

České vysoké učení technické v Praze
Fakulta strojní

Ústav řídicí a přístrojové techniky



Bakalářská práce

Datová analýza podobnosti vývoje cen populárních kryptoměn

Vypracovala: Kateřina Hynková
Vedoucí práce: Ing. Matouš Cejnek Ph.D.
Rok: 2021

Prohlášení

Prohlašuji, že jsem svou bakalářskou práci vypracovala samostatně a použila jsem pouze podklady (literaturu, projekty, SW atd.) uvedené v příloženém seznamu.

V Praze dne

.....
Kateřina Hynková

Poděkování

Chtěla bych vyjádřit poděkování svému vedoucímu Ing. Matoušovi Cejnkovi Ph.D., za jeho rady, ochotu dohlížet na tuto práci a čas, který této práci věnoval.

Abstrak

Analýza souvislosti mezi jednotlivými obchodovatelnými komoditami je klíčová pro jejich predikci. Kryptoměny jsou obchodovatelné komodity postavené na technologii blockchain, které v posledních letech velmi nabývají na popularitě. Datová analýza je v posledních letech velmi využívaný obor, hlavně díky velkému množství nově vznikajících informací všude okolo nás. V tomto článku je úkolem představit analýzu čtyřech vybraných kryptoměn pomocí různých druhů zkoumání podobnosti. Konkrétně byly použity tyto čtyři metody: Korelace, Křížová korelace, Fourierova analýza a Empirický režim rozkladu.

Klíčová slova

Kryptoměny, Korelace, Frekvenční spektrum, Fourierova analýza, Autokorelace, Křížová korelace, Empirický režim rozkladu, Datová analýza v Pythonu.

Abstract

Analysis of the relationship between other tradable commodities is key to their prediction. Cryptocurrencies are tradable commodities based on blockchain technology, which have become very popular in recent years. Data analysis is a widely used field in recent years, mainly due to the large amount of emerging information all around us. In this paper, the service is represented by four selected cryptocurrencies using different types of similarity studies. Specifically, these four methods are: Correlation, Cross-correlation, Fourier analysis and Empirical decomposition mode.

Keywords

Cryptocurrencies, Correlation, Frequency spectrum, Fourier analysis, Autocorrelation, Cross-correlation, Empirical mode decomposition, Data analysis with Python

Obsah

Seznam obrázků	viii
Úvod	1
1 Doposud provedené výzkumy	3
2 Trhy podobné kryptoměnám	5
2.1 Forex	5
2.2 Akcie	5
2.3 NFT	6
2.4 Predikční trhy	7
3 Kryptoměny	9
3.1 Úvod do kryptoměn	9
3.2 Blockchain	9
3.3 Peer-to-peer	10
3.4 Populární kryptoměny	10
3.4.1 Bitcoin	12
3.4.2 Ethereum	13
3.4.3 Litecoin	13
3.4.4 Binance Coin	14
3.4.5 XRP	14
3.4.6 Dogecoin	15
4 Datová analýza	17
4.1 Datová věda	17
4.2 Vytěžování dat	18
4.2.1 Rozhodovací strom	18
4.2.2 Neuronové sítě	19
4.2.3 Regresní analýza	21
5 Metodologie	23
5.1 Korelace	23
5.1.1 Pearsonův korelační koeficient	24
5.1.2 Spearmanův korelační koeficient	24
5.1.3 Granularita dat	24
5.2 Posunutá korelace	25
5.3 Fourierova analýza	25
5.3.1 Rychlá Fourierova transformace	26
5.4 Hilbert-Hungova transformace	26
5.4.1 Empirický režim rozkladu	26
5.5 Python a jeho uživatelské rozhraní	28
5.5.1 NumPy	28

5.5.2	SciPy	28
5.5.3	Matplotlib	29
5.5.4	IPython	29
5.5.5	Pandas	29
6	Praktická část	31
6.1	Motivace	31
6.2	Získání dat z virtuální směnárny	31
6.3	Korelace	33
6.4	Posunutá korelace v čase	34
6.5	Fourierova analýza	36
6.6	Empirický režim rozkladu	37
6.7	Získané výsledky	38
6.7.1	Posunutá korelace	38
6.7.2	Korelace	38
6.7.3	Fourierova analýza	39
6.7.4	Empirický režim rozkladu	41
	Závěr	43
	Bibliografie	45
	Přílohy	51
A	První příloha	51
B	Druhá příloha	55
B.1	Posunutá korelace - hodinová granularita	55
B.2	Posunutá korelace - denní granularita	59
C	Třetí příloha	63
D	Čtvrtá příloha	66

Seznam obrázků

2.1	Historicky první tweet	7
3.1	Celková tržní kapitalizace všech kryptoměn	11
3.2	Celkové zastoupení jednotlivých kryptoměn v procentech	11
3.3	Celková tržní kapitalizace všech kryptoměn	16
5.1	Korelační koeficient	23
6.1	Nastavení automatického stahování dat z on-line směnárny a nastavení daných granularit	32
6.2	Stáhnutí konkrétního souboru a určení, jakou granularitu požadují	32
6.3	Stáhnutí více csv souborů najednou, pro práci s korelacemi	32
6.4	Kód pro vytvoření korelační tabulky a jednotlivých korelačních koeficientů	33
6.5	Kód pro vytvoření posunuté korelace - první část	34
6.6	Kód pro vytvoření posunuté korelace - druhá část	35
6.7	Kód pro vytvoření FFT	36
6.8	Kód pro vytvoření EMD signálů - část první	37
6.9	Kód pro vytvoření EMD signálů - část druhá	37
6.10	Kód pro vytvoření EMD signálů - část třetí	38
6.11	Rychlá fourierova transformace vybraných kryptoměn, hodinová granularita	40
6.12	Rychlá fourierova transformace vybraných kryptoměn, denní granularita	40
13	Závislost XRP na dolaru-hodinové záznamy	51
14	Závislost XRP na dolaru-denní záznamy	52
15	Závislost BNB na dolaru-hodinové záznamy	52
16	Závislost BNB na dolaru-denní záznamy	52
17	Závislost ETH na dolaru-hodinové záznamy	53
18	Závislost ETH na dolaru-denní záznamy	53
19	Závislost BTC na dolaru-hodinové záznamy	53
20	Závislost BTC na dolaru-denní záznamy	54
21	Autokorelace kryptoměny BNB na BNB, hodinová granularita	55
22	Autokorelace kryptoměny BTC na BTC, hodinová granularita	55
23	Autokorelace kryptoměny ETH na ETH, hodinová granularita	56
24	Autokorelace kryptoměny XRP na XRP, hodinová granularita	56
25	Společná posunutá korelace BTC a BNB, hodinová granularita	56
26	Společná posunutá korelace ETH a BNB, hodinová granularita	57
27	Společná posunutá korelace ETH a BTC, hodinová granularita	57
28	Společná posunutá korelace XRP a BNB, hodinová granularita	57
29	Společná posunutá korelace XRP a BTC, hodinová granularita	58

30	Společná posunutá korelace XRP a ETH, hodinová granularita	58
31	Autokorelace kryptoměny BNB na BNB, denní granularita	59
32	Autokorelace kryptoměny BTC na BTC, denní granularita	59
33	Autokorelace kryptoměny ETH na ETH, denní granularita	60
34	Autokorelace kryptoměny XRP na XRP, denní granularita	60
35	Společná posunutá korelace BTC a BNB, denní granularita	60
36	Společná posunutá korelace ETH a BNB, denní granularita	61
37	Společná posunutá korelace ETH a BTC, denní granularita	61
38	Společná posunutá korelace XRP a BNB, denní granularita	61
39	Společná posunutá korelace XRP a BTC, denní granularita	62
40	Společná posunutá korelace XRP a ETH, denní granularita	62
41	Rychlá fourierova transformace XRP s hodinovými záznamy	63
42	Rychlá fourierova transformace XRP se záznamy jednotlivých dní . .	63
43	Rychlá fourierova transformace BNB se záznamy jednotlivých hodin .	64
44	Rychlá fourierova transformace BNB se záznamy jednotlivých dní . .	64
45	Rychlá fourierova transformace ETH se záznamy jednotlivých hodin .	64
46	Rychlá fourierova transformace ETH se záznamy jednotlivých dní . .	65
47	Rychlá fourierova transformace BTC se záznamy jednotlivých hodin .	65
48	Rychlá fourierova transformace BTC se záznamy jednotlivých dní . .	65
49	Empirický režim rozkladu kryptoměny ETH- záznam jednotlivých hodin	66
50	Empirický režim rozkladu kryptoměny ETH- záznam jednotlivých dní	67
51	Empirický režim rozkladu kryptoměny XRP- záznam jednotlivých ho- din	68
52	Empirický režim rozkladu kryptoměny XRP- záznam jednotlivých dní	69
53	Empirický režim rozkladu kryptoměny BNB- záznam jednotlivých ho- din	70
54	Empirický režim rozkladu kryptoměny BNB- záznam jednotlivých dní	71
55	Empirický režim rozkladu kryptoměny BTC- záznam jednotlivých hodin	72
56	Empirický režim rozkladu kryptoměny BTC- záznam jednotlivých dní	73

Úvod

Znalost dat a jejich shromažďování podstatně ovlivnilo spoustu odvětví, například sociologii, lékařství, informatiku či ekonomii. Pomocí zkoumání dat vypovídajících o předešlých událostech se analytici snaží zjistit budoucí vývoj v daném odvětví. Obor pro dosahování těchto úkolů se nazývá prediktivní analytika. Pod tento termín spadají algoritmy z různorodých odvětví, a to od statistiky přes umělou inteligenci či strojové učení až po znalosti z matematiky. Určení přesného modelu pomocí dat může zapříčinit velké zisky v dané oblasti, a hlavně díky tomu je tento výzkum v současnosti populární.

V této práci se bude používat Explorační analýza dat (zkráceně EDA, z anglického názvu Exploratory data analysis), což je v oblasti statistiky metoda využívající se pro hledání hypotéz, které by byly vhodné na otestování, jejich hlubší průzkum pak testuje konfirmační analýza (Confirmatory data analysis). EDA je z pravidla první krok, který se s daty provádí a který do dalších analýz přináší velké množství poznatků. V neposlední řadě pak data připravená k dalšímu zpracování. Tyto body o EDA ve své knize Explorační analýza dat popsal John Tukey:

- Navrhnout hypotézy o příčinách pozorovaných jevů.
- Ověřit předpoklady statistických metod, které se použijí.
- Podložit výběr vhodných statistických nástrojů a technik

V této práci bude použit k datové analýze programovací jazyk Python, který je jedním z nejpopulárnějších jazyků pro analýzu dat ve vědách. Jako nedílnou součástí se využívá Jupiter notebook, kde probíhá spojování dat a analytických kódů. Následně se zde vytváří živé zprávy, které je možno dále upravovat a analyzovat. Jedním z hlavních důvodů, proč je Python v tomto odvětví tak populární, je široká škála existujících knihoven. Tyto knihovny obsahují více než 130 000 balíčků, které obsahují nástroje pro manipulaci a automatizaci databází, vědecké výpočty a další funkce.

Kryptoměny se staly globálním fenoménem hlavně díky velkému rozvoji v oblasti elektronického obchodování, blockchainových technologií a finančního průmyslu. Kryptoměny od svého vzniku zaznamenávají výraznou fluktuaci svých hodnot, dosud není znám vzorec poklesu a růstu jejich ceny, přestože bylo objeveno mnoho systémů predikce pohybu trhu s kryptoměnou, díky kterým by investorům usnadňovalo rozhodování o nákupu a prodeji. K predikci vývoje cen kryptoměn na základě znalosti údajů z historie byly použity tradiční algoritmy učení pod dohledem. Dále byl testován vliv rozpoznání dopadu zpráv na pohyb cen, což může investorovi poskytnout výhodu nákupu a prodeje. Další varianta pro predikci cen kryptoměny

by mohla být metoda s využitím novinek a historických cenových údajů. V tomto případě se analyzuje schopnost zpravodajských dat předpovídat cenové výkyvy u zvolených druhů kryptoměn.

Díky průlomům v oblasti technologií, umělé inteligence, strojového učení a dalších oblastí se velmi rozšiřuje využití v oblasti ekonomie. Díky analýze dat lze předpokládat vývoj akcií a dalších prostředků a získává tak uznání mezi mnoha finančními institucemi. Díky tomu se tato analýza stala z kvalitativního nástrojem kvantitativním.

Kapitola 1

Doposud provedené výzkumy

Výzkumy ukazují, že předpovídání ceny kryptoměn lze provádět pomocí hypotézy efektivního trhu (Efficient Market Hypothesis), dále se používá pro danou hypotézu zkratka EMH. Teorie uvádí všechny dostupné informace o trhu, díky tomu se nazývá efektivní [1]. Akcie a kryptoměny se na trhu obchodují za jejich reálnou hodnotu, což znemožňuje překonat trh nebo jeho načasování. Nové informace o akcích se rychle začleňují do tržní ceny, proto investoři nakupují riskantní investice [2].

Ačkoli je EMH v dnešní době velmi používaná finanční teorie, je často zpochybňovaná a velmi kontroverzní. Lidé, kteří zpochybňují danou hypotézu tvrdí, že je zbytečné hledat podhodnocené akcie a snažit se předpovídat trendy trhu pomocí fundamentální či technické analýzy [3]. To hlavně kvůli neschopnosti dané analýzy upravené o riziko (alfa), tudíž vnitřní informace trhu mohou vyústit v nadměrné výnosy upravené podle daného rizika [3].

Funkčnost EMH má velké množství důkazů, ale jsou zde i výjimky, například investoři jako je Warren Buffett dlouhodobě poráží trh, což není z definice EMH možné [4]. Dále kritici často ukazují na krach akciového trhu v roce 1987, při kterém Dow Jones Industrial Average (DJIA) klesl za jeden den o více než 20%, což se uvádí jako důkaz toho, že ceny akcií se mohou výrazně odchylovat od svých reálných hodnot [5].

Využití strojového učení ve financích bylo studováno již několik desetiletí, ale výzkum zaměřený jen k predikci ceny kryptoměny je velmi omezený. To se v poslední době mění a na Bitcoin se vztahuje velká pozornost. Hlavně v poslední době přitahuje značnou pozornost v oblasti ekonomiky, datové analýzy a počítačové vědy díky svým vlastnostem, jako jsou šifrovací technologie peněžních jednotek [6].

Ve své studii s názvem Correlation dynamics in the cryptocurrency market based on dimensionality reduction analysis popisuje Chun-Xiao Nie svůj výzkum ohledně korelace na trhu kryptoměn. V této práci se popisuje časově závislá sekvence korelační matice, která je následně transformována do matice vzdálenosti. Byla zde sestavena Frobeniova matice vzdálenosti korelačních koeficientů na trhu kryptoměn. Následně proběhla transformace korelační struktury do vektoru nastaveného v euklidovském prostoru. Tato matice byla uvažována jako vysokodimenzionální vektor. Následná analýza ukázala údaje o korelačním koeficientu. Matice by podle této studie mohla být vhodně zabudována do podprostoru, kde je rozměr podprostoru menší než počet prvků matice, to znamená menší než počet druhů kryptoměn. Dimenze

matice byla následně pomocí vybraného algoritmu zmenšena, tím byla zjištěna podobnost některých korelačních matic a daného koeficientu. Na základě srovnání s indexem CCI30 bylo zjištěno, že existuje vztah mezi složkami v síti kNN a stavem trhu. Oblast mezi jednotlivými složkami pak odpovídá významným změnám v průměru [7].

V knize *Collective behavior of cryptocurrency price changes* je taktéž pojednáváno o kryptoměnách a jejich vývoji. Analyzuje se zde vzájemná korelace mezi cenovými změnami různých kryptoměn pomocí metod teorie náhodných matic a minimálních rozpětí stromů. Bylo zjištěno, že matice křížové korelace nevykazuje přítomnost ve vzájemných korelacích. Ve srovnání s jinými finančními trhy, bylo zjištěno, že většina vlastních čísel ve spektru korelační matice nesouhlasí s předpovědí náhodných matic, několik vlastních čísel se ale odchyluje podle očekávání. Překlenovací strom křížových korelací kryptoměn odhaluje odlišné struktury, které se jeví jako stabilní. Kolektivní chování přítomné na trhu kryptoměn pak může být užitečné pro konstrukci portfolia kryptoměn i pro budoucí výzkum v této oblasti [8].

V publikaci s názvem *Extreme correlation in cryptocurrency market* byla studována strukturu závislosti párového srovnání deseti největších kryptoměn roku 2018, jmenovitě Bitcoin, Dash, Dogecoin, Ethereum, Litecoin, Monero, Namecoin, Novacoin, Peercoin a Ripple. Byl odhadnut maximální korelační koeficient pomocí datové analýzy. Nálezy odhalují jasné vzorce výrazně vysoké bivariační závislosti u některých z nejzákladnějších a nejrozšířenějších kryptoměn. To znamená, že extrémní korelace nesouvisí s volatilitou trhu s kryptoměnami jako takovými, ale s trendem na trhu [9].

Kapitola 2

Trhy podobné kryptoměnám

2.1 Forex

Forex (FX), celým jménem devizový trh, je globální mimoburzovní či decentralizovaný trh zaměřený na obchody s měnami a měnovými deriváty. FX udává kurzy pro každou světovou měnu, zahrnuje do ní všechny známé aspekty nákupu, prodeje a směny za aktuálně stanovené ceny. V odvětví obchodu je to podle jeho objemů největší trh na světě. Nemá žádné centralizované umístění, tudíž je to spíše elektronická síť bank, makléřů a jednotlivých obchodníků.

V případě obchodování na tomto trhu jsou měny uvedeny v párech, to znamená, že se dané měny porovnávají. Na devizovém trhu se měny obchodují s loty zvanými mikro, mini a standardní loty. Mikro lot má v dané měně hodnotu 1000, mini lot nabývá 10 000 jednotek a běžný lot je 100 000. Při obchodování na elektronickém forexovém trhu se obchody odehrávají v nastavených měnových blocích a může se obchodovat s jakoukoli velikostí. Forexový trh je otevřen 24 hodin denně, pět dní v týdnu napříč hlavními finančními centry po celém světě. To znamená, že měny se dají nakupovat nebo prodávat kdykoli během týdne [10].

2.2 Akcie

Akcie jsou cenné papíry, které představují část vlastnictví dané společnosti. Vlastník akcií má právo k podílu aktiv společnosti, a díky tomu k ziskům rovnajícím se akciím, které vlastní (kapitálu). Akcie se nakupují a prodávají převážně na burzách a jsou součástí portfolia individuálních investorů. Korporace prodávají akcie za účelem získání financí pro provoz svých firem.

Vlastnictví je určeno počtem akcií, které osoba vlastní, v poměru k počtu akcií v oběhu. Majitel akcií má právo hlasovat na valných hromadách, dostávat dividendy, což jsou zisky společnosti a právo prodat nakoupené akcie někomu jinému [11].

Akcie jsou vydávány společností za účelem navyšování kapitálu. Rozdíl mezi dluhopisem a akcií je hlavně v tom, že majitel dluhopisu má nárok na úroky a splácení jistin. Akcionáři jsou naproti tomu poslední v řadě a v případě bankrotu často nedostávají nic nebo jen setiny na dolar. To znamená, že akcie jsou ze své podstaty riskantnější investice než dluhopisy [12].

2.3 NFT

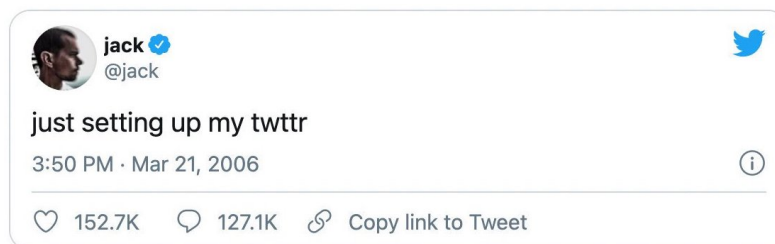
NFT je zkratka pro anglické slovní spojení „non-fungible token“, v češtině se používá výraz nezastupitelný token. Je to jedinečná, nezaměnitelná položka, což je hlavní rozdíl mezi NFT a kryptoměnami. Například bitcoin lze vyměnit za dolar či jiný bitcoin. NFT si lze představit jako unikátní exemplář, který byl vydán pouze jeden. Tento druh majetku vznikl v roce 2017 v závislosti se zvyšující se poptávkou po kryptoměnách. Jako první NFTs se označují takzvané CryptoKitties (virtuální kočky), které byly možné zakoupit za ethereum a poté se o ně starat na způsob podobný Tamagoči. Rekordní prodej zaznamenala CryptoKittie s názvem Dragon, která byla prodána v přepočtu na dolary za 170 tisíc amerických dolarů [13].

Pod pojem NFTs lze označit prakticky cokoli s nějakou digitální podobou, například hudba, text, či obraz. Důležitým prvkem je potvrzení o jejich vlastnictví a unikátnosti, které je zapsané na platformě blockchain, nejčastěji se v těchto finančních trzích využívá Ethera. Velkou výhodou je anonymní zápis, který zabraňuje plagiátorství a krádežím, na druhou stranu, pokud majitel ztratí svojí digitální peněženku, ke svým NFTs už nikdy nebude mít přístup, jako tomu je i u kryptoměn. Jako příklad může posloužit sportovní sběratelská karta s konkrétním sportovcem, která lze vyměnit za kartu jiného sportovce, ale hodnota jednotlivých karet se může s časem nezávisle měnit.

Nákup těchto tokenů probíhá na online tržištích, jedno z aktuálně nejvyužívanějších je OpenSea, kde je možno nakupovat digitální umělecká díla či gify. Nákupy probíhají jak za fixně stanovenou cenu od prodejce, tak formou aukce za nejvyšší příhoz. Většina těchto obchodů se prodává pomocí měny ethereum [13].

NFT je v současnosti spojováno především s virtuálním sběratelstvím umění. Nejčastěji otázka ohledně tohoto trendu spočívá v zajištění autorských práv těchto děl. Majitel totiž není schopen zabránit stažení obrázku jiným uživatelem, který si ho může uložit, v tomto principu NFT ale nespočívá. Důležitý faktor je jediný oficiální vlastník zapsaný v blockchainové databázi, kde si může tuto informaci kdokoli ověřit. Majitel díky tomu se svým dílem nakládá podle uvážení, například může dané dílo prodat a transakci zapsat do blockchainu, tím se vlastnické právo převede na někoho jiného [14].

Na konci března byl prodán jako NFT první tweet, který publikoval spoluzakladatel a šéf Twitteru Jack Dorsey, vydražil se za více než 2,9 milionu dolarů, kde 5% z celkové částky jde na poplatky obchodující platformě. Jedná se o zprávu ve znění „just setting up my twttr“ z roku 2006, která je označována jako první příspěvek na Twitteru, kterou Dorsey nabídl v prosinci 2020 na aukční platformě v.cent.co k prodeji za kryptoměnu ethereum s garancí zápisu na blockchain, díky čemuž si veřejnost bude moci ověřit vlastnictví osoby, která si tweet zakoupila [14].



Obrázek 2.1: Historicky první tweet

2.4 Predikční trhy

Predikční trhy, někdy nazývané taktéž sázkové trhy, jsou trhy, které se obchodují na burze a byly vyvinuty za účelem obchodování s výsledky nadcházejících se událostí. Podle tržní ceny lze určovat, jak se bude daná událost podle davu vyvíjet. Smlouva o predikčním trhu se obchoduje mezi 0% a 100%. Tento systém je navržen tak, aby vybíral informace o tématech, které jsou v dané době populární. Obchodníci obchodují se smlouvami a výplaty jsou závislé na budoucím výsledku. Schopnost predikčního trhu shlukovat informace a dělat přesné předpovědi je založena na hypotéze efektivního trhu, která uvádí, že ceny aktiv odrážejí dostupné informace [15].

Tyto online tržiště mají oproti jiným typům predikcí výhodu v získávání finančních pobídek, díky tomu se rychle začlení nové informace a následně s nimi nelze manipulovat. Jednou z hlavních nevýhod je neznalost davu dané problematiky. Díky tomuto nedostatku mohou být odpovědi výrazně zkreslené. Druhý mechanismus, jak obchodovat s tímto trhem je myšlenka, že budou obchodníci hledat místa, kde se dav mýlí [16].

PredictIt je jeden z nejpopulárnějších online predikčních trhů, původně pochází z Nového Zélandu. Jeho funkce spočívá v záměně informací o politických a finančních událostech. Tato společnost byla na trh uvedena v roce 2014. Tento trh využívá zdvojenou aukci za účelem prodeje akcií pro aktuální událost, kterou má na svém trhu. Znamená to, že uživatel, který předpovídá, že se daná událost v budoucnu stane, musí proti sobě mít osobu, která předpokládá pravý opak. Tyto předpovědi jsou pak shlukovány na trhu, náklady na správu jsou účtovány poplatkem činící 10% ze zisku převyšujícího původní investici a dalších 5% poplatku je účtováno za výběr částky. PredictIt nabízí program sdílení dat pro členy akademické komunity. PredictIt má více než 160 datových partnerů, včetně výzkumných pracovníků přidružených k Massachusettskému technologickému institutu, Harvardské univerzitě, Univerzitě Yale, Dukově univerzitě, Pennsylvánské univerzitě, Kalifornské univerzitě, Berkeley, Michiganské univerzitě a Univerzitě ve Virginii. [16].

Kapitola 3

Kryptoměny

3.1 Úvod do kryptoměn

Kryptoměny jsou virtuální či digitální systémy, umožňující spolehlivé online platby. Je to systém zabezpečený pomocí kryptografie, což znemožňuje jeho padělání. Většina kryptoměn je založena na technologii blockchain, což je distribuovaná účetní kniha vynucená různorodou sítí počítačů. Oproti klasickým měnám nejsou obvykle kryptoměny ovlivňovány různými vládními zásahy či manipulací.

Krypto je značení pro různé šifrovací algoritmy a techniky, které chrání hlavní knihy uvnitř systému, například šifrování eliptické křivky, páry veřejného a soukromého klíče a hashovací funkce [17].

První vynalezenou kryptoměnou založenou na blockchainu byl bitcoin, v současné době je stále nejpoblárnější a má největší hodnotu. Existuje tisíce dalších, alternativních kryptoměn s různými specifikacemi a funkcemi. Některé z nich jsou klony bitcoinu, zatímco jiné jsou zcela nově vzniklé systémy, které byly postaveny od nuly, na nových základech [18].

Mezi další kryptoměny na stejném základu jako je bitcoin patří například litecoin, peercoin či namecoin. Celková tržní hodnota všech existujících kryptoměn nabývá hodnoty 214 miliard dolarů, kde bitcoinu zaujímají největší podíl a to 68% z celkové hodnoty [19].

3.2 Blockchain

Blockchain je soubor elektronicky uložených informací v počítačovém systému. Jeho struktura je většinou v podobě tabulky nebo databáze. Menší skupiny nebo jednotlivci nejčastěji používají tabulky k ukládání menšího množství informací [20]. Oproti tomu databáze je navržena pro podstatně větší množství informací, které je možno rychle a systematicky vyhledávat více uživateli najednou.

Pomocí blockchainu jsou informace shromažďovány společně ve skupinách, takzvaných blocích, které obsahují sady informací. Bloky mají dané úložné kapacity a po jejich zaplnění jsou zacykleny do dříve vyplněného bloku a tím vytváří řetězec dat známých jako blockchain. Systém tak tedy vytváří nevratnou časovou osu dat,

za podmínky, že je vkládán do bloků v decentralizované podobě. V případě, že je blok zcela zaplněn, stane se součástí této časové osy a dostane své časové razítko [21].

Blockchainový operační systém využívá blockchain jako podpůrný systém, který běží na pozadí počítačového systému. Tento operační systém zachycuje všechny příkazy a transakce ze zařízení uživatele. Následně na blockchainu dochází k jejich ověřování a následnému zaznamenávání.

3.3 Peer-to-peer

Peer-to-peer (P2P) někdy používáno také sousloví klient-klient, je označení pro decentralizovaný typ počítačových sítí, kde spolu komunikují dva uživatelé přímo mezi sebou, bez jakéhokoli zprostředkovatele. Architektura P2P nezná pojem server a všechny uzly nacházející se v síti si jsou rovnocenné. Tato síť tedy slouží k vzájemnému vyměňování dat.

P2P ekonomika je alternativou ke kapitalistické ekonomice, kdy software s otevřeným zdrojovým kódem funguje společně s běžným komerčním softwarem. Příkladem typu klient-klient může být například Uber či Airbnb, které slouží jako alternativy k taxislužbám a podnikání v hotelnictví. Tyto společnosti poskytují zprostředkovatelské služby, včetně sítě propojující kupující a prodejce a zpracovávají platby. Důležitou součástí je využívání výhradně soukromého dodavatele k poskytnutí služeb přímo zákazníkům. U P2P tedy není do transakce zapojována třetí strana, díky tomu se zde vyskytuje větší riziko, že produkt nebude mít očekávanou kvalitu, nebo že kupující nezaplatí danou částku [22].

Platformy online tržišť a různé další vyhledávače velmi usnadňují sdílení, filtrování a shromažďování informací o kupujících a prodejcích a tím se usnadňuje činnost P2P. Tato ekonomika proto může být úspěšná tam, kde existují technologické nástroje, které jednotlivcům zlehčují správu jejich vlastního podnikání a snižují komparativní výhodu specializace.

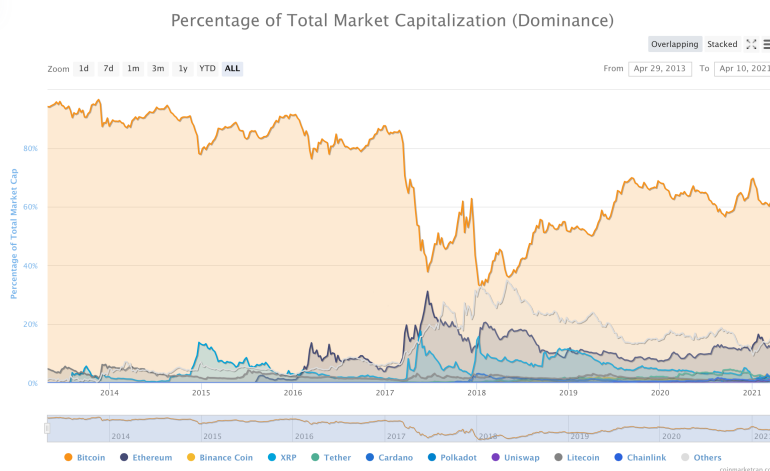
3.4 Populární kryptoměny

Na přelomu roku 2020 a 2021 zaznamenala většina kryptoměn významný růst svých hodnot a v současnosti mohou konkurovat i nejhodnotnějším firmám na světě. Celková tržní kapitalizace všech dosud známých kryptoměn činí k 10. 4. 2021 2,060,213,578,984 USD [19]. Což je řádově stejně vysoká částka jako jsou hodnoty nejlépe oceňovaných firem světa, jako jsou Apple, Microsoft či Facebook. Hodnota společnosti Apple činila ke konci roku 2020 přes 2,230,000,000,000 amerických dolarů, u Microsoftu se pak celková tržní kapitalizace pohybovala okolo 1,930,000,000,000 USD a Facebook nabýval hodnot 1,540,000,000,000 USD [23].



Obrázek 3.1: Celková tržní kapitalizace všech kryptoměn

U prvního grafu je možno vidět růst všech dosud známých kryptoměn v rozmezí osmi let. Významný nárůst je možno pozorovat především v posledních třech letech. Největší skok se pak připisuje začátku roku 2021.



Obrázek 3.2: Celkové zastoupení jednotlivých kryptoměn v procentech

V grafu výše je možno pozorovat zastoupení jednotlivých, populárních kryptoměn v provnání celkové tržní kapitalizaci. Je jasně viditelné, že nejdominantnější je měna bitcoin a v průběhu let nebyla nikým převýšena. Její dominance se ovšem díky vzniku nových kryptoměn postupně snižuje.

3.4.1 Bitcoin

Bitcoin je nejstarší a nejpobulárnější kryptoměna založená na platformě blockchain. V současné době má ze všech dosud známých kryptoměn nejvyšší hodnotu, která k 5.4.2021 dosahovala 1 134 571 883 958 amerických dolarů. Z celkové hodnoty všech kryptoměn představuje bitcoin 56,38%. Existují tisíce alternativních kryptoměn s podobnými funkcemi a specifikacemi. Některé z nich jsou klony bitcoinu, zatímco jiné jsou nově vzniklé měny, které byly vybudovány od nuly. Bitcoin nabízí příslib nižších transakčních poplatků než tradiční měny [24]. Je to první z digitálních měn, která začala používat technologii peer-to-peer k usnadnění plateb. Popularita této měny spustila vlnu dalších stovek kryptoměn, označovaných souhrnně jako altcoiny.

Bitcoin byl popsán v roce 2008 jednotlivcem či skupinou známou pod pseudonymem Satoshi Nakamoto, identita této osoby, která tuto technologii vytvořila není doposud stále známa. Následně v roce 2009 byla spuštěna do oběhu. Nové mince jsou těžařům vydávány pravidelně, ale jejich množství postupně klesá [24].

K 30. lednu 2021 je vytěženo 18 614 806 mincí. Způsob těžení je jedním z hlavních rozdílů od nekrytých měn. Klasické měny se v centralizovaných bankovních systémech uvolňují tempem odpovídajícím růstu zboží a tím se určuje cenová stabilita. Naopak decentralizovaný systém jako je bitcoin, nastavuje rychlost uvolňování předem podle algoritmu. Cena jednoho bitcoinu je v současnosti velmi vysoká a za poslední dobu zaznamenala velký nárůst. V lednu roku 2021 dosahovala přibližně 800 00 korun českých za jednu minci [25].

Bitcoinový systém je sbírka uzlů (počítačů), které provozují bitcoinový kód a ukládají jeho blockchain. V každém bloku se nachází uskupení transakcí. Na všech uzlech se nachází stejný seznam transakcí a bloků a díky tomu je systém transparentní a není možnost tento systém podvést. Transakce je možno živě sledovat. K hackerskému útoku by tedy mohlo dojít pouze, pokud by daný člověk ovládal 51% výpočetní síly, která tvoří bitcoin. V lednu roku 2021 bitcoin obsahoval přibližně 12 000 uzlů a toto číslo se bude dále navyšovat, tudíž je jakýkoliv útok na tuhle minci vysoce nepravděpodobný [24].

Zůstatky bitcoinových tokenů se skladují pomocí veřejných či soukromých klíčů, což jsou řetězce písmen a čísel, které jsou spojeny šifrovacím algoritmem, který byl použit při vytvoření daného klíče. Veřejný klíč pak slouží jako adresa, která je známá a lze na ní odeslat bitcoiny.

Naopak soukromý klíč je chráněný a má funkci autorizace bitcoinových přenosů. Bitcoinové klíče se mnohdy mylně zaměňují s bitcoinovou peněženkou, což je zařízení usnadňující obchod s bitcoiny. Pomocí této penženky je možno, aby daní uživatelé sledovali informace o jeho zakoupených mincích [24].

3.4.2 Ethereum

Ethereum bylo vyvinuto v roce 2013 programátorem Vitalikem Buterinem a síť byla do provozu následně uvedena v roce 2015. Do oběhu bylo vloženo 72 000 000 mincí. Ethereum je taktéž programovací jazyk, který má schopnost spouštět skripty Turing-complete, a díky tomu i decentralizované aplikace [26].

Ethereum je platforma založená na blockchainu, která se pokouší řešit problémy s centralizací tím, že pracuje autonomně na základě definovaných pravidel. Umožňuje, aby data uživatele zůstaly do značné míry skryté a bez centralizovaného ovládání. Informace daného uživatele se nachází na globální síti stovek počítačů nazvaných uzly, které vlastní a provozují anonymní správci [26]. Kdokoli může snadno využívat zdroje dostupné v síti Ethereum.

Blockchain slouží jako úložiště všech potřebných transakcí v síti a pomáhá rozhodovat s ověřováním identity uživatele, ukládají se zde platební údaje. Proto se pro irelevantní informace využívá datový kontejner podporovaný IPFS, který umožňuje plynulý tok všech druhů dat. Po bitcoinu je to nejznámější a největší kryptoměna. Významnou výhodou je jeho rychlá ziskatelnost, je možné jej nakoupit během 15 sekund, další výhodou je mnohem větší množství, které se v oběhu nachází [26].

Ether je kryptoměna této platformy, která funguje jako prostředek platby pro přispěvatele do sítě. Je to token provádějící danou platbu, umožňuje síti zpracovat změny, které uživatel provedl. Každá akce provedená v síti ethereum nebo v decentralizované aplikaci vyžaduje jiný čas a množství výpočetního výkonu. Čím větší výkon a delší čas je, tím se etherový poplatek navyšuje [26]. Tímto způsobem se ether liší od většiny digitálních měn, jako je bitcoin a jemu podobné altcoiny. Další odlišností je konečné množství vytěžených mincí, ether takové omezení nemá. Ročně se vytěží 18 000 000 etherů. V současné době je obtížné přesně určit, kolik mincí v oběhu vlastně existuje.

3.4.3 Litecoin

Litecoin byl spuštěný v roce 2011, je to kryptoměna založená na podobném principu jako je bitcoin. Kryptoměna byla vyvinuta bývalým inženýrem společnosti Google, který se jmenuje Charlie Lee. Byl spuštěn s cílem zlepšit nedostatky nacházející se u bitcoinu. Je to měna typu peer-to-peer a je založena na síti s plně decentralizovaným, otevřeným zdrojem. Není kontrolována žádným ústředním orgánem [27]. Jako algoritmus využívá Proof-of-work scrypt.

Oproti bitcoinu je litecoin schopný generovat čtyřikrát víc svých bloků, to znamená, že je nový blok vytvářen každých 2,5 minuty. Díky tomu je limit mincí Litecoinu 84 000 000. Což je v porovnání s bitcoinem velká výhoda. Na druhou stranu díky využívajícímu algoritmu scrypt vyžaduje Litecoin těžební hardware, kvůli kterému je potřeba mnohem více výpočetního výkonu. V současné době je v oběhu 50 000 000 mincí [19].

3.4.4 Binance Coin

Je to kryptoměna vydaná burzou Binance a její zkratka je BNB, funguje od roku 2017. Binance coin se používá na platformě Binance, která slouží hlavně k rychlé směně mezi různými druhy kryptoměn. Platforma je schopna zvládnout až 1,5 milionu příkazů za jednu sekundu. Nevýhoda platformy Binance je, že neumožňuje převody mezi tradičními měnami. Domovské tokeny BNB se primárně využívají jako prostředek pro hrazení poplatků, držitelé těchto mincí mají výhodnější směnné podmínky, a to hlavně poplatkové slevy [28]. Binance coin je spuštěn na blockchainu Ethereum a jeho limit je striktních 200 milionů BNB tokenů. Polovina z této hodnoty byla vyhrazena na pře prodej tokenů.

Všechny nové kryptoměny, které budou uvedené na burze Binance budou součástí tržiště pro obchodování s různými zavedenými a novými virtuálními tokeny. Některé platformy streamují videa Uplive, kde se promítají aktuální kurzy různých mincí. Oproti jiným kryptoměnám je Binance coin podporována mobilní aplikací a debetní kartou od společnosti VISA Monaka. Binance má v plánu zpětně odkupovat mince, podle svého plánu chce použít 20% svých zisků na zpětný odkup BNB tokenů a pak je vypálí a zničí, dokud nebudou zpětně odkoupeny maximálně 50 procenty, což je 100 milionů BNB tokenů. Tento proces tedy ponechá v oběhu pouze 100 milionů tokenů BNB, což mu pomůže udržet podstatnou hodnotu [28]. V polovině dubna 2018 tým Binance oznámil, že dokončil vypalování 2 220 314 tokenů BNB, zhruba 30 milionů USD [19].

3.4.5 XRP

XRP je kryptoměnou sítě s názvem XRP Ledger, která funguje v otevřeném softwaru v síti peer-to-peer. Hlavním vývojářem této měny je společnost Ripple. Vznikla v roce 2012 a do oběhu bylo vpuštěno 100 miliard jednotek XRP, jejich počet ale klesá, což znamená, že XRP bude čím dál méně. Nevzniká postupným těžením, ale 100% XRP bylo stvořeno již při prvním spuštění sítě [29]. Toto množství se rozdělilo na 100 miliard tokenů, každý jednotlivý token se dělí na šest desetinných míst. Nejmenší možný díl je tedy 0.000001, této jednotce se říká drop. Z celkových 100 miliard XRP vlastní Ripple 60% a z toho 55% je uzamčeno v časových kontraktech v síti XRP Ledger, kde se každý měsíc uvolňuje jedna miliarda. Ripple jednou za měsíc využije z každé miliardy část na podporu XRP ekosystému a zbytek uzamkne na dalších 55 měsíců do escrow. Cena jednoho XRP se od jeho vypuštění do oběhu pohybovala mezi 0,004 a 0,006 USD a zaznamenala nárůst v roce 2017 [30].

XRP Ledger je síť s otevřeným decentralizovaným kódem, neexistuje žádný server, který by kontroloval či evidoval dílčí transakce, to provádějí uzly a validátory. Uzlem a validátorem v síti XRP Ledger se může stát kdokoliv, kdo si stáhne příslušný software a spustí ho. Kromě jednotlivců provozují validátory i velké instituce jako Microsoft, či Massachusettský technologický institut. Uzly a validátory obsahují Ledger, který je synonymem pro blockchain u Bitcoinu [29]. Přibližně každých pět sekund se zapisuje nový blok, který obsahuje mimo jiné zůstatky všech účtů a poslední transakce. Potvrzení transakcí probíhá shodou všech v síti a transakce proběhne a zapíše se. Validátory transakce se třídí chronologicky. Potvrzení trvá přibližně 2 sekundy a poplatky za provedení jsou minimální.

XRP Ledger není platforma jen pro XRP, ale obsahuje i řadu dalších funkcí jako je vydávání dalších digitálních aktiv či používání platebních kanálů. Transakce s XRP jsou řádově rychlejší a levnější než ty bitcoinové či etherové. XRP Ledger nemá možnost provádět 51% útoků a publikovat alternativní verzi historie, tím je platforma bezpečnější. K uzavření a zápisu nového Ledgeru dochází každých pět sekund a tyto zápisy jsou neměnné, důležitým faktem je také, že XRP platforma je šetrnější k životnímu prostředí, protože na svůj provoz spotřebuje o mnohem méně elektrické energie [29].

3.4.6 Dogecoin

Dogecoin je kryptoměna, která byla vymyšlena na konci roku 2013 z recese během několika hodin. Vyvinuli ho softwaroví inženýři Billy Markus a Jackson Palmer. Jejich účel bylo vytvořit platební systém, který bude bez jakýchkoliv bankovních poplatků [31]. O této kryptoměně se začalo mluvit na začátku roku 2021, kdy zaznamenala významný nárůst. V dubnu 2021 pak došlo k velmi prudkému růstu, což mohlo zapříčinit vstup burzy kryptoměn Coinbase na burzu, přestože tato platforma neposkytovala obchodování s Dogecoiny [32]. Ke dni 16. dubna dosáhla tato měna zatím svého historického maxima, a to 0,45 USD, což bylo o 400% více než v předchozím týdnu. Během 24 hodin se prodalo za 70 miliard USD [32]. K polovině dubna se tedy tržní kapitalizace Dogecoinu blíží k 50 miliardám dolarů, a to z ní činí pátou nejlépe hodnocenou kryptoměnu. Za zmínku taktéž stojí, že se její hodnota meziročně zvýšila o více než 7 000%.

Tento výrazný nárůst během 24 hodin se připisuje Elonu Muskovi a internetovým memes, před takto výrazným nárůstem totiž Musk přidal tweet, který obsahoval obrázek psa se slovy: „Doge štěká na Měsíc“. Zájem o tuto měnu pak přispěl v polovině dubna 2021 k výpadku kryptoměnového systému Robinhood na platformě elektronického obchodování způsobenému bezprecedentní poptávkou a vyvolal obavy odborníků z blížící se spekulativní bubliny na trhu kryptoměn [32].

V týdnu od 10. do 14. května zaznamenala tato kryptoměna opět obrovský nárůst. Tento vývoj se připisuje Elonu Muskovi, který prohlásil, že spolupracuje s Doge devs na zlepšování efektivity transakcí. Musk je dlouhodobě obviňovaný z ovlivňování trhu, jeho obranou je, že by jeho komentáře ohledně dogecoinu neměly být brány vážně. Tento týden avšak opakovaně prohloubil propad trhu s kryptoměnami, když prohlásil, že nebude možno platit za produkty firmy Tesla pomocí bitcoinu, protože ho znepokojuje jeho dopad na životní prostředí. Doge coin je v současnosti oceňovaný jako čtvrtá nejhodnotnější kryptoměna a to po bitcoinu, etheru a binance coinu. Jeho celková tržní kapitalizace je 70 miliard dolarů [33].



Obrázek 3.3: Celková tržní kapitalizace všech kryptoměn

Z grafu je možno vyčíst vývoj hodnoty Dogecoinu v průběhu roku 2020 a 2021. Výrazný nárůst zaznamenala tato měna až v první a druhé čtvrtině roku 2021. Největší růst je pak připisován konci měsíce dubna, kdy se jeho hodnota sedminásobila. Někteří lidé tento rapidní růst připisují za příčinu Elonu Muskovi. A to kvůli jeho tweetu, zveřejněnému v dubnu, kde popisuje, že Dogecoin rapidně poroste. Po sdílení tohoto příspěvku byl zaznamenán enormní nákup této doposud relativně neznámé kryptoměny.

Kapitola 4

Datová analýza

Analýza dat je široká oblast statistiky, která pracuje na principu shromažďování dat za cílem objevit důležité informace a tím ovlivnit budoucí rozhodování [34]. Využívá se jak ve vědě, tak i v aplikovaných oblastech a její důležitost se stále zvyšuje, hlavně díky digitalizaci a automatizaci v mnoha odvětvích. Datová analýza se dělí na tři části, které jsou:

- Statistická analýza dat, která se dále skládá ze sumarizace dat, hledání v datech dosud neznámých souvislostí a testování statistické hypotézy.
- Business intelligence, což je analýza dat komerčních organizací, zaměřená na monitorování a řízení činnosti v daných subjektech.
- Vytěžování dat a k tomu příbuzná datová věda, jsou oblasti zabývající se informacemi získanými v určitém čase. Tato data se pak využívají v prediktivním modelování, čímž se chápe predikce chování lidí, či jiných jednotek a následné zpracování velkých dat na jejichž základě se tvoří aplikace. Tomuto odvětví se budeme dále zabývat v této práci [35].

4.1 Datová věda

Tento vědní obor využívá různé procesy, algoritmy a systémy pro získávání znalostí a poznatků z dat ve strukturovaných i nestrukturovaných podobách. Toto odvětví sjednocuje statistiku, analýzu dat, strojové učení a související metody za účelem naučit se chápat a analyzovat informace o datech. Při těchto procesech využívá mnoha oborů jako například matematiky, statistiky, informatiky a matematické informatiky.

4.2 Vytěžování dat

Tato analytická metoda funguje na principu získávání složitých informací z dat. Její součástí je získávání znalostí z databází. Tato schopnost se využívá hlavně v marketingu, ale i ve vědecké sféře, například při analyzování genetických informací či při pozorování dění na internetu za účelem odhalit potencionální kriminální činnost.

4.2.1 Rozhodovací strom

Princip této techniky spočívá v identifikaci objektů do přehledných tříd. Jako příklad lze uvést tabulku, kde jednotlivé sloupce jsou atribut. Nedílnou součástí tohoto odvětví je algoritmus. Rozhodovací strom se musí nejprve vytvořit z množiny daných objektů, které musí někdo (učitel, jiný algoritmus) zařadit do skupin skupina se obvykle označuje jako závislý člen a zapisuje se do tabulky do posledního sloupce. Jedná se tedy o učení s učitelem. Každý uzel stromu představuje rozhodování podle vybrané vlastnosti objektu. Z takového uzlu vede konečný počet hran, díky tomu je potřeba dané vlastnosti nejprve diskretizovat, například z reálných čísel do intervalů. Atribut pro kořenový uzel se vybere takový, aby od sebe zkoumané moduly co nejvíce odlišil, je využívána entropie [36].

Supervizované učení

U tohoto učení se používá taktéž název učení s učitelem. Algoritmus je charakteristický trénovacími daty, což je většinou informace skládající se z jedné dvojice na vstupu a daného výstupu. Funkce na konci trénovacích dat jsou při regresi spojité hodnoty, při klasifikaci predikují tato data značení dané třídy na vstupu. Úloha algoritmu učení s učitelem je predikovat výstupní hodnotu pro každý funkční vstup, zpracovat a zobecnit příkaz. Jako příklad může sloužit situace, kde se datová sada skládá z počtu obyvatel měst v České republice za posledních 30 let, a za úkol je zjistit, kolik obyvatel budou mít tato města za čtyři roky. Výsledek je pak analyzován z dat vygenerovaných v minulosti, kterými v tomto případě jsou počet obyvatel, měst a rok [37].

Algoritmy strojového učení

Algoritmy strojového učení jsou části kódu, které analyzují a hledají význam ve složitých datových sadách. Každý algoritmus je konečnou sadou jasně daných instrukcí, podle kterých následně počítač dosahuje zadaného úkolu, což je většinou vytvoření vzorce, který se pak aplikuje k predikování a kategorizaci informací. Algoritmy strojového učení aplikují parametry skládající se z trénovacích dat. Postupným rozšiřováním těchto dat, se stává algoritmus přesnějším. Nejčastěji se slučují technikami strojového učení, které se dělí na tři části, a to na supervizované učení, učení bez dozoru a zpětnovazební učení [38]. K předpovídání finálních požadavků, nestandardních datových bodů, či zjišťování podobností využívají tyto algoritmy regrese a klasifikace. Tato technika je vhodná k použití, pokud uživatel ví, jak by měl vypadat výsledek. Existuje zde zpětná vazba od uživatele pro lepší automatizaci. Tento typ predikce je vhodný pro algoritmičké obchodování a bude dále využíván v této práci [38].

4.2.2 Neuronové sítě

Umělá neuronová síť je model využívaný v umělé inteligenci, která je založena na základu chování neuronových sítí v mozku. Tato síť je určena pro zpracovávání dat, je složena z umělých neuronů, které jsou navzájem spojené a dochází v nich k předání informací, a pak se informace transformují ke svojí potřebě. Struktura neuronu je taková, že má neomezené množství vstupů, výstup je však vždy pouze jeden, s mnoha větvemi [39].

Neuronová síť je umělá inteligenční síť, která kopíruje chování a funkce v lidském mozku. Nejprve se zavedou datové řetězce do sítě, poté se samotný počítač učí přivedené vzory, to pak umožňuje vybírat správné odpovědi a analyzovat trendy. Tato síť se využívá při úkolech, kde se klade důraz na určitý stupeň učení a rozpoznávání vzorců [39]. Těžba dat je analýza trendů ze shluku informací zkoumaného odvětví.

Mezi nejvyužívanější modely umělých neuronových sítí patří vrstevnaté neuronové sítě typu zpětného šíření, které byly navrženy Rumelhartem a Hintonem v roce 1986. Skládají se ze vstupní vrstvy neuronů, které si předávají signály z vnějšího prostředí, několika mezilehlých vrstev, které se někdy označují jako skryté vrstvy neuronů a nejsou nijak přímo napojené na vnější prostředí a výstupní vrstvy neuronů, která předává signál zpracovaný sítí [40]. Neurony uvnitř jednotlivých vrstev mění svou aktivitu paralelně, po jednotlivých vrstvách od vstupu až na výstup.

Vrstevnaté neuronové sítě se naučí vhodně reagovat na vstupy pomocí trénovacích dat ve tvaru: vstupní vzor/požadovaný výstup, což je proces strojového učení s učitelem. Zároveň se adaptují váhy spojů mezi jednotlivými neurony sítě, od výstupní vrstvy směrem ke vstupní. Cílem je najít množinu vah sítě, která bude dávat pro všechny vstupy stejné výstupy dané sítě. Úloha neupřesňuje aktivitu skrytých neuronů. Vhodné chování sítě pak specifikuje cílová funkce, což je odchylka mezi skutečným a požadovaným výstupem sítě pro každý vzor [40]. V průběhu učení se hodnota cílové funkce zmenšuje.

Elmanova síť vzniká přidáním kontextové vrstvy neuronů, ty opakují pomocné vstupy i výstupy sítě skrytých neuronů, které byly vypočteny pro předchozí vzor. Tato funkce slouží jako typ vnitřní paměti, a díky tomu je umožněno zpracovávat posloupnosti variabilní délky. Na podobném principu jsou pak založeny i další, pokročilejší varianty rekurentních neuronových sítí [40].

Neuron

Základní funkční jednotku neuronových sítí představuje takzvaný formální neuron, přijímající enormní množství přichozích informací v jednom čase. Zpracované informace se předávají dále pomocí jednoho velmi členitého výstupu. Formální neuron tedy počítá hodnotu vnitřního potenciálu, která je výsledkem součtu vstupních složek a hodnoty prahu. Díky synaptickým spojům jsou neurony propojené a tím se vytváří neuronová síť. Orientace synaptického spoje je hodnocena reálnou hodnotou, která se nazývá váha. V dnešní době je popsána široká škála neuronů, od velmi jednoduchých, které jsou založeny na principu nespojitě přenosové funkce až po velmi složité, které popisují detailní chování neuronů [41]. Jedním z

nejvíce využívaným modelem je popsán McCullochem a Pittsem:

$$\sum_{i=1}^N (w_i x_i) + \Theta \quad (4.1)$$

Kde x_i označuje vstup neuronu, Y charakterizuje výstup.

Symbol w_i má význam synaptické váhy.

Theta je práh a $S(x)$ charakterizuje význam přenosové funkce neuronu.

Dopředná neuronová síť

Je to typ neuronové sítě, spojení zde necestuje ve smyčce, ale v jediné směrové cestě. To je hlavní rozdíl mezi dopřednou a opakující se neuronovou sítí. V opakující se síti lze pohyb systému dopředu i dozadu [42]. Dopředná neuronová síť je nejběžněji používaným typem neuronové sítě, hlavně díky svojí jednoduchosti a konfiguraci. Využívá se při těžbě dat a dalších oblastech učení, kde je požadována predikce dějů. Tato síť se značí svojí velkou efektivitou, hlavně kvůli koncovým uživatelům, kteří umožňují přístup k čím dál většímu množství experimentálních dat. Stejně jako při počítání průměru bude přesnější výsledek dosažen využitím velkého počtu testovacích událostí. Například pravděpodobnost převrácení čísla jedna na šestibokou matrici je 16,667 %; požadováno ale bude stovky nebo tisíce simulací, než bude vypočítaný průměr potvrzen pomocí skutečných dat [42]. Nejjednodušší druh této neuronové sítě je jednovrstvá síť s perceptrony, je složena z jedné vrstvy výstupních uzlů a vstupy se přivádí přímo na výstup závažím. Váhy a vstupy se sčítají v každém uzlu, pokud hodnota nabývá prahové hodnoty, obvykle nula, neuron si propůjčuje aktivovanou hodnotu, která se z pravidla rovná jedné. Neurony, které mají tuto schopnost se nazývají umělé neurony či lineární prahové jednotky.

Perceptron

Perceptron je nejjednodušší model dopředné neuronové sítě, obsahující pouze jeden neuron. U strojového učení s učitelem má funkci algoritmu. Jednovrstevný perceptron je možno využít pouze u množin, které jsou lineárně separovatelné. Větší škálu využití je pak možné najít u vícevrstevnatého perceptronu. Hlavní úkol zde má binární klasifikátor této funkce, který rozhoduje zda vstup, který je charakterizován vektorem čísel bude součástí konkrétní třídy [43]. Jedná se o typ lineárního klasifikátoru, který predikuje dění díky lineárnímu, tj. klasifikačnímu algoritmu, který dělá své předpovědi na základě funkce lineárního prediktoru kombinujícího sadu vah s vektorem prvku. Nejjednodušší perceptron je binární klasifikátor, který mapuje vektor vstupů na výstupní funkci.

$$f(x) = \begin{cases} 1 & \text{pro } \omega \cdot x + b > 0 \\ 0 & \text{jindy} \end{cases} \quad (4.2)$$

Vektor vah je značen ω , b je konstanta.

Finální impuls neuronu udává vážený součet

$$\xi = \sum_{i=1}^N \omega \cdot x_i - \theta \quad (4.3)$$

Tento vztah se označuje jako celkový potenciál neuronu, na který reaguje perceptron (neuron) výstupní odezvou $Z=S(\xi)$, kde S je označení pro přenosovou funkci [43].

Vstupy

Jednotka x zpravidla označuje výstupy jiných neuronů nebo impulsy z vnějšího prostředí. Neuron získává z těchto prvků informace v podobě číselné hodnoty.

Váhy

Priorita a důležitost konkrétních vstupních spojů se udává pomocí synaptických vah ω_i . Jsou to reálné hodnoty, jejichž vektor je kolmice na rozhodovací hranice. Váhy tudíž mohou být nastaveny na mnoho řešení, aby odpovídaly správnému výsledku [43].

Práh

Je považován za speciální druh váhy, je to spojení vedoucí do imaginárního neuronu s výstupem o hodnotě -1, díky tomu je $\omega_0 = \theta$ a $x_0 = 1$, kde následně dostáváme vztah

$$\xi = \sum_{i=0}^N \omega_i \cdot x_i = \omega \cdot x \quad (4.4)$$

4.2.3 Regresní analýza

Regresní analýza je statistická metoda, díky které se odhaduje hodnota náhodné veličiny, která se nazývá závislá proměnná na základě předchozí znalostí jiných veličin. Využívá se v situacích, kdy je potřeba určit kvantitativní proměnné závislé na jedné nebo více proměnných, které se označují pojmem regresory. Předem je určeno, jaká proměnná bude nezávislá, takzvaná vysvětlující a která bude označována jako závislá, které se říká odezva. Tato analýza se využívá za účelem popsat vzájemné závislosti pomocí vyhovujícího matematického modelu [44].

V závislosti na množství nezávislých proměnných se dělí na modely jednoduché či vícenásobné regrese. Jednoduchá regrese se zabývá pouze závislostí na jednom regresoru, kdežto vícenásobná regrese řeší závislost na větším množství regresorů. Tento typ analýzy lze taktéž dělit na lineární a nelineární modely, podle typu jejich funkce. Podstata regresní analýzy je stanovit nejlépe vyhovující model či matematickou rovnici, která bude vhodně charakterizovat závislost x na y [44]. Dále je jejím úkolem stanovit nejlepší odhad zkoumaných vzorků, určit zda vytvořený model za příčiní přesnější odhad nezávisle proměnné veličiny, oproti využití běžného průměru [45].

Kapitola 5

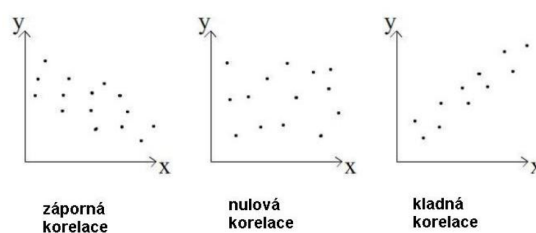
Metodologie

5.1 Korelace

Korelace je nejznámější a nejčastěji využívaný model v oboru statistiky, kde princip spočívá v porovnávání dvou na sobě závislých veličin. Znamená to, že pokud se mění jedna veličina, změnu zaznamená i druhá a naopak. Pokud se korelace u dvou zkoumaných vzorků prokáže, je pravděpodobné, že na sobě veličiny závisejí, nelze však posoudit, jestli je jedna z nich příčinou a druhá následkem, čímž se liší od regresní analýzy [46]. Z toho lze říci, že korelace zkoumá závislost, ale není možné rozhodnout o kauzalitě. Korelace se hojně využívá v oblasti statistiky a bude dále využívána v této práci.

Ve statistice pojem korelace znamená lineární vztah mezi dvěma veličinami, například x a y . Míru korelace vyjadřuje korelační koeficient, který nabývá hodnot v rozmezí mezi $+1$ a -1 . Čím blíže je hodnota korelačního koeficientu jedné nebo mínus jedné, tím je vztah silnější. Hodnoty kolem nuly naopak ukazují, že mezi sebou dané vzorky nemají žádný vztah [47].

- Negativní korelace: Čím vyšší hodnotu naplňuje jedna veličina, tím více klesá hodnota druhé veličiny [46].
- Nulová korelace: Tato korelace je velmi slabá až nulová a k tomuto jevu dochází, pokud mezi daným příklady nedochází k žádné asociaci. Mezi veličinami tedy není nalezena žádná lineární závislost [47].
- Pozitivní korelace: Hodnoty jedné funkce mají stejný směr, jako hodnoty druhé a naopak. To znamená že rostou obě funkce.



Obrázek 5.1: Korelační koeficient

5.1.1 Pearsonův korelační koeficient

Je to koeficient využívající se ve statistice a datové analýze k určení lineární korelace mezi dvěma nebo více soubory dat. Výsledek se pohybuje vždy v intervalu +1 a -1 Tato kovariance má schopnost odrážet pouze lineární korelaci proměnných a ignoruje proto mnoho dalších typů vztahů nebo korelací [48].

Vztah pro Pearsonův korelační koeficient se značí kovariancí dvou proměnných, která je dělena součinem jejich směrodatných odchylek. Ve vzorci je zahrnut takzvaný produktový moment, což je v podstatě průměr součinu náhodných proměnných upravených podle střední hodnoty.

$$\rho_{X,Y} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}, \quad (5.1)$$

Pearsonův korelační koeficient se značí řeckým písmenem ρ a cov je zkratka pro kovarianci, σ_X je deviace hodnoty X a σ_Y naopak deviací hodnoty Y [48].

5.1.2 Spearmanův korelační koeficient

Tato korelace je téměř totožná s Pearsonovou korelací, zásadním rozdílem mezi nimi je, že Spearmanova korelace se zabývá monotónními vztahy, ať už jsou lineární nebo ne. Pokud neexistují žádné opakované hodnoty dat, dojde k dokonalé Spearmanově korelaci, která nabývá +1 nebo -1, kde každá proměnná funkce je dokonalou monotónní funkcí té druhé [49].

Vztah pro Spearmanův korelační koeficient

$$r_s = \rho_{rg_X, rg_Y} = \frac{\text{cov}(rg_X, rg_Y)}{\sigma_{rg_X} \sigma_{rg_Y}}, \quad (5.2)$$

Kde cov je označení pro kovarianci hodnot daných proměnných, σ_{rg_X} a σ_{rg_Y} pak charakterizuje standardní odchylky řadových proměnných.

Veličina X je označována jako nezávislá proměnná a Y jako závislá. Pokud má Y tendenci se zvyšovat a zároveň narůstá i X, Spearmanův korelační koeficient je vyhodnocen jako kladný. Pokud naopak veličina Y klesá, ale X se naopak zvyšuje, je Spearmanův korelační koeficient záporný. Pokud vyjde Spearmanova korelace nulová, znamená to, že neexistuje žádný ukazatel toho, zda hodnoty X či Y rostou nebo klesají. Pokud jsou X a Y dokonale monotónně příbuzné, Spearmanovy korelační koeficienty se stanou 1. Dokonale monotónně rostoucí vztah znamená, že pro jakékoli dva páry datových hodnot X_i, Y_i a X_j, Y_j mají $X_i - X_j$ a $Y_i - Y_j$ vždy stejné znaménka. Z dokonale monotónního klesajícího vztahu vyplývá, že tyto rozdíly mají vždy opačná znaménka [49].

5.1.3 Granularita dat

Granularita nebo také zrnitost, je úroveň hloubky do které se budou data reprezentovat. Jednoduše lze granularita popsat jako úroveň podobnosti. Také se na základě zrnitosti určuje dělitelnost daného systému. Jemnozrné systémy s vysokou granularitou jsou rozděleny na větší počet menších částí. Hrubo-zrný systém má menší počet větších objektů. Jako konkrétní příklad lze uvést měření hmotnosti předmětu v gramech, což bude granulárnější než měření hmotnosti stejného objektu

v tunách [50]. Tento koncept je důležitý při zkoumání dat pomocí jejich analýzy a bude využíváno v této práci.

Granularita se také využívá k popisu rozdělení dat. Data s nízkou granularitou jsou rozdělena do malého počtu polí, zatímco data s vysokou granularitou jsou rozdělena do většího počtu konkrétnějších polí. Například záznam o fyzických vlastnostech osoby s vysokými daty může mít samostatná pole pro výšku, hmotnost, věk, pohlaví, barvu vlasů, barvu očí a podobně [50]. Zatímco záznam s nízkými údaji sice zaznamená stejné informace do menšího počtu obecnějších polí a ještě nižší záznam uvádí všechny informace v jednom poli. Díky větší granularitě jsou data flexibilnější tím, že umožňují zpracovávat konkrétnější části dat samostatně, ale vyžadují větší výpočetní prostředky.

5.2 Posunutá korelace

Posunutá korelace, někdy taktéž nazývaná jako křížová korelace je mírou podobnosti dvou řad v závislosti na posunutí jednoho zkoumaného prvku vůči druhému. Tato metoda je využívána k hledání dlouhého signálu pro kratší funkci. Podstatnou složkou této kapitoly je také autokorelace, což je vzájemná korelace signálu se sebou samým, vrchol této funkce musí vždy obsahovat zpoždění nuly a jeho velikost bude energie signálu.

Autokorelace náhodných složek je jev, označovaný jako porušení Gauss-Markovova požadavku pro možnost odhadu regresních parametrů metodou nejmenších čtverců [51]. Při autokorelaci je vytvořena nenulová kovariance, která je charakterizována matematickým vztahem:

$$\Sigma = \sigma^2 \cdot I_n, \quad (5.3)$$

Kde σ^2 je dosud neznámý rozptyl a I_n je charakterizována jako jednotková matice n -tého řádu [51].

V regresní analýze jsou hodnoty dány závislostí teoretické funkce i a náhodné složky i . Hodnota náhodné složky je pak určována rozdílem mezi skutečnou a teoretickou hodnotou.

5.3 Fourierova analýza

Fourierova transformace je založená na principu integrální transformace a převádí signál mezi časově a frekvenčně závislými veličinami. Následně se vyjadřuje harmonickými signály, to znamená sinusovými a cosinusovými funkcemi. Jedná se tedy o exponenciální funkce [52].

Fourierova analýza vyjadřuje dané funkce jako součet periodických složek a z nich následného získávání signálu. Je to matematická operace, která mění osu X signálu z času na frekvenci [52]. Během této matematické operace se funkce nahradí diskretizovanými protějšky. Tento úkon se nazývá diskrétní Fourierova transformace (zkráceně DFT). DFT je základ při numerických výpočtech a je často využívaný

hlavně díky rychlému výpočetnímu algoritmu, který se nazývá Rychlá Fourierova transformace, anglicky pak Fast Fourier Transform (FFT) [54].

5.3.1 Rychlá Fourierova transformace

Vztah pro DFT je definován pomocí vzorce:

$$y[k] = \sum_{n=0}^{N-1} e^{-2\pi j \frac{kn}{N}} x[n], \quad (5.4)$$

Přímé vyhodnocení těchto sum by bylo časově velmi náročné, díky DFT algoritmům je ušetřeno velké množství práce. Pro inverzní DFT pak platí vztah:

$$x[k] = \frac{1}{N} \sum_{n=0}^{N-1} e^{2\pi j \frac{kn}{N}} y[n]. \quad (5.5)$$

Inverzní DFT a původní DFT jsou stejné, jen se liší znaménka u exponentu a koeficientu $1/N$, proto kterýkoli algoritmus je možné snadno modifikovat i pro výpočet inverzní DFT. Vstupní signál DFTZ je zkrácen a díky tomu lze vymodelovat násobení nekonečného signálu funkcí obdélníkového okna [53].

5.4 Hilbert-Hungova transformace

Hilbert-Huangova transformace, z anglického názvu Hilbert–Huang transform, je metoda, která umožňuje rozklad signálu do funkcí vnitřního režimu. Označuje se zkráceně písmeny IMF. Pomocí rozkladu signálu se získají okamžitá frekvenční data. Tato metoda umí dobře pracovat s nestacionárními a nelineárními daty [54].

5.4.1 Empirický režim rozkladu

Empirický režim rozkladu je základní částí HHT, tato metoda funguje na základě rozkladu signálu na různé komponenty. Často je připodobňována k jiným analytickým metodám, jako jsou například Fourierova transformace či vlnková transformace [54]. EMD, kde tato zkratka pochází z anglického Empirical Mode Decomposition, je možná rozložit jakoukoliv složitou sadu dat na konečný a často malý počet komponent. Tyto komponenty tvoří úplný a téměř ortogonální základ pro původní signál [55].

První funkce vnitřního režimu značená zkratkou IMF zpravidla nese neoscilující, vysokofrekvenční prvky, které se značí jako náhodné šumy. Odstrašení šumů probíhá díky vyhlazovacím algoritmům [55]. Vzhledem k tomu, že rozklad je založen na charakteristické časové škále dat, lze jej využít i u nelineárních a nestacionárních procesů.

IMF je definována jako funkce, která musí splnit dané požadavky, kterými jsou:

- Počet extrémů a počet nulových přechodů musí být stejný, nebo se musí lišit maximálně o jeden.

- V kterémkoliv bodě je střední hodnota nadefinovaná jako místní maximum a nula je charakterizovaná jako místní minimum.

Hilbertova spektrální analýza zkoumá MMF, kde konečným výsledkem je frekvenční a časové rozdělení amplitudy signálu (nebo energie), označené jako Hilbertovo spektrum. To umožňuje identifikaci lokalizovaných prvků. MMF představuje jednoduchý oscilační režim jako protějšek jednoduché harmonické funkce. Místo konstantní amplitudy a frekvence v jednoduchém harmonickém komponentu může mít MMF podél časové osy proměnnou amplitudu a frekvenci [56]. Postup extrakce MMF se nazývá prosévání. Proces prosévání je následující:

- Určí se všechny místní extrémy v testovacích datech.
- Připojí se všechny místní maxima pole a kubická spline čára jako horní obálka.
- Opakujte postup pro místní minima k výrobě spodní obálky.
- Horní a dolní obálka by měla pokrývat všechna data mezi nimi.

Metoda EMD byla vyvinuta tak, aby mohla být data zkoumána v adaptivním časově-frekvenčně-amplitudovém prostoru pro nelineární a nestacionární signály [56]. Metoda EMD vstupní signál rozloží na několik funkcí vnitřního režimu (IMF). Pro tento vztah je charakteristická rovnice:

$$I(n) = \sum_{m=1}^M IMF_m(n) + Res_M(n) \quad (5.6)$$

kde $I(n)$ je označení pro vícesložkový signál. IMF_m je funkce vnitřního režimu a $Res_M(n)$ představuje zbytek odpovídající M vnitřní množiny [56].

5.5 Python a jeho uživatelské rozhraní

V této práci bude použit k datové analýze jeden z nejpoužívanějších programovacích jazyků, kterým je Python. Díky jeho jednoduchému uživatelskému rozhraní je vhodný pro analýzu dat ve vědách. V posledních letech se díky vylepšené podpoře knihoven Pythonu, hlavně knihovny Pandas stal silnou alternativou pro úkoly manipulující s daty. Jako součást se používá Jupiter notebook, kde probíhá spojování dat a analytických kódů [55]. Následně se zde vytváří živé zprávy, které je možno dále upravovat a analyzovat. Jedním z hlavních důvodů, proč je Python v tomto odvětví tak populární, je široká škála existujících knihoven. Tyto knihovny obsahují více než 130 000 balíčků, které obsahují nástroje pro manipulaci a automatizaci databází, vědecké výpočty a další funkce. Níže jsou uvedené některé knihovny, které byly využity v této práci.

5.5.1 NumPy

Nezkráceně Numerical Python, je balíček vyžadovaný pro vysoce výkonné vědecké výpočty a analýzu dat. Je to základ, na kterém budou postaveny další nástroje a knihovny v této práci [58]. Zde jsou některé z věcí, které poskytuje:

- Numpy: je to vícerozměrné pole poskytující aritmetické operace
- Standardní matematické funkce pro rychlé operace s celou řadou dat bez nutnosti psát smyčky
- Nástroje pro čtení / zápis dat pole na disk a práci se soubory mapovanými do paměti
- Lineární algebra, generování náhodných čísel a Fourierova transformace
- Nástroje pro integraci kódu napsaného v jazycích C, C++ a Fortran

Numerická knihovna Pythonu má také v neposlední řadě významné statistické funkce, například `np.corrcoef()`, které vracejí matici Pearsonových korelačních koeficientů, to znamená, že lze začít importem NumPy a definováním dvou polí NumPy. Jedná se o instance třídy `ndarray`. Jako příklad poslouží dva subjekty značené X a Y, použije se zde `np.arange()` k vytvoření pole X celých čísel mezi 10 a 20. Potom pomocí `np.array()` vytvoříte druhé pole y obsahující libovolná celá čísla [59]. Jakmile máte dvě pole stejné délky, můžete zadat příkaz `np.corrcoef()`. S touto funkcí se bude pracovat následně v této části práce.

5.5.2 SciPy

Je to kolekce balíčků, zabývající se řadou různých standardních problémových domén ve vědeckých výpočtech. Skládá se z dalších balíčků, kterými jsou například:

- `scipy.integrate`: numerické integrační rutiny, které se zabývají řešením diferenciálních rovnic
- `scipy.linalg`: rutiny lineární algebry a maticové rozklady, které jsou uvedeny v `numpy.linalg`

- `scipy.optimize`: optimalizátory funkcí (minimalizátory) a algoritmy hledání kořenů
- `scipy.signal`: nástroje pro zpracování signálu
- `scipy.sparse`: řídké matice a řídicí řešitelé lineárního systému
- `scipy.special`: obal kolem SPECFUN, Fortranské knihovny implementující mnoho běžných matematických funkcí, jako je funkce gama
- `scipy.stats`: standardní spojitá a diskrétní rozdělení pravděpodobnosti (funkce hustoty, vzorkovače, funkce spojité distribuce), různé statistické testy a podrobnější statistiky
- `scipy.weave`: nástroj pro použití vloženého kódu C ++ k urychlení výpočtů pole[58].

5.5.3 Matplotlib

Je to nejoblíbenější knihovna v Pythonu pro vytváření grafů a dalších dvou-rozměrných vizualizací dat. Tato knihovna dobře spolupracuje s IPythonem, a díky tomu umožňuje pohodlné interaktivní prostředí pro vykreslování a zkoumání dat.

5.5.4 IPython

Tato součást standardní vědecké sady nástrojů Pythonu spojuje vše dohromady. Poskytuje významné a produktivní prostředí pro interaktivní a průzkumné výpočty. Jedná se o vylepšené prostředí Pythonu navržené k urychlení psaní, testování a ladění kódu Pythonu. Je to zvláště užitečné pro interaktivní práci s daty a vizualizaci dat pomocí matplotlib. IPython obvykle také obsahuje tyto položky:

- Notebook typu HTML podobný Mathematice pro připojení k IPythonu prostřednictvím webového prohlížeče
- Konzole GUI založená na rozhraní Qt s inline vykreslením, úpravami více řádků a zvýrazněním syntaxe
- Infrastruktura pro interaktivní paralelní a distribuované výpočty

5.5.5 Pandas

Tato knihovna poskytuje čtení datových struktur a funkce navržené pro rychlou a snadnou práci se strukturovanými daty. Tato funkce umožňuje Pythonu být výkonným a produktivním prostředím pro analýzu dat. Primárním objektem v pandas, který bude použit v této práci je DataFrame, dvourozměrná tabulková, sloupcově orientovaná datová struktura s popisky řádků i sloupců [58].

Pandas kombinuje výkonné maticové výpočetní funkce NumPy s různými možnostmi manipulace s daty v tabulkách a databázích. Poskytuje funkce indexování, které usnadňují přetváření, rozdělování, provádění agregací a výběr podmnožin dat. Pandas je primární nástroj, který bude v této práci použit.

Pro finanční analýzy je pandas vhodná díky vysoce výkonému funkci čtení časových řad a nástrojům vhodných pro práci s finančními daty. I když se jedná o knihovnu v Pythonu, je velmi podobná jazyku R, protože je to jedno z nejpoužívanějších prostředí pro analýzu dat s otevřeným zdrojovým kódem [58].

Kapitola 6

Praktická část

6.1 Motivace

Datová analýza a hledání podobností zkoumaných veličin je v současnosti velmi populární obor. Hlavně díky pokroku v oblasti technologií, je možnost dostat se k mnohem větším objemům dat. Získávání potřebných data se také navyšuje díky popularitě ukládání si informací na cloudy. Během posledních několika let se množství volně dostupných dat dramaticky zvýšilo, a proto je možnost tyto data testovat a vylepšovat pomocí algoritmů.

6.2 Získání dat z virtuální směnárny

Data použitá v této práci byla získána z internetové stránky: <https://cryptocandledata.com>. Jako zkoumané kryptoměny byly vybrány Bitcoin, XRP, Ethereum a Binance coin. Zkoumaný časový údaj je úsek mezi červencem roku 2020 a květnem 2021. Byly zkoumány hodinové a denní záznamy pro každou kryptoměnu. Následně byla tato data převedena do csv souborů a byl vytvořen kód, který je schopen generovat tato data za účelem dalšího využití.

```

import json
import csv
import time
import math
import sys

import requests

import pandas as pd

def get_data(pair, interval, number_days, exchange="binance"):
    """
    pair: napr: ETHBTC
    exchange: napr: binance
    interval: granularita (jak často), napr: 1d, 1h
    number: počet svíček (funkce by měla složit více dotazu, pokud je number > 500)
    """
    URL = "https://cryptocandledata.com/api/candles"
    MAX_COUNT = 4

    data = []
    done = False

    end_date = int(time.time())
    start_date = int(end_date - (number_days * 3600 * 24))

    while not done:
        payload = {
            "exchange": "binance",
            "tradingPair": pair,
            "interval": interval,
            "startDateTime": start_date,
            "endDateTime": end_date
        }

        r = requests.get(URL, params=payload)

        data += r.json()["candles"]

        start_date = int(data[-1]["timestamp"]) // 1000

        if start_date >= end_date - (1 * 3600 * 24):
            done = True

    df = pd.DataFrame(data)
    df = df.drop_duplicates(subset='timestamp', keep="first")

    # plt.plot(df["timestamp"].values)
    # plt.show()

    filename = "data/{}-{}.csv".format(pair, interval)

    try:
        df.to_csv(filename, index=False)
    except Exception as e:
        print(e)

```

Obrázek 6.1: Nastavení automatického stahování dat z on-line směnárny a nastavení daných granularit

Stahnutí konkrétního souboru

```

PAIR = "BTCUSDT"
INTERVAL = "1d"

get_data(PAIR, INTERVAL, 100)

```

Obrázek 6.2: Stahnutí konkrétního souboru a určení, jakou granularitu požadují

Stahnutí množiny souborů

```

pairs = [
    "BTCUSDT",
    "ETHUSDT",
    "BNBUSDT",
    "XRPUSDT",
    "XDGUSDT"
]

intervals = [
    "1h", "1d"
]

for pair in pairs:
    for interval in intervals:
        print(pair, interval)

        done = False
        while not done:
            try:
                get_data(pair, interval, 300)
                done = True
            except Exception as e:
                print(e)
                print("Fail, trying again")
                time.sleep(15)
            time.sleep(2)

print("Done")

```

Obrázek 6.3: Stahnutí více csv souborů najednou, pro práci s korelacemi

6.3 Korelace

V této části byl vytvořen kód, který ze stažených dat vytvoří pomocí příkazů v obrázku níže korelační tabulku. Z té lze vyčíst vzájemný vztah mezi dvěma zvolenými měnami, které se obecně označí x a y . V diagonální části tabulky se nacházejí hodnoty $+1$, což značí, že kód funguje správně, protože vzniká porovnání dvou totožných měn. Což znamená, že se vlastně porovnává veličina x a x . Pokud se hodnota x bude měnit, je pravděpodobné, že se změní i hodnota y a naopak. Nelze však posoudit, zda je jedna veličina příčina a druhá následek.

Korelace

```

l: n = len(pairs)
results = np.zeros((n, n, len(intervals)))

for fn1 in files:
    for fn2 in files:
        name1 = os.path.split(fn1)[-1].split(".")[0]
        name2 = os.path.split(fn2)[-1].split(".")[0]

        pair1, interval1 = name1.split("-")
        pair2, interval2 = name2.split("-")

        if interval1 == interval2:
            data1 = pd.read_csv(fn1)
            data2 = pd.read_csv(fn2)

            start = max(data1["timestamp"].values[0], data2["timestamp"].values[0])
            end = min(data1["timestamp"].values[-1], data2["timestamp"].values[-1])

            data1 = data1[data1["timestamp"] >= start]
            data1 = data1[data1["timestamp"] <= end]

            data2 = data2[data2["timestamp"] >= start]
            data2 = data2[data2["timestamp"] <= end]

            corr = np.corrcoef(data1["open"], data2["open"])[0,1]

            i, j, k = pairs.index(pair1), pairs.index(pair2), intervals.index(interval1)
            results[i,j,k] = corr

pretty_pairs = [pair.replace("USD", "") for pair in pairs]
for interval in intervals:
    print()
    print(interval)
    k = intervals.index(interval)
    data = results[:, :, k]

    data = np.round(data, 3)

    table_content = r"""
\begin{table}[H]
\centering
\begin{tabular}{|c|c|c|}
\hline
"""

    table_content += " & " + " & ".join(pretty_pairs) + r" \\ \hline" + "\n"

    for pair, line in zip(pretty_pairs, data[:]):
        table_content += pair + " & " + " & ".join(map(str, list(line))) + r" \\ " + "\n"

    table_content += r"""
\hline
\end{tabular}
\end{table}
"""

    print(table_content)

```

Obrázek 6.4: Kód pro vytvoření korelační tabulky a jednotlivých korelačních koeficientů

6.4 Posunutá korelace v čase

V určování pravděpodobností a statistice je křížová korelace označována jako vztah mezi dvěma náhodnými vektory X a Y . Narozdíl od korelace náhodného vektoru X , kde je vztah pouze mezi samotnou položkou X . Následně se vytvoří k tomuto prvku korelační matice X .

Křížová korelace v Pythonu je nástroj, který vypočítává míru závislosti dvou časových řad. Základem této metody je určit vhodný časový posun časové řady vůči té druhé. Informace z csv souboru jsou převedeny do datového rámce „pandas“, kde jsou pomocí nástrojů v knihovně „pandas“ připravena potřebná data. Pro vykreslení časových řad se pak využívá funkcí matplotlib. Pomocí kódu a funkce crosscorr je možno vypočítat korelace mezi dvojicí časových řad, kde jedna z nich obsahuje zpoždění.

```
import glob, os

import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
import emd
```

```
files = [
#     "BNBUSDT-1d.csv",
#     "BTCUSDT-1d.csv",
#     "ETHUSDT-1d.csv",
#     "XRPUSDT-1d.csv",
    "BNBUSDT-1h.csv",
    "BTCUSDT-1h.csv",
    "ETHUSDT-1h.csv",
    "XRPUSDT-1h.csv"
]
```

Obrázek 6.5: Kód pro vytvoření posunuté korelace - první část

```
for idx1, fn1 in enumerate(files):
    for idx2, fn2 in enumerate(files):

        name1 = os.path.split(fn1)[-1].split(".")[0]
        name2 = os.path.split(fn2)[-1].split(".")[0]

        pair1, interval1 = name1.split("-")
        pair2, interval2 = name2.split("-")

        if interval1 == interval2 and idx1 >= idx2:

            x = pd.read_csv(fn1)["open"]
            y = pd.read_csv(fn2)["open"]

            N = 1500 # 1500 pro hodiny, 30 pro dny
            corrs_L = np.zeros(N)
            corrs_P = np.zeros(N)

            corrs_L[0] = np.corrcoef(x, y)[0,1]
            corrs_P[0] = np.corrcoef(y, x)[0,1]

            for n in range(1,N):
                corrs_L[n] = np.corrcoef(x[n:], y[:-n])[0,1]
                corrs_P[n] = np.corrcoef(y[n:], x[:-n])[0,1]

            corrs = list(corrs_L[::-1]) + list(corrs_P[1:])

            plt.figure(figsize=(10, 6))
            plt.title("Korelace {} proti posunutému {} (granularita {})".format(pair1, pair2, interval1))
            plt.plot(range(-N+1, N), corrs)
            plt.axvline(0, linewidth=1, color='r')
            plt.ylim(0, 1.05)
            plt.tight_layout()
            plt.show()
```

Obrázek 6.6: Kód pro vytvoření posunuté korelace - druhá část

6.5 Fourierova analýza

Cílem v tomto druhu určování podobnosti je, vzít periodickou časovou řadu a následně ji rozložit na jednodušší periodické funkce. Z matematické věty formulované Josephem Fourierem, lze libovolnou periodickou funkci rozložit bez ohledu na to, jak složitá či jednoduchá je. Následně lze vytvořit Fourierovu řadu.

Fourierova analýza je funkce součtu sinusů a cosinusů z daných komponent, je to proces získávání spektra frekvencí $H(f)$ zahrnujících časovou řadu $h(t)$, a tím je realizována Fourierovou transformací (FT). Fourierova analýza převádí časovou řadu z původní domény na reprezentaci ve frekvenční doméně a naopak.

Metoda FT vyjádří funkci, jako součet periodických složek, následně se z těchto složek získá signál. Funkce i FT se nahrazují jejich diskretními protějšky a tím vzniká diskretní fourierova transformace (DFT). Následně se využívá výpočetní algoritmus DFT, který je charakteristický svojí rychlostí a nazývá se rychlá fourierova transformace z anglického fast fourier transform (FFT).

Na obrázku přiloženém níže, lze pozorovat kód, který byl po načtení dat z jednotlivých csv souborů využit na převod dané časové řady na jednodušší signál.

```
[18]: import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
from scipy.fft import fft, fftfreq

df = pd.read_csv('BTCUSD1-1d.csv')
print(df)

   close    high    low    open  timestamp  volume
0  9232.00  9292.00  9080.10  9138.08  1593561600000  38488.528699
1  9086.54  9261.96  8940.00  9231.99  1593648000000  45725.168076
2  9058.26  9125.00  9037.47  9086.54  1593734400000  28943.420177
3  9135.46  9190.00  9040.04  9057.79  1593820800000  26441.968484
4  9069.41  9145.24  8893.03  9135.00  1593907200000  34073.653627
..  ..  ..  ..  ..  ..  ..
296 51125.14 52131.85 47500.00 51690.95 1619136000000 132230.780719
297 50047.84 51166.22 48657.14 51110.56 1619222400000 55361.512573
298 49066.77 50567.91 46930.00 50047.84 1619308800000 58255.645004
299 54001.39 54356.62 48753.44 49066.76 1619395200000 86310.802124
300 54450.84 54933.74 53222.00 54001.38 1619481600000 18224.565617

[301 rows x 6 columns]

[20]: T = 3600
y = df["open"].values
N = len(y)

yf = fft(y)
xf = fftfreq(N, T)[:N//2]

zoom = 100
plt.plot(xf[:zoom], 2.0/N * np.abs(yf[0:N//2][:zoom]))
plt.grid()
plt.show()
```

Obrázek 6.7: Kód pro vytvoření FFT

6.6 Empirický režim rozkladu

Reprezentace signálu je implementována jako superpozice jednodušších komponent, které se nazývají funkce vnitřního režimu (IMF). Zařazují se pomocí algoritmů, které jsou založené na empirickém režimu dekompozice (EMD) a iterativním filtrování (IF). Zkratka EMD pochází z anglického názvu Empirical Mode Decomposition. Tato metoda je schopna řešit nelineární a nestacionární procesy, které vygeneruje signál. Časově-frekvenční analýza a následná extrakce funkcí může dopomoci k řešení daných problémů v mnoha oblastech výzkumu [54].

Při řešení nelineárních a nestacionárních časových řad mohou standardní moduly, jako je Fourierova transformace či vlnková transformace způsobovat lineární dekompozici. Kromě toho mají výše uvedené metody potíže s poskytováním přesné časově-frekvenční reprezentace dat kvůli již zmiňovanému Heisenbergovu principu neurčitosti [55]. Ze všech těchto důvodů, bylo navrženo několik metod ke zvýšení přesnosti časově-frekvenční reprezentace vytvořené dříve zmíněnými metodami, jako je Short Time Fourier Transform (STFT), Synchronously Squeezed Wavelet Transform nebo ConceFT metoda.

V Pythonu se pro implementaci EMD využívá balíček EMD 0.4.0, kde se pomocí příkazu `pip install emd` a dále pomocí `import emd` získávají potřebná data. Následně se provádí samotná analýza. Balíček EMD (Empirical Mode Decomposition) obsahuje funkce Pythonu pro analýzu nelineárních a nestacionárních oscilačních časových řad. EMD implementuje řadu prosévacích algoritmů, okamžitých frekvenčních transformací, konstrukce výkonového spektra a analýzy jednoho cyklu funkcí. Tyto implementace jsou podporovány online dokumentací obsahující řadu praktických návodů.

```
n = 2
plt.plot(imf[:,n])
plt.show()

plt.plot(IF[:,n])
plt.show()
```

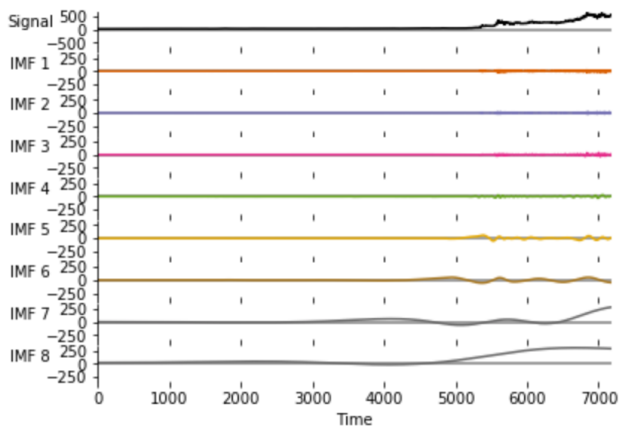
Obrázek 6.8: Kód pro vytvoření EMD signálů - část první

```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
import emd

df = pd.read_csv('BNBUSDT-1d.csv')
print(df)
```

Obrázek 6.9: Kód pro vytvoření EMD signálů - část druhá

```
x = df["open"].values
imf = emd.sift.sift(x)
emd.plotting.plot_imfs(imf, scale_y=True, cmap=True)
```



```
IP, IF, IA = emd.spectra.frequency_transform(imf, sample_rate, 'hilbert')
```

Obrázek 6.10: Kód pro vytvoření EMD signálů - část třetí

6.7 Získané výsledky

V této části budou uvedeny výsledky čtyř používaných metod k analýze podobnosti vývoje cen vybraných kryptoměn. Jmenovitě byly v této práci použity 4 typy: korelace, posunutá korelace, fourierova analýza a empirický režim rozkladu.

6.7.1 Posunutá korelace

Pomocí metody posunuté korelace byly zaznamenány podobnosti vývoje u etherea a bitcoinu. Kryptoměna XRP se svým vývojem úplně odlišovala od ostatních jak v denní tak v hodinové granularitě. Je vysoce pravděpodobné, že tato metoda je pro daný typ dat nevyhovující díky tomu, že se křížová korelace využívá hlavně u oscilujících dat. Data používaná v této práci jsou ve většině případů rostoucí hlavně díky rapidnímu nárůstu hodnoty kryptoměn.

V příloze jsou uvedeny grafy jednotlivých kryptoměn na sebe navázaných s posunutím 1500 hodin a následně 30 dní. Dále byla vytvořena autokorelace jednotlivých měn, kde je možné pozorovat podobný vývoj bitcoinu a etherea. Znamená to, že když křivka v grafu u bitcoinu roste, tak u etherea také a naopak.

6.7.2 Korelace

V této části práce byly vytvořeny grafy jednotlivých kryptoměn a následně byly pomocí kódu, který byl napsán v programovacím jazyce Python, vypočítány jednotlivé korelační koeficienty. Výsledky je možno nalézt v tabulce níže. Na diagonální ose lze pozorovat koeficienty o hodnotě jedna, což ověřuje správnost zvolené metody, protože byly proti sobě položeny dvě totožné měny a data proto musí mít korelační koeficient stoprocentní.

Tabulka 6.1: Tabulka korelačních koeficientů

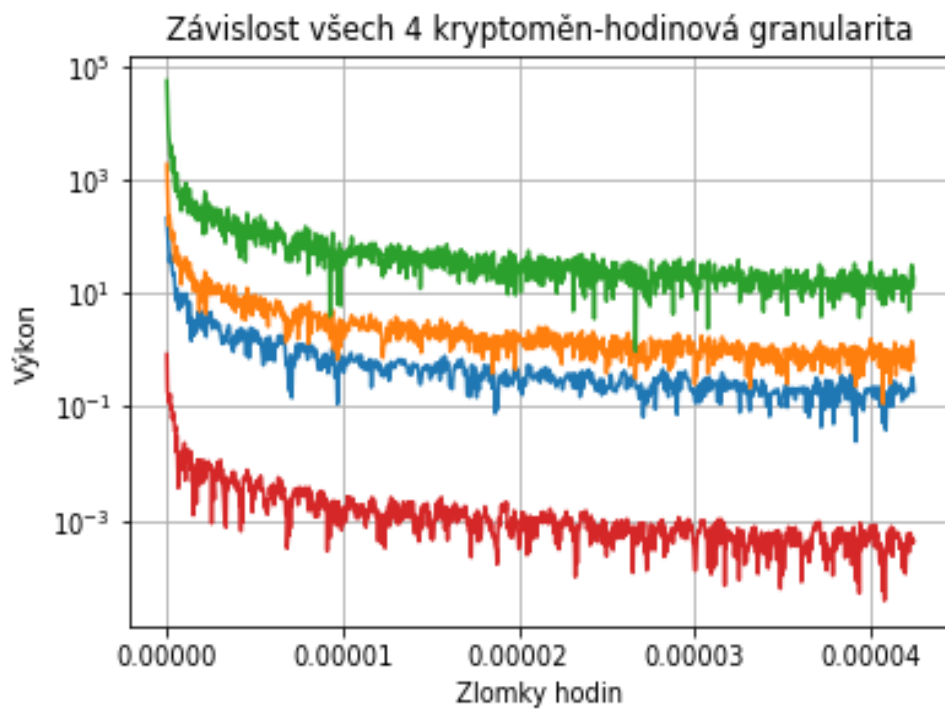
	XRP	BNB	ETH	BTC
XRP	1.0	0.857	0.725	0.665
BNB	0.857	1.0	0.868	0.843
ETH	0.725	0.868	1.0	0.977
BTC	0.665	0.843	0.977	1.0

6.7.3 Fourierova analýza

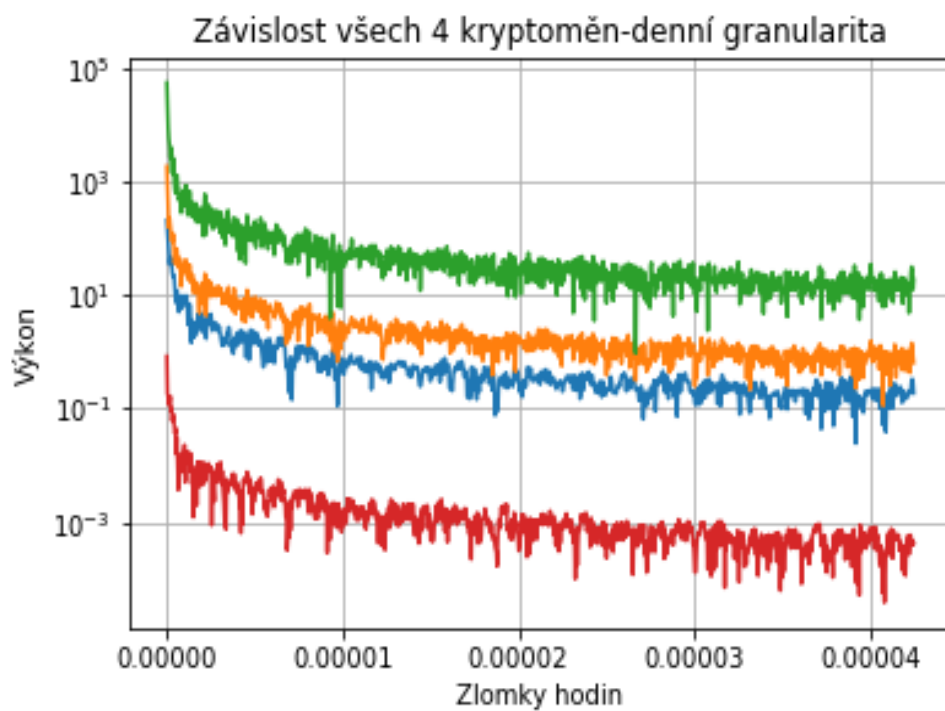
V části, kde byla data převáděna na signál pomocí rychlé fourierovy transformace byla zkoumaná suma výkonu. Z grafů, které byly vygenerované v Pythonu lze pozorovat dané chování:

- Pro kryptoměnu s názvem **XRP**, kde byla vstupní data vygenerovaná po jednotlivých **hodinách** v určitém časovém úseku, byla suma výkonu pro frekvence v rozmezí od 0.00002 Hz do 0.00004 Hz vyčtena z grafu jako: $10^{-2.92}$ až $10^{-4.42}$.
- Pro kryptoměnu s názvem **XRP**, kde byla vstupní data vygenerovaná po jednotlivých **dnech** v určitém časovém úseku, byla suma výkonu pro frekvence v rozmezí od 0.00002 Hz do 0.00004 Hz vyčtena z grafu jako: $10^{-1.6}$ až $10^{-3.8}$.
- Pro kryptoměnu s názvem **Binance coin**, kde byla vstupní data vygenerovaná po jednotlivých **hodinách** v určitém časovém úseku, byla suma výkonu pro frekvence v rozmezí od 0.00002 Hz do 0.00004 Hz vyčtena z grafu jako: $10^{-0.3}$ až $10^{-1.64}$.
- Pro kryptoměnu s názvem **Binance coin**, kde byla vstupní data vygenerovaná po jednotlivých **dnech** v určitém časovém úseku, byla suma výkonu pro frekvence v rozmezí od 0.00002 Hz do 0.00004 Hz vyčtena z grafu jako: 10^1 až 10^0 .
- Pro kryptoměnu s názvem **Ethereum**, kde byla vstupní data vygenerovaná po jednotlivých **hodinách** v určitém časovém úseku, byla suma výkonu pro frekvence v rozmezí od 0.00002 Hz do 0.00004 Hz vyčtena z grafu jako: $10^{0.5}$ až $10^{-0.66}$.
- Pro kryptoměnu s názvem **Ethereum**, kde byla vstupní data vygenerovaná po jednotlivých **dnech** v určitém časovém úseku, byla suma výkonu pro frekvence v rozmezí od 0.00002 Hz do 0.00004 Hz vyčtena z grafu jako: $10^{1.61}$ až 10^1 .
- Pro kryptoměnu s názvem **Bitcoin**, kde byla vstupní data vygenerovaná po jednotlivých **hodinách** v určitém časovém úseku, byla suma výkonu pro frekvence v rozmezí od 0.00002 Hz do 0.00004 Hz vyčtena z grafu jako: $10^{1.8}$ až 10^0 .
- Pro kryptoměnu s názvem **Bitcoin**, kde byla vstupní data vygenerovaná po jednotlivých **dnech** v určitém časovém úseku, byla suma výkonu pro frekvence v rozmezí od 0.00002 Hz do 0.00004 Hz vyčtena z grafu jako: $10^{2.95}$ až $10^{2.6}$.

Z daných grafů uvedených níže lze vyčíst, že křivky jsou si podobné svojí strukturou, avšak nelze nalézt žádné dominantní frekvence, které by společně vystupovaly a pojednávaly o nějaké vyšší korelaci jak v hodinové tak denní granularitě.



Obrázek 6.11: Rychlá fourierova transformace vybraných kryptoměn, hodinová granularita



Obrázek 6.12: Rychlá fourierova transformace vybraných kryptoměn, denní granularita

6.7.4 Empirický režim rozkladu

Jako čtvrtá metoda určování podobnosti vybraných kryptoměn byl zvolen empirický režim rozkladu. Princip této metody spočívá v rozkladu 1-D algoritmu do vícerozměrného signálu. Pomocí sepsaného kódu uvedeného výše se spustily iterace, které opakovaly předepsané kroky, dokud nebyly průměrné obálky dostatečně blízko nule. Tento proces byl implementován pomocí `emd.sift.sift`.

U kryptoměn ethereum a bitcoin byla zaznamenána na koncích těchto signálů podobnost ve většině iterací, jak u hodinových tak denních granularit. XRP opět nevykazovalo nijak zvlášť výraznou podobnost s dalšími vybranými vzorky, a to hlavně u denní granularity. BNB mělo podobně jako ethereum a bitcoin rostoucí trend na konci zkoumaného iteračního signálu.

Závěr

Cílem této práce bylo vytvořit kód pro určení podobnosti cen pro vybrané kryptoměny. V první části práce byly popsány teoretické aspekty týkající se kryptoměn a datové analýzy. Následně bylo popsáno uživatelské rozhraní programovacího jazyka Python a jeho uživatelských balíčků.

V praktické části byl následně vytvořen kód pro výpočet korelací mezi jednotlivými kryptoměnami. Tyto výsledky byly vloženy do korelační tabulky.

Dále byla v této práci zkoumána metoda posunuté korelace. Z výsledků lze říci, že je tato metoda pro zkoumané vzorky nevyhovující. Především díky aspektům, že budoucí cena kryptoměn neustále roste a trend vývoje křivky je tedy čím dál více narůstající. Posunutou korelaci je vhodné využívat u dat, která oscilují a snižují i zvyšují své hodnoty.

Analýza zkoumaná pomocí Rychlé Fouriérové transformace sice zaznamenala podobné výchyly v signálu a jejich strukturu. Bližší podobnosti však nebyly vypořovány ani u jedné z vybraných granularit.

Čtvrtá metoda určování podobnosti vybraných kryptoměn, konkrétně empirický režim rozkladu neukázal větší podobnosti. V iterativním rozkladu signálu vnitřní složky IMF byly frekvence uspořádány od nejvyšších po nejnižší. Podobný vývoj tohoto rozkladu bylo možno pozorovat u etherea, bitcoinu a binance coinu. Naopak XRP se opět výrazně lišilo. Všechny kryptoměny, ale zaznamenaly u zkoumaného iteračního signálu rostoucí trend.

Bibliografie

- [1] Jeffrey S. Abarbanell, Brian J. Bushee. "Fundamental Analysis, Future Earnings, and Stock Prices." Accounting Research Center, Booth School of Business, University of Chicago, Wiley, 1997. ISBN 00218456, 1475679X.
- [2] Suresh a.s. "A study on fundamental and technical analysis." International Journal of Marketing, Financial Services & Management Research, 2013.
- [3] Yang Bing, Jian Kun Hao, Si Chang Zhang. "Stock Market Prediction Using Artificial Neural Networks." Advanced Engineering Forum, 2012.
- [4] Burton G. Malkiel. "Efficient Market Hypothesis." Palgrave Macmillan, London. ISBN 978-1-349-20213-3.
- [5] *Here's What Warren Buffett Thinks About The Efficient Market Hypothesis - Business Insider.* <https://www.businessinsider.com/warren-buffett-on-efficient-market-hypothesis-2010-12>. (4.1.2021).
- [6] Chun-Xiao Nie: Correlation dynamics in the cryptocurrency market based on dimensionality reduction analysis. School of Statistics and Mathematics, Zhejiang Gongshang University, Hangzhou 310018, China. (55.4.2020)
- [7] Darko Stotic, Dusan Stotic, Teresa B. Ludermir, Tatijana Stoti. "Collective behavior of cryptocurrency price changes." Centro de Informática, Universidade Federal de Pernambuco, Av. Luiz Freire s/n, 50670-901, Recife, PE, Brazil.
- [8] Konstantinos Gillas, Stelios Bekiros, Costas Skriopoulos. "Extreme Correlation in Cryptocurrency Markets." Departamento de Estatística e Informática, Universidade Federal Rural de Pernambuco, Rua Dom Manoel de Medeiros s/n, Dois Irmãos, 52171-900, Recife, PE, Brazil.
- [9] *Forex Trading: A Beginner's Guide.* <https://www.investopedia.com/articles/forex/11/why-trade-forex.asp>. (16.1.2021).
- [10] Robert J. Barro: "The Stock Market and Investment. The Review of Financial Studies, Volume 3, Issue 1, January 1990, Pages 115–131.
- [11] *Stock.* <https://www.investopedia.com/terms/s/stock.asp>. (20.1.2021).

- [12] *NFTs, explained.*
<https://www.theverge.com/22310188/nft-explainer-what-is-blockchain-crypto-art-faq>. (20.4.2021)
- [13] *Digitální koláž za 1,5 miliardy korun aneb co je nový fenomén zvaný NFT a proč teď hýbe nejen kryptosvět.*
<https://www.czechcrunch.cz/2021/03/digitalni-kolaz-za-15-miliardy-korun-aneb-co-je-novy-fenomen-zvany-nft-a-proc-ted-hybe-nejen-kryptosvetem/> (3.5.2021)
- [14] *PredictIt.*
<https://www.predictit.org>. (5.5.2021)
- [15] *Prediction market.*
<https://en.wikipedia.org/wiki/Predictionmarket>. (5.5.2021)
- [16] Petr Tesař. "Kryptoměny". Vysoká škola ekonomická v Praze, 2018. ISBN: 978-80-7400-722-4.
- [17] *Cryptocurrency Price Prediction Using Deep Learning.*
<https://towardsdatascience.com/cryptocurrency-price-prediction-using-deep-learning-70cfca50dd3a>. (5.5.2021).
- [18] *Cryptocurrency.*
<https://www.investopedia.com/terms/c/cryptocurrency.asp> (5.1.2021).
- [19] *Global Charts, CoinMarketCap.*
<https://coinmarketcap.com/charts/>. (27. 2. 2020).
- [20] *Blockchain.*
<https://www.blockchain.com>. (20.3.2021).
- [21] *Blockchain.*
<https://www.investopedia.com/terms/b/blockchain.asp>. (3.3.2021).
- [22] *Peer-to-Peer (P2P) Economy,*
<https://www.investopedia.com/terms/p/peertopeer-p2p-economy.asp>. (14.1.2021).
- [23] *Největší firmy světa v roce 2020 – 2021, dle jejich tržní kapitalizace.*
<https://akciovník.cz/nejvetsi-firmy-sveta-v-roce-2020-2021-dle-jejich-trzni-kapitalizace-top-100/> (10.4.2020).
- [24] *Global Cryptocurrency Charts*
<https://coinmarketcap.com/charts/>(5.4.2021).
- [25] *Bitcoin - aktuální a historické ceny kryptoměny Bitcoin, graf vývoje ceny kryptoměny Bitcoin.*
<https://www.kurzy.cz/komodity/bitcoin-graf-vyvoje-ceny/>. (5.2.2021).

- [26] *Ethereum*.
<https://ethereum.org/en/what-is-ethereum/>. (6.2.2021).
- [27] *Litecoin*.
<https://litecoin.com/en/individuals>. (6.2.2021).
- [28] *Binance*.
<https://www.binance.com/en>. (6.2.2021).
- [29] Ikuya Takashima: "Ripple: The Ultimate Guide to the World of Ripple XRP, Ripple Investing, Ripple Coin, Ripple Cryptocurrency, Cryptocurrency."Space Independent Publishing Platform, 2018. ISBN 9781986181617
- [30] *Ripple Price Index*.
<https://cointelegraph.com/xrp-price-index>. (3.1.2021).
- [31] *Dogecoin*.
<https://www.coindesk.com/price/dogecoin>. (4.4.2021).
- [32] *Dogecoin*.
<https://www.coinbase.com/price/dogecoin>. (4.4.2021).
- [33] *What Is Dogecoin, How to Say It, and Why It's No Longer a Joke*.
<https://www.wsj.com/articles/what-is-dogecoin-how-to-say-it-and-why-its-no-longer-a-joke-thanks-elon-11612820776>. (20.4.2021).
- [34] *What are the characteristics of data analysis?*
<https://www.quora.com/What-are-the-characteristics-of-data-analysis>. (24.4.2021).
- [35] *Analýza dat*.
<https://cs.wikipedia.org/wiki/Analýzadat> (8.4.2021).
- [36] *Datová analýza*.
<https://cw.fel.cvut.cz/old/media/courses/a3m33ui/prednasky/ui-02-dectrees.pdf>. (3.4.2021).
- [37] *Rozdíly mezi supervizovaným učením a nepozorovaným učením*.
<https://cs.weblogographic.com/differences-between-supervised-learning-and-unsupervised-learning-15856>. (6.4.2021).
- [38] *Algoritmy strojového učení*.
<https://azure.microsoft.com/cs-cz/overview/machine-learning-algorithms/overview>. (1.3.2021).
- [39] *Neuronové sítě a jejich využití*
<http://www.statsoft.cz/file1/PDF/Statsoftneuronovesite.pdf> (17. 3. 2021).

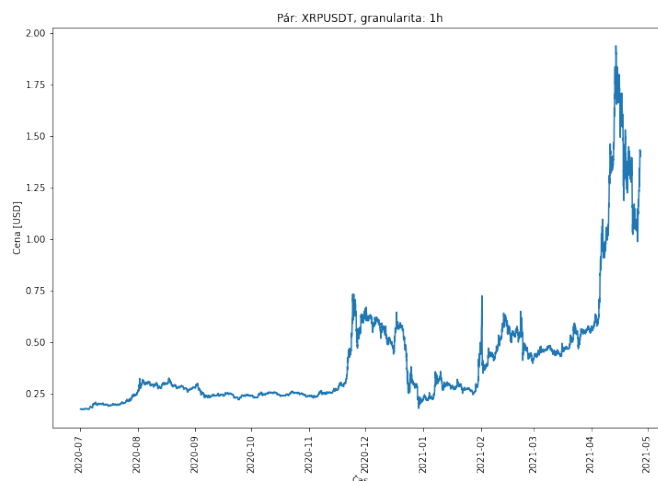
- [40] Terrence L. Fine. ”*Feedforward Neural Network Methodology.*” Springer Science Business Media, 2006. ISBN 0387226494, 9780387226491.
- [41] *McCulloch–Pitts Neuron — Mankind’s First Mathematical Model Of A Biological Neuron.*
<https://towardsdatascience.com/mcculloch-pitts-model-5fdf65ac5dd1>. (5.5.2021).
- [42] *Dopředné neuronové sítě.*
<https://portal.matematickabiologie.cz/index.php?pg=analyza-a-hodnoceni-biologicky-ch-dat-umela-intelligence-neuronove-site-perceptrony-dopredne-neuronove-site>. (10.3.2021).
- [43] *What is Perceptron: A Beginners Tutorial for Perceptron.*
<https://www.simplilearn.com/what-is-perceptron-tutorial>. (19.3.2021).
- [44] *Úvod do regresní analýzy.*
<http://www.statsoft.cz/file1/PDF.pdf> (19.3.2021)
- [45] *Regresní analýza.*
<https://meloun.upce.cz/docs/lecture/chemometrics/slidy/65model.pdf>. (5.4.2021).
- [46] *Correlation.*
<https://www.investopedia.com/terms/c/correlation.asp>. (3.4.2021).
- [47] *Correlation and regression.*
<https://www.bmj.com/about-bmj/resources-readers/publications/statistics-square-one/11-correlation-and-regression>. (3.4.2021).
- [48] *Pearsonův korelační koeficient.*
<https://mathstat.econ.muni.cz/media/12657/pearcor.pdf>. (4.4.2021).
- [49] *Spearman’s Rank-Order Correlation.*
<https://statistics.laerd.com/statistical-guides/spearman-rank-order-correlation-statistical-guide.php>. (6.4.2021).
- [50] *Time Granularity.*
<https://www.sciencedirect.com/topics/computer-science/granularity>. (5.5.2021).
- [51] *Cross-correlation.*
<https://en.wikipedia.org/wiki/Cross-correlation> (5.5.2021).
- [52] *Fourier Transforms With scipy.fft: Python Signal Processing.*
<https://realpython.com/python-scipy-fft/> (20.4.2021).
- [53] *Frequency and the Fast Fourier Transform.*
<https://www.oreilly.com/library/view/elegant-scipy/9781491922927/ch04.html>. (7.5.2021).

- [54] *Hilbert–Huang transform*.
<https://en.wikipedia.org/wiki/HilbertHuangtransform>. (12. 7. 2021).
- [55] *NCL: EEMD: Extended Empirical Mode Decomposition*
<https://www.ncl.ucar.edu/Applications/eemd.shtml>. (12. 7. 2021).
- [56] Tai-Yi Chu, Wen-Cheng Huang: *Application of Empirical Mode Decomposition Method*. Hushan Reservoir in Taiwan
- [57] *Fourier Transforms*.
<https://docs.scipy.org/doc/scipy/reference/tutorial/fft.html>. (17.4.2021).
- [58] Wes McKinney. *Python for Data Analysis*.
O'Reilly Media, Inc., 2013. ISBN : 9781449319793.
- [59] *NumPy, SciPy, and Pandas: Correlation With Python*.
<https://realpython.com/numpy-scipy-pandas-correlation-python/pearson-correlation-coefficient>. (16.4.2021).
- [60] *EMD: Empirical Mode Decomposition and Hilbert-Huang Spectral Analyses in Python*.
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7610596/>. (12. 7. 2021).

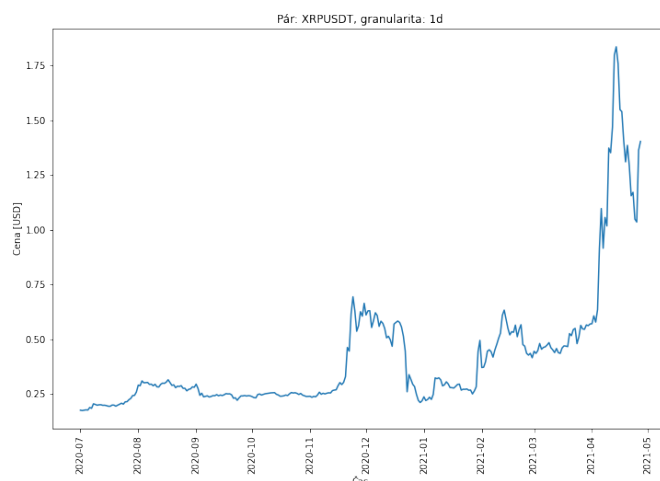
Přílohy

A První příloha

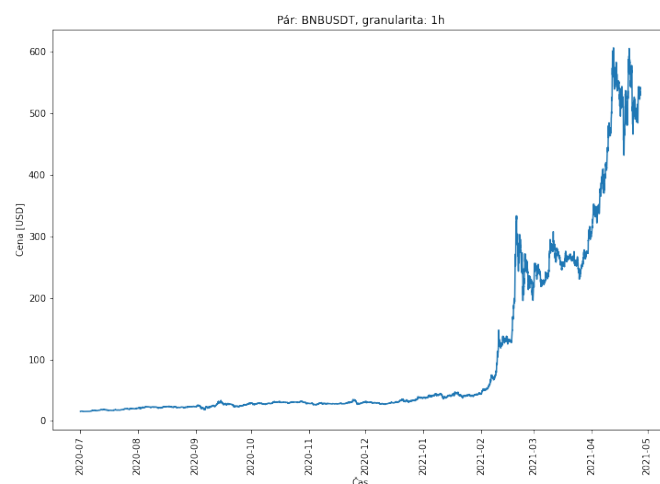
V této části jsou uvedeny grafy vývoje vybraných kryptoměn, a to konkrétně XRP, Binance coinu, Ethera a Bitcoinu vůči americkému dolaru. Z těchto příloh je možno vyčíst hodinové a denní záznamy zvolených kryptoměn v rozmezí od července roku 2020, do poloviny května 2021.



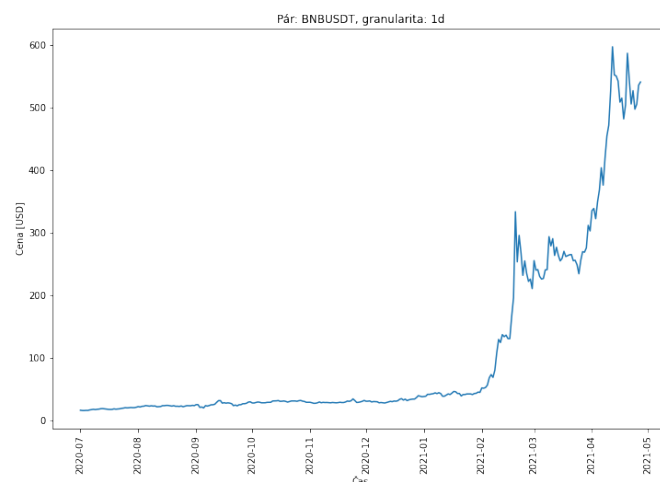
Obrázek 13: Závislost XRP na dolaru-hodinové záznamy



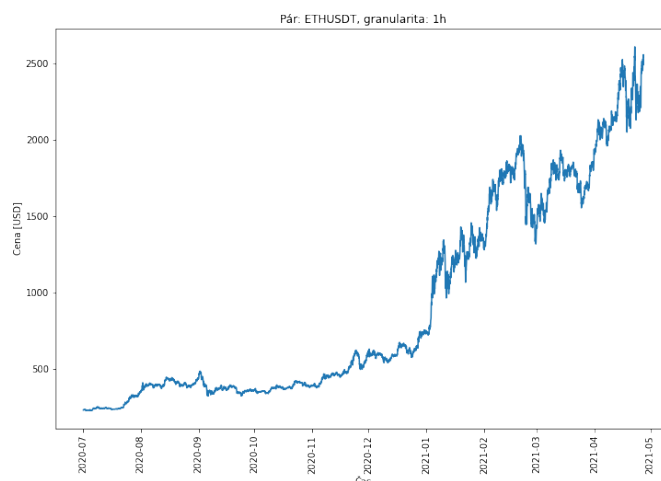
Obrázek 14: Závislost XRP na dolaru-denní záznamy



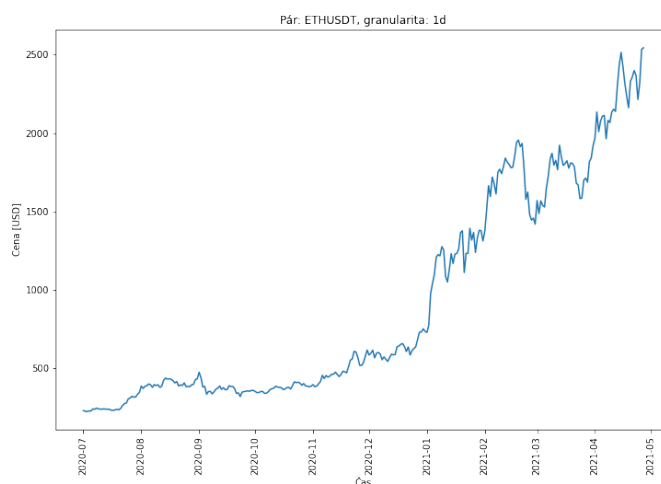
Obrázek 15: Závislost BNB na dolaru-hodinové záznamy



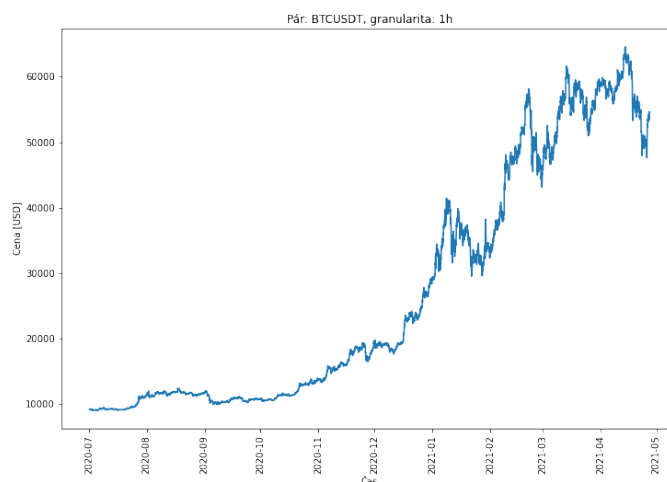
Obrázek 16: Závislost BNB na dolaru-denní záznamy



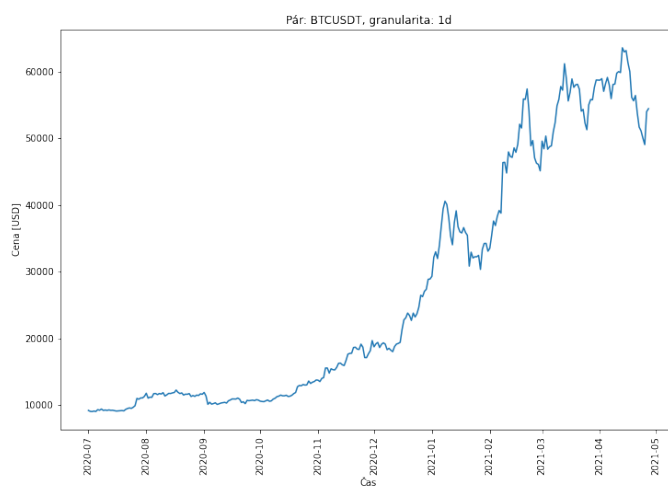
Obrázek 17: Závislost ETH na dolaru-hodinové záznamy



Obrázek 18: Závislost ETH na dolaru-denní záznamy



Obrázek 19: Závislost BTC na dolaru-hodinové záznamy

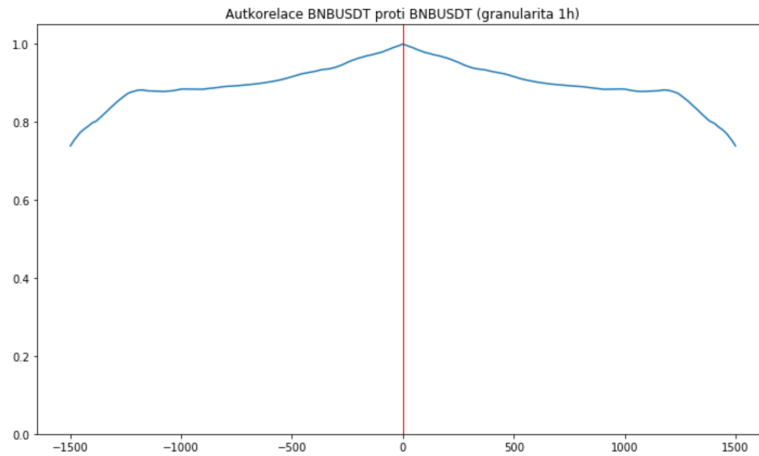


Obrázek 20: Závislost BTC na dolaru-denní záznamy

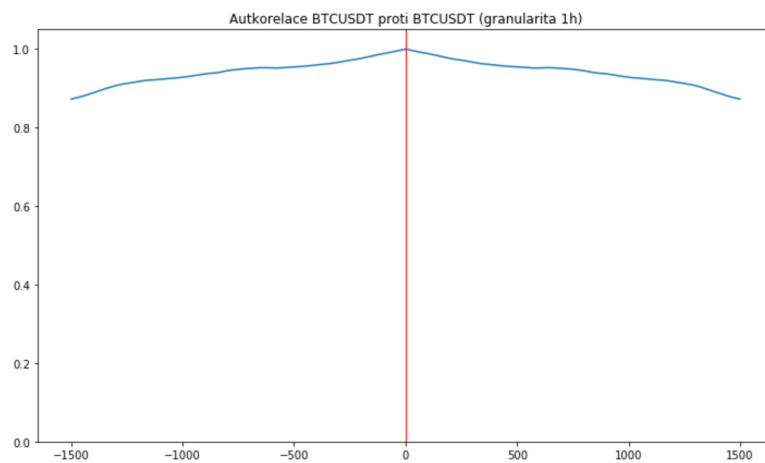
B Druhá příloha

B.1 Posunutá korelace - hodinová granularita

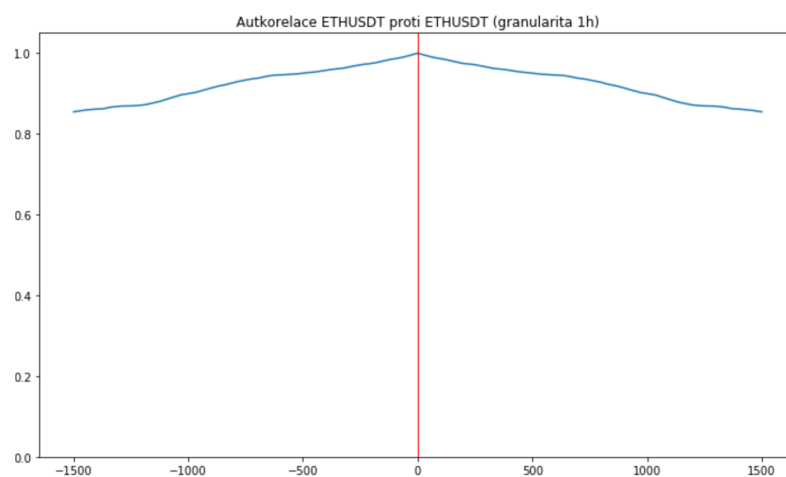
V této části přílohy je možno v grafech s hodinovou granularitou pozorovat posunutou korelaci o 1500 hodin, což je v přepočtu na dny přibližně 60 dnů. Jak je možné pozorovat v grafech níže, kryptoměna bitcoin a ethereum si jsou velmi podobné a ve stejných časových jednotkách mají podobný vývoj (pokles, růst). Naopak XRP má naprosto odlišný vývoj pozorovaného trendu.



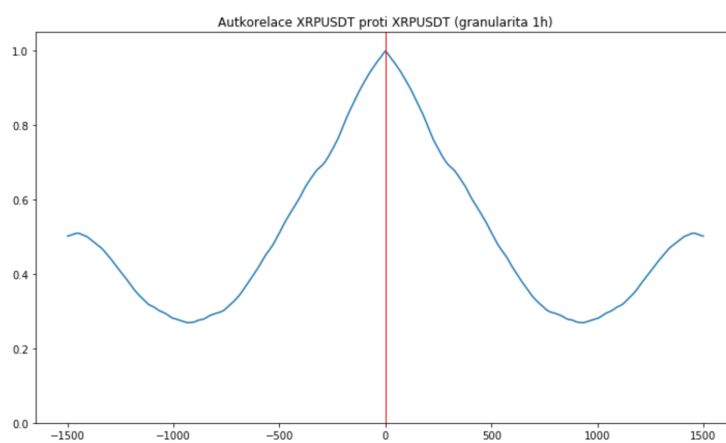
Obrázek 21: Autokorelace kryptoměny BNB na BNB, hodinová granularita



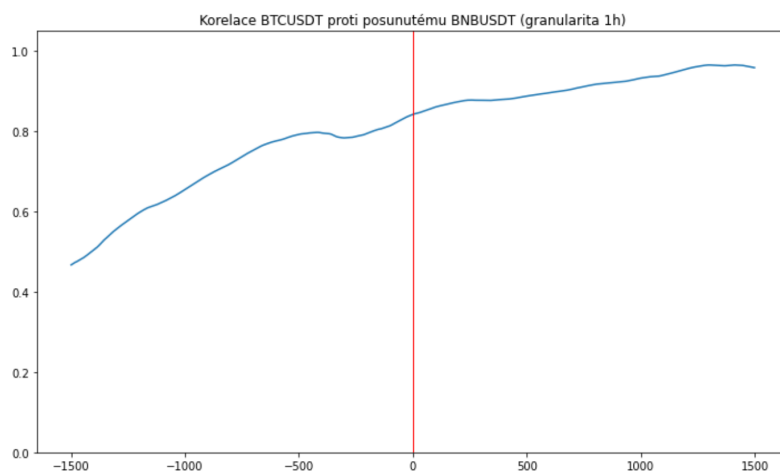
Obrázek 22: Autokorelace kryptoměny BTC na BTC, hodinová granularita



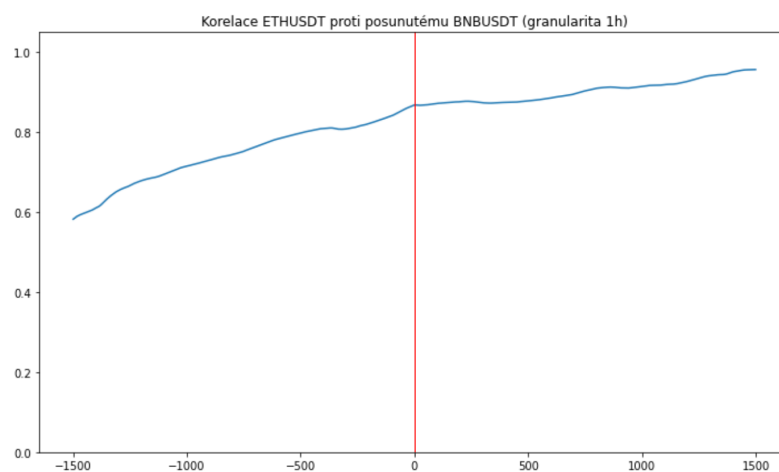
Obrázek 23: Autokorelace kryptoměny ETH na ETH, hodinová granularita



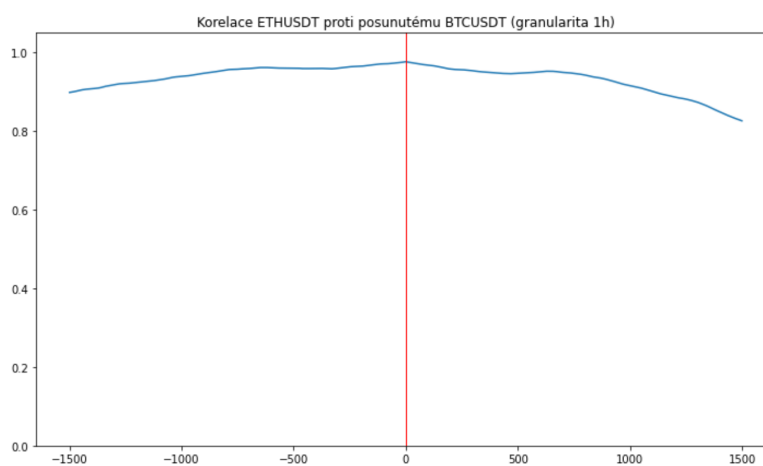
Obrázek 24: Autokorelace kryptoměny XRP na XRP, hodinová granularita



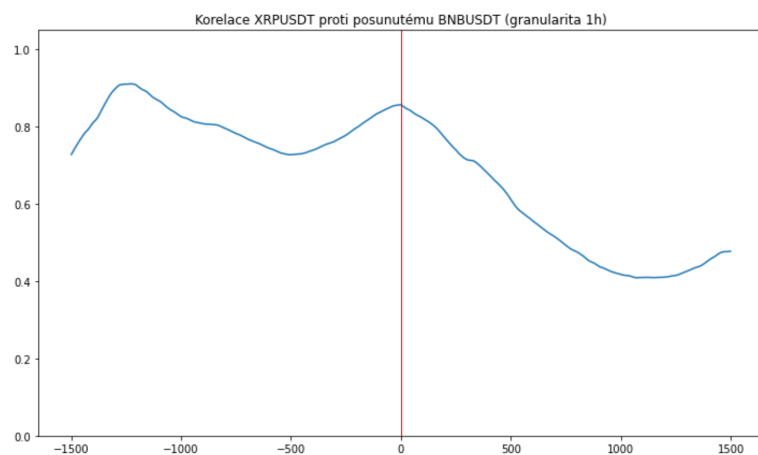
Obrázek 25: Společná posunutá korelace BTC a BNB, hodinová granularita



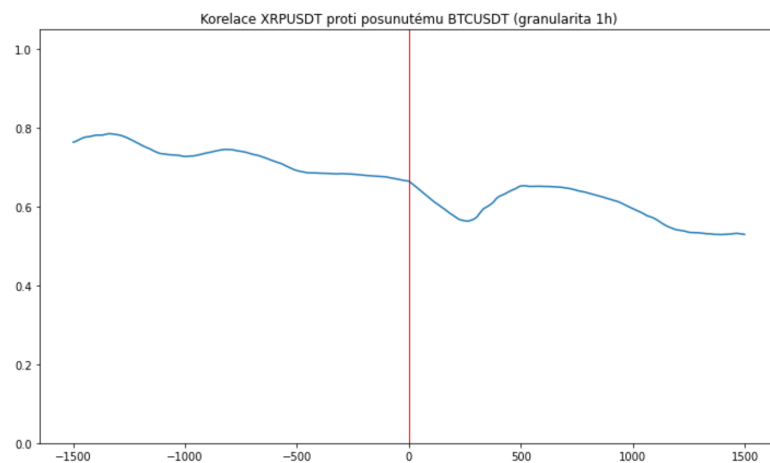
Obrázek 26: Společná posunutá korelace ETH a BNB, hodinová granularita



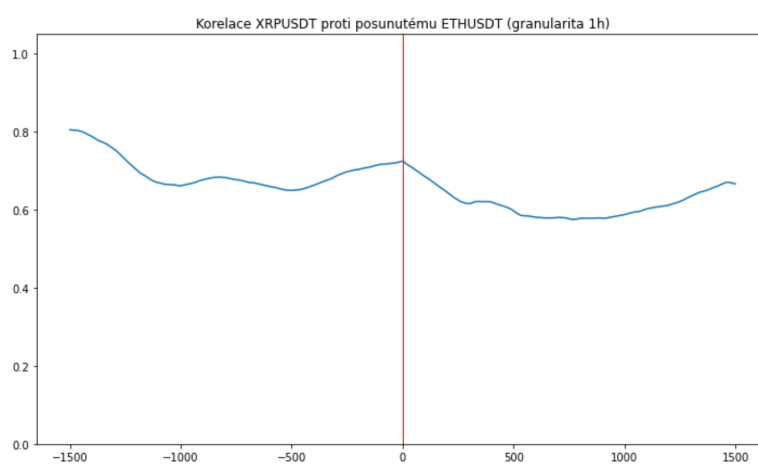
Obrázek 27: Společná posunutá korelace ETH a BTC, hodinová granularita



Obrázek 28: Společná posunutá korelace XRP a BNB, hodinová granularita



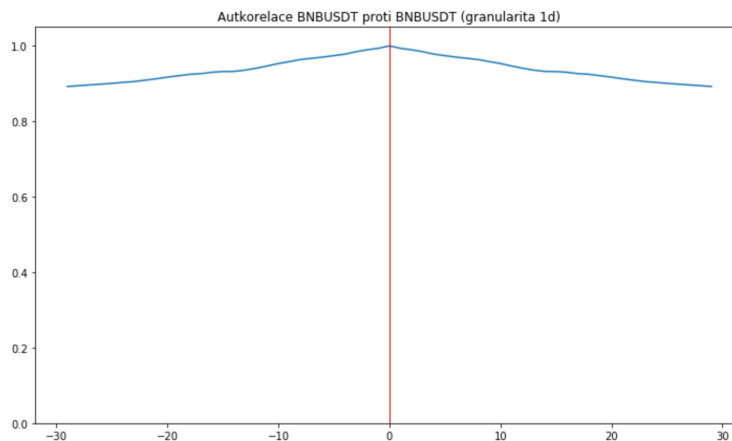
Obrázek 29: Společná posunutá korelace XRP a BTC, hodinová granularita



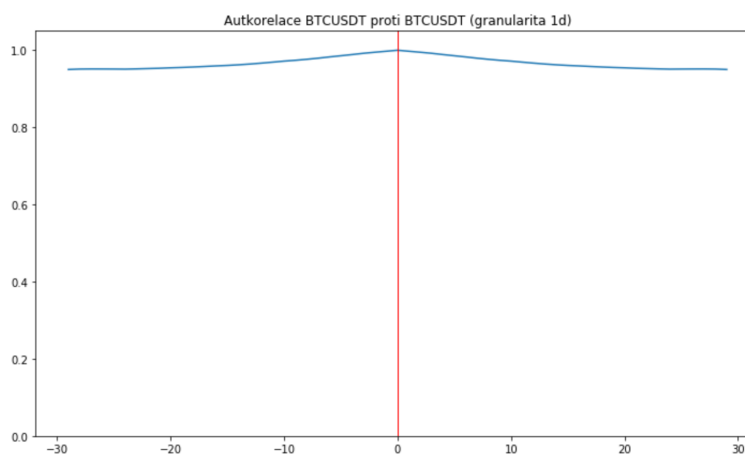
Obrázek 30: Společná posunutá korelace XRP a ETH, hodinová granularita

B.2 Posunutá korelace - denní granularita

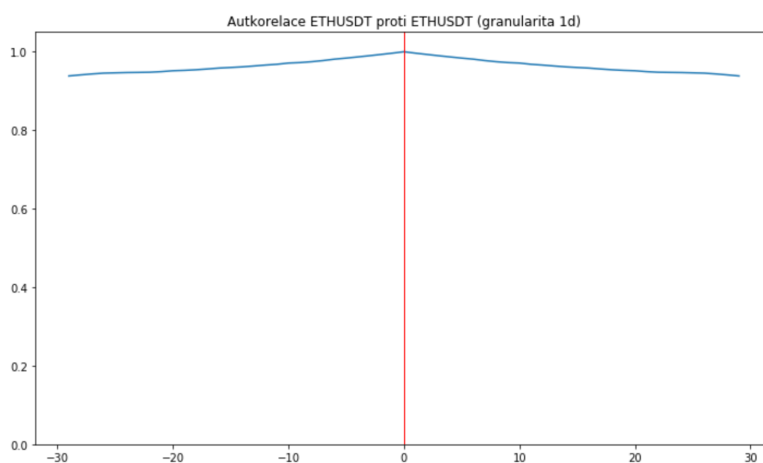
V této příloze lze nalézt grafy, kde byla za danou granularitu určena velikost jeden den. Posunutí bylo nastaveno na 30 dnů. Jako u hodinové granularity, lze i v této části práce pozorovat velmi podobný vývoj kryptoměn ethereum a bitcoin. Poklesne-li bitcoin, klesá i ethereum, naopak při růstu jedné kryptoměny, narůstá i druhá. Naopak XRP se vyvíjelo nezávisle na ostatních kryptoměnách.



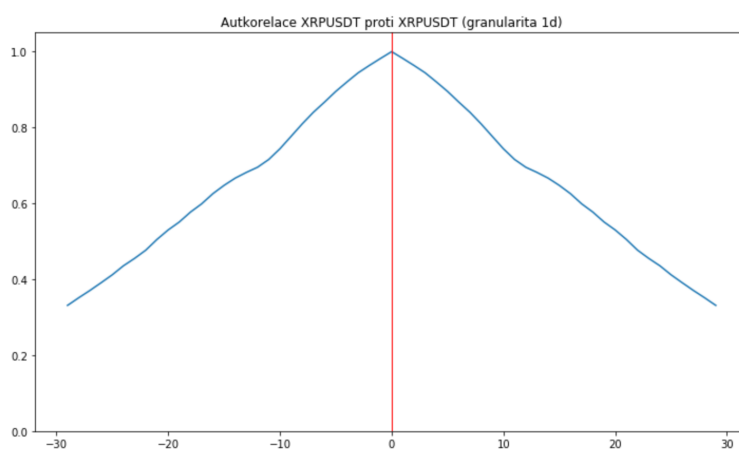
Obrázek 31: Autokorelace kryptoměny BNB na BNB, denní granularita



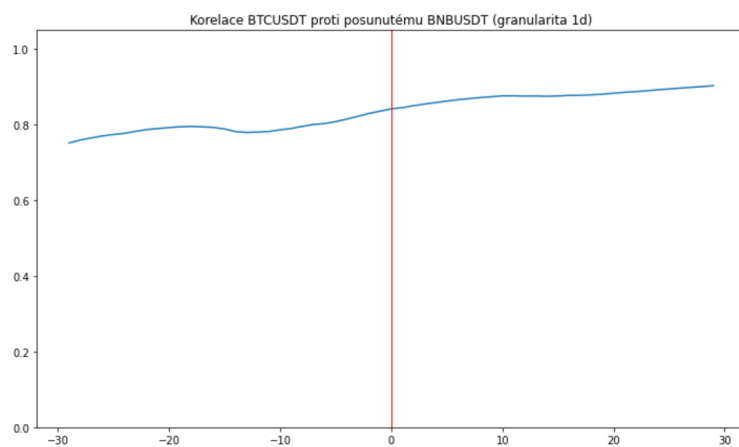
Obrázek 32: Autokorelace kryptoměny BTC na BTC, denní granularita



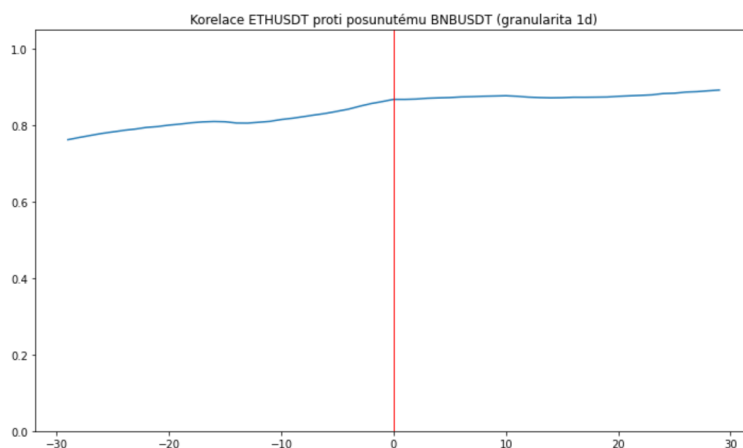
Obrázek 33: Autokorelace kryptoměny ETH na ETH, denní granularita



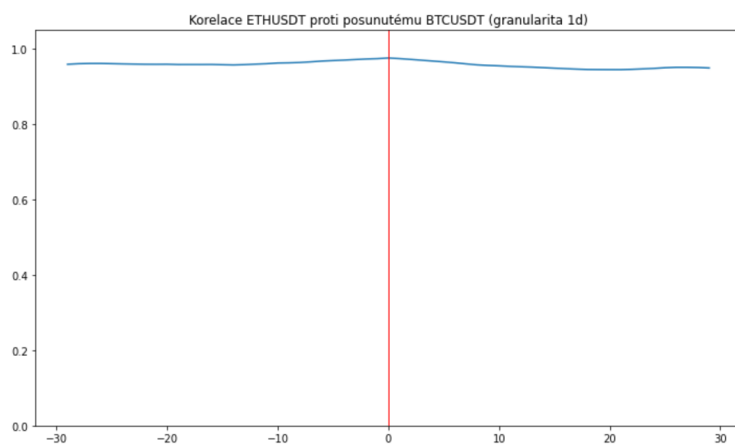
Obrázek 34: Autokorelace kryptoměny XRP na XRP, denní granularita



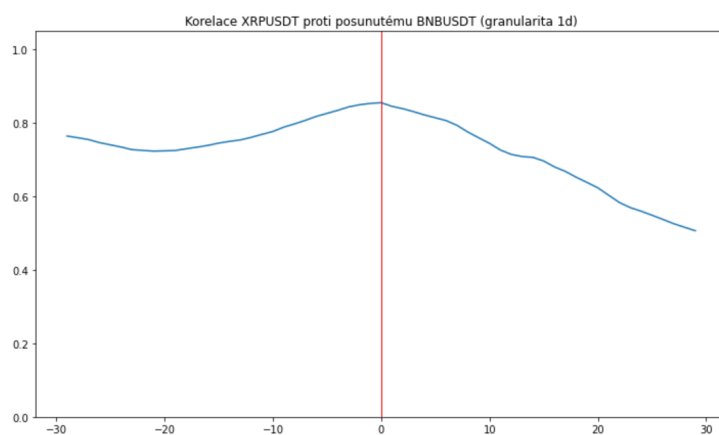
Obrázek 35: Společná posunutá korelace BTC a BNB, denní granularita



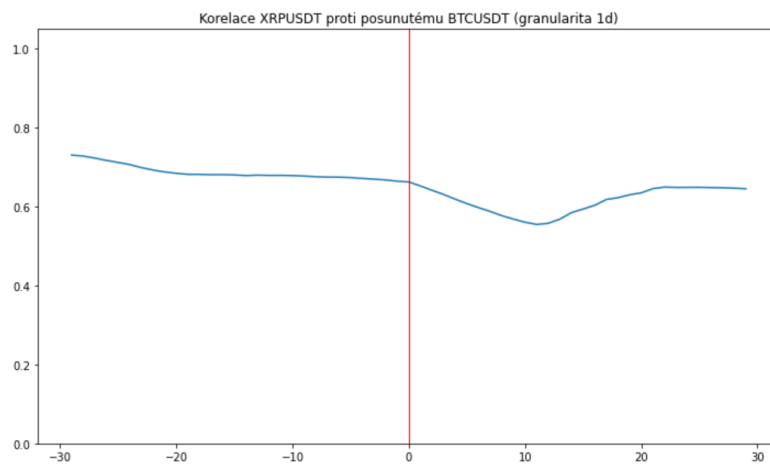
Obrázek 36: Společná posunutá korelace ETH a BNB, denní granularita



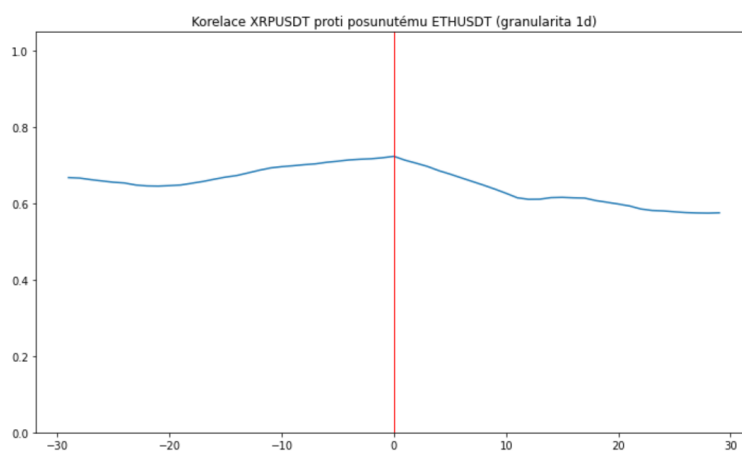
Obrázek 37: Společná posunutá korelace ETH a BTC, denní granularita



Obrázek 38: Společná posunutá korelace XRP a BNB, denní granularita



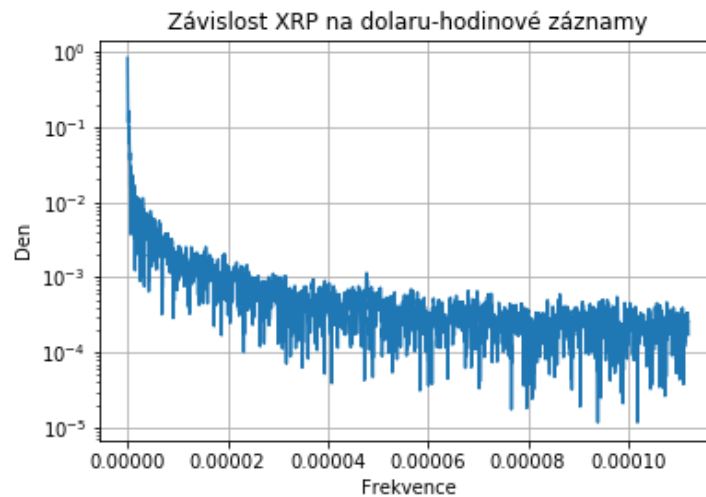
Obrázek 39: Společná posunutá korelace XRP a BTC, denní granularita



Obrázek 40: Společná posunutá korelace XRP a ETH, denní granularita

C Třetí příloha

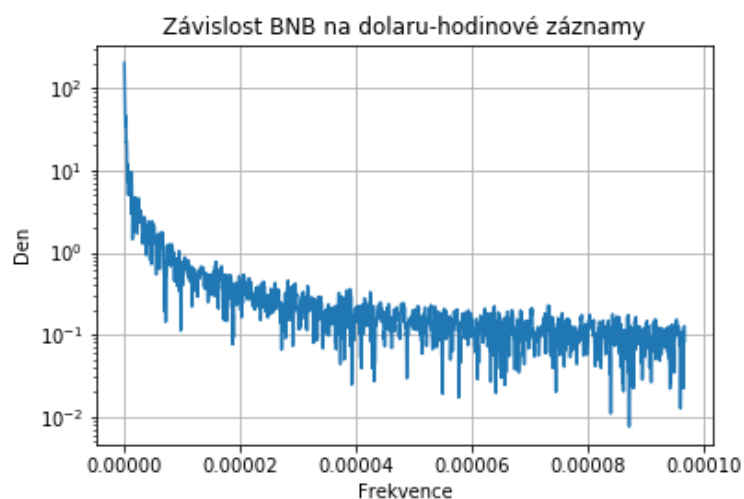
V této části jsou uvedeny grafy vývoje ceny pomocí Fourierovy transformace. Probíhá zde analýza signálu. Programovací jazyk Python používá k této operaci modul `scipy.fft`. Pro lepší čitelnost grafů se následně osa y převádí do logaritmických souřadnic.



Obrázek 41: Rychlá fourierova transformace XRP s hodinovými záznamy



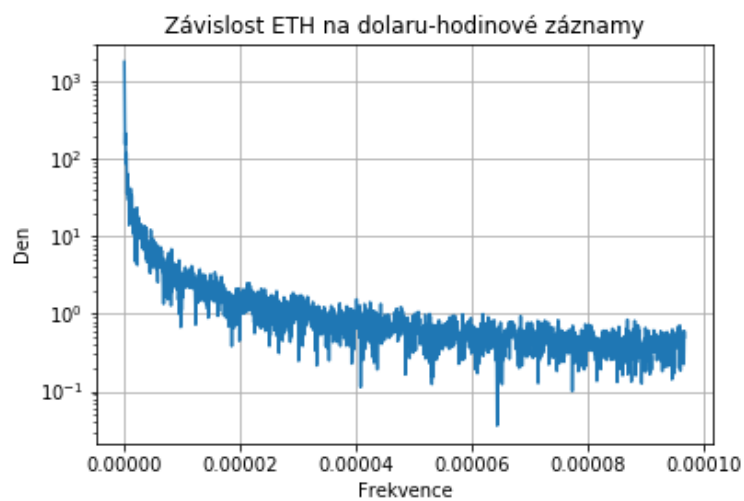
Obrázek 42: Rychlá fourierova transformace XRP se záznamy jednotlivých dní



Obrázek 43: Rychlá fourierova transformace BNB se záznamy jednotlivých hodin



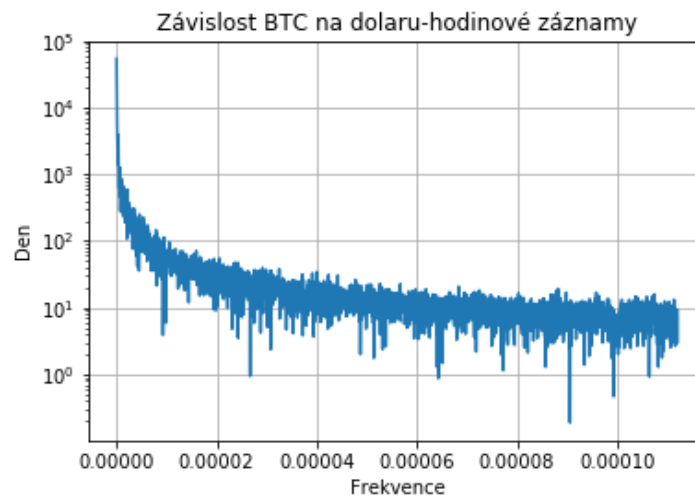
Obrázek 44: Rychlá fourierova transformace BNB se záznamy jednotlivých dní



Obrázek 45: Rychlá fourierova transformace ETH se záznamy jednotlivých hodin



Obrázek 46: Rychlá fourierova transformace ETH se záznamy jednotlivých dní



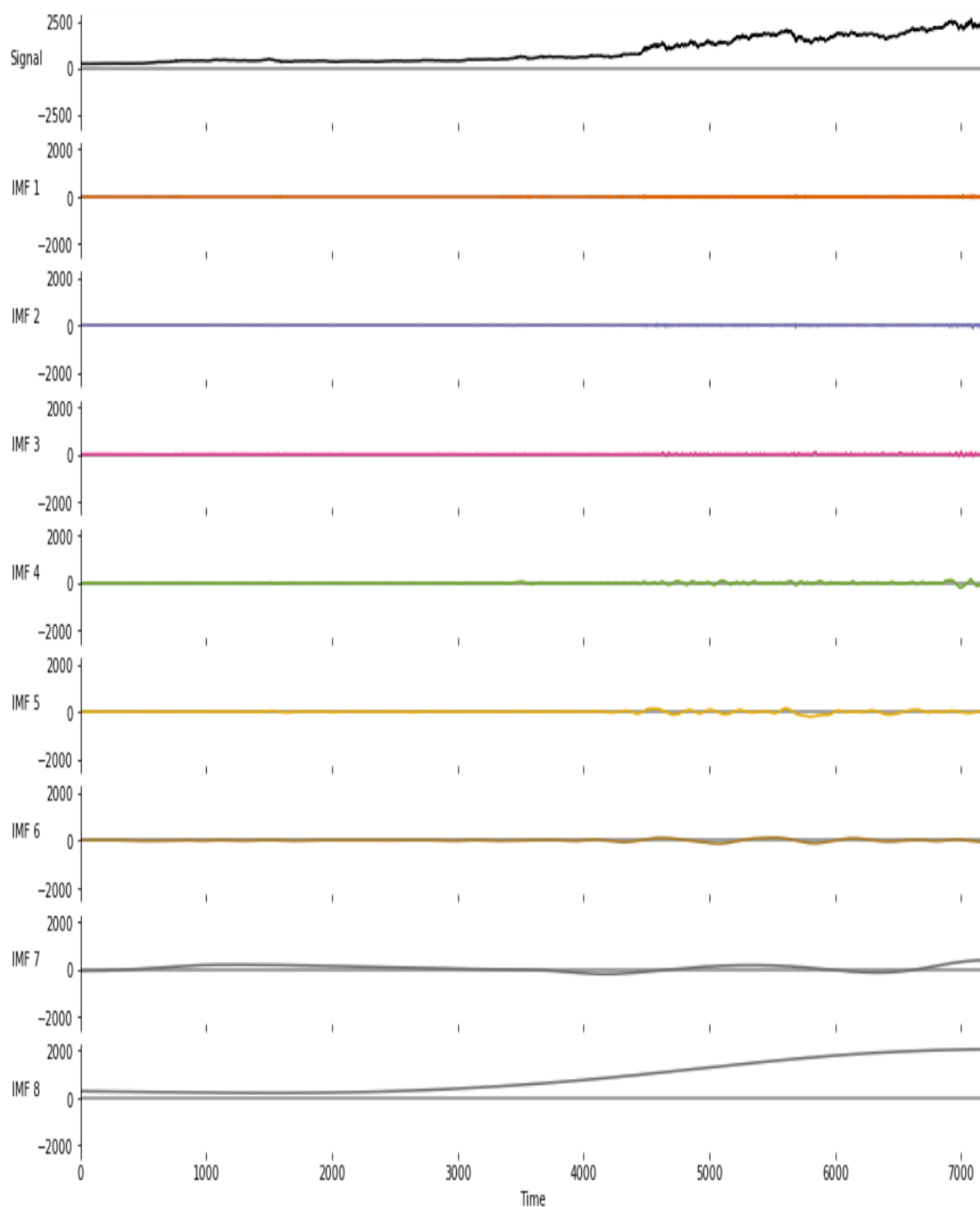
Obrázek 47: Rychlá fourierova transformace BTC se záznamy jednotlivých hodin



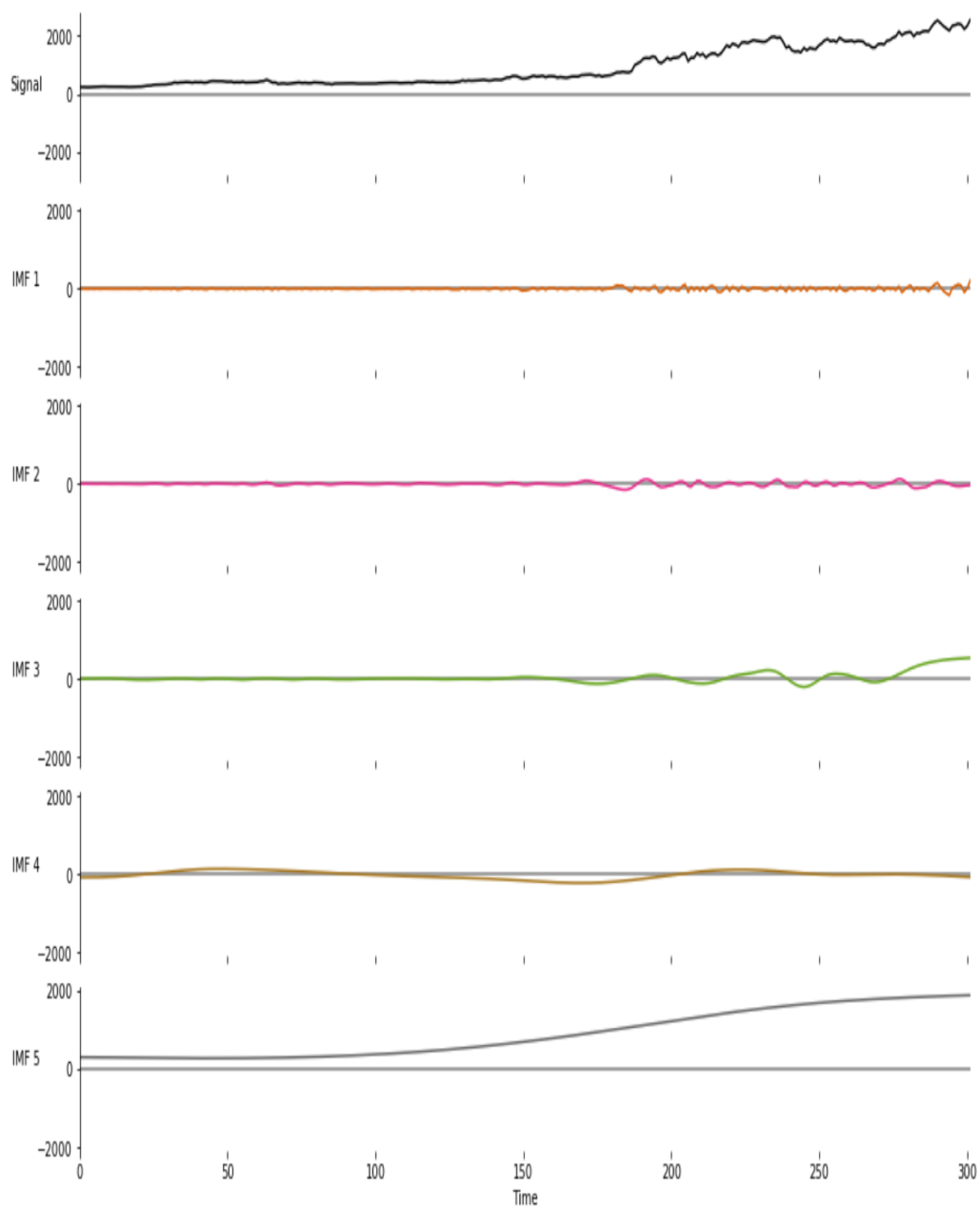
Obrázek 48: Rychlá fourierova transformace BTC se záznamy jednotlivých dní

D Čtvrtá příloha

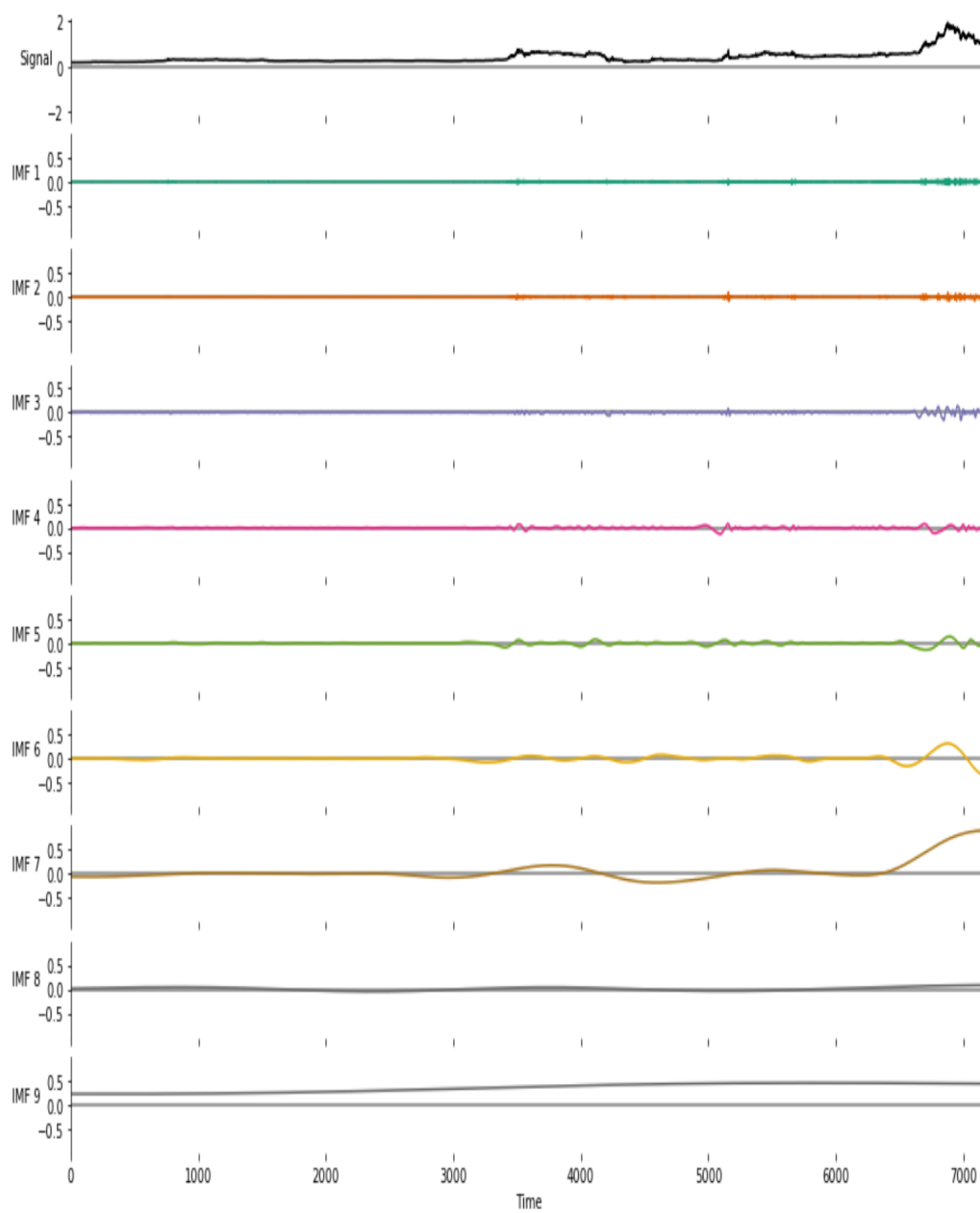
V této části jsou uvedeny grafy vývoje vybraných kryptoměn a to konkrétně XRP, Binance coinu, Ethera a Bitcoinu vůči americkému dolaru pomocí empirického režimu rozkladu.



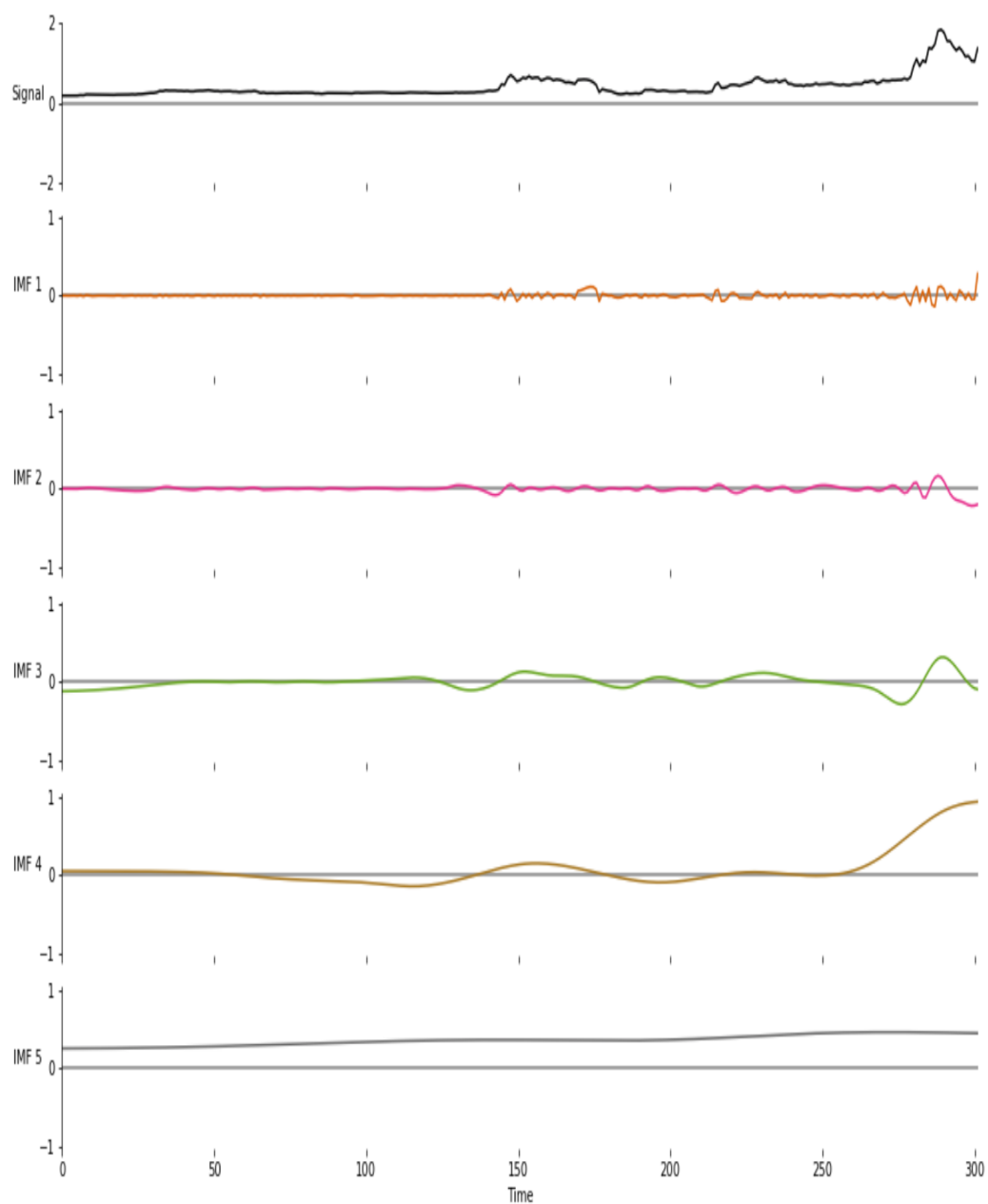
Obrázek 49: Empirický režim rozkladu kryptoměny ETH- záznam jednotlivých hodin



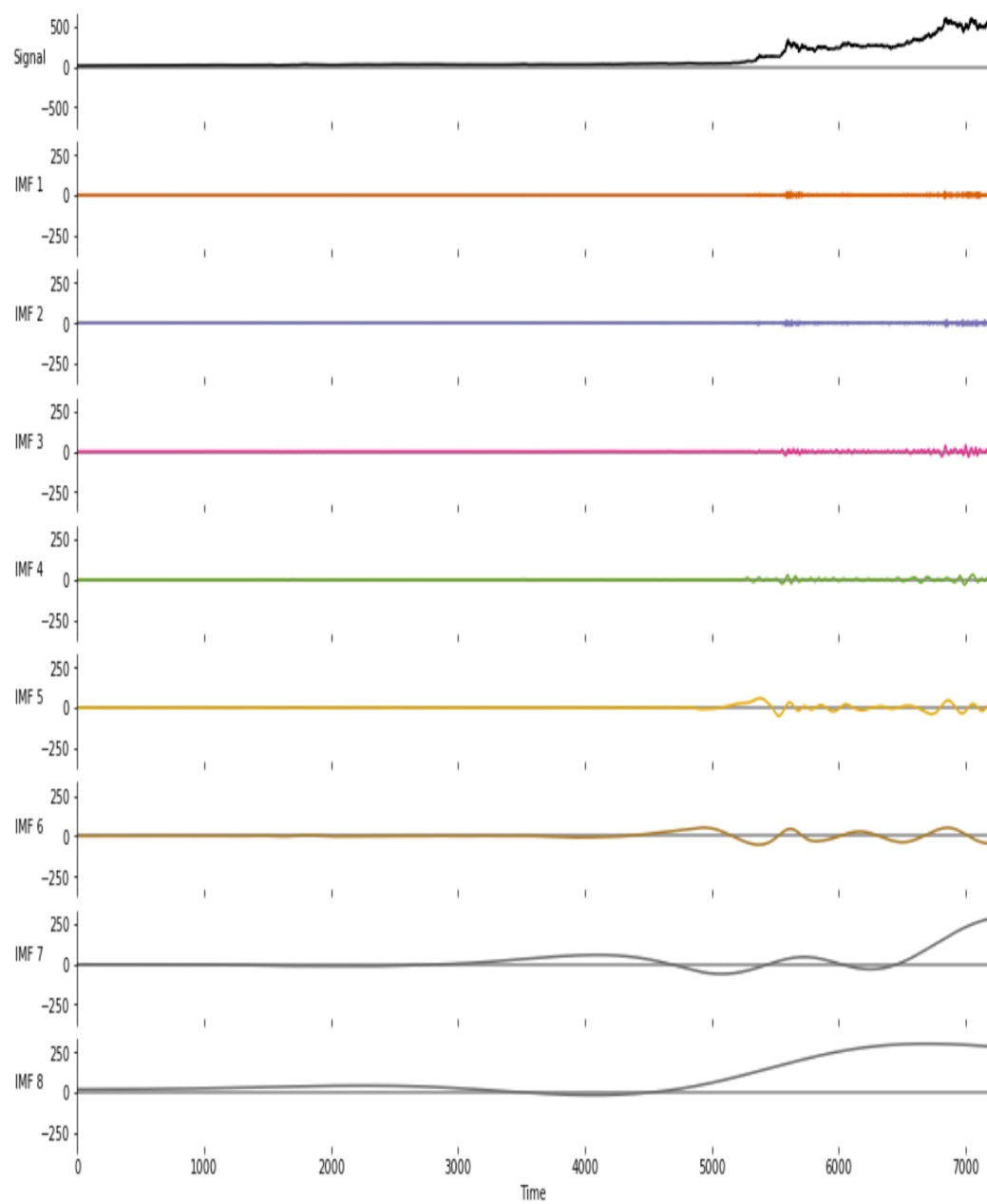
Obrázek 50: Empirický režim rozkladu kryptoměny ETH- záznam jednotlivých dní



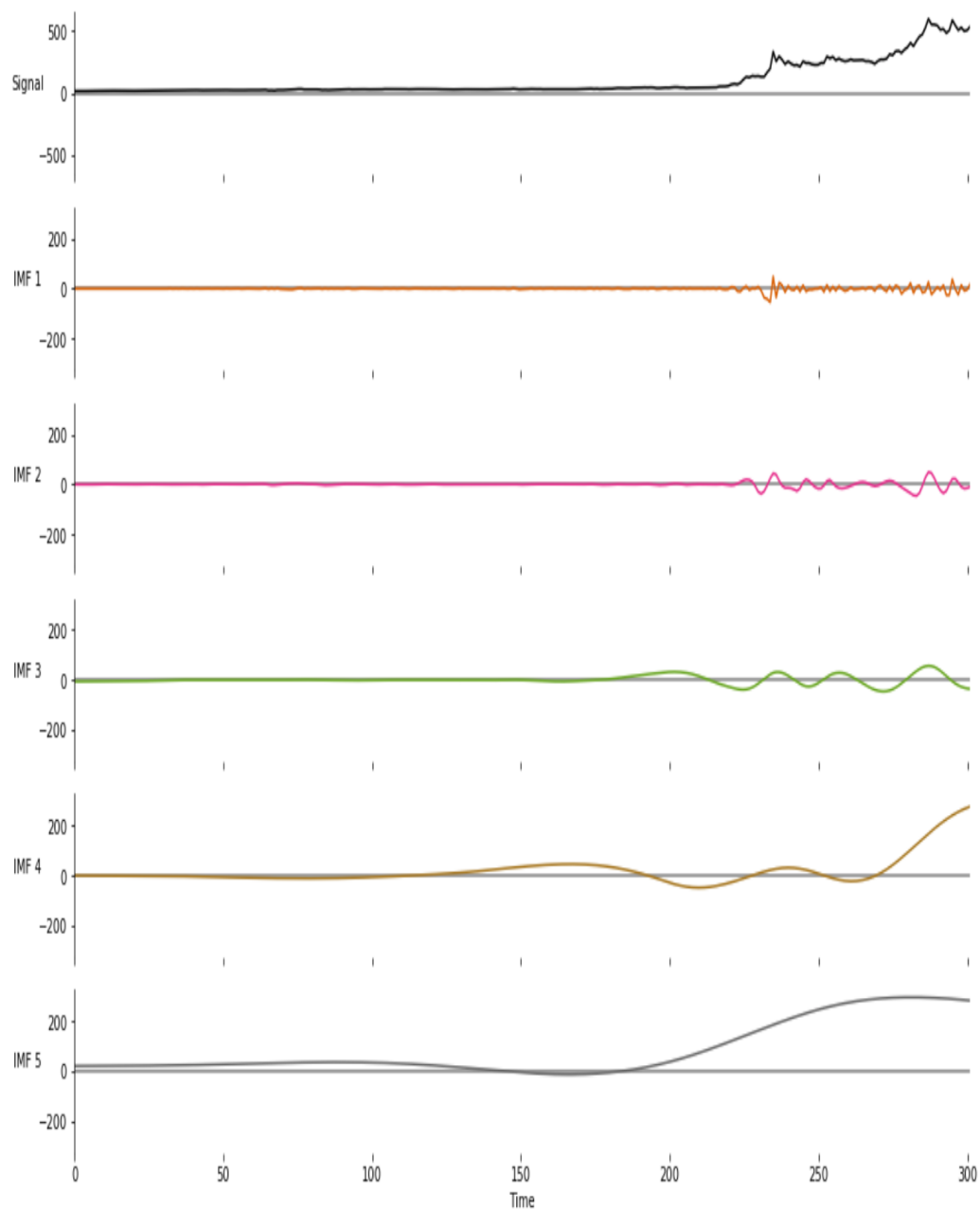
Obrázek 51: Empirický režim rozkladu kryptoměny XRP- záznam jednotlivých hodin



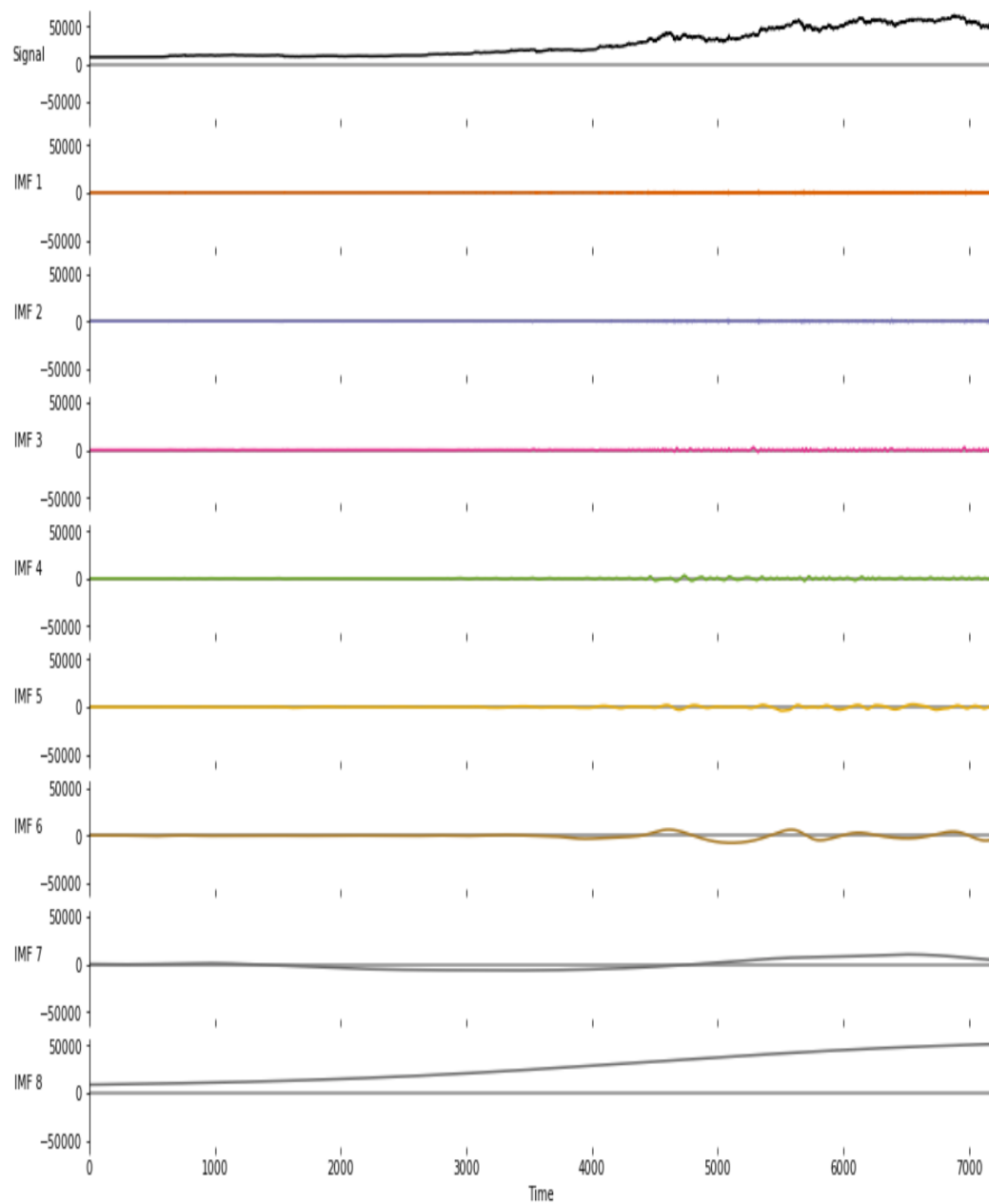
Obrázek 52: Empirický režim rozkladu kryptoměny XRP- záznam jednotlivých dní



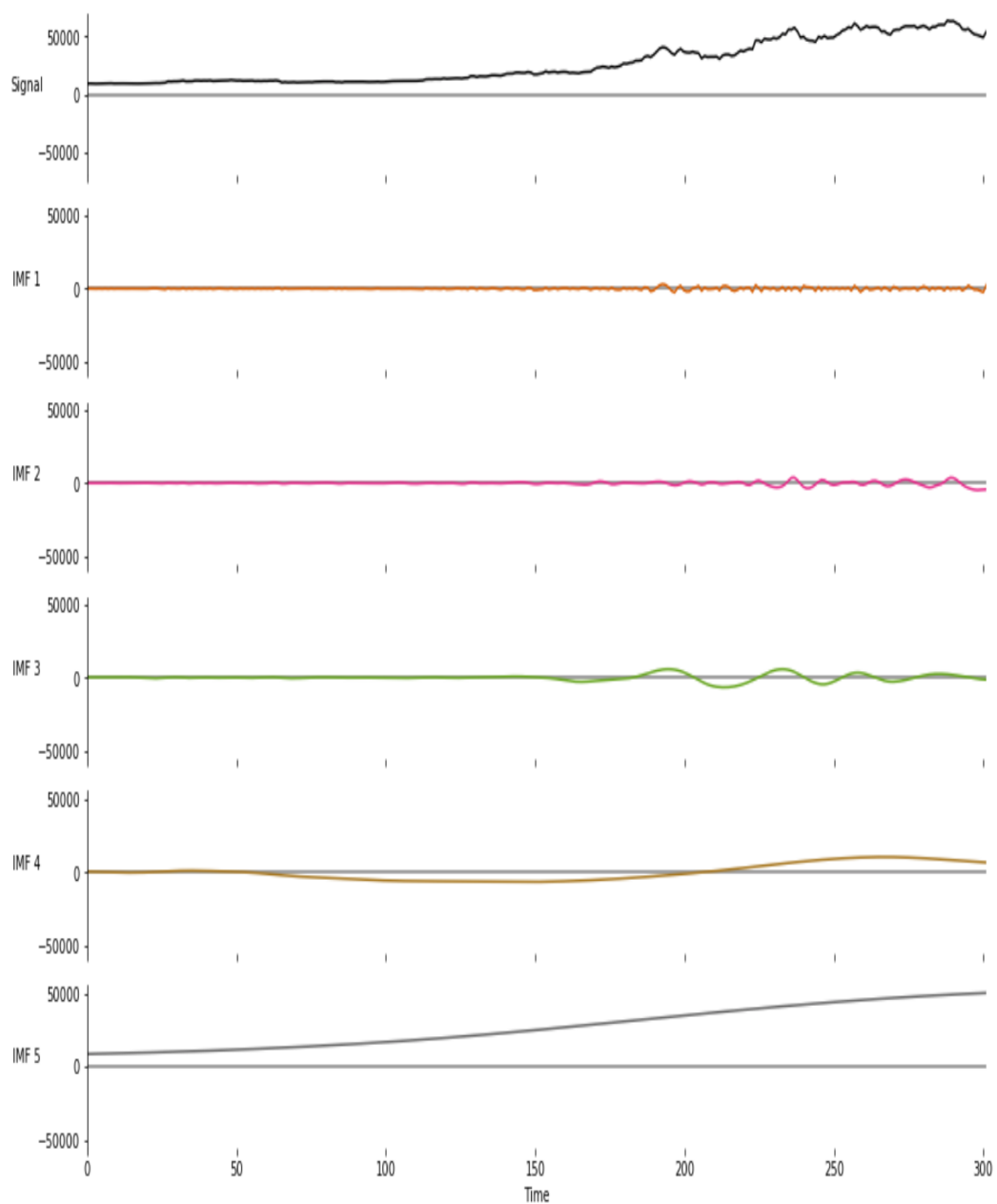
Obrázek 53: Empirický režim rozkladu kryptoměny BNB- záznam jednotlivých hodin



Obrázek 54: Empirický režim rozkladu kryptoměny BNB- záznam jednotlivých dní



Obrázek 55: Empirický režim rozkladu kryptoměny BTC- záznam jednotlivých hodin



Obrázek 56: Empirický režim rozkladu kryptoměny BTC- záznam jednotlivých dní