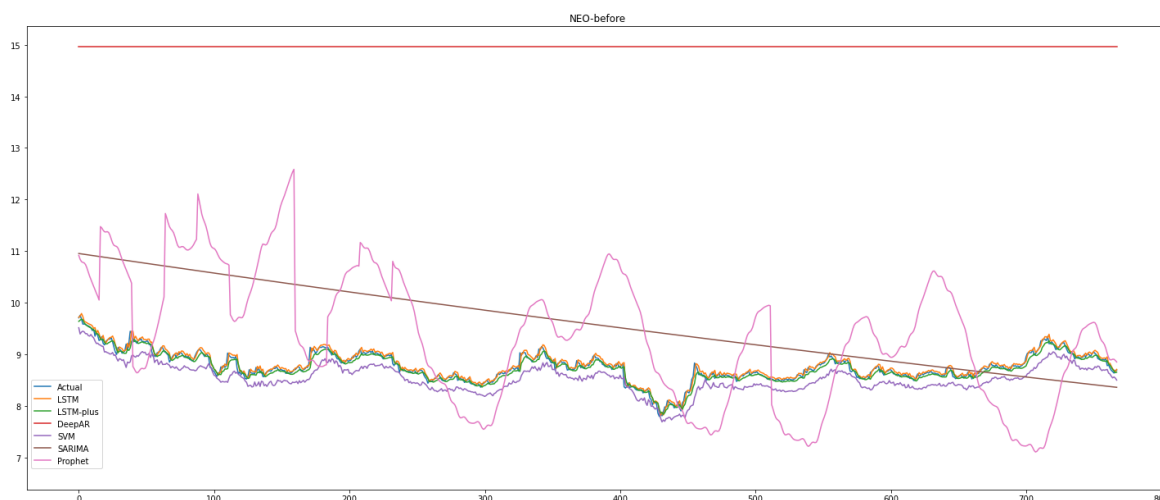
Posledným testovaným algoritmom strojového učenia je LSTM. Už na prvý pohľad zaznamenal daný model pre všetky dataset-y vysokú presnosť predpovede. Predikcie sú blízko skutočným hodnotám a model dobre chápe spôsob, akým sa cena mení.

Daný model LSTM predstavuje jednu bunkovú sieť nasledovanú jednou vrstvou dense¹³. 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

¹³ **Dense layer** – 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.

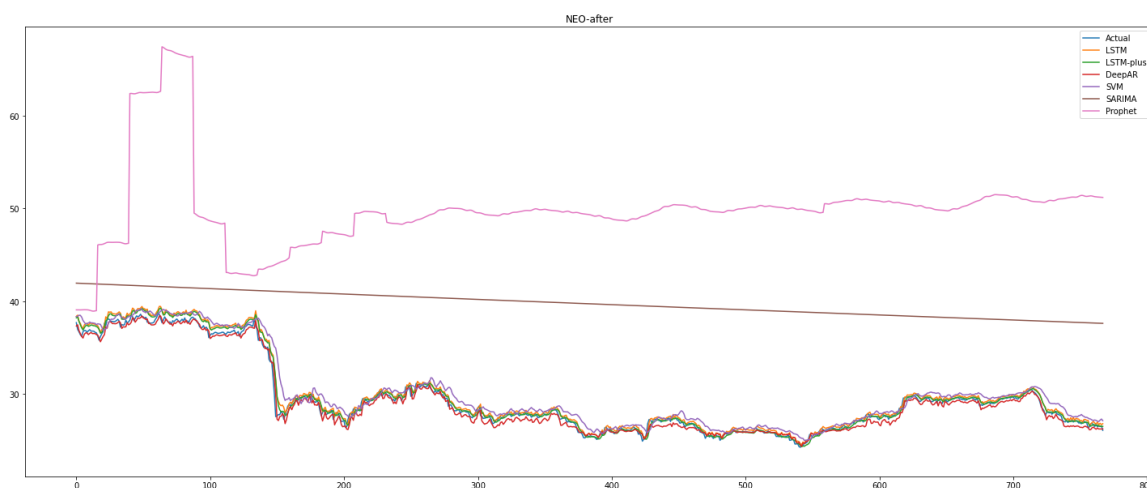
predstavovala v našich modeloch na nami analyzovaných údajoch druhá bunka LSTM, ktorá vracia sekvencie podstatné spomalenie tréningového času a tým aj výpočtovú náročnosť, pričom sme nezaznamenali zvýšenie výkonu výsledného modelu.



Obr. 32 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 Tab. 2 až Tab. 5.



Obr. 33 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 uzatváracej ceny a hodnotou predikovanou modelom. Vzhľadom na faktor kvadratury je náchylnejší na outlier-ov (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 dataset-ov.

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

Tab. 2 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 benchmark-ovej hodnoty NEO počas pandémie aj pre dataset-y Ethereum a Litecoin pred pandemiou, kde práve pri Litecoin-e dosiahol takmer totožnú hodnotu s benchmark-om pre daný dataset, ktorým je pri MSE LSTM+ model. Pre ostatné dataset-y 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 dataset-y 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+.

Tab. 3 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 benchmark-u nastali zmeny pre dataset Litecoin pred pandémiou a NEO počas pandémie, pre ktoré si navzájom vymenili prvé a druhé miesto modely DeepAR a LSTM+.

Tab. 4 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.

Tab. 5 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 informácii 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 dataset-mi, 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.

Preto práve modely LSTM a LSTM+ a nimi nadobudnuté hodnoty budeme ďalej využívať ako benchmark. LSTM+ pre nás však bude predstavovať hlavný model, keďže subjektom skúmania je Bitcoin a práve tento model dosiahol pre dané aktívum najlepšie hodnoty, schopnosť porozumenia vývoju ceny v čase. Keďže predikcie zahrňujúce textové údaje budú na základe datasetu tweet-ov analyzované na dataset-e pre Bitcoin počas pandémie, pre dané modely budeme využívať LSTM+ ako základný model predikcie podľa vývoja minulých uzatváracích cien.

4.2 Sentimentálna analýza – modely analyzujúce text

Pri obchodovaní finančne špecifického aktíva akým je Bitcoin, investor nemá k dispozícii takmer žiadne informácie na analýzu, či tvorbu obchodného systému okrem cenovej histórie a objemu obchodovania v jednotlivých časových horizontoch.

Rovnako investovanie, či tradeovanie na takomto trhu na základe fundamentálnej analýzy je do značnej miery limitované, nakoľko nemáme k dispozícii žiadne oficiálne informácie z centrálnych bánk, predstaviteľov riadiacich entít, výsledkov hospodárenia a podobne, ktoré sú využívané pri tradičných menových, dlhopisových, či akciových trhoch ako základ pri tvorbe fundamentálnej analýzy, daná skutočnosť plynie práve z decentralizovaného charakteru Bitcoin-u.

Jedinou možnosťou pre rozšírenie modelu postaveného na historickom vývoji ceny Bitcoin-u ako podkladového aktíva je využiť informácie dostupné prostredníctvom analýzy sentimentu na trhu.

Vzhľadom na predošlé výsledky budú všetky modely využívajúce sentimentálnu analýzu postavené na základe algoritmu LSTM+. Časť modelu analyzujúca text využíva algoritmy:

- GloVe,
- VADER.

4.2.1 GloVe

Autori algoritmu GloVe Pennington, Socher a Manning (2014) poskytujú na GloVe web stránke <https://nlp.stanford.edu/projects/glove/> dataset-y pred-trénovaných vektorov slov pre Wikipediu, Gigaword, Common Crawl a Twitter, ktoré sú prístupné verejnosti na základe licencie PDDL v1.0 (Open Data Commons).

Nami využívaný dataset zameraný na vektory slov pre Twitter je k dispozícii ako 25, 50, 100 a 200 dimenzionálny. Súbor obsahuje slová a nim prislúchajúce vektory pre rozličné dimenzie, kde 100 dimenzionálny dataset predstavuje 100 dimenzií pre každé slovo v dataset-e. Dataset obsahuje dve miliardy tweet-ov obsahujúcich 27 miliárd špecifických znakov, symbolov a unikátnych 1 193 515 slov, kde všetky slová mimo daného dataset-u nadobúdajú hodnotu nulového vektora. Vzhľadom na štruktúru EmbeddingLayer (vkladaciu vrstvu), ktorá predstavuje maticu veľkosti dataset-u (slov) vynásobenú počtom dimenzií,

z hľadiska výpočtovej náročnosti v našej analýze vytvárame 25 dimenzionálny GloVe model, na ktorý teda využívame 25d časť zo súboru glove.twitter.27B dostupného na <http://nlp.stanford.edu/data/glove.twitter.27B.zip>.

```
In [11]: # zaokrúhliť dátum nadol na najbližšiu hodinu
def shrink(date):
    return date.replace(microsecond=0,second=0,minute=0) + timedelta(hours=1)
```

Obr. 34 Zaokrúhlenie jednotlivých tweet-ov nadol na najbližšiu hodinu

Zdroj: Vlastné spracovanie v Python 3.8.0.

Keďže finálny model predpovede vychádza z dvoch odlišných dataset-ov historickej ceny a tweet-ov, je potrebné zosúladiť časové hľadisko za účelom použiteľnosti predikcie. Vývoj postavený na minulých cenách predstavuje hodinové uzatváracie ceny pre jednotlivé dni v dataset-e, avšak tweet-y obsahujú časový údaj pre deň datasetu, hodinu, minútu a sekundu, kedy boli uverejnené. Tým jednotlivý tweet predstavuje na základe sentimentálnej analýzy ohrozený konkrétny sentiment pre čas kedy bol publikovaný.

Zaokrúhlením dátumovej informácie dataset-u tweet-ov nadol na najbližšiu hodinu dostaneme požadovaný tvar, kde nadobudneme tweet-y rozdelené podľa hodiny, ku ktorej sú najbližšie, čím predstavujú sentiment k danej uzatváraciej hodinovej cene. Daná konverzia je potrebná pre oba algoritmy sentimentálnej analýzy.

```
In [13]: # aplikuj embedding a zorad dátum
tweets_df["date"] = tweets_df["date"].apply(shrink)
tweets_df["embedding"] = tweets_df.apply(embed,axis=1)
tweets_df
```

Out[13]:	user_followers	date	text	embedding
0	8534.0	2021-02-11 00:00:00	Blue Ridge Bank shares halted by NYSE after #b...	[5218.13511622412, 552.4998659421594, -2265.21...
1	6769.0	2021-02-11 00:00:00	🤖 Today, that's this #Thursday, we will do a "...	[4006.3798120700726, 810.1367675477169, -1672....
2	128.0	2021-02-11 00:00:00	Guys evening, I have read this article about B...	[81.2747060720203, 12.589200126314626, -32.112...
3	625.0	2021-02-11 00:00:00	\$BTC A big chance in a billion! Price: 1487264...	[262.9843497962444, 38.14590486640833, -156.29...
4	1249.0	2021-02-11 00:00:00	This network is secured by 9 508 nodes as of t...	[827.3227724310777, 159.30761073314238, -227.7...
...	...	...	...	...
1999529	529.0	2021-12-29 20:00:00	OlegBarca's Stream №547! Live in few mins!\n...	[324.69768362595653, 24.986974983292555, -133....
1999530	1117.0	2021-12-29 20:00:00	Crypto: MistToken\nSymbol: #MistSwap \nMistSwa...	[638.4036980761531, 90.20963401090903, -311.92...
1999531	26.0	2021-12-29 20:00:00	Just 🏠 for 3x , what about you ? \n\n@AccelDef...	[14.575599582019176, 3.2205335235392507, -9.45...
1999532	67.0	2021-12-29 20:00:00	@Bitboy_Crypto A Sinari Trailer Video has been...	[40.172756546191444, 5.460549678151626, -14.87...
1999533	11162.0	2021-12-29 20:00:00	CNBC has a crypto special tonight... you know ...	[6799.942565108972, 293.672634526479, -3405.32...

Obr. 35 Zoradený výstupný dataset tweet-ov podľa GloVe

Zdroj: Vlastné spracovanie v Python 3.8.0.

Obr. 35 obsahuje už zoradené príspevky pre dátum/čas, ktorý je v súlade s hodinovými uzatváracími cenami. Rovnako už obsahuje v časti embedding hodnoty pre daný príspevok na základe algoritmu GloVe a vybraných 25 dimenzií.

Využitie pred-trénovaných GloVe embeddings predstavuje rozšírenie prístupu Bag of Words¹⁴, či Tf-Idf¹⁵ ako dávame modelu 25 rozmerov údajov o význame tweet-u. Aby sa veci nekomplikovali, hodnota každého tweet-u predstavuje priemer zo slov, ktoré obsahuje a následne sa k údajom o tweet-och pridá vážený súčet, pričom váženou metrikou je počet follower-ov, ktorá určuje vplyv jednotlivých používateľov, ktorí publikujú tweet-y.

Tab. 6 *Prehľad výsledkov jednotlivých dimenzií model GloVe*

dátum	close (usd)	tweet00	...	tweet23	tweet24
29/12/2021 16:00	47727.43	1.571432e+05	...	9.003687e+04	9.262889e+04
29/12/2021 17:00	47554.91	6.834538e+06	...	3.579332e+06	4.260165e+06
29/12/2021 18:00	47531.47	5.671593e+06	...	2.836463e+06	3.592987e+06
29/12/2021 19:00	47148.61	1.203812e+06	...	6.934903e+05	7.884608e+05
29/12/2021 20:00	47369.97	1.701538e+06	...	1.003373e+06	1.089736e+06

Zdroj: Vlastné spracovanie

Keďže vstupné údaje pre optimalizáciu hyperparametrov sú oproti modelom postavených iba na historických cenách rozšírené o 25 dimenzií (tweet00 až tweet24), výsledný finálny dataset predstavuje z hľadiska veľkosti nárast voči iba cenovému modelu o viac ako 33 násobok. Časové obmedzenie definované pri cenových modeloch aplikované za účelom zabránenia overfitting-u na úrovni 2 hodín, bolo nutné pre modely rozšírené o dáta sentimentu predĺžiť na 6 hodín.

¹⁴ **Bag of Words (BoW)** – text je pri spracovaní prirodzeného jazyka reprezentovaný ako vrece slov, pričom sa nehľadí na gramatiku, slovosled, ale zameriava sa na mnohorakosť. Model potrebuje poznať slovo na základe predefinovaného slovníka, ako aj množstvo výskytu poznaných slov.

¹⁵ **Tf-Idf (term frequency-inverse document frequency)** - je miera používaná v oblasti získavania informácií (IR) a strojového učenia, ktorá dokáže kvantifikovať dôležitosť alebo relevantnosť reprezentácií reťazcov (slová, frázy atď.) v dokumente medzi zbierkou dokumentov.

```

In [20]: # cieľ minimalizovať loss function
study = optuna.create_study(direction="minimize")
study.optimize(objective,n_trials=100,timeout=21600) # stop po 6 hodinách
# ponechaj najlepší pokus
best_trial = study.best_trial
best_trial

_learning_rate : -2.6286385629765000}. Best is trial 93 with value: 0.00025088968675947.
[I 2022-01-18 23:35:20,511] Trial 97 finished with value: 0.000986288883723,
log2_batch_size': 7, 'epochs': 61, 'neurons': 95, 'dense': 87, 'dropout': 0.0,
learning_rate': -2.9215102021101136}. Best is trial 93 with value: 0.00025088968675947.
[I 2022-01-18 23:36:18,059] Trial 98 finished with value: 0.007042206358164,
log2_batch_size': 7, 'epochs': 65, 'neurons': 94, 'dense': 8, 'dropout': 0.0,
arning_rate': -2.7251937080052104}. Best is trial 93 with value: 0.00025088968675947.
[I 2022-01-18 23:37:06,606] Trial 99 finished with value: 0.001692597870714,
log2_batch_size': 7, 'epochs': 69, 'neurons': 86, 'dense': 10, 'dropout': 0.0,
learning_rate': -2.571615482972475}. Best is trial 93 with value: 0.00025088968675947.

Out[20]: FrozenTrial(number=93, values=[0.00025088968675947], datetime_start=datetime(2022, 1, 18, 23, 32, 27, 263364), datetime_complete=datetime(2022, 1, 18, 23, 37, 06, 606), distributions={'log2_batch_size': IntUniformDistribution(high=8, low=4, step=1), 'epochs': IntUniformDistribution(high=100, low=2, step=1), 'dense': IntUniformDistribution(high=100, low=6, step=1), 'dropout': UniformDistribution(high=0.5, low=0.0), 'log_learning_rate': UniformDistribution(high=-1.0, low=-4.0)}, user_attrs={}, system_attrs={}, intermediate_values={}, state=State.COMPLETE, value=None)

```

Obr. 36 Optimalizácia hyperparametrov GloVe

Zdroj: Vlastné spracovanie v Python 3.8.0.

V porovnaní s modelmi postavenými iba na historických cenách, predikované hodnoty využívajúce informácie sentimentu trhu predstavujú väčší rozptyl, kedy model predpokladá výrazné nárasty, poklesy ceny, ktoré v skutočnosti nenastali. Daná situácia plynie z faktu, že je model príliš závislý na sentimente, kde musíme brať do úvahy skutočnosť, že ak niekto presadzuje vo svojich príspevkoch výrazne negatívny alebo pozitívny názor na Bitcoin, automaticky to nemusí znamenať, že si daný užívateľ nakoniec Bitcoin aj kúpi, resp. ho predá.

Modely využívajúce sentiment plynú z rozličného dataset-u z hľadiska obdobia, pre ktoré máme dáta k dispozícii. Dataset historických cien pre Bitcoin počas pandémie obsahuje údaje od začiatku roka 2020, pričom údaje o sentimente máme k dispozícii od februára 2021. S cieľom zvýšiť závislosť modelu od cien a tým obmedziť precitlivosť modelu na údaje o sentimente, pre učenie modelu sme tak využívali celý pandemický dataset vývoja cien, v porovnaní so skráteným dataset-om so začiatkom totožným s dataset-om tweet-ov.

Týmto sme obmedzili časť vysokú volatilitu predikovaných hodnôt. Obr. 37 predstavuje výslednú predikciu modelu GloVe v porovnaní s cenami dosiahnutými trhom.



Obr. 37 Výsledné predikcie ceny vs. reálne ceny model GloVe

Zdroj: Vlastné spracovanie v Python 3.8.0.

4.2.2 VADER

VADER je súčasťou balíka NLTK pre Python, ktorý slúži na spracovanie prirodzeného jazyka. Pre aplikáciu algoritmu VADER je rovnako ako v prípade modelu GloVe nutné zaokrúhliť dataset tweet-ov, ako aj predĺžiť čas optimalizácie. Pri modeli VADER máme štyri embeddings (tweet00 až tweet03), čím je dataset pre optimalizáciu hyperparametrov menší ako pri algoritme GloVe, to nám umožňuje skrátenie časového obmedzenia trénovania modelu na 3 hodiny.

Out[11]:

	user_followers	date	text	tweet00	tweet01	tweet02	tweet03
0	8534.0	2021-02-11 00:00:00	Blue Ridge Bank shares halted by NYSE after #b...	0.0	0.855	0.145	0.296
1	6769.0	2021-02-11 00:00:00	🤖 Today, that's this #Thursday, we will do a "...	0.0	0.728	0.272	0.825
2	128.0	2021-02-11 00:00:00	Guys evening, I have read this article about B...	0.0	0.793	0.207	0.572
3	625.0	2021-02-11 00:00:00	\$BTC A big chance in a billion! Price: \487264...	0.0	0.859	0.141	0.316
4	1249.0	2021-02-11 00:00:00	This network is secured by 9 508 nodes as of t...	0.0	0.895	0.105	0.402
...	...	...	...	...	...	...	...
1999529	529.0	2021-12-29 20:00:00	OlegBarca's Stream №547! Live in few mins!\n...	0.0	0.338	0.662	0.323
1999530	1117.0	2021-12-29 20:00:00	Crypto: MistToken\nSymbol: #MistSwap \nMistSwa...	0.106	0.894	0.0	0.416
1999531	26.0	2021-12-29 20:00:00	Just 🤖 for 3x , what about you ? \n\n@AccelDef...	0.0	0.207	0.793	0.7
1999532	67.0	2021-12-29 20:00:00	@Bitboy_Crypto A \$inari Trailer Video has been...	0.0	0.454	0.546	0.25
1999533	11162.0	2021-12-29 20:00:00	CNBC has a crypto special tonight... you know ...	0.0	0.803	0.197	0.402

Obr. 38 Zoradený výstupný dataset tweet-ov podľa VADER

Zdroj: Vlastné spracovanie v Python 3.8.0.

Obr. 38 predstavuje začiatok a koniec zoradeného dataset-u tweet-ov s informáciami o množstve follower-ov, ako aj výsledné hodnoty algoritmu VADER, kde stĺpec tweet00

predstavuje hodnoty pre negatívny sentiment, tweet01 neutrálny sentiment, tweet02 pozitívny sentiment a tweet03 zložené skóre (compound score).

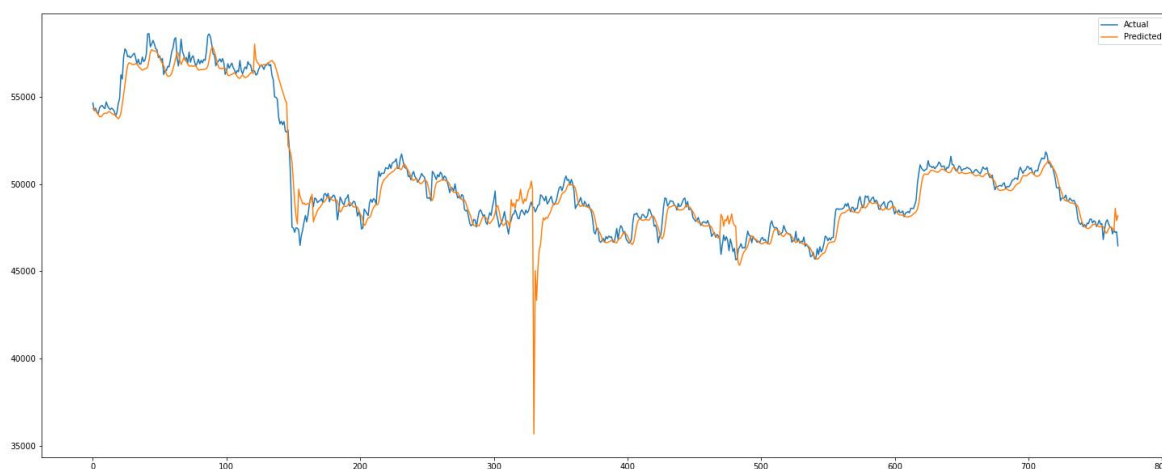
Zložené skóre je vyjadrené ako súčet valenčných hodnôt každého slova v lexike, upraví sa podľa pravidiel algoritmu VADER a následne sa normalizuje tak, aby bolo medzi hodnotami -1 (najviac negatívne) a +1 (najviac pozitívne). Takéto hodnotenie predstavuje jednorozmernú mieru sentimentu.

Pre správne fungovanie modelu VADER je potrebné nastaviť zložené skóre pre negatívny, neutrálny a pozitívny sentiment, čím algoritmus bude vedieť, do ktorej kategórie na základe skóre by mal byť tweet zaradený. Najčastejšie využívanými hodnotami sú:

- pozitívny sentiment: compound score > 0.5 ,
- neutrálny sentiment: $(-0.5 \leq \text{compound score} \leq 0.5)$,
- negatívny sentiment: compound score > -0.5 .

Skóre modelu pre pozitívny, negatívny, neutrálny sentiment predstavuje pomer proporcií textu tweet-u, ktorý patrí do daných kategórií, čím by sa súčet týchto kategórií mal rovnať 1.

Po definovaní finálneho zloženého skóre jednotlivých tweet-ov výpočtom algoritmu VADER je daná hodnota vynásobená počtom follower-ov používateľa, ktorý príspevok pridal, čím kvantifikujeme dosah daného sentimentu na okolie. Výstupom sentimentálnej analýzy je vážená hodnota sentimentu, ktorá rozširuje údaje o historických cenách pre zlepšenie predikcií.



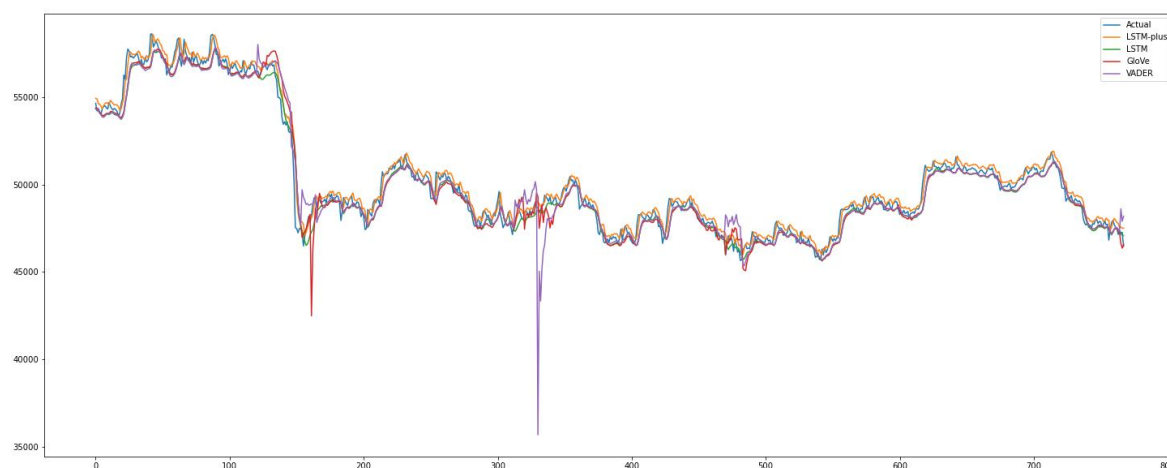
Obr. 39 *Výsledné predikcie ceny vs. reálne ceny model VADER*

Zdroj: Vlastné spracovanie v Python 3.8.0.

Obr. 39 poukazuje na podobné výsledky, ktoré sme nadobudli využitím modelu GloVe, kde väčšinu času model dosahuje dobré predpovede, avšak niekedy očakáva vysokú volatilitu, ktorá však na trhu reálne nenastala. Pre výsledné zhodnotenie výkonnosti modelov využívajúcich údaje o sentimente je potrebné kvantifikovať jednotlivé metriky podobne ako pri porovnaní modelov postavených iba na minulých cenách.

4.2.3 Porovnanie modelov sentimentálnej analýzy s modelmi historických cien

Pri pohľade na vývoj predikcií benchmark-ových modelov bez dát sentimentu v porovnaní s modelmi, ktoré dané dáta obsahovali zobrazených na Obr. 40, vidieť veľmi dobré predikcie pri väčšine predpovedí, kde modely využívajúce sentiment pri predpovediach zaznamenali neočakávané prepady a vrcholy, ktoré sme pri modeloch obsahujúcich iba cenové informácie nezaznamenali.



Obr. 40 Porovnanie modelov sentimentálnej analýzy s modelmi historických cien

Zdroj: Vlastné spracovanie v Python 3.8.0.

Dané skutočnosti potvrdzujú aj hodnoty metrík pre jednotlivé modely nachádzajúce sa v Tab. 7, ktorých výsledky tvrdia, že model využívajúci historické dáta rozšírené o dáta sentimentu analyzované pomocou algoritmu Unsupervised learning-u strojového učenia GloVe prekonal pri všetkých metrikách algoritmus analyzujúci text pomocou lexikónu VADER. Využitie extra dát predstavujúcich práve dáta zo sociálnej siete Twitter za účelom zvýšenia presnosti predikcií modelov však nebolo úspešné, nakoľko modely LSTM a LSTM+ bez extra dát textu zaznamenali podstatne lepšie hodnoty chybovosti.

Táto skutočnosť by tak mohla znamenať, že časť sentimentálnej analýzy a extra dáta príspevkov sťažujú modelom robiť presné predpovede.

Tab. 7 Porovnanie hodnôt metrík modelov sentimentu a historických cien

BTC_after	LSTM	LSTM+	GloVe	VADER
MSE	4.6211723	2.790392	9.6222	10.5152
RMSE	0.775703	0.602771	5.18683	6.15553
MAE	0.6314443	0.403306	4.39841	4.56318
MAPE	0.968 %	0.617 %	3.205 %	3.882 %

Zdroj: Vlastné spracovanie

Dané tvrdenie však nemusí byť pravdivé nakoľko pri bližšom pohľade na presné aktuálne hodnoty (reálne dosiahnuté trhom) uzatváracích cien v testovacom období a ceny predikované modelmi, najmä model GloVe prekonáva všetky ostatné modely presnosťou predpovedaných cien v porovnaní so skutočnými, resp. rozdielom medzi predpovedanými a reálnymi hodnotami. To znamená, že horšie hodnoty modelu GloVe pri jednotlivých metrikách sú zapríčinené práve danými niekoľkými výraznými odchýlkami od skutočných hodnôt, kde daná občasná avšak výrazná volatilita ovplyvnila celkové ohodnotenie modelu.

Out[10]:

		Actual	LSTM-plus	LSTM	GloVe	VADER
date						
28/11/2021	0:00	54660.03	54951.992188	54356.531250	54390.527	54323.188
28/11/2021	1:00	54220.21	54910.964844	54344.578125	54379.508	54330.420
28/11/2021	2:00	54357.95	54618.332031	54216.378906	54243.688	54228.305
28/11/2021	3:00	54137.31	54616.433594	54093.808594	54172.773	54095.062
28/11/2021	4:00	54049.56	54468.980469	53999.019531	54057.812	53993.395
...		...	...	...	...	...
29/12/2021	19:00	47148.61	47785.546875	47390.578125	47415.234	47457.184
29/12/2021	20:00	47369.97	47518.386719	47251.101562	47268.035	47325.164
29/12/2021	21:00	47230.00	47580.496094	47167.886719	46633.426	48618.652
29/12/2021	22:00	47274.61	47510.972656	47121.839844	46363.940	47904.460
29/12/2021	23:00	46464.66	47509.050781	47087.542969	46570.950	48199.440

768 rows × 5 columns

Obr. 41 Začiatok a koniec testovacieho dataset-u reálne a predikované hodnoty modelov

Zdroj: Vlastné spracovanie v Python 3.8.0.

Jestvujúci výsledok najmä modelu GloVe predstavuje vysoký potenciál pre reálne využitie, nakoľko charakteristika modelu a to občasná vysoká volatilita môže byť zakomponovaná a do určitej miery redukovaná finálnym obchodným systémom. Rovnako je pre nás vysoko pozitívne vidieť relatívne dobre fungovať model postavený na viacerých vstupoch (uzatváracia cena a sentiment) v porovnaní s modelom umelej inteligencie s jediným vstupom.

Pri modeloch pokročilých neurónových sietí strojového učenia a konkrétne našich benchmark-och postavených na algoritme LSTM, extrémne nízke výsledne hodnoty odchýlky môžu predstavovať skutočnosť, kde model sa naozaj naučil chápať vrodené vlastnosti vývoja ceny na trhu (ako sa cena tvorí v čase) pre dané aktívum, alebo menej žiaduci výsledok kde model sa na základe minimalizácie odchýlky naučí danú metriku optimalizovať prostredníctvom tzv. pattern spotting (rozpoznanie vývoja, vzorov).

Model s takouto vlastnosťou týmto predstavuje stroj, ktorý vykonáva technickú analýzu, čiže stanovuje svoju predpoveď na základe pochopenia vzorov, ktoré sa naučil na tréningových dátach.

Modely rozšírené o druhý vstup v našom prípade o informáciu o sentimente na trhu v danú hodinu však dávajú do interakcie naučené vlastnosti vývoja sentimentu v porovnaní s cenou, čím tvoria predikciu. Týmto je podstatne obmedzená tvorba modelu iba na základe technickej analýzy.

4.3 Investičné stratégie

Aby sme vedeli finálne odpovedať na otázku, či je možné využitím našich modelov dosahovať zisky, je potrebné ich aplikovať pri tvorbe investičnej stratégie prostredníctvom obchodného systému, ako predikčné modely.

Pre porovnanie rozličných prístupov tvorby investičnej stratégie, obchodných systémov a nimi dosiahnutých výsledkov tvoríme investičné stratégie postavené na prahových hodnotách, škálovacích hodnotách a prostredníctvom preformulovania riešeného problému predikcií ceny pomocou algoritmov Reinforcement Learning-u.

Medzi základné parametre, ktoré platia pre všetky obchodné systémy za účelom dosiahnutia porovnateľnosti výstupných hodnôt výkonnosti patrí štartovacia suma, resp. množstvo dostupného kapitálu, ktorý má obchodný systém k dispozícii počas celého trvania obchodovania vo výške 50 000 USD. Ďalej je to definovanie výšky poplatkov za kúpu/predaj

BTC zmenárňou vo výške 0,1 % (pre obchod). Ďalšími nastaveniami zobrazenými na Obr. 42 je definovanie počiatočného stavu, kde množstvo USD je rovné veľkosti kapitálu a množstvo BTC je rovné 0. Konečný stav predstavuje predanie všetkých BTC a výstupom výkonnosti je rozdiel množstva USD na konci v porovnaní so začiatkom obdobia v percentuálnom vyjadrení. Rovnako má systém nastavené aby pri kúpe/ predaji menil výšku USD a BTC podľa aktuálneho obchodu.

```
In [9]: # parametre pre obchodnú stratégiu
fee = 1e-3 # https://www.binance.com/en/fee/trading
start = 50000 # $50,000
prediction_size = test_Y.shape[0]
actual = df["close"].values[-prediction_size-1:]

In [10]: # simuluj obchodnú stratégiu
def simulate(function, predictions):
    # štart so začiatkovou sumou peňazí a žiadnym Bitcoinom
    money = start
    bitcoin = 0
    # simuluj každú predikciu
    for i in range(len(predictions)):
        # vypočítaj koľko potrebujeme kúpiť
        buy = function(actual[i], predictions[i], money, bitcoin)

        # nedovoľ utratiť viac ako máme peňazí alebo predat viac BTC ako vlastníme
        maximum = money / ((1 + fee) * actual[i])
        minimum = -bitcoin
        buy = max(minimum, buy)
        buy = min(maximum, buy)

        # kúp Bitcoin a utrať peniaze
        bitcoin += buy
        money -= (1 + fee) * actual[i] * buy
    # predaj všetky BTC na konci a vypočítaj celkovú hodnotu
    money += bitcoin * actual[-1]
    return money
```

Obr. 42 Nastavenie obchodnej stratégie modelov sentimentu a historických cien

Zdroj: Vlastné spracovanie v Python 3.8.0.

Pre hodnotenie výkonnosti obchodných systémov a tým jednotlivých modelov je nevyhnutné brať do úvahy spomínaný vývoj na trhu počas testovacieho obdobia. Obr. 41 zobrazuje hodnotu BTC na začiatku testovacieho obdobia vo výške 54 660.03 BTC/USD a na konci s hodnotou 46 464.66 BTC/USD. Daný rozdiel predstavuje 14.99 % pokles ceny a výsledne hodnoty obchodných systémov predstavujú výsledky pre fázu trhu, ktorý je v poklese.

4.3.1 Obchodný systém – prahové hodnoty (thresholds)

Tento obchodný systém je postavený na obchodovaní prostredníctvom predikcií modelu, ktoré sú následne porovnávané s tzv. thresholds, čiže prahovými hodnotami nastavenými obchodným systémom ako parametre. Na základe prahových hodnôt systém vie, či predikovaná hodnota modelom pre nasledujúce obdobie predstavuje pre systém signál na kúpu, predaj alebo držanie pozície bezo zmeny, prostredníctvom vyjadrenia rozdielu medzi predikovanou hodnotou a hodnotou súčasnou.

Nastavenými prahovými hodnotami sú hladiny 0.1 %, 0.5 %, 1 %, 5 %, 10 %, a 50 %. Dané hladiny pre systém znamenajú nasledovné akcie na príklade:

1. ak 1 % threshold a je predikovaný cenový pokles/nárast o 0.5 % = držím pozíciu bezo zmeny,
2. ak 0.5 % threshold a je predikovaný nárast ceny o 0.6 % = nakupuj za 100 % USD,
3. ak 0.1 % threshold a je predikovaný cenový pokles o 0.6 % = predaj na 0 % BTC.

```
In [14]: # testuj pre rôzne prahové hodnoty
threshold_df = pd.DataFrame(index=(0.001,0.005,0.01,0.05,0.1,0.5),columns=models.keys())
for bound in threshold_df.index:
    for model in predictions_df:
        threshold_df.loc[bound,model] = (simulate(threshold(bound * 2 / 100, bound / 100)
threshold_df
```

```
Out[14]:
```

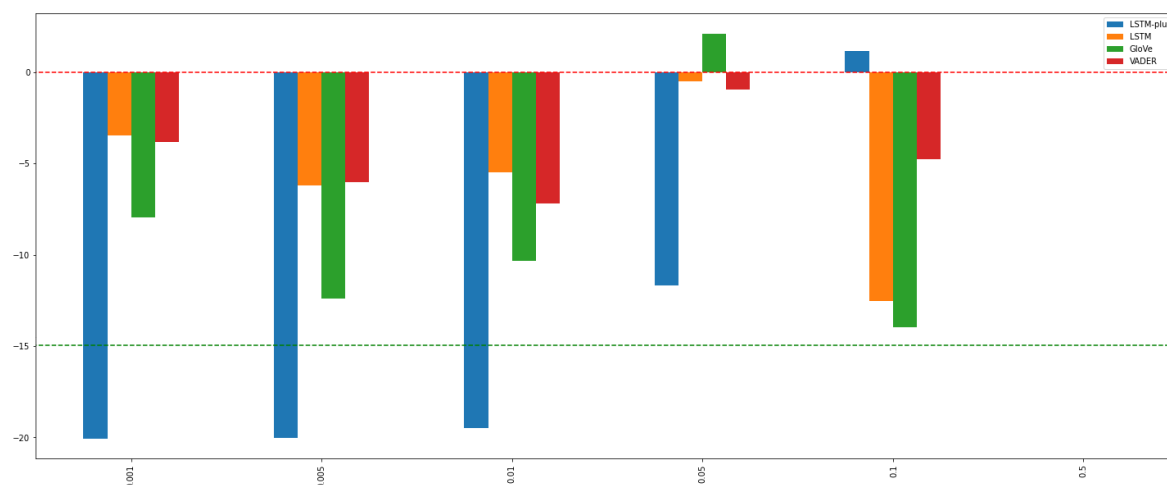
	LSTM-plus	LSTM	GloVe	VADER
0.001	-20.055701	-3.45776	-7.935850	-3.809948
0.005	-20.026182	-6.214239	-12.397502	-6.022255
0.01	-19.484097	-5.505055	-10.334888	-7.187519
0.05	-11.675421	-0.491442	2.110317	-0.968144
0.10	1.126818	-12.549935	-13.969750	-4.787431
0.50	0.0	0.0	0.000000	0.000000

Obr. 43 Obchodný systém prahové hodnoty

Zdroj: Vlastné spracovanie v Python 3.8.0.

Ak už model vie aké predikované hodnoty pre neho znamenajú akú akciu, resp. aký signál, poslednou neznámou je výška investovanej sumy pri jednotlivých možnostiach. Keďže tento systém vychádza z predpokladu správnosti predikovaných hodnôt, model za

účelom maximalizovania zisku a minimalizovania straty pozná rovnako iba tri možnosti, ktorými sú predaj všetko, nákup za všetky prostriedky alebo drž pozíciu.



Obr. 44 Obchodný systém prahové hodnoty výsledky

Zdroj: Vlastné spracovanie v Python 3.8.0.

Obr. 44 predstavuje grafické znázornenie výsledkov obchodného systému pre jednotlivé analyzované modely, ktorých presné percentuálne zhodnotenie/znehodnotenie je možné vidieť na Obr. 43. Grafické zobrazenie rovnako obsahuje v tomto prípade dva benchmark-y, kde prvým je prekonanie trhu, čiže vývoj aktíva BTC za dané obdobie, počas ktorého nastal takmer 15 % pokles. Druhým benchmark-om je prekonanie nulovej hodnoty, čiže dosiahnutie zisku. Výstupom obchodného systému je, že všetky modely s výnimkou LSTM+ pre prahové hodnoty 0.1 %, 0.5 % a 1 % prekonali trh. Zisk, čiže dosiahnutie výslednej sumy vyššej ako 50 000 USD po započítaní všetkých nákladov spojených s obchodovaním v danom období podľa systému dosiahli iba modely LSTM+ pre prahovú hodnotu 10 % vo výške +1.13 % a model GloVe pri prahovej hodnote 5 % výsledných +2.11 %.

Ziskovosť daných modelov pri vyšších prahových hodnotách predstavuje fakt, že modely s nižšími hodnotami pre signál obchodujú častejšie, čím sa ich náklady na obchodovanie rovnako zvyšujú, a pri danom systéme tak nedokázali zabezpečiť zisk.

4.3.2 Obchodný systém – škálovacie hodnoty

Tento obchodný systém je postavený na rozličných investičných škálach. Systém neustále prepočítava aktuálnu veľkosť pozície, ktorú porovnáva s nasledujúcou predikovanou hodnotou.

Jednotlivé škály, na ktorých testujeme modely predstavujú hodnoty 1, 2, 5, 10, 20, 50 a 100. Dané hodnoty predstavujú násobky predikovaného rozdielu medzi aktuálnou cenou a cenou predpovedanou na základe modelu pre nasledujúce obdobie. Inherentnou vlastnosťou tohto systému sú vyššie celkové náklady na poplatky, nakoľko systém neustále nakupuje alebo predáva aktívum. Systém by držal pozíciu iba v prípade, že by model predikoval úplne rovnakú cenu v nasledujúcom období ako je cena súčasná.

Týmto systém tak pracuje nasledovným spôsobom:

1. model predpovedá 0.5 % nárast a hodnota škálovania je 10 = nakupuj BTC za 5 % celkovej investičnej sumy,
2. model predpokladá zápornú percentuálnu zmenu = predaj všetko,
3. hodnotu škálovania je 5 a model predpovedá 1 % nárast ale počas predchádzajúcich obchodov model investoval na súčasnú pozíciu 10 % z celkovej investičnej sumy = signál modelu je vlastniť 5 % v BTC avšak súčasná pozícia je 10 %, čiže predaj 50 % z pozície BTC za USD.

```
In [15]: # testuj pre rozličné škálovacie faktory
kelly_df = pd.DataFrame(index=(1,2,5,10,20,50,100), columns=models.keys())
for scale in kelly_df.index:
    for model in predictions_df:
        kelly_df.loc[scale,model] = (simulate(kelly(scale),np.array(predict
kelly_df
```

```
Out[15]:
```

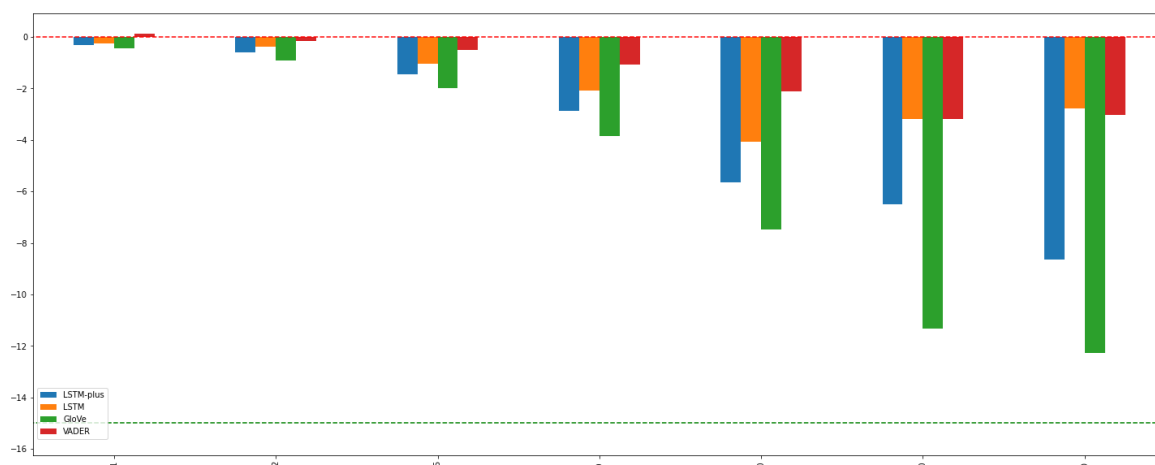
	LSTM-plus	LSTM	GloVe	VADER
1	-0.315467	-0.253474	-0.431066	0.135655
2	-0.587443	-0.388969	-0.897862	-0.147836
5	-1.460042	-1.048125	-1.970933	-0.514550
10	-2.872652	-2.094586	-3.859341	-1.077915
20	-5.64577	-4.062536	-7.478890	-2.109188
50	-6.511422	-3.185623	-11.334483	-3.186975
100	-8.635303	-2.770589	-12.257937	-3.028573

Obr. 45 Obchodný systém škálovacie hodnoty

Zdroj: Vlastné spracovanie v Python 3.8.0.

Výsledný výkon vyjadrený v percentách zobrazený na Obr. 45 a graficky na Obr. 46, opisuje skutočnosť, že všetky modely na všetkých škálach prekonal vývoj trhu. Rovnako však poukazujú na skutočnosť, že všetky modely zaznamenali stratu okrem modelu VADER pre škálu 1 na úrovni +0.136 %. Tento prístup tvorby obchodného systému pre nás

nepredstavuje prístup s potenciálom na ďalšie využitie, nakoľko najnižšie hodnoty straty (zisku) dosiahli systémy s najnižšou škálou. Priama úmera platí pri zvyšovaní škály a zvyšovaní straty, čo znamená, že modely s vyššou sadzbou investovania nedokážu správnym investovaním pokryť zvyšujúce sa náklady na poplatky z obchodu plynúce z objemu jednotlivých obchodov.



Obr. 46 Obchodný systém škálovacie hodnoty výsledky

Zdroj: Vlastné spracovanie v Python 3.8.0.

4.3.3 Reinforcement learning

Pre využitie viacerých možností tvorby obchodného systému s cieľom nájdenia najlepšieho riešenia ďalej využívame možnosť preformulovať nami riešený problém pomocou algoritmov Reinforcement learning-u strojového učenia.

Súčasný prístup zahŕňa využitie umelej inteligencie na predpovedanie budúcej ceny podkladového aktíva, na základe ktorej následne obchodný systém (algoritmus) vyhodnotí predpoveď a vykoná akciu, čím stratová funkcia modelu predpovedí nie je prepojená s finálnou ziskovosťou obchodného systému investičnej stratégie.

Využitím Reinforcement learning algoritmov výstupom obchodného systému bude predstavovať jediné číslo, ktoré povie systému aké množstvo Bitcoin-ov má systém nakúpiť alebo predat'. Metrikou na vyhodnotenie týchto algoritmov je výsledný zisk obchodného systému, čím je ziskovosť priamo prepojená s predikovanými hodnotami modelov.

Pri aplikovaní obchodných systémov budeme využívať modelové architektúry, ktoré sme vyvinuli s podobnými nastaveniami pre porovnanie systémov s a bez sentimentálnej

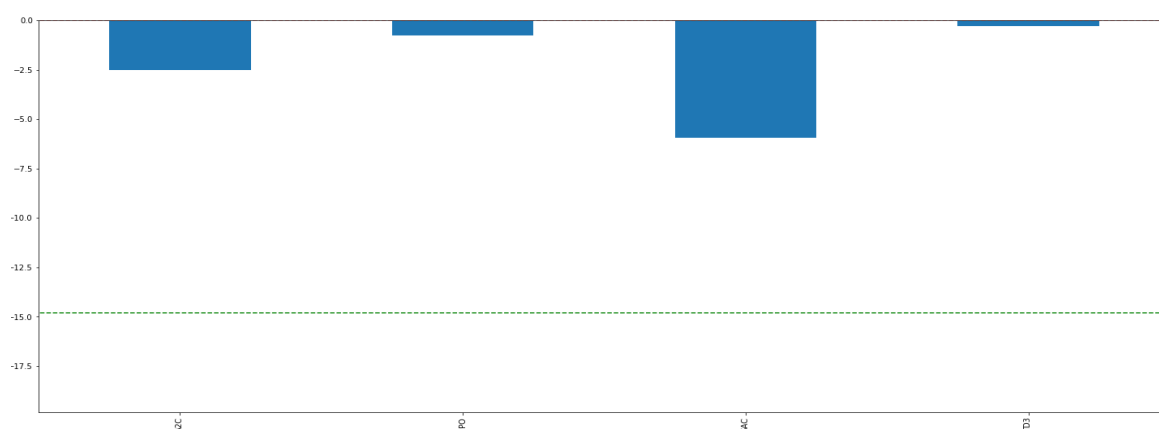
analýzy. Predspracovanie dataset-ov je úplne rovnaké ako pri tréňovaní predošlých modelov pre zabezpečenie konzistentnosti a porovnateľnosti.

Tvorba Reinforcement learning modelov od nuly je pomalá a neefektívna, preto pri tvorbe rozličných modelov využitím viacerých algoritmov daného učenia využívame OpenAI Gym. OpenAI Gym je prostredie pre vývoj a testovanie učiacich sa agentov. Je zamerané a najvhodnejšie práve pre algoritmy Reinforcement learning-u, pričom obsahuje viaceré vopred definované prostredia pre ich skúšku a aplikovanie. Taktiež umožňuje tvorbu simulácií a porovnávaní viacerých prístupov, nastavení (OpenAI, 2022).

Gym tak predstavuje open source knižnicu Python-u a od svojho vydania predstavuje daná knižnica štandard v oblasti vývoja a porovnávania RL algoritmov. Daná knižnica nám umožní využiť Stable Baselines učiacu sa knižnicu postavenú ako súčasť OpenAI Gym. Stable Baselines konkrétne obsahuje súbor spoľahlivých implementácií RL algoritmov, ktoré nám umožnia nastaviť vlastné obchodné prostredie na spustenie modelov postavených na algoritmoch A2C, PPO, SAC, TD3 (gymlibrary, 2022).

Algoritmy A2C, PPO, SAC a TD3 predstavujú tzv. Actor Critic (aktér-kritické) RL algoritmy, ktoré zahŕňajú výhody modelov založených na hodnotách a politike, čo zvyšuje stabilitu tréningu a zefektívňuje výpočtový proces.

Tieto modely nevyžadujú veľa ladenia hyperparametrov, pretože to má obmedzený účinok na učenie a pôvodné hyperparametre dosahujú požadované výsledky, čím ušetríme množstvo času, keďže tréňovanie RL algoritmov je samotne vysoko náročné na čas, výkon, a preto sme pri tréňovaní týchto modelov využívali 20 epoch s výberom najlepšieho modelu a nie modelu posledného.

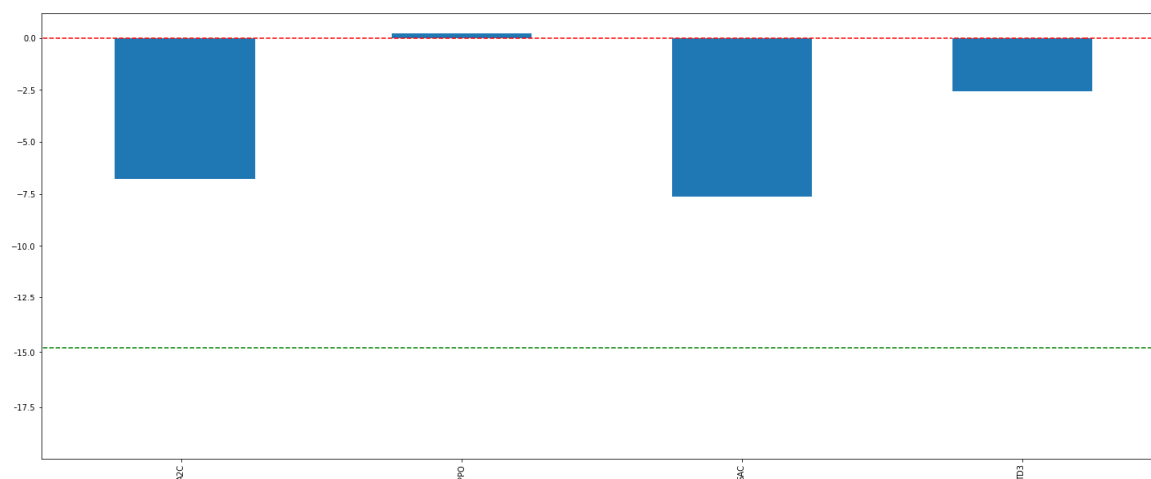


Obr. 47 Obchodný systém RL modely A2C, PPC, SAC, TD3 historické ceny

Zdroj: Vlastné spracovanie v Python 3.8.0.

Algoritmy RL strojového učenia fungujú na princípe odmeňovania agentom, ktorý je daný do prostredia modelu. Nastavením parametrov odmeňovania učíme model akým požadovaným spôsobom sa má správať. Pre zabezpečenie aby bolo pre model výhodnejšie obchodovať ako fungovať bez obchodu, stanovené hodnoty odmeny pre obchody pozitívne ovplyvňujúce bilanciu sú s hodnotou 1, negatívne -1 a odmena ak sa neuskutočnil obchod na úrovni 0. Výstupom modelu je pre každý krok číslo medzi 0 a 1 reprezentujúce aké množstvo peňazí má model investovať z celkového kapitálu, ktorý má k dispozícii.

Obr. 47 zobrazuje finálnu výkonnosť obchodných systémov postavených na algoritmoch RL strojového učenia učiacimi sa výhradne z historického vývoja cien základového aktíva. Kde žiaden z modelov nedosiahol zisk aj keď výrazne prekonal trh. Algoritmus A2C so stratou na úrovni -2.506989 % a SAC -5.936217 % boli výrazne prekonané modelom s algoritmom PPO -0.767201 % a TD3 -0.302845 %.



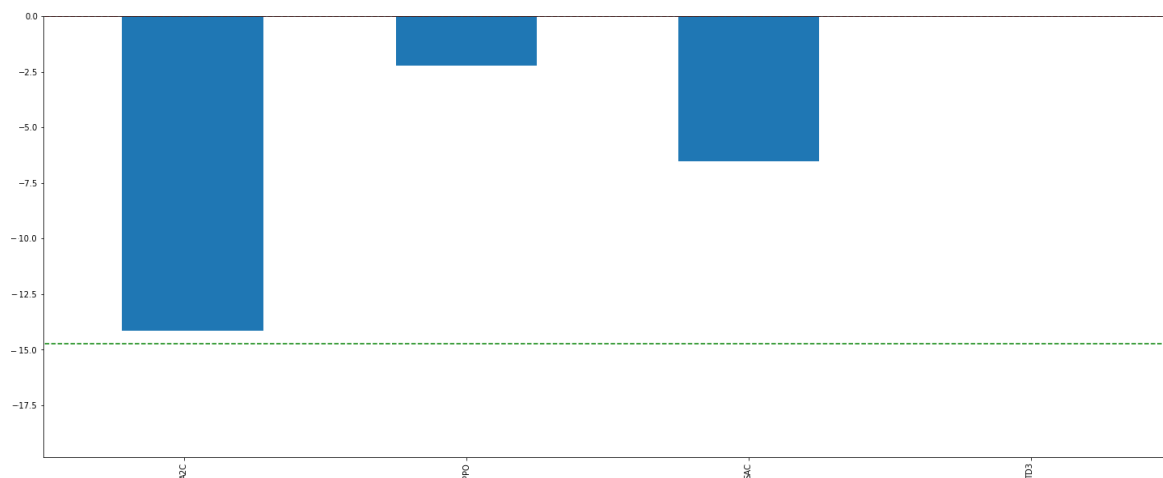
Obr. 48 Obchodný systém RL modely A2C, PPC, SAC, TD3 historické ceny + GloVe

Zdroj: Vlastné spracovanie v Python 3.8.0.

Obr. 48 poukazuje na zmeny vo výkonnosti obchodných systémov, ktorých vstupné hodnoty historického vývoja ceny sú rozšírené analýzou sentimentu prostredníctvom algoritmu GloVe. Kde ako jediný obchodný systém Reinforcement learning-u s využitím algoritmu PPO dosiahol zisk v sledovanom období na úrovni 0.230467 %. Ostatné modely zaznamenali stratu vo výške A2C -6.768233 %, SAC -7.628983 % a TD3 -2.558267 %.

Obchodné systémy využívajúce pre sentimentálnu analýzu algoritmus VADER zobrazené na Obr. 49 zaznamenali záporné hodnoty pre A2C -14.622127 %, PPO -2.212554 %, SAC -6.505936 % a algoritmus TD3 s výstupom bez obchodu na úrovni 0 %.

PPO a TD3 dosiahli spomedzi algoritmov RL strojového učenia pre všetky skúmané vzorky najlepšie výstupné hodnoty, preto dané algoritmy využijeme pre finálne testovanie na nových dátach.

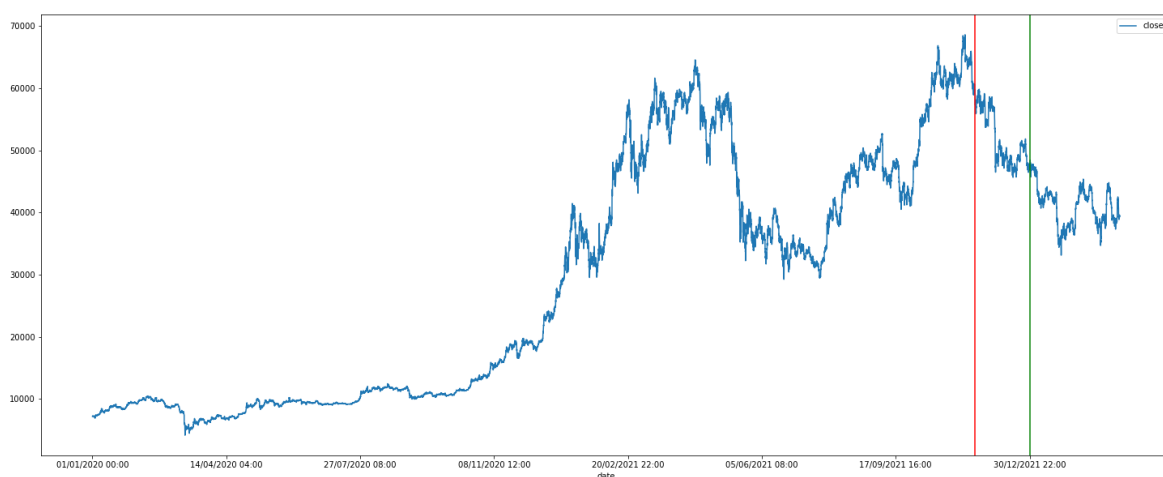


Obr. 49 Obchodný systém RL modely A2C, PPC, SAC, TD3 historické ceny + VADER

Zdroj: Vlastné spracovanie v Python 3.8.0.

4.3.4 Porovnanie najlepších systémov a modelov na nových dátach

Pre konečné zhodnotenie výkonnosti aplikujeme modely, ktoré dosiahli najlepšie výsledky v obchodných systémoch na aktualizované dataset-y vývoja ceny Bitcoin-u ako aj sentimentu prostredníctvom tweet-ov.



Obr. 50 Finálny aktualizovaný dataset historických cien Bitcoin-u 01.01.20 – 11.03.22

Zdroj: Vlastné spracovanie v Python 3.8.0.

Aktualizovaný dataset historických cien predstavuje pre modely rovnakú tréningovú vzorku v porovnaní s predošlými modelmi postavenými na dataset-e pre covidové ceny Bitcoin-u predstavujúcu obdobie od 01.01.2020 0:00 až 27.11.2021 23:00.

Predošlý testovací dataset, ktorý nemal model pre učenie k dispozícii za obdobie od 28.11.2021 0:00 až 31.12.2021 0:00 pri finálnych modeloch predstavuje validačný dataset, ktorý je na Obr. 50 ohraničený červenou a zelenou čiarou. Vytvorenie samostatného a takto dlhého validačného dataset-u nám umožní dosiahnuť lepšie výsledky pre modely citlivé na optimalizáciu hyperparametrov, čím sa však aj zvýši výpočtová náročnosť.

Výsledkom aktualizácie dát sú nové dáta predstavujúce testovací dataset za obdobie od 31.12.2021 1:00 až 11.03.2022 0:00, čiže 1680 unikátnych hodinových uzatváracích cien. Testovací dataset rovnako predstavuje fázu poklesu na trhu Bitcoin-u, nakoľko prvou uzatváracou cenou dataset-u je 47 121.94 BTC/USD a poslednou 39 289.99 BTC/USD, čo predstavuje 16.621% pokles ceny.

Tab. 8 Rozdelenie testovacích dát na 10 týždňových testovacích prostredí

Č.	Testovacie prostredie (okno) pre obdobie	% zmena počas obdobia
0.	31.12.2021 1:00 – 07.01.2022 0:00	-9.197 %
1.	07.01.2022 1:00 – 14.01.2022 0:00	-0.604 %
2.	14.01.2022 1:00 – 21.01.2022 0:00	-3.96 %
3.	21.01.2022 1:00 – 28.01.2022 0:00	-6.553 %
4.	28.01.2022 1:00 – 04.02.2022 0:00	+0,813 %
5.	04.02.2022 1:00 – 11.02.2022 0:00	+15.797 %
6.	11.02.2022 1:00 – 18.02.2022 0:00	-6.036 %
7.	18.02.2022 1:00 – 25.02.2022 0:00	-6.272 %
8.	25.02.2022 1:00 – 04.03.2022 0:00	+7.886 %
9.	04.03.2022 1:00 – 11.03.2022 0:00	-5.349 %

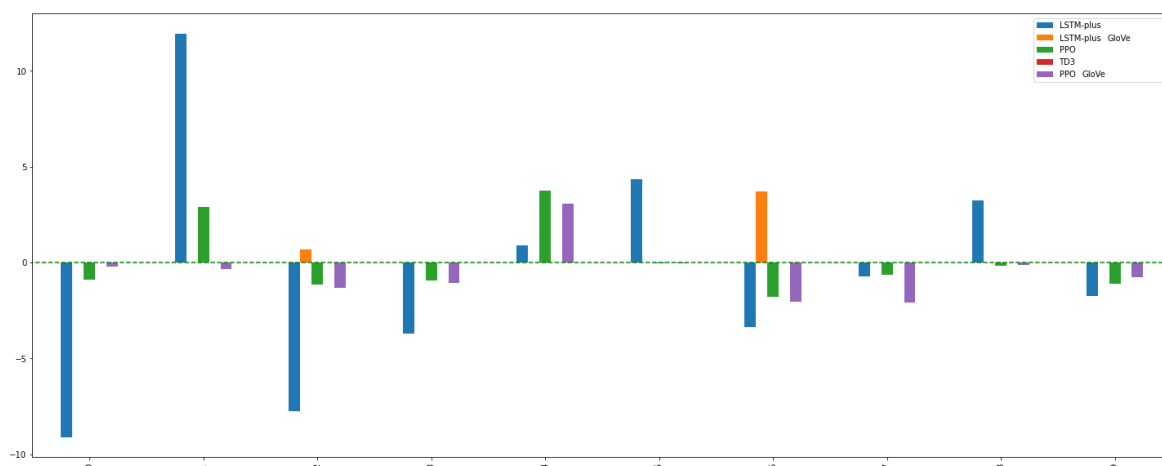
Zdroj: Vlastné spracovanie v Python 3.8.0.

Pri pohľade na Obr. 50 je zrejmé, že vývoj ceny počas testovacieho obdobia zaznamenal výrazné výkyvy a tým aj rozličné mini-fázy trhu. Na základe danej skutočnosti a s cieľom lepšieho porozumenia výkonnosti jednotlivých modelov si testovacie dáta

rozdelíme na 10 týždňových testovacích prostredí (okien) ako sú zobrazené v Tab. 8, ktorých výkonnosť budeme samostatne pozorovať a finálne hodnotenie modelu na základe metriky ziskovosti bude predstavovať súčet výkonnosti za jednotlivé týždne.

Finálne porovnanie bude vykonané pre modely, ktoré dosiahli v predošlej analýze najslubnejšie výsledky, medzi ktoré patria:

1. PPO (historické ceny),
2. TD3 (historické ceny),
3. LSTM+ prahové hodnoty (historické ceny),
4. LSTM+ prahové hodnoty a GloVe (historické ceny + sentimentálna analýza),
5. PPO a GloVe (historické ceny + sentimentálna analýza).



Obr. 51 Grafické vyjadrenie ziskovosti finálnych modelov podľa analyzovaných týždňov

Zdroj: Vlastné spracovanie v Python 3.8.0.

Výsledné hodnoty jednotlivých finálnych investičných stratégií využitím špecifických nastavení obchodného prostredia (systému) postaveného na nami vytrénovaných modeloch, ktoré používajú algoritmy strojového učenia zaznamenali značné rozdiely.

Investičné stratégie postavené na modeloch Reinforcement learning-u využívajú iba historické ceny nedosiahli požadované výsledky. Algoritmus TD3 na základe vývoja aktualizovaných cien dospel k záveru, že najvýhodnejšie bude pre model nevykonať ani jeden obchod, čím síce prekonal trh, avšak nezhodnotil alokované finančné prostriedky.

Model PPO na rozdiel od TD3 fungoval správne a zaznamenal na základe predikcií obchodovanie počas všetkých skúmaných okien. Pri pohľade na vývoj percentuálnych zmien trhu jednotlivých analyzovaných týždňov v porovnaní s vývojom ziskovosti v daných oknách modelom PPO nie je možné pozorovať ich prepojenie, čiže model nezaznamenal priamoúmerné správanie sa ziskovosti a vývoja ceny Bitcoin-u na trhu. Aj keď model zaznamenal v týždni jedna zisk vo výške +2.888942 % a týždni štyri +3.734982 %, ostatné okná predstavovali pre model stratu, čo malo za dôsledok finálnu ziskovosť modelu vo výške -0.105059 %.

Out[16]:

	LSTM-plus	LSTM-plus GloVe	PPO	TD3	PPO GloVe
0	-9.103854	0.000000	-0.893623	0.0	-0.221008
1	11.947396	0.000000	2.888942	0.0	-0.353658
2	-7.745354	0.666445	-1.159233	0.0	-1.333290
3	-3.706739	0.000000	-0.938159	0.0	-1.071733
4	0.890215	0.000000	3.734982	0.0	3.058910
5	4.326025	0.000000	-0.057048	0.0	-0.026631
6	-3.354547	3.712718	-1.772022	0.0	-2.064753
7	-0.726285	0.000000	-0.649774	0.0	-2.082813
8	3.242703	0.000000	-0.151867	0.0	-0.147432
9	-1.732089	0.000000	-1.107258	0.0	-0.779790

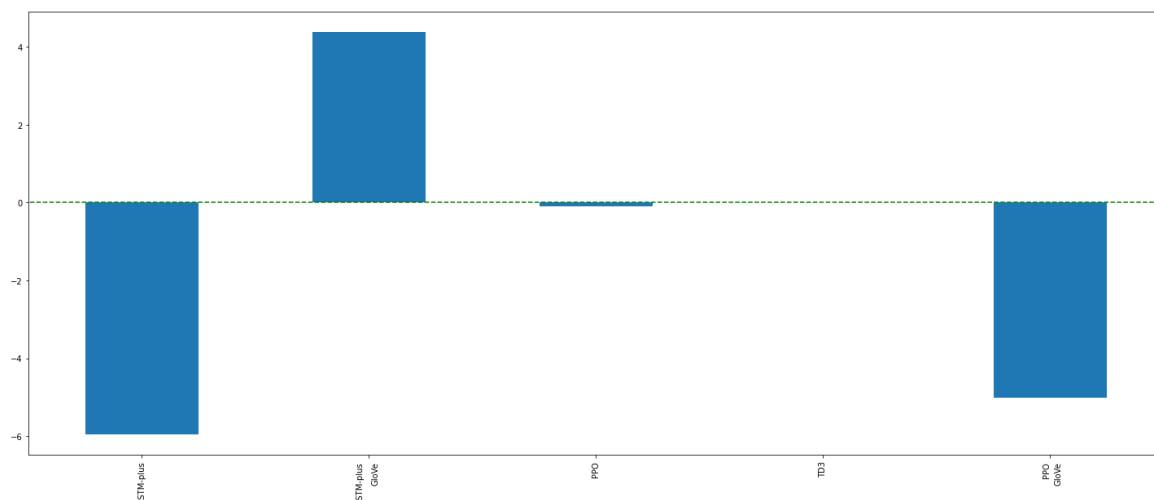
Obr. 52 % vyjadrenie ziskovosti finálnych modelov podľa analyzovaných týždňov

Zdroj: Vlastné spracovanie v Python 3.8.0.

Model PPO rozšírený o algoritmus GloVe analyzujúci sentiment rovnako nezaznamenal prepojenie vývoja predikcií s vývojom ceny na trhu, avšak rozšírenie modelu RL o sentiment malo na zvýšenie ziskovosti opačný efekt, kde daná investičná stratégia zaznamenala v testovacom dataset-e stratu vo výške -5.022198 %, ktorá síce prekonala trh, avšak výrazne zaostávala za modelom PPO.

Zaujímavé hodnoty zaznamenala investičná stratégia využívajúca Supervised learning algoritmus strojového učenia LSTM+, ktorý modeloval predikcie na základe minulého vývoja ceny Bitcoin-u v spojení s obchodným systémom postaveným na prahových hodnotách. Táto stratégia dosiahla najvyššiu volatilitu v ziskovosti v rámci skúmaných období, kde v týždni číslo jedna zaznamenala zisk na úrovni +11.947396 %, či stratu v nultom týždni vo výške -9.103854 %. Výsledkom modelu je najvyššia finálna strata

-5.962529 %. Stratégia síce prekonala trh avšak vývoj ziskovosti modelu bol ovplyvnený vývojom ceny Bitcoin-u na trhu čo znamená, že model nedokázal trénovaním porozumieť vývoju ceny ale naučil sa predikovať hodnoty na základe vývojových vzorov minulého správania sa ceny Bitcoin-u.



Obr. 53 Porovnanie hodnôt ziskovosti modelov sentimentu a historických cien

Zdroj: Vlastné spracovanie v Python 3.8.0.

Poslednou stratégiou a zároveň jedinou, ktorá dosiahla celkový zisk vo výške +4.379163 %, vyjadrený ako agregovanú hodnotu vývoja počas jednotlivých sledovaných období je stratégia využívajúca algoritmus LSTM+ na trénovanie dát minulých cien rozšírená o algoritmus GloVe a analýzu sentimentu na trhu s obchodným systémom postaveným na prahových hodnotách.

Tab. 9 Porovnanie hodnôt ziskovosti modelov sentimentu a historických cien

Bitcoin 31.12.2021 1:00 - 11.03.2022 0:00	% finálna ziskovosť
LSTM+ prahové hodnoty (historické ceny)	-5.962529 %
LSTM+ prahové hodnoty a GloVe (historické ceny + sentimentálna analýza)	+4.379163 %
PPO (historické ceny)	-0.105059 %
TD3 (historické ceny)	0.000000 %
PPO a GloVe (historické ceny + sentimentálna analýza)	-5.022198 %

Zdroj: Vlastné spracovanie

5 Diskusia

Algoritmické obchodovanie ako plne automatizovaný nástroj predstavuje výkonný, avšak rovnako aj veľmi riskantný prostriedok investičnej stratégie investora, pretože algoritmy nedokážu brať do úvahy všetky údaje, ktoré ľudia majú, a nikdy sa nedokážu pripraviť na všetky nepredvídané okolnosti.

COVID-19 bol skvelým príkladom, pretože z trhov boli vymazané obrovské sumy peňazí, ale algoritmy pokračovali ako zvyčajne. Zakódovanie takéhoto druhu správania sa trhov do modelu je takmer nemožné, keďže ani ekonomickí a finanční odborníci sa nevedeli zhodnúť na tom čo nás čaká, nakoľko celosvetová pandémia predstavovala špecifickú situáciu, ktorá v minulosti nenastala, resp. ju nebolo možné porovnať s minulosťou, kvôli inherentným vlastnostiam definujúcim súčasný digitálny a globálny trh.

Jediným spôsobom ako je možné pre nie inštitucionálneho investora postaviť svoju investičnú stratégiu na algoritmickom obchodovaní, je definovať svoje základné požiadavky, ktoré musia vychádzať z obmedzení, ktoré má, nastaviť testovacie prostredie a vyskúšať celý rad obchodných algoritmov, ktoré nakupujú alebo predávajú na základe špecifických kritérií. Až potom sa s porovnaním algoritmov, modelov a jednotlivých obchodných systémov, investičných stratégií môže rozhodnúť, aké veľké riziko algoritmické obchodovanie predstavuje a aký výsledok, potenciál môže očakávať.

Práve takýmto prístupom, využitím rozličných algoritmov strojového učenia, ich rôznym kombinovaním spolu s kombináciou vstupných dát modelov obsahujúcich iba historický vývoj ceny Bitcoin-u alebo aj dáta analýzy sentimentu sú stanovené prístupy tvorby s najlepšimi výslednými hodnotami, resp. najväčším potenciálom pre aplikáciu pri tvorbe obchodného systému, investičnej stratégie postavenej na inovatívnom FinTech prístupe umelej inteligencie.

Výstupom práce je tak investičná stratégia postavená na strojovom učení, ktorá dosiahla najlepšie výsledky počas analyzovaného obdobia, medzi testovanými modelmi, prístupmi. Touto stratégiou je obchodný systém s parametrami pre obchodovanie postavenými na prahových hodnotách, modelom Supervised learning-u LSTM+ využitím na učenie sa prostredníctvom dataset-u historického vývoja ceny a algoritmom Unsupervised learning-u GloVe analyzujúcim dataset sentimentu, ktorý predstavuje rozširujúcu vstupnú zložku pre investičnú stratégiu.

Táto finančne inovatívna investičná stratégia využívajúca umelú inteligenciu pre rozhodovanie o kúpe, predaji alebo držaní pozície zaznamenala za obdobie finálne analyzovaných desiatich týždňov zisk vo výške +4.379163 %, čím prekonal trh o viac ako 21 %.

Je nutné poznamenať, že už pri trénovaní základných modelov, väčšina vytvorených modelov prekonal trh, čiže základný benchmark. Tým, že oba testovacie dataset-y boli ovplyvnené poklesom na trhu a Bitcoin počas nich zaznamenal vysoké straty, ďalším benchmark-om, ktorý je logický je prekonanie hladiny 0 %. Pre investora pri tvorbe obchodného systému pre konkrétne investičné aktívum je automatickým benchmark-om práve vývoj trhu, čo predstavuje hodnotu, ktorú musí obchodný systém svojimi rozhodnutiami po zaplatení všetkých obchodných poplatkov prekonať. Ak daný trh neprekona znamená to, že je pre investora výhodnejšie nakúpiť toto aktívum a držať ho bez akýchkoľvek ďalších akcií. Modely, ktoré dosiahli ziskovosť okolo hladiny 0 % nepovažujeme za dobré, pretože pri 0 % ziskovosti síce priamo investor nezaznamenal stratu, ale po započítaní inflácie, ktorá je najmä v dnešnej situácii enormne vysoká, investor stráca kúpnu silu. Rovnako musíme pri definovaní očakávanej (cielenej) ziskovosti modelu brať okrem inflácie do úvahy aj oportunitné náklady a ďalšie náklady na kapitál.

Výstupom stratégie využívajúcej GloVe LSTM+ model a tým aj analýzy tejto práce je fakt, že rozšírením vstupného súboru údajov o údaje sentimentu plynúce z príspevkov na sociálnej sieti Twitter spracované algoritmom GloVe sme dosiahli výrazné zvýšenie výkonnosti modelu pri využití metriky ziskovosti pri aplikácii na reálne podmienky trhu.

GloVe LSTM+ prekonal model LSTM+ pri *ceteris paribus*, čiže s rovnakými parametrami pre učenie sa na historických dátach ako aj nezmeneným obchodným systémom prahových hodnôt o 10.341692 %. Model GloVe dosahoval výsledky s vysokým potenciálom už pri základnom nastavení na prvotnom dataset-e aj napriek tomu, že bol regresnými metrikami prekonaný modelmi, ktoré neobsahovali dáta sentimentu. Aj keď model dosahoval presnejšie hodnoty predikovaných cien v porovnaní so skutočnými ako porovnávané modely, bol ovplyvnený občasnými vysokými výkyvmi v cene, ktoré reálne nenastali.

Pri tvorbe finálnych modelov, obchodných stratégií sme danú vlastnosť modelov využívajúcich sentiment obmedzili spôsobom, kde model neberie do úvahy krátkodobé zmeny vyššie ako 10 % voči cene za predchádzajúce obdobie pri tvorbe signálov. Tento prístup v reálnom tradeovaní (obchodovaní) by mohol byť reprezentovaný tzv. stop-lossom,

čiže vopred definovanou hodnotou, ktorá ak nastane na trhu, pozícia je automaticky predaná, čím si investor riadi svoju maximálnu možnú stratu. Rovnaký princíp predstavuje príkaz take profit, kedy je definovaná hladina zisku, pri ktorej je pozícia automaticky predaná so ziskom. Využitím daného obmedzenia vyjadreného ako absolútna hodnota, čiže $\pm 10\%$ dosiahol GloVe LSTM+ model svoju maximálnu výkonnosť pre prahovú hodnotu na úrovni 5% .

Výstupom daných poznatkov práce je tak nie len fakt, že sentiment pozitívne pri správnom nastavení ovplyvňuje výkonnosť investičného modelu, ale hovorí nám to aj o skutočnosti, že modely využívajúce výhradne historický vývoj ceny pre učenie a tvorbu predikcii, napriek extrémne dobrým výstupným hodnotám chyby pri regresných metrikách, s využitím nástrojov na zamedzenie over-fitting-u, a práve so zameraním sa pri tvorbe modelov a nastavení hyperparametrov na schopnosť zovšeobecňovania modelov, ďalej pri nízkych dosiahnutých rozdieloch presnosti modelov pre tréningové a testovacie dáta, dané modely nedokázali dosiahnuť zisk pri aplikovaní na reálne podmienky trhu. Táto skutočnosť plynie z faktu, že ani umelá inteligencia a algoritmy strojového učenia, ktoré využívajú ako vstupné dáta iba vývoj ceny neboli schopné v našej aplikácii nájsť informáciu vo vývoji ceny, ktorou by pochopili akým spôsobom je cena na trhu investičných aktív tvorená. Ak je teda vývoj ceny náhodný, dané modely sa naučili na základe dostupného obdobia vzorce správania sa ceny a ich výsledná výkonnosť je priamo ovplyvňovaná skutočným vývojom ceny podkladového aktíva.

Rozšírením modelov o dáta sentimentu jednotlivé modely, algoritmy strojového učenia hľadajú vzťah medzi nájdenými vzorcami vývoja ceny a sentimentom, ktorý bol v danom období na trhu, čo umožnilo kombinovaným modelom dosiahnuť vyššiu mieru zovšeobecnenia. Pri pohľade na vývoj predikovaných hodnôt modelu GloVe LSTM+ v rozmedzí desiatich skúmaných týždňov nie je možné pozorovať žiaden vzťah medzi predikovanými hodnotami, ziskovosťou počas jednotlivých období v porovnaní s trhom.

Prínosom dizertačnej práce sú tak relevantné výsledky pre vedu a prax, ktoré môžeme zhrnúť do nasledujúcich bodov:

1. využitím FinTech nástrojov umelej inteligencie konkrétne algoritmov strojového učenia pre tvorbu inovatívneho prístupu investovania a investičnej stratégie kryptomenového investora sme popísaným postupom dokázali investičnou stratégiou na reálnych údajoch prekonať trh a dosiahnuť zisk,

2. nami vytvorené modely boli pozitívne ovplyvnené rozšírením vstupných údajov predikčného modelu o dáta sentimentu na trhu Bitcoin-u,
3. modely s využitím iba vstupných údajov historického vývoja ceny dosahovali extrémne malú odchýlku predikcii od skutočnosti, avšak nedokázali dosiahnuť zisk pri aplikovaní cez obchodný systém,
4. modely historického vývoja ceny rovnako zaznamenali priamu úmeru medzi výkonnosťou modelu a skutočným vývojom ceny podkladového aktíva,
5. práca odporúča kryptomenovému investorovi opísaný model GloVe LSTM+, ktorý dosiahol najsľubnejšie výsledky.

Definovaním investičnej stratégie s najväčším potenciálom na základe skúmaného problému však úloha investora pri aplikovaní investičnej stratégie nekončí. Je potreba ďalšieho výskumu prostredníctvom aplikácie investičnej stratégie GloVe LSTM+ do finálneho obchodného prostredia, kde sa bude prístup ďalej ladiť podľa špecifik, ktoré nastanú na základe každodenného obchodovania a rozličných situáciách na trhu.

Model GloVe LSTM+ dosiahol zisk, avšak pri 8 týždňoch z celkovo analyzovaných 10, nezaznamenal žiadnu zmenu v ziskovosti. Svoj zisk teda dosiahol konzervatívnou, nízko rizikovou stratégiou. Aké vlastnosti modelu zabránili podieľať sa na rastu hodnoty Bitcoin-u v týždni 5, ktorý bol na úrovni +15.797 %? Akým spôsobom sa bude správať model počas fázy rastu, konsolidácie na trhu, či pri konkrétnych špecifických udalostiach? Aké adresné riešenia budú nutné aby sa zabezpečila maximalizácia ziskovosti pri obmedzení rizika? Nebude ohraničenie na úrovni +-10 % znižovať potenciál na vysoko volatilnom trhu kryptomeny Bitcoin z dlhodobého hľadiska? Na všetky tieto otázky potrebuje investor najst' odpoveď a je to možné iba každodenným fungovaním modelu, neustálou kontrolou výkonnosti, úpravou špecifik, ktoré týmto procesom nastanú.

Pri tomto procese však bude potrebné zabezpečiť live dáta ako ceny Bitcoin-u, tak aj dát sentimentu. Dáta pre vývoj ceny je možné čerpať napríklad z platformy CryptoCompare, kde však najkratší voľne dostupný časový horizont sú denné dáta, pre použitie hodinových uzatváracích cien prostredníctvom prepojenia platformy cez API s investičným modelom, je nutné využiť mesačný platený prístup k danej databáze.

Rovnako je potrebné dolovať dáta o aktuálne zverejnených tweet-och od spoločnosti Twitter prostredníctvom ich developerského účtu, čo však bude predstavovať zvýšenie výpočtovej náročnosti.

Záver

Práca poskytuje metodický návod akým spôsobom tvoriť, analyzovať, validovať modely využívajúce umelú inteligenciu využitím algoritmov strojového učenia pre tvorbu FinTech inovatívnej investičnej stratégie.

Na príklade konkrétneho rozhodovacieho procesu nie inštitucionálneho kryptomenového investora vysvetľuje relevantné prístupy tvorby modelov strojového učenia na základe dát historického vývoja ceny podkladového aktíva Bitcoin, ktoré rozširuje o využitie umelej inteligencie pri definovaní sentimentu trhu analýzou textu.

Poskytuje finálne zhodnotenie výkonnosti, využiteľnosti vytvorených modelov, ktoré predstavujú základ pri tvorbe investičnej stratégie investora v porovnaní s trhom, ako aj navzájom jednotlivých prístupov riešenia skúmaného problému medzi sebou.

Predmetné navrhované stratégie opisujú možné riziká a prístupy k využitiu inovatívnej umelej inteligencie pri investovaní na finančných trhoch, čím tvoria analytickú podporu pre investičné rozhodovanie investora.

Obchodný systém s parametrami pre obchodovanie postavenými na prahových hodnotách, modelom Supervised learning-u LSTM+ využitím na učenie sa prostredníctvom dataset-u historického vývoja ceny a algoritmom Unsupervised learning-u GloVe analyzujúcim dataset sentimentu, ktorý predstavuje rozširujúcu vstupnú zložku pre investičnú stratégiu, zaznamenal najväčší potenciál pre využitie investora obchodujúceho kryptomeny.

Aj keď výsledky práce negarantujú dosiahnutie pozitívnych výsledkov pri aplikovaní na rozličné fázy, situácie trhu, finálna investičná stratégia GloVe LSTM+ dokázala v skúmanom období výrazne prekonať trh a dosiahnuť zhodnotenie investičnej sumy.

Výstupom práce a popísaním krokov tvorby FinTech inovatívneho investičného prístupu investičnej stratégie pre skúmaného investora, je splnenie hlavného ako aj jednotlivých čiastkových cieľov dizertačnej práce.

Zoznam použitej literatúry

1. AHMADI, E. – JASEMI, M. – MONPLAISIR, L. – NABAVI, M. A. – MAHMOODI, A. – JAM, P. A. 2018. *New efficient hybrid candlestick technical analysis model for stock market timing on the basis of the Support Vector Machine and Heuristic Algorithms of Imperialist Competition and Genetic*. In: Expert Systems with Applications, 94 (1), s. 21–31. doi: 10.1016/j.eswa.2017.10.023.
2. AMAZON. 2019. *DeepAR Forecasting Algorithm* [online]. 2019 [cit. 2022-05-18]. Dostupné na: <https://docs.aws.amazon.com/sagemaker/latest/dg/deepar.html>.
3. ANNETT, M. – KONDRACK, G. 2008. *A Comparison of Sentiment Analysis Techniques: Polarizing Movie Blog*. In: Advances in Artificial Intelligence, 5032, s. 25–35. doi: 10.1007/978-3-540-68825-9_3.
4. ARIYO, A. A. – ADEWUMI, A. O. – AYO, C. K. 2014. *Stock Price Prediction Using the ARIMA Model*. In: KSim-AMSS 16th International Conference on Computer Modelling and Simulation, 16 (1), s. 106–112. doi: 10.1109/UKSim.2014.67.
5. ASSIS, C. 2015. *Elon Musk tweet about 'new product line' boosts Tesla shares* [online]. 2015 [cit. 2022-05-18]. Dostupné na internete: < <https://www.marketwatch.com/story/elon-musk-tweet-about-new-product-line-boosts-tesla-shares-2015-03-30>>.
6. BENGIO, Y. 2012. *Practical recommendations for gradient-based training of deep architectures*. In Neural Networks: Tricks of the Trade, 7700, s. 437–478, doi: 10.48550/arXiv.1206.5533.
7. BHARDWAJ, A. – NARAYAN, Y. – DUTTA, M. 2015. *Sentiment analysis for Indian stock market prediction using Sensex and nifty*. In: Procedia Computer Science, 70 (1), s. 85–91. doi: 10.2307/2284112.
8. BITCOIN.ORG. 2021. Bitcoin Core [online]. 2021 [cit. 2021-02-26]. Dostupné na internete: < <https://bitcoin.org/en/download>>.
9. BLOOMBERG. 2021. *Musk tweet featuring broken-heart emoji and Linkin Park meme is latest to hit Bitcoin's price* [online]. 2021 [cit. 2022-05-18]. Dostupné na internete: < <https://fortune.com/2021/06/04/elon-musk-twitter-bitcoin-price-crypto-meme/>>.

10. BOX, B. – GEORGE, E. P – GWILYM, M. J. – REINSEL, G. C – LJUNG, G. M. 2015. *Time Series Analysis: Forecasting and Control*. In: Journal of the American Statistical Association, 342 (68), s. 493–494. doi: 10.1109/BigData.2015.7364089.
11. BRITO, J. – CASTILLO, A. 2013. *BITCOIN A Primer for Policymakers*. In: Mercatus Center George Mason University [online]. 2013 [cit. 2021-02-26]. Dostupné na internete: <
https://www.mercatus.org/system/files/Brito_BitcoinPrimer.pdf>.
12. CASTRO, M. – LISKOV, B. 2002. *Practical Byzantine Fault Tolerance and Proactive Recovery*. In: ACM Transactions on Programming Languages and System [online]. 20 (4), s. 398–461. [cit. 2021-02-26]. Dostupné na internete: <
<http://www.pmg.csail.mit.edu/papers/bft-tocs.pdf>>.
13. CAVALCANTEA, R. C. – BRASILEIROB, R. C. – SOUZAB, V. F. – NOBREGAB, J. P. – OLIVEIR, A. I. 2016. *Computational intelligence and financial markets: a survey and future directions*. In: Expert Systems with Applications, 55 (1), s. 194–211. doi: 10.1016/j.eswa.2016.02.006.
14. CHEN, J. 2010. *SVM application of financial time series forecasting using empirical technical indicators*. In: ICINA information networking and automation, 10 (1), s. 1–77. doi: 10.1109/ICINA.2010.5636430.
15. CHEN, K. – ZHOU, Y. – DAI, F. 2015. *A LSTM-based method for stock returns prediction: A case study of China stock market*. In: International Conference on Big Data (Big Data), s. 2823–2824. doi: 10.1109/BigData.2015.7364089.
16. CHIANG, M. – ZHANG, T. 2016. *Fog and IoT: An Overview of Research Opportunities*. In: IEEE Internet of Things Journal, 3 (6), s. 854–864. doi: 10.1109/JIOT.2016.2584538.
17. COLAH, C. 2015. *Understanding LSTM Networks*. In: Stanford [online] 2015 [cit. 2021-02-26]. Dostupné na internete: <
https://web.stanford.edu/class/cs379c/archive/2018/class_messages_listing/content/Artificial_Neural_Network_Technology_Tutorials/OlahLSTM-NEURAL-NETWORK-TUTORIAL-15.pdf>.
18. CONSOLI, S. – NEGRI, M. – TEBBIFAKHR, A. – TOSETTI, E. – TURCHI, M. 2022. *Comparability of segmented line regression models*. In: International Conference on Machine Learning, Optimization, and Data Science, s. 308–323, doi: 10.1007/978-3-030-95467-3_23.

19. CORTES, C. – VAPNIK, V. 1995. *Support-vector networks*. In: Machine learning, 20 (3), s. 273–297. doi: 10.1007/BF00994018.
20. CRAN. 2021. *Package ‘prophet’* [online]. 2021 [cit. 2022-05-18]. Dostupné na internete: < <https://cran.r-project.org/web/packages/prophet/prophet.pdf>>.
21. CRUZ, E. M. – PYRCZ, M. J. 2021. *Tuning machine learning dropout for subsurface uncertainty model accuracy*. In: Journal of Petroleum Science and Engineering, 205, doi: 10.1016/j.petrol.2021.108975.
22. CRYPTODATADOWNLOAD. 2022. *BINANCE EXCHANGE DATA* [online]. 2022 [cit. 2022-05-18]. Dostupné na internete: <https://www.cryptodatadownload.com/data/binance/>.
23. DELOITTE. 2016. *What is a blockchain?* [online]. 2021 [cit. 2021-02-26]. Dostupné na internete: < <https://www2.deloitte.com/content/dam/Deloitte/uk/Documents/Innovation/deloitte-uk-what-is-blockchain-2016.pdf>>.
24. DING, X. – ZHANG, Y. – LIU, T. – DUAN, J. 2015. *Deep Learning for Event-Driven Stock Prediction*. In: Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence (IJCAI 2015) [online] [cit. 2021-02-26]. Dostupné na internete: < <https://www.ijcai.org/Proceedings/15/Papers/329.pdf>>.
25. DI PERSIO, L. – HONCHAR, O. 2016. *Artificial Neural Networks architectures for stock price prediction: comparisons and applications*. In: International Journal of Mathematics and Computers in Simulation [online] [cit. 2021-02-26]. Dostupné na internete: < <https://www.naun.org/main/NAUN/circuitssystemssignal/2016/b482005-303.pdf>>.
26. DI PERSIO, L. – HONCHAR, O. 2017. *Recurrent neural networks approach to the financial forecast of Google assets*. In: International Journal of Mathematics and Computers in Simulation [online] [cit. 2021-02-26]. Dostupné na internete: < <https://www.naun.org/main/NAUN/mcs/2017/a042002-041.pdf>>.
27. DROBETZ, W. – OTTO, T. 2021. *Empirical asset pricing via machine learning: evidence from the European stock market*. In: Journal of Asset Management volume, 22, s. 507–538. doi: 10.1057/s41260-021-00237-x.
28. EUROPARL.EUROPA. 2018. *Cryptocurrencies and blockchain* [online]. 2017 [cit. 2021-02-26]. Dostupné na internete: < <https://www.europarl.europa.eu/cmsdata/150761/TAX3%20Study%20on%20cryptocurrencies%20and%20blockchain.pdf>>.

29. FACEBOOK. 2017. *PROPHET Forecasting at scale*. [online]. 2017 [cit. 2022-05-18]. Dostupné na: <https://facebook.github.io/prophet/>.
30. FU, T.-CH. – CHUNG, F.-L. – LUK, R. W. P. – NG, C.-M. 2005. *Preventing Meaningless Stock Time Series Pattern Discovery by Changing Perceptually Important Point Detection*. In: International Conference Fuzzy Systems and Knowledge Discovery, s. 27–29. doi: 10.1007/11539506_146.
31. FUJIMOTO, S. – HOOFF, H. – MEGER, D. 2018. *Addressing Function Approximation Error in Actor-Critic Methods*. In: Proceedings of the 35 th International Conference on Machine Learning, Stockholm, Sweden, PMLR 80 [online]. 2018 [cit. 2022-05-18]. Dostupné na internete: < <https://arxiv.org/abs/1802.09477>>.
32. GAO, P. – ZHANG, R. – YANG, X. 2020. *The Application of Stock Index Price Prediction with Neural Network*. In: Mathematical and Computational Applications, 25 (3), 53, doi: 10.3390/mca25030053.
33. GERS, F. A. – SCHMIDHUBER, J. – CUMMINS, F. 2000. *Learning to forget: continual prediction with LSTM*. In: Ninth International Conference on Artificial Neural Networks, 12 (10), s. 850–855. doi: 10.1049/cp:19991218.
34. GEORGOULA, I. – POURNARAKIS, D. – BILANAKOS, C. – SOTIROPOULOS, D. N. – GIAGLIS, G. M. 2015. *Using time-series and sentiment analysis to detect the determinants of bitcoin prices*. In: Mediterranean Conference on Information Systems (MCIS) [online]. 13 s. [cit. 2021-02-26]. Dostupné na internete: < <https://aisel.aisnet.org/cgi/viewcontent.cgi?referer=&httpsredir=1&article=1017&context=mcis2015> >.
35. GOMBER, P. – ZIMMERMANN, K. 2018. *Algorithmic Trading in Practice*. In: The Oxford Handbook of Computational Economics and Finance [online]. 22 s. [cit. 2021-02-26]. Dostupné na internete: < <https://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199844371.001.0001/oxfordhb-9780199844371-e-12?print=pdf> >.
36. GRAVES, A. – LIWICKI, M. – SCHMIDHUBER, J. – FERNÁNDEZ, R. – BERTOLAMI, R. – BUNKE, H. 2009. *A Novel Connectionist System for Unconstrained Handwriting Recognition*. In: IEEE Transactions on Pattern Analysis and Machine Intelligence, 31 (5), s. 855–868. doi: 10.1109/TPAMI.2008.137.

37. GREATLEARNING. 2020. *What is Word Embedding | Word2Vec | GloVe* [online]. 2020 [cit. 2022-05-18]. Dostupné na internete: < <https://www.mygreatlearning.com/blog/word-embedding/>>.
38. GROBYS, K. – JUNTILA, J. – KOLARI, J. W. – SAPKOTA, N. 2021. *On the stability of stablecoins*. In: *Journal of Empirical Finance*, 64 (1), s. 207-223. doi: 10.1016/j.jempfin.2021.09.002.
39. GUGGENHEIM INVESTMENTS. 2019. *Dow Jones Industrial Average Historical Trends* [online]. 2019 [cit. 2022-05-18]. Dostupné na internete: < <https://www.guggenheiminvestments.com/mutual-funds/resources/interactive-tools/dow-jones-historical-trends>>.
40. GYMLIBRARY. 2022. *Environment Creation* [online]. 2022 [cit. 2022-05-18]. Dostupné na: < https://www.gymlibrary.ml/content/environment_creation/>.
41. HAARNOJA, T. – PONG, V. – HARTIKAINEN, K. – ZHOU, A. – DALAL, M. – LEVINE, S. 2018. *Soft Actor Critic—Deep Reinforcement Learning with Real-World Robots* [online]. 2018 [cit. 2022-05-18]. Dostupné na: < <https://bair.berkeley.edu/blog/2018/12/14/sac/>>.
42. HASBROUCK, J. – SAAR, G. 2013. *Low-latency trading*. In: *Journal of Financial Markets*, 16 (4), s. 646–679. doi: 10.1016/j.finmar.2013.05.003.
43. HIRANSHA, M. – GOPALAKRISHNAN, E. A. – MENON, V. K. – SOMAN, K. P. 2018. *NSE Stock Market Prediction Using Deep-Learning Models*. In: *Procedia Computer Science*, 132 (1), s. 1351–1362. doi: 10.1016/j.procs.2018.05.050.
44. HUTTO, C. J. – GILBERT, E. E. 2014. *VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text..* In: *Proceedings of the 8th International Conference on Weblogs and Social Media*, 8 (1), s. 216–225. [cit. 2022-05-18]. Dostupné na: https://www.researchgate.net/publication/275828927_VADER_A_Parsimonious_Rule-based_Model_for_Sentiment_Analysis_of_Social_Media_Text.
45. KABBANI, T. – USTA, F. E. 2022. *Predicting The Stock Trend Using News Sentiment Analysis and Technical Indicators in Spark*. In: *Artificial Intelligence Department of Lusi Computer Science, Istanbul, Turkey*, [online]. 2022 [cit. 2022-05-18]. Dostupné na internete: < <https://arxiv.org/abs/2201.12283>>.
46. KAGGLE. 2022. *Bitcoin Tweets* [online]. 2022 [cit. 2022-05-18]. Dostupné na: <https://www.kaggle.com/datasets/kaushiksuresh147/bitcoin-tweets>.

47. KHAN, Z. F. – KANNAN, A. 2014. *Intelligent Approach for Segmenting CT Lung Images Using Fuzzy Logic with Bitplane*. In: Journal of Electrical Engineering and Technology, 9 (4), s. 1426–1436. doi: 10.5370/JEET.2014.9.4.1426.
48. KILIMCI, Z. H. 2020. *Sentiment Analysis Based Direction Prediction in Bitcoin using Deep Learning Algorithms and Word Embedding Models* In: International Journal of Intelligent Systems and Applications, 8 (2), s. 60–65. doi: 10.18201/ijisae.2020261585.
49. KIM, S. – KIM, H. 2016. *A new metric of absolute percentage error for intermittent demand forecasts*. In: International Journal of Forecasting, 32 (3), s. 669–679. doi: 10.1016/j.ijforecast.2015.12.003.
50. KINGMA, D. P. – BA, J. 2017. *Adam: A Method for Stochastic Optimization*. In: International Conference for Learning Representations, 3, s. 15. doi: 10.48550/arXiv.1412.6980
51. KOBER, J. – BAGNELL, J. A. – PETERS, J. 2013. *Reinforcement learning in robotics: a survey*. In: The International Journal of Robotics Research, 32 (11), s. 1238–1274. doi: 10.1177/0278364913495721.
52. KUHN, M. – JOHNSON, K. *Applied Predictive Modeling*. Switzerland: Springer, 2013. 613 s. ISBN 978-1461468486.
53. LAMPORT, L. – SHOSTAK, R. – PEASE, M. 1982. *The Byzantine Generals Problem*. In: ACM Transactions on Programming Languages and System [online]. 4 (3), s. 382–401. [cit. 2021-02-26]. Dostupné na internete: <<https://lamport.azurewebsites.net/pubs/byz.pdf>>.
54. LEONG, K. – SUNG, A. 2018. *Forecasting stock market index daily direction: a bayesian network approach*. In: Expert Systems with Applications, 105 (1), s. 11–22. doi: 10.18178/ijimt.2018.9.2.791.
55. LI, Y. – ZHENG, Z. – DAI, H - N. 2020. *Enhancing Bitcoin Price Fluctuation Prediction Using Attentive LSTM and Embedding Network* In: Applied Sciences 10(14):4872, s. 1–17. doi: 10.3390/app10144872.
56. MADAN, I. – SALUJA, S. – ZHAO, A. 2015. *Automated Bitcoin Trading via Machine Learning Algorithms*. In: Stanford, [online] [cit. 2021-02-26]. Dostupné na internete: <<http://cs229.stanford.edu/proj2014/Isaac%20Madan,%20Shaurya%20Saluja,%20Aojia%20Zhao,Automated%20Bitcoin%20Trading%20via%20Machine%20Learning%20Algorithms.pdf>>.

57. MAKRIDAKIS, S. – SPILIOTIS, E. – ASSIMAKOPOULOS, V. 2018. *Statistical and Machine Learning forecasting methods: Concerns and ways forward*. In: PLoS ONE, 13 (3), s. 141–156. doi: 10.1371/journal.pone.0194889.
58. MALAGRINO, L. S. – ROMAN, N. T. – MONTEIRO, A. M. 2018. *Did the Profitability of Momentum and Reversal Strategies Decline with Arbitrage Costs After the Turn of the Millennium?* In: The Journal of Portfolio Management, 41 (3), s. 70–83. doi: 10.1016/j.eswa.2018.03.039.
59. MASTELIC, T. – OLEKSIK, A. – CLAUSSEN, H. – BRANDIC, I. 2015. *Cloud Computing: Survey on Energy Efficiency*. In: ACM Computing Surveys, 47 (2):33, doi: 10.1145/2656204.
60. MATTA, M. – LUNESU, I. – MARCHESI, M. 2018. *The Predictor Impact of Web Search Media on Bitcoin Trading Volumes*. In: Proceedings of the 7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management, 1 (2), s. 620–626 doi: 10.5220/0005618606200626.
61. MCGREGOR, J. 2018. *Elon Musk's April Fools' tweets were 'not a joking matter,' experts say* [online]. 2018 [cit. 2022-05-18]. Dostupné na internete: <<https://www.washingtonpost.com/news/on-leadership/wp/2018/04/03/elon-musks-april-fools-tweets-were-not-a-joking-matter-experts-say/>>.
62. MCNALLY, S. – ROCHE, J. – CATON, S. 2018. *Predicting the price of bitcoin using machine learning* In: Proceedings - 26th Euromicro international conference on parallel, distributed, and network-based processing, s. 339–343. doi: 10.1109/PDP2018.2018.00060.
63. MEESAD, P. – RASEL, R. I. 2013. *Predicting stock market price using Support Vector Regression*. In: International Conference on Informatics, Electronics and Vision, s. 1–6. doi: 10.1109/ICIEV.2013.6572570.
64. MENCULINI, L. – MARINI, A. – PROIETTI, M. – GARINEI, A. – BOZZA, A. – MORETTI, C. – MARCONI, M 2021. *Comparing Prophet and Deep Learning to ARIMA in Forecasting Wholesale Food Prices* In: MDPI, 3(3), 644-662. doi: 10.3390/forecast3030040.
65. NAKAMOTO, S. 2008. *Bitcoin: A Peer-to-Peer Electronic Cash System*. In: bitcoin.org [online] [cit. 2021-02-26]. Dostupné na internete: <<https://bitcoin.org/bitcoin.pdf>>.

66. NASSIRTOUSSI, A. K. – AGHABOZORGI, S. – WAH, T. Y. – NGO, D. C. L. 2014. *Text mining for market prediction: A systematic review*. In: Expert Systems with Applications, 41 (16), s. 7653–7670. doi: 10.1016/j.eswa.2014.06.009.
67. NBS. 2021. *Algoritmické obchodovanie* [online]. 2021 [cit. 2021-02-26]. Dostupné na internete: < <https://www.nbs.sk/sk/dohlad-nad-financnym-trhom/fintech/algoritmicke-obchodovanie>>.
68. NBS. 2021. *Alternatívne spôsoby platieb* [online]. 2021 [cit. 2021-02-26]. Dostupné na internete: < <https://www.nbs.sk/sk/dohlad-nad-financnym-trhom/fintech/alternativne-sposoby-platieb>>.
69. NBS. 2021. *Automatizované poradenstvo (robo-advice)* [online]. 2021 [cit. 2021-02-26]. Dostupné na internete: < <https://www.nbs.sk/sk/dohlad-nad-financnym-trhom/fintech/automatizovane-poradenstvo-robo-advice>>.
70. NBS. 2021. *Biometria* [online]. 2021 [cit. 2021-02-26]. Dostupné na internete: < <https://www.nbs.sk/sk/dohlad-nad-financnym-trhom/fintech/biometria>>.
71. NBS. 2021. *Crowdfunding* [online]. 2021 [cit. 2021-02-26]. Dostupné na internete: < <https://www.nbs.sk/sk/dohlad-nad-financnym-trhom/fintech/crowdfunding>>.
72. NBS. 2021. *FinTech* [online]. 2021 [cit. 2021-02-26]. Dostupné na internete: < <https://www.nbs.sk/sk/dohlad-nad-financnym-trhom/fintech>>.
73. NBS. 2021. *InsurTech poistenie* [online]. 2021 [cit. 2021-02-26]. Dostupné na internete: < <https://www.nbs.sk/sk/dohlad-nad-financnym-trhom/fintech/insurtech-poistenie>>.
74. NBS. 2021. *Kryptoaktíva a Initial Coin Offerings (ICOs)* [online]. 2021 [cit. 2021-02-26]. Dostupné na internete: < <https://www.nbs.sk/sk/dohlad-nad-financnym-trhom/fintech/kryptoaktiva-a-initial-coin-offerings-icos>>.
75. NBS. 2021. *Mobilná peňaženka* [online]. 2021 [cit. 2021-02-26]. Dostupné na internete: < <https://www.nbs.sk/sk/dohlad-nad-financnym-trhom/fintech/mobilna-penazenka>>.
76. NBS. 2021. *Umelá inteligencia a "big data"* [online]. 2021 [cit. 2021-02-26]. Dostupné na internete: < <https://www.nbs.sk/sk/dohlad-nad-financnym-trhom/fintech/umela-inteligencia-a-big-data>>.
77. NITHA, K. P. – SIVAKUMARI, S. 2021. *Sentimental Analysis for Sensex Stocks using Vader Algorithm- An approach for Investment recommendations in Indian stock market*. In: Turkish Online Journal of Qualitative Inquiry (TOJQI), 9 (2), s.1700-1707.

78. OPENAI. 2017. *OpenAI Baselines: ACKTR & A2C* [online]. 2017 [cit. 2022-05-18]. Dostupné na: <<https://openai.com/blog/baselines-acktr-a2c/>>.
79. OPENAI. 2018. *Soft Actor-Critic* [online]. 2018 [cit. 2022-05-18]. Dostupné na: <<https://spinningup.openai.com/en/latest/algorithms/sac.html>>.
80. OPENAI. 2018. *Twin Delayed DDPG* [online]. 2018 [cit. 2022-05-18]. Dostupné na: <<https://spinningup.openai.com/en/latest/algorithms/td3.html>>.
81. OPENAI. 2022. *OpenAI About* [online]. 2022 [cit. 2022-05-18]. Dostupné na: <<https://openai.com/about/>>.
82. OPTUNA. 2022. *optuna.samplers.TPESampler* [online]. 2022 [cit. 2022-05-18]. Dostupné na internete: <<https://optuna.readthedocs.io/en/stable/reference/generated/optuna.samplers.TPESampler.html>>.
83. PAI, P. -F. – HSU, M. -F. – WANG, M. -C. 2011. *A support vector machine-based model for detecting top management fraud*. In: Knowledge-Based Systems, 24 (2), s. 314–321. doi: 10.1016/j.knosys.2010.10.003.
84. PANO, T. – KASHEF, R. 2020. *A complete vader-based sentiment analysis of bitcoin (BTC) tweets during the ERA of COVID-19*. In: Big Data and Cognitive Computing, 4(4):33. doi: 10.3390/bdcc4040033.
85. PATEL, J. – SHAH, S. – THAKKAR, P. – KOTTECHA, K. 2015. *Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques*. In: The Journal of Wealth Management, 18 (4), s. 75–89. doi: 10.1016/j.eswa.2014.07.040.
86. PEDRYCZ, W. – CHEN, S. M. *Sentiment Analysis and Ontology Engineering: An Environment of Computational Intelligence*. Switzerland: Springer, 2016. 456 s. ISBN 978-3-319-30317-8.
87. PIRBHULAL, S. – POMBO, N. – FELIZARDO, V. – GARCIA, N. – SODHRO, A. H. – MUKHOPADHYAY, S. C. 2019. *Towards Machine Learning Enabled Security Framework for IoT-based Healthcare*. In: International Conference on Sensing Technology (ICST), 13 (1), s. 1–6. doi: 10.1109/ICST46873.2019.9047745.
88. PONGDATU, G. – PUTRA, Y. H. 2018. *Seasonal Time Series Forecasting using SARIMA and Holt Winter's Exponential Smoothing*. In: IOP Conference Series Materials Science and Engineering, 407 (1), 6 s. doi: 10.1088/1757-899X/407/1/012153.

89. PREM SANKAR, C. – VIDYARAJ, R. – SATHEESH KUMAR, K. 2015. *Trust based stock recommendation system— a social network analysis approach*. In: Elsevier Masson SAS, 46 (1), s. 299–305. doi: 10.1016/j.procs.2015.02.024.
90. PWC. 2021. *This is InsurTech's moment. Will insurers seize the opportunity?* [online]. 2021 [cit. 2021-02-26]. Dostupné na internete: < <https://www.pwc.com/us/en/industries/financial-services/library/insurtech-innovation.html>>.
91. PYTHON-TWITTER. 2020. *A Python wrapper around the Twitter API* [online]. 2020 [cit. 2022-05-18]. Dostupné na internete: < <https://python-twitter.readthedocs.io/en/latest/>>.
92. RAUCHS, M. – GLIDDEN, A. – GORDON, B. – PIETERS, G. C. 2018. *Distributed Ledger Technology Systems: A Conceptual Framework*. In: SSRN Electronic Journal. 110 s. doi: 10.2139/ssrn.3230013.
93. ROONDIWALA, M. – PATEL, H. – VARMA, S. 2017. *Predicting Stock Prices Using LSTM*. In: International Journal of Science and Research, 6 (4), s. 1754–1756. doi: 10.21275/ART20172755.
94. RUTA, D. 2014. *Automated trading with machine learning on big data*. In: BigData Congress, 3 (1), s. 259–268. doi: 10.1109/BigData.Congress.2014.143.
95. SAINI, G. – SHOBANA, M. 2022. *CRYPTOCURRENCY PRICE PREDICTION USING PROPHET AND ARIMA TIME SERIES*. In: International Research Journal of Modernization in Engineering Technology and Science, [online]. s. 1–8. [cit. 2022-05-18]. Dostupné na internete: <https://www.irjmets.com/uploadedfiles/paper/issue_4_april_2022/21042/final/fin_irjmets1650284252.pdf>.
96. SALINAS, D. – FLUNKERT, V. – GASTHAUS, J. – JANUSCHOWSKI, T. 2020. *DeepAR: Probabilistic forecasting with autoregressive recurrent networks*. In: International Journal of Forecasting, 36 (3), s. 1181–1191. doi: 10.1016/j.ijforecast.2019.07.001.
97. SCHUEFFEL, P. 2016. *Taming the Beast: A Scientific Definition of Fintech*. In: SSRN Electronic Journal, 4 (4), s. 32–54. doi: 10.2139/ssrn.3097312.
98. SCHULMAN, J. – WOLSKI, F. – DHARIWAL, P. – RADFORD, A. – KLIMOV, O. 2017. *Proximal Policy Optimization Algorithms*. In: OpenAI Gym [online]. 2017 [cit. 2022-05-18]. Dostupné na internete: < <https://arxiv.org/pdf/1707.06347.pdf>>.

99. SCIKIT-LEARN. 2022. *sklearn.preprocessing.PowerTransformer* [online]. 2022 [cit. 2022-05-18]. Dostupné na: <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.PowerTransformer.html>.
100. SHAH, D. – CAMPBELL, W. – ZULKERNINE, F. 2018. *A Comparative Study of LSTM and DNN for Stock Market Forecasting*. In: International Conference on Big Data (Big Data), s. 4148–4155. doi: 10.1109/BigData.2018.8622462.
101. SHAH, D. – ISAH, H. – ZULKERNINE, F. 2019. *Stock Market Analysis: A Review and Taxonomy of Prediction Techniques*. In: International Journal of Financial Studies, 7 (2), s. 26–48. doi: 10.3390/ijfs7020026.
102. SHIN, K. -S. – LEE, T. S. – KIM, H. -J. 2005. *An application of support vector machines in bankruptcy prediction model*. In: Expert Systems with Applications, 28 (1), s. 127–135. doi: 10.1016/j.eswa.2004.08.009.
103. STANFORD. 2014. *GloVe: Global Vectors for Word Representation* [online]. 2014 [cit. 2022-05-18]. Dostupné na internete: < <https://nlp.stanford.edu/projects/glove/>>.
104. STATISTA. 2022. *Bitcoin (BTC) price per day from October 2013 to July 8, 2022* [online]. 2021 [cit. 2021-07-08]. Dostupné na internete: < <https://www.statista.com/statistics/326707/bitcoin-price-index/>>.
105. SZABO, N. 1996. *Smart Contracts: Building Blocks for Digital Markets*. [online]. [cit. 2021-02-26]. Dostupné na internete: < https://www.fon.hum.uva.nl/rob/Courses/InformationInSpeech/CDROM/Literature/LOTwinterschool2006/szabo.best.vwh.net/smart_contracts_2.html>.
106. ŠTERBA, J. – HILOVSKA, K. 2018. *The implementation of hybrid ARIMA neural network prediction model for aggregate water consumption prediction. Aplimat- Journal of Applied Mathematics*. In: Aplimat – Journal of Applied Mathematics, [online]. 3 (3), s. 123–130. [cit. 2021-02-26]. Dostupné na internete: < https://www.aplimat.com/files/Journal_volume_3/Number_3.pdf>.
107. TAYLOR, A. 2018. *In a Nutshell - Bitcoin and Cryptocurrencies*. In: Scott & Wilkinson [online]. 2018 [cit. 2021-02-26]. Dostupné na internete: < <https://www.scott-wilkinson.com/blog/bitcoin-and-cryptocurrencies/>>.
108. THEBUSINESSRESEARCHCOMPANY. 2020. *Fintech Global Market Opportunities And Strategies* [online]. 2020 [cit. 2021-02-26]. Dostupné na internete: < <https://www.thebusinessresearchcompany.com/report/fintech-market>>.

109. VARSHNEY, M. – RATURI, S. – VERMA, J. 2020. *A Survey on Cloud Computing and Security Challenges to Cloud Computing*. In: International Journal of Computer Applications, 175 (29), s. 29–33. doi: 10.5120/ijca2020920830.
110. WANG, W. – SU, C. – DUXBURY, D. 2022. *The conditional impact of investor sentiment in global stock markets: A two-channel examination*. In: Journal of Banking & Finance, 138. doi: 10.1016/j.jbankfin.2022.106458.
111. WANG, B – ZHANG, X. 2021. *Deep Learning Applying on Stock Trading* [online]. 2021 [cit. 2022-05-18]. Dostupné na internete: < https://cs230.stanford.edu/projects_spring_2021/reports/74.pdf>.
112. WANG, Y. 2014. *Stock price direction prediction by directly using prices data: an empirical study on the KOSPI and HSI*. In: International Journal of Business Intelligence and Data Mining, 9 (2), s. 145–160. doi: 10.1504/IJBIDM.2014.065091.
113. WORLDBANKGROUP. 2017. *Distributed Ledger Technology (DLT) and Blockchain* [online]. 2017 [cit. 2021-02-26]. Dostupné na internete: < <https://documents1.worldbank.org/curated/en/177911513714062215/pdf/122140-WP-PUBLIC-Distributed-Ledger-Technology-and-Blockchain-Fintech-Notes.pdf>>.
114. XIE, Q. 2011. *The optimization of share price prediction model based on support vector machine*. In: Automation and Systems Engineering, s. 1–4. doi: 10.1109/ICCSE.2011.5997714.
115. XIONG, R. – NICHOLS, E. P. – SHEN, Y. 2016. *Deep Learning Stock Volatility with Google Domestic Trends*. In: Cornell University, [online]. 2016 [cit. 2021-02-26]. Dostupné na internete: < <https://arxiv.org/pdf/1512.04916.pdf> >.
116. YANG, B. 2020. *Deep Reinforcement Learning for Automated Stock Trading* [online]. 2020 [cit. 2022-05-18]. Dostupné na internete: < <https://towardsdatascience.com/deep-reinforcement-learning-for-automated-stock-trading-f1dad0126a02>>.
117. YANG, H. – LIU, X. Y. – ZHONG, S. – WALID, A. 2020. *Deep Reinforcement Learning for Automated Stock Trading: An Ensemble Strategy* In: SSRN, s.9. doi: 10.2139/ssrn.3690996.
118. ZHANG, G. P. – QI, M. 2005. *Neural network forecasting for seasonal and trend time series*. In: European Journal of Operational Research, 160 (2), s. 501–514. doi: 10.1016/j.ejor.2003.08.037.

119. ZHANG, J. – CUI, S. – XU, Y. – LI, Q. – LI, T. 2017. *A novel data-driven stock price trend prediction system*. In: *Expert Systems with Applications*, 97 (1), s. 60–69. doi: 10.1016/j.eswa.2017.12.026.
120. ZHENG, A. – CASARI, A. 2018 *Feature Engineering for Machine Learning: Principles and Techniques for Data Scientists*. O'Reilly Media; 1st edition, 2018. 218 s. ISBN 978-1491953242.
121. ZHENG, S. 2011. *Gradient descent algorithms for quantile regression with smooth approximation*. In: *International Journal of Machine Learning and Cybernetics*, 2, doi: 10.1007/s13042-011-0031-2.
122. ZHENG, Z. – XIE, S. – DAI, H-N. – CHEN, W. – CHEN, X. – WENG, J. – IMRAN, M. 2020. *An Overview on Smart Contracts: Challenges, Advances and Platforms*. In: *Future Generation Computer Systems*, 105 (1), s. 475-491. doi: 10.1016/j.future.2019.12.019%202019.