



**VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ**

BRNO UNIVERSITY OF TECHNOLOGY

**FAKULTA INFORMAČNÍCH TECHNOLOGIÍ**

FACULTY OF INFORMATION TECHNOLOGY

**ÚSTAV INFORMAČNÍCH SYSTÉMŮ**

DEPARTMENT OF INFORMATION SYSTEMS

**VYUŽITÍ STROJOVÉHO UČENÍ PRO PREDIKCI VÝVOJE  
TRHU**

USE MACHINE LEARNING TO PREDICT FUTURE MARKET PRICES

**BAKALÁŘSKÁ PRÁCE**

BACHELOR'S THESIS

**AUTOR PRÁCE**

AUTHOR

**MICHAL KLHŮFEK**

**VEDOUCÍ PRÁCE**

SUPERVISOR

**Ing. MARTIN HOLKOVIČ**

BRNO 2017

**Vysoké učení technické v Brně - Fakulta informačních technologií**

Ústav informačních systémů

Akademický rok 2016/2017

**Zadání bakalářské práce**

Řešitel: **Klhůfek Michal**

Obor: Informační technologie

Téma: **Využití strojového učení pro predikci vývoje trhu**  
**Use Machine Learning to Predict Future Market Prices**

Kategorie: Informační systémy

Pokyny:

1. Nastudujte kurzový devízový finanční trh.
2. Seznamte se s možnostmi aplikování strojového učení v oblasti predikci vývoje budoucích cen na trhu. Zaměřte se na algoritmy pro klasifikaci dat (např. algoritmus k-nejbližších sousedů).
3. Navrhněte systém pro predikci vývoje trhu použitím techniky strojového učení.
4. Implementujte systém navržený v bodě 3.
5. Otestujte vytvořený systém na vybraných finančních trzích.
6. Zhodnoťte dosažené výsledky.

Literatura:

- REJNUŠ, Oldřich. Finanční trhy. 4., aktualiz. a rozš. vyd. Praha: Grada, 2014, 760 s. Partners. ISBN 978-80-247-3671-6
- ZEZULA Pavel, AMATO Giuseppe, DOHNAL Vlastislav a BATKO Michal. Similarity Search The Metric Space Approach, Springer US, 2006, 220 s., ISBN 978-0-387-29146-8
- MITCHELL Tom. M. Machine Learning, McGraw Hill, 1997, 414 s., ISBN 0070428077

Pro udělení zápočtu za první semestr je požadováno:

- Body 1 až 3 zadání.

Podrobné závazné pokyny pro vypracování bakalářské práce naleznete na adrese

<http://www.fit.vutbr.cz/info/szz/>

Technická zpráva bakalářské práce musí obsahovat formulaci cíle, charakteristiku současného stavu, teoretická a odborná východiska řešených problémů a specifikaci etap (20 až 30% celkového rozsahu technické zprávy).

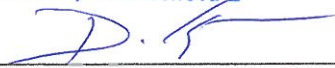
Student odevzdá v jednom výtisku technickou zprávu a v elektronické podobě zdrojový text technické zprávy, úplnou programovou dokumentaci a zdrojové texty programů. Informace v elektronické podobě budou uloženy na standardním nepřepisovatelném paměťovém médiu (CD-R, DVD-R, apod.), které bude vloženo do písemné zprávy tak, aby nemohlo dojít k jeho ztrátě při běžné manipulaci.

Vedoucí: **Holkovič Martin, Ing.**, UIFS FIT VUT

Datum zadání: 1. listopadu 2016

Datum odevzdání: 17. května 2017

**VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ**  
Fakulta Informačních technologií  
Ústav informačních systémů  
612 66 Brno, Božetěchova 2

  
doc. Dr. Ing. Dušan Kolář  
vedoucí ústavu

## Abstrakt

Práce pojednává o systému pro predikci vývoje trhu na základě dat, která byla získána na zkoumaném trhu v minulosti. Cíl byl kladen na využití metod technické analýzy k vytvoření co nejpřesnějšího odhadu chování trhu v budoucnosti. Data získaná z aktuálního stavu trhu jsou pomocí algoritmů na klasifikaci dat z oblasti strojového učení vzájemně porovnávána s historickými hodnotami trhu. Na základě jednotlivých algoritmů byl implementován software, který se snaží nalézt co nejpodobnější shodu dvou skupin dat. Testování probíhalo na sadě dat, která reprezentovala minulý vývoj trhu a bylo pozorováno, na kolik je celkový systém výkonný.

## Abstract

This thesis discusses a market prediction system based on the data obtained from the historic transaction. The main goal was to use the techniques of technical analysis to create a more accurate estimation of market behavior in the future. The data obtained from the current state of the market are compared with the historical market values using the algorithms for the classification of data from the field of learning. Based on individual algorithms, the software was designed to try to match the two sets of data as closely as possible. Testing took place on a dataset that represented the past market enthusiasm, and how much the overall system is performing.

## Klíčová slova

finanční trhy, technická analýza, Dowova teorie, klasifikace dat, Jaccardův koeficient

## Keywords

financial markets, technical analysis, Dow theory, data classification, Jaccard index

## Citace

KLHŮFEK, Michal. *Využití strojového učení pro predikci vývoje trhu*. Brno, 2017. Bakalářská práce. Vysoké učení technické v Brně, Fakulta informačních technologií. Vedoucí práce Ing. Martin Holkovič

# Využití strojového učení pro predikci vývoje trhu

## Prohlášení

Prohlašuji, že jsem tuto bakalářskou práci vypracoval samostatně pod vedením pana Ing. Martina Holkoviče. Uvedl jsem všechny literární prameny a publikace, ze kterých jsem čerpal.

.....

Michal Klhůfek

18. května 2017

## Poděkování

Děkuji Ing. Martinu Holkovičovi za vedení a podporu při vzniku mé bakalářské práce, za poskytnutí podnětných a cenných rad při vzájemných konzultacích, které umožnily, aby tato práce mohla vzniknout.

# Obsah

<b>1</b>	<b>Úvod</b>	<b>3</b>
<b>2</b>	<b>Finanční trhy</b>	<b>4</b>
2.1	Cenová reprezentace . . . . .	5
2.1.1	Čárový graf . . . . .	5
2.1.2	Sloupcový graf . . . . .	6
2.1.3	Svíčkový graf . . . . .	7
2.2	Technická analýza . . . . .	7
2.2.1	Vzory v grafické analýze . . . . .	8
2.2.2	Trendy v technické analýze . . . . .	11
2.2.3	Nevýhody technické analýzy . . . . .	12
2.3	Dowova teorie . . . . .	13
2.4	Trendová analýza . . . . .	15
2.4.1	Technické indikátory . . . . .	15
2.4.2	Trendové indikátory . . . . .	16
2.4.3	Cenové indikátory . . . . .	17
2.4.4	Objemové indikátory . . . . .	17
2.5	Fundamentální analýza . . . . .	18
<b>3</b>	<b>Klasifikace dat</b>	<b>19</b>
3.1	Vyhledávání na základě měření vzdáleností . . . . .	19
3.1.1	Jaccardův koeficient . . . . .	20
3.1.2	Rozšíření Jaccardova koeficientu . . . . .	21
3.1.3	Rozdíl vzdáleností . . . . .	23
3.1.4	Manhattanova vzdálenost . . . . .	24
3.1.5	Euclideánská vzdálenost . . . . .	24
3.1.6	Minkowskiho vzdálenost . . . . .	24
3.2	Shrnutí jednotlivých metod . . . . .	25
<b>4</b>	<b>Návrh a implementace metod pro analyzování trhu</b>	<b>26</b>
4.1	Příprava dat z trhu . . . . .	26
4.2	Normalizace dat z trhu . . . . .	28
4.3	Jednotlivé metody . . . . .	28
4.3.1	Postupy pro nalezení automatické odchylky . . . . .	29
<b>5</b>	<b>Testování a vyhodnocení jednotlivých metod</b>	<b>33</b>
5.1	Etapy testování . . . . .	34
5.1.1	První etapa testování . . . . .	34

5.1.2	Druhá etapa testování . . . . .	35
5.1.3	Třetí etapa testování . . . . .	37
5.2	Možnosti dalšího testování . . . . .	37
5.2.1	Rozšíření pro analýzu historického trendu na základě výkyvů trendu	37
5.2.2	Rozšíření pro analýzu nejen konečné uzavírající hodnoty kurzu . . .	38
5.3	Shrnutí dosažených výsledků z testování . . . . .	38
<b>6</b>	<b>Závěr</b>	<b>39</b>
	<b>Literatura</b>	<b>40</b>

# Kapitola 1

## Úvod

Při obchodování na finančních trzích se v současné době dá využít celá řada metod a postupů, které se snaží investorům vhodně pomoci predikovat budoucí vývoj směru trendu na trhu. Mezi ty nejvýznamnější patří analýzy pomocí technické nebo fundamentální analýzy. Každá z těchto metod nabízí rozdílný úhel pohledu na danou problematiku a popisuje, jak co nejefektivněji odhadnout budoucí chování trendu na trhu. Při technické analýze je největší důraz kladen na samotná data a jejich zkoumání s ohledem do minulosti, kdežto u fundamentální analýzy je veškerý důraz kladen na příčiny, které vedly ke změnám kurzů.

Tato práce se zabývá analýzou historických dat pro predikci budoucího směru vývoje trhu za využití principů technické analýzy. Samotná analýza je prováděna na datech, která obsahují hodnoty kurzů zaznamenaných v minulosti. Ke zkoumání se využívá sady algoritmů a jejich modifikací, za účelem nalezení co nejpodobnější skupiny dat vůči aktuálnímu trendu. Algoritmy jsou využívány z oblasti klasifikace dat, pro účely nalezení co nejpodobnějšího vzorku dat, který by odpovídal aktuálnímu trendu.

Cílem této práce je ověřit teorii, zda se pomocí analýzy minulých tvarů trendů dají adekvátně předpovídat situace na trhu a analyzovat budoucí vývoj trendu. Současně se zabývá zjištěním, která metoda pro analýzu dat je nejefektivnější. Je tvořena následujícími kapitolami:

- **2. kapitola** - Popisuje a zkoumá základní principy fungování finančních trhů s největším důrazem kladeným na obchodování prováděné na Forexu. Dále je zde podrobně rozepsáno, jakým způsobem se provádí samotná technická analýza, a jak se výsledná data vyhodnocují.
- **3. kapitola** - Zabývá se popisem jednotlivých metod pomocí nichž se budou data následně porovnávat a vyhodnocovat. Jsou zde zastoupeny metody jak pro porovnávání metod na základě vzájemné podobnosti, tak i metody pro hledání K-nejbližších sousedů.
- **4. kapitola** - Popisuje jakým způsobem jsou jednotlivé metody implementovány a jakým způsobem celý systém analyzování daných dat funguje.
- **5. kapitola** - Názorně zobrazuje, jakým způsobem bylo prováděno testování a které metody dosahují nejvhodnějších výsledků. Kromě toho také obsahuje část, která se zabývá doporučeními pro budoucí vylepšení daných metod.

## Kapitola 2

# Finanční trhy

Finanční systém představuje část ekonomického systému, kde se odehrává pohyb finančních zdrojů. Samotný systém je tvořen třemi základními prvky, které reprezentují: finanční instituce, finanční dokumenty a finanční trhy. Pod pojmem finanční instituce jsou označovány instituce, které vytvářejí finanční dokumenty, se kterými se nadále obchoduje na finančních trzích. Jedná se například o banky či spořitelny. Finanční dokumenty nám poté představují všechny formy cenných papírů jako jsou například depozita, dluhopisy či akcie. Finanční trhy naopak představují místa, kde dochází k samotnému přesunu finančních prostředků od subjektů nakupujících k subjektům prodávajícím. Z důvodu zachování stability je tento trh velmi regulován, a to jak na národní, tak mezinárodní úrovni. Díky tomu je zajištěna jednak stabilita trhu, ale i ochrana všech, kteří na trh vstupují [4].

Samotná teorie finančních trhů se zabývá využíváním principů finančního systému. Jedná se o mechanismy, které zajišťují řadu důležitých funkcí, z nichž nejdůležitější je právě přemístování finančních prostředků od přebytkových subjektů přímo k subjektům deficitním, které pro ně mají efektivnější využití. Díky tomu je výsledná hodnota tvořena právě na základě nabídky a poptávky. To je reprezentováno cenami, za které jsou jednotliví obchodníci ochotni prodávat a nakupovat. Finanční trhy řadíme do několika kategorií podle jejich typu. Jedná se o devizové, úvěrové, komoditní a kapitálové trhy. Všechny typy jsou považovány za nedílnou součást ekonomického systému, kde zajišťují jeho obsluhu a dokonce se podílejí i na jeho fungování. V této práci je největší důraz kladen právě na obchodování prováděné na devizových trzích.

### Forex

Při obchodování na devizových trzích se využívá označení obchodování na forexovém trhu. Tento trh vznikl v roce 1973 a jeho zkratka znamená FOReign EXchange, v překladu trh pro směnu cizích měn [9]. Valná většina obchodů je zde prováděna vůči US dolaru s tím, že nejvíce obchodované měny jsou euro, japonský jen, britská libra a švýcarský frank. S veškerými měnami je obchodováno v párech, přičemž první měna se nazývá základní a druhá kótovací. Jako příklad lze uvést měnový pár v podobě EUR/USD. Při kurzu 1.35 to znamená, že za 1 euro koupíme 1.35 dolaru.

Obchodování na forexovém trhu probíhá pomocí lotů. Pod pojmem lot se označuje objem transakcí o hodnotě 100 000 jednotek, se kterými zrovna chceme obchodovat. Investorům je ale umožněno obchodovat i s minimálním kapitálem pomocí tzv. mikrolotů. Mikroloty představují 1/100 lotu. Pokud investor nakoupí 1 lot EUR/USD, znamená to tedy 100 000 krát aktuální kurz dané měny.



Samotné obchodování je pak možné provádět 24 hodin denně, kromě víkendů. Zde je nutné zdůraznit, že forexový trh má i své hlavní obchodní hodiny, ve kterých jsou objemy provedených transakcí řádově vyšší. Tyto obchodní hodiny nám říkají, kdy jsou jednotlivé měnové páry nejaktivnější, protože velké finanční centra právě provádějí svoje transakce. Mezi nejvýznamnější finanční centra ovlivňující dané měnové páry patří například Londýnské. To má největší vliv na obchody prováděné vůči libře. Mezi další patří například New Yorkské obchodní centrum, kde jsou transakce prováděny vůči dolaru. Hlavní hodiny pro obchodování vůči britské libře jsou od 8:00 do 16:00, u dolaru jsou hlavní obchodní časy od 13:00 do 21:00 [3].

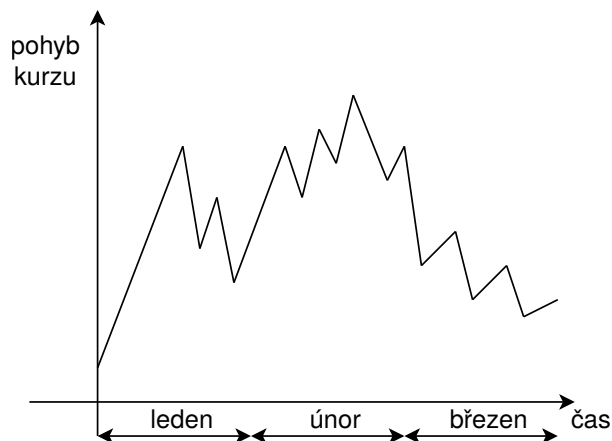
Nejvhodnější doba pro obchodování nastává právě ve chvíli, kdy se jednotlivé časy dvou finančních center díky rozdílným časovým pásmům překrývají. Tím vzniká ideální prostor pro provádění obchodů mezi těmito dvěma měnami. Vzhledem k tomu, že trh není centralizovaný, nastává zde problém, který spočívá v tom, že nemáme k dispozici celkové počty objemů provedených obchodů [2].

## 2.1 Cenová reprezentace

Pro cenovou analýzu vývoje jednotlivých kurzů se v praxi často využívá grafické znázornění v podobě zanesení dat do grafu. Můžeme se setkat s celou řadou grafů, které se mezi sebou liší jednotlivými způsoby znázornění objemů dat, či přehledností znázornění vývoje pohybu kurzu. Důvodů, proč se k reprezentaci cen kurzů využívá právě zanesení do grafu je celá řada. Především se jedná o názornější znázornění dat na základě minulých pohybů na trhu. Jednotlivé grafické metody se poté vzájemně liší právě způsobem a mírou znázornění samotných dat. Díky tomu jsme dále schopni se v technické analýze zabývat zkoumáním vývoje kurzu na základě podobnosti grafů.

### 2.1.1 Čárový graf

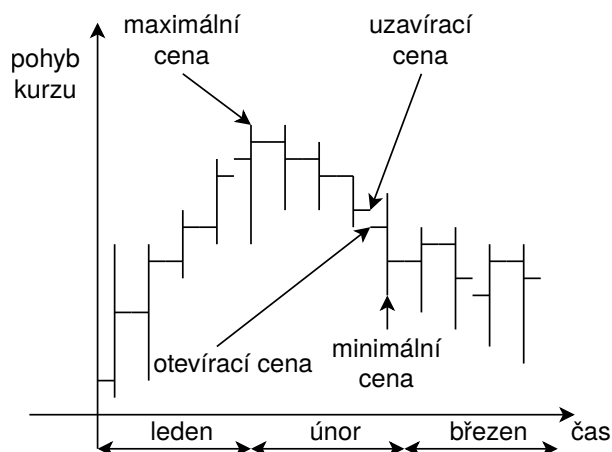
Nezákladnějším druhem grafu je čárový graf. Sestavení grafu probíhá primárně podle jedné hodnoty, kterou je otevírací či zavírací kurz. Existuje však více možností. Možnosti jsou například součet maximální a minimální hodnoty kurzu v daném časovém úseku, který je podělen dvěma. Jiným přístupem je součet maximální, minimální, otevírací a uzavírací hodnoty kurzu poděleno čtyřmi. Velká výhoda této metody spočívá v jednoduchosti jejího čtení a následného vyhodnocování ukazatelů, jakým směrem by se trh mohl dále vyvíjet. Podstatnou nevýhodou zde je, že při grafické reprezentaci chybí názorná informace o maximálních, minimálních, otevíracích a uzavíracích kurzech. Graf zobrazuje pouze jednu hodnotu, na jejímž základě vykresluje daný trend. Graf je znázorněn na obrázku 2.1.



Obrázek 2.1: Čárový graf s rozložením na tři měsíce

### 2.1.2 Sloupcový graf

Sloupcové grafy patří k velmi oblíbeným a používaným grafům využitelným dále v technické analýze. Pomocí různě vysokých horizontálních úseček je zde znázorněn vývoj kurzu. Vynesená data obsahují informace o uzavíracích, otevíracích, maximálních či minimálních kurzech. Reprezentace jednotlivých informací je prováděna pomocí zanesení minimální hodnoty kurzu na spodní konec úsečky a maximální hodnoty na horní začátek dané úsečky. Dále umístěním malé čárky vlevo (otevírající kurz), nebo vpravo (uzavírající kurz). Sloupcový graf přináší vyšší vypovídající hodnotu díky tomu, že jsou do něj zakreslovány až 4 hodnoty pro každý časový údaj, což přináší investorovi cenné informace, jak se dále na trhu zachovat. Bohužel negativním dopadem je snížení přehlednosti grafu a zvýšení náročnosti na způsob jeho využití. Investor se díky tomu už nerozhoduje pouze na základě jedné hodnoty, ale je nucen vzít v potaz i ostatní faktory, což vede ke složitějšímu rozhodování. Grafické znázornění grafu je na obrázku 2.2.

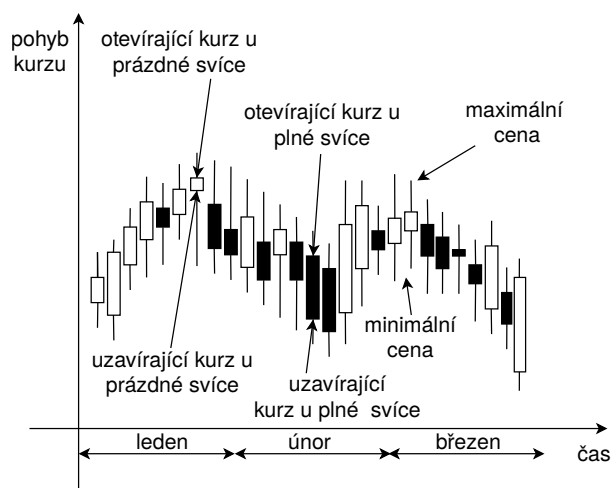


Obrázek 2.2: Sloupcový graf s rozložením na tři měsíce

### 2.1.3 Svíčkový graf

Svíčkový graf patří mezi nejstarší metody v technické analýze. Jeho původ nalezneme v Japonsku již v 18. století. Pracuje se stejnými daty jako sloupcové grafy. Jedná se o otevírací, uzavírací, maximální a minimální hodnoty kurzu. Jedna časová perioda je v grafu znázorněna jedinou svící, která je tvořena dvěma knoty a tělem. Samotné knoty poté udávají rozpětí mezi maximálním a minimálním kurzem. Tělo svíce odpovídá rozpětí mezi otevíracím a uzavíracím kurzem. Pokud je tělo svíce vyplněné jedná se o situaci, kdy kurz klesal a otevírací kurz je vyšší než kurz uzavírací [4].

V praxi se můžeme setkat i s jinými grafickými reprezentacemi a ne pouze s vyplněnými a prázdnými svícemi, nicméně ale v následujícím schématu je zvolena právě tato notace. Velká výhoda následujícího grafu spočívá v tom, že nabízí graficky přívětivější znázornění situace na trhu. Pokud vznikají prázdné svíce, na trhu vládne optimismus a dochází k nákupům. Vyplněné svíce spíše reflektují situaci, kdy na trhu vládne pesimismus a většina investorů spíše prodává. Pro názornost je graf znázorněn na obrázku 2.3.



Obrázek 2.3: Svíčkový graf s rozložením na tři měsíce

## 2.2 Technická analýza

Jedná se o metodu, která zkoumá historický vývoj kurzů na trhu a dle toho se snaží predikovat její budoucí vývoj. Díky svým nástrojům se snaží pokrýt všechny potřebné a pozorovatelné veličiny, které se na finančních trzích objevují. Mezi nejdůležitějšími veličinami jsou například trendy, ceny či objemy daných finančních instrumentů, či sentiment investorů. Samotná technická analýza je považována za jednu z nejstarších a stoletími prověřenou analytickou metodu pro studium vývoje kurzu. Je postavena na praktických zkušenostech obchodníků, které byly získány při obchodování na finančních trzích. Při vykreslování jsou využívány údaje o vývoji kurzů, indexů a objemech provedených obchodů [14].

Techničtí analytici, kteří se zabývají zkoumáním vývoje předešlého kurzu, jsou přesvědčeni, že právě studiem dosavadního vývoje kurzů lze získat informace, které nám napomohou v predikci jejich dalšího vývoje. Své rozhodnutí zakládají na informacích o kurzech, síle trhu, směru a rozpětí objemu provedených obchodů. Nejdůležitějším cílem je však správné

načasování. Na základě správného načasování nákupů a prodejů se investorovi podaří nakupovat kolem tržních minim a prodávat kolem tržních maxim. Jejich samotné investiční strategie poté spočívají v nákupech za nízké kurzy a prodeji za kurzy vysoké [13].

Na samotný vývoj kurzů má pak zcela zásadní vliv struktura nabídky a poptávky, která je však ovlivněna celou řadou faktorů. Ne vždy obchodník jedná na trhu podle racionálních faktorů, jako jsou například ekonomické statistiky, účetní výkazy apod. Jeho chování je z části řízeno i iracionálně, a tím pádem nejde přesně předpovídat. V technické analýze se však předpokládá, že psychologické faktory jsou již z větší míry zahrnuty právě v nabídce a poptávce, které utvářejí hodnotu výsledného kurzu. Z tohoto důvodu se tomu analytik nevěnuje a soustředí se až na důsledky, které se projevují růstem, klesáním či stagnací kurzu [4].

### 2.2.1 Vzory v grafické analýze

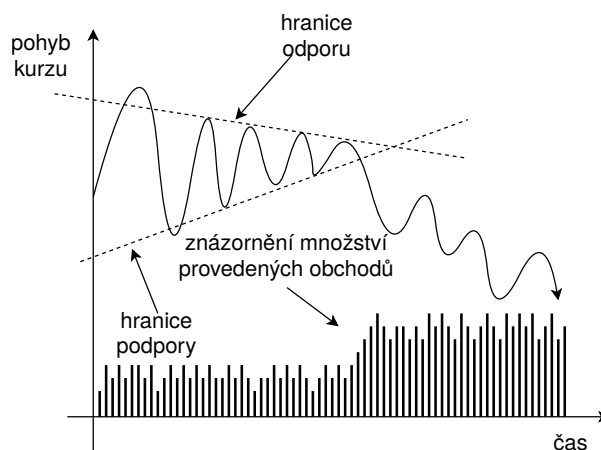
Studiem grafů, znázorňujících minulý vývoj kurzů bylo odhaleno, že se u nich do jisté míry jednotlivé vzory a obrazce opakují. Tyto opakující se vzory lze nazvat standardizované formace. Na jejich vytváření mají podstatný vliv právě investoři, kteří své očekávání promítají do změn nabídky a poptávky, což vede ke změnám kurzu. Stejně vzory a formace se dají nalézt také na jiných trzích, protože investoři mají tendenci se chovat v obdobných situacích velmi podobně. U všech grafických formací však hraje důležitou roli i objem obchodů, nejen hodnota kurzů. Právě díky informaci o celkovém množství obchodů víme, zda je růst opravdu doprovázen i zvyšujícím se počtem obchodů.

#### Konsolidační formace

Po období prudkého růstu nebo poklesu kurzů se vyskytují konsolidační formace. Díky nim dochází k dočasnému zastavení a korekci vývoje kurzu. Tato změna trvá do doby než je formace opuštěna. Poté dochází buď k návratu do původního trendu, nebo se začíná vytvářet trend opačný [4].

#### Trojúhelníky

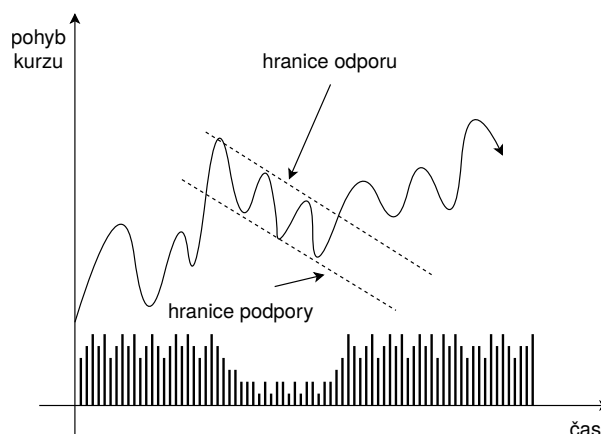
S trojúhelníky se setkáváme v době, kdy kurzy kolísají ve velmi malém vzájemném rozdílu, což je způsobeno tím, že hranice odporu a hranice podpory sevrou kurzy mezi sebe. Obě hranice se díky sevřenému tvaru protínají v jednom bodě a vytvářejí tím tvar připomínající trojúhelník. Během samotného trvání trojúhelníku klesají objemy obchodů. Výrazně začínají stoupat až po opuštění trojúhelníku. Graf je znázorněn na obrázku 2.4.



Obrázek 2.4: Formace tvaru trojúhelníku

## Vlajky

Vlajky nastávají v době, kdy pohyb kurzu je svírán v úzkém rozmezí mezi hranicemi podpory a hranicemi odporu, které tvoří dvě rovnoběžné linie. Jednotlivé vlajky mohou vznikat jak v rostoucích, tak i v klesajících trendech. V průběhu trvání trendu samotné obchody postupně klesají a po opuštění formace dochází opět k jejich postupnému růstu. Nejčastěji se vlajky vytvářejí v průběhu trvání platného trendu, a to v jeho polovině růstové fáze. Grafické znázornění grafu je na obrázku 2.5.



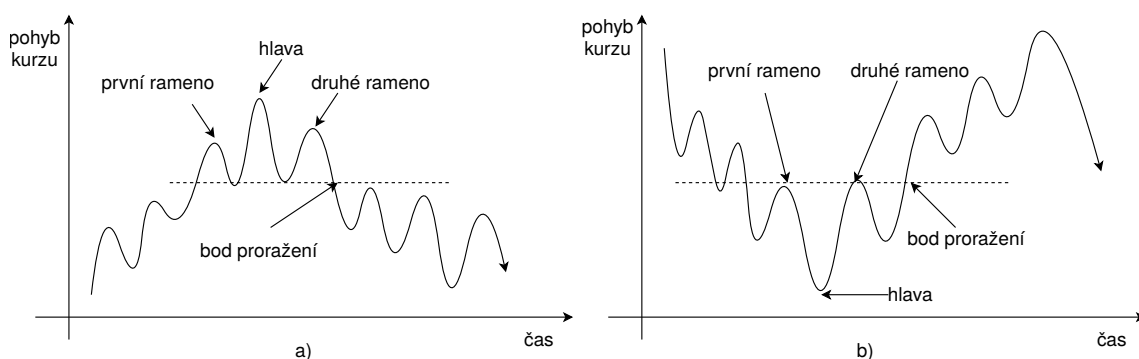
Obrázek 2.5: Formace tvaru vlajky

## Reverzní formace

V době, kdy se trh dostává během svého vývoje do oblastí vrcholů, či se blíží svému dnu, začínají se vytvářet reverzní formace, které upozorňují na změnu trendu. Tyto vznikající formace si kladou za cíl nás upozornit, že v případě formací vznikajících u vrcholů se jedná o překoupený trh, kdežto u situací kolem dna se jedná o přeprodaný trh, a proto je vhodné očekávat změnu trendu.

## Hlava ramena

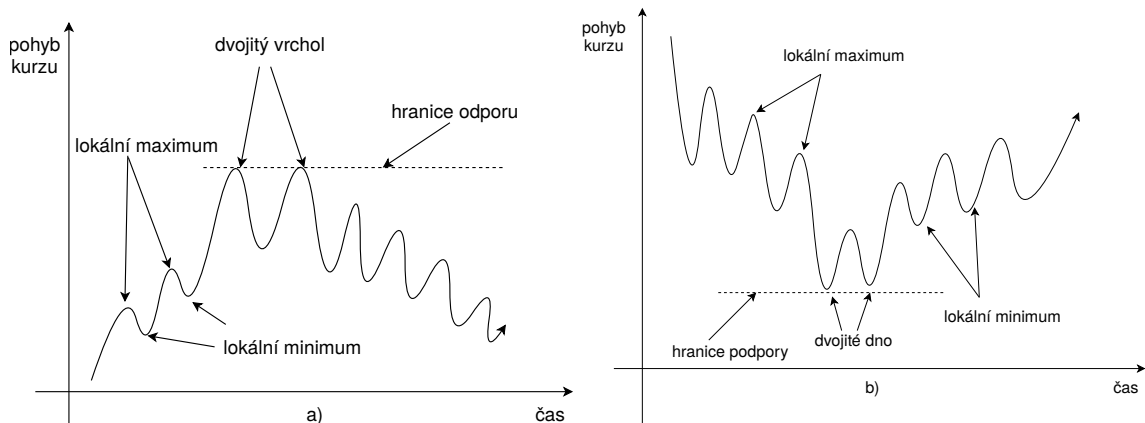
Jedná se o nejznámější formaci, která se utváří buď v době, kdy se kurzy blíží svému lokálnímu dnu, nebo když se blíží svému lokálnímu maximu. Tvořená je jednou hlavou a dvěma menšími rameny. Formace hlavy a ramena nám může vzniknout jak pro medvědí, tak i pro býčí trend. Například pro medvědí trend je situace následná. Brzy po vytvoření ramene má hodnota kurzu tendenci mírně klesnout, což je ale doprovázeno opětovným růstem, čímž dochází k vytvoření hlavy. V následujícím kroku opět dochází k mírnému poklesu, což je ale doprovázeno růstem a vytvořením pravého ramene. K porušení trendu poté nastává ve chvíli, kdy se trend zlomí a dochází k proražení jednoho z ramen. V případě býčího trendu je situace opačná. Pro názornost je graf znázorněn na obrázku 2.6.



Obrázek 2.6: a) medvědí trend; b) býčí trend

## Dvojité vrcholy, dna

Při této formaci se vytváří dva vrcholy nebo dvě dna za sebou pomocí odrazu kurzů od hranice podpory nebo od hranice odporu. Pro případ dvou vrcholů je situace následující. Nejprve kurz vystoupá až k hranici odporu, kde následně ustoupí zpět, aby se mohl k hranici odporu znovu vrátit. Poté začíná docházet ke klesání, které je známkou toho, že klesající trend započal. U situace se dvěma dny je postup velmi podobný. Nejprve kurz klesne až k hranici podpory, kde dochází k mírnému růstu, následuje však návrat k hranici podpory. Po jeho odražení dochází k mírnému růstu což značí, že stoupavý trend započal. Grafické znázornění jednotlivých metod je na následujícím obrázku 2.7.



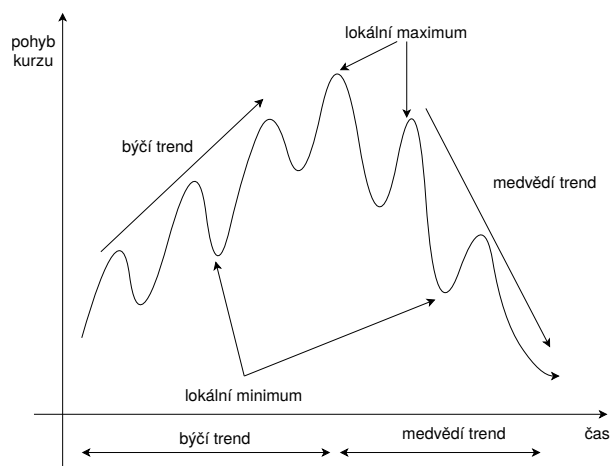
Obrázek 2.7: a) Formace dvojitého vrcholu; b) Formace dvojitého dna

## 2.2.2 Trendy v technické analýze

Jedna z hlavních myšlenek, ze kterých technická analýza vychází jsou právě trendy, ve kterých se dané kurzy vždy pohybují. Ty mají daný směr ve kterém i určitou dobu setrvávají. Ke změnám v trendech dochází právě na základě nabídky a poptávky. Techničtí analytici věří, že na základě studia předešlých vývoje kurzů se dají změny trendů předpovídat. Jednotlivé trendy se poté liší ve své délce trvání a rozlišují se do několika kategorií. Primární trendy mohou trvat několik měsíců až let. Sekundární, neboli střednědobé trendy mohou trvat několik týdnů až měsíců. Velmi krátkodobé trendy se pohybují v rámci několika dnů. U trendů intradenních se pohybujeme dokonce v řádu hodin či minut [4].

Základní typy trendů jsou rostoucí, klesající a stagnující trend. Rostoucí trend bývá označován také jako býčí a je charakterizován rostoucími lokálními maximy a lokálními minimy provedených transakcí za dané časové období. Vychází se z toho, že každé další lokální maximum by mělo být výše než předcházejí lokální maximum, a také každé lokální minimum by mělo být výše než lokální minimum předcházející. Klesající trend bývá někdy označován jako medvědí, kdy každé následující maximum je níže než maximum předcházející.

Investor by měl umět sledovat nejen aktuální trend, ale i chování trhu v rámci měsíců a na základě toho přizpůsobit všechny své nástroje a metody právě tomu. Samotná volba trhu je výrazně ovlivněna osobními preferencemi investora, cíli které sleduje, druhy investic o které má zájem a objemem investovaného kapitálu. V průběhu obchodování se často stává, že investor své rozhodnutí mění a délka sledovaného trendu se v rámci investičního horizontu mění. K určování trendu slouží mnohé metody jako jsou například klouzavé průměry či trendové linie. Grafické znázornění grafu je na obrázku 2.8.



Obrázek 2.8: Přehled trendů u technické analýzy

Základní motto technické analýzy zní: „Následuj trend, trend je tvůj přítel“.[4]

Problémy však nastávají u situací, kdy trend nejde na trhu zcela přesně rozpoznat. Zde začínají některé metody technické analýzy selhávat. Samotné trendy se v průběhu času mezi sebou střídají s tím, že se mezi sebou liší silou a dobou trvání. Seskupení několika trendů do jednoho se nazývá vzor, nebo také formace a ty mají tendenci se často opakovat. Na technickém analytikovi poté je takové formace nalézt a správně vyhodnotit.

Reakce na aktuální změny trhu nejsou vždy provedeny okamžitě, protože ne všichni investoři mají stejný přístup k relevantním informacím. To také vede k pomalejším změnám. Pokud jsou informace pozitivní, trh roste, pokud jsou negativní, trh klesá. Tyto informace však technického analytika nezajímají a on zaměřuje svou pozornost především na sledování vývoje hodnoty kurzu. Předmětem zkoumání je tedy vždy trend sám o sobě, nikoliv příčiny jeho vzniku. Za dobu existence technické analýzy byly vyvinuty desítky metod a nástrojů pro technickou analýzu, přičemž každá přináší své klady i zápory. Například grafické metody, technické indikátory a mnohé další [4].

Vždy je ze začátku nezbytné posoudit o jaký charakter trhu se jedná a jestli jde výrazným směrem nahoru či dolů. V případě silného trhu se jedná o trh, který trenduje. Pokud se kurz pohybuje v úzkém pásu lze hovořit o tom, že trh stagnuje.

### 2.2.3 Nevýhody technické analýzy

Při zpracování dat technické analýzy dochází k problémům zejména proto, že nelze očekávat přesné opakování historie. Z tohoto důvodu nikdy nenastanou úplně stejné vzory a obrazce i napodruhé, což bývá jeden z hlavních aspektů, které se technické analýze nejvíce vyčítají. Zejména to, že je až přespříliš zahleděná do minulosti. Důvodem je, že lidské reakce na vzniklé události jsou podobné, ne však zcela stejné. Lidský faktor investorů také může způsobit to, že některé signály jsou špatně interpretovány, což poté může vést ke ztrátám. Ovšem některé metody nabízejí více způsobů interpretace, což může vést k mylným výsledkům. Nutné je vzít v úvahu také fakt, že zde dochází ke zpoždování správného rozpoznání trendu, což může přinášet další potenciální ztráty. Problém je v tom, že pokud byl systém úspěšný někdy v minulosti neznamená to, že musí svůj úspěch zopakovat i v budoucnosti.



Samotný trend se může kdykoli změnit, bez jakéhokoliv varování, což se nedá přesně předpovědět. Z tohoto důvodu se ve světě technické analýzy nepohybujeme s danou jistotou a přesnými informacemi. Kvůli podstatě fungování technické analýzy nejsme schopni posoudit, zda jsou hodnoty kurzů nadhodnocené či podhodnocené. Nelze tedy doporučit nákup podhodnocených, či prodej pouze nadhodnocených titulů.

Velkému zájmu se technická analýza těší ze strany praktiků, kteří ji s oblibou využívají v praxi, méně však už ze strany akademiků, kde je vyšší důraz kladen na teoretické modely, které v praxi nejsou efektivně využívány, kvůli jejich spíše ideálnímu charakteru. Nejčastější důvod, proč jí akademici nevěří je, že předpokládá opakování nějakého chování, tedy co se již jednou stalo v minulosti. Raději preferují hypotézy, které mají platnost spíše jen na efektivním trhu. Techničtí analytici však preferují technickou analýzu a její části jako například hranice podpory a odporu [18].

## 2.3 Dowova teorie

Na přelomu 19. a 20. století pan Charles Henry Dow začal vytvářet svou teorii pro obchodování na burze. Po jeho smrti se na zdokonalování teorie podílela celá řada dalších autorů, ale pan Dow je díky své původní práci nazýván otcem moderní technické analýzy. Samotná teorie je považována za nejstarší a nejznámější teorii z oblasti technické analýzy. Klade si za cíl identifikovat trendy na akciových trzích a správně je určovat [14].

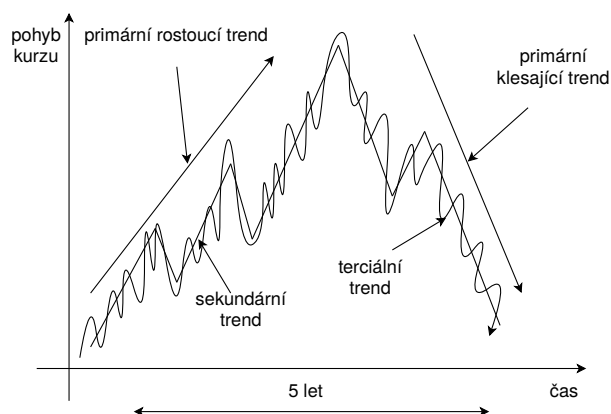
Podstata teorie je založená na myšlence, že většina kurzů se na trhu zpravidla pohybuje v souladu s hlavním trendem, který je na trhu nastolený. Pokud se na trhu vyskytuje hlavní trend, trvá do doby, než je výrazněji změněn směr kurzu. Největší důraz je kladen na identifikaci hlavního trendu trhu, popřípadě na jeho změny [7].

U Dowovy teorie se setkáváme se třemi základními typy trendů, které jsou dále zobrazeny na obrázku 2.9.

**Primární trend** Jedná se o nejdéle trvající trend a díky tomu i hlavní nastoupený trend na trhu. Délka trvání se pohybuje v rozmezí od jednoho roku až po několik let.

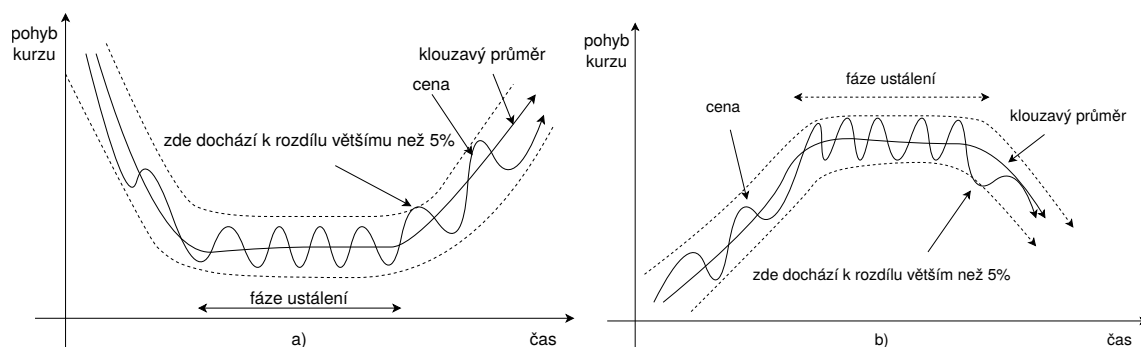
**Sekundární trend** Tento trend se od primárního trendu výrazně liší například v délce trvání, která se většinou pohybuje od 3 týdnů až do několika měsíců. Celkový pohyb trendu je charakterizován opačným pohybem oproti primárnímu trendu v rozpětí od 33% do 66%.

**Terciální trend** Doba trvání se pohybuje v rozmezí několika hodin až do 3 týdnů. Díky tomu je pro dlouhodobé investování nejméně důležitý a lze na něho nahlížet pouze jako na malé nevýznamné vlnky v rámci hlavního proudu.



Obrázek 2.9: Znázornění primárního, sekundárního a terciálního trendu

Dowova linie nám slouží k vymezení cenového pohybu trvajících od tří týdnů do několika měsíců. Při tomto pohybu kurzu pozorujeme jak nedochází k vychýlení od svého klouzavého průměru více než o 5%. Díky této linii je možno upozornit na změny trendu, které vedou ke změně na býčí či na medvědí. Pokud dojde k porušení horního pásma začíná se jednat o fázi akumulace, která indikuje počátek býčího trendu. V opačném případě, pokud by došlo k proražení dolního pásma, nastává fáze distribuční, což indikuje počátek medvědího trendu. Pro názornost je metoda zobrazena na následujícím obrázku 2.10.



Obrázek 2.10: a) Dowova linie předpovídající budoucí býčí trend; b) Dowova linie předpovídající budoucí medvědí trend

Dowova teorie také zohledňuje celkový objem provedených obchodů. Rostoucí počet nakupujících na trhu se projevuje u býčího trendu vysokým počtem obchodů. Pro zachování stejně rostoucího trendu je potřeba zachování stejného nákupního impulzu. Při výrazném snížení počtu nákupů dochází k ochlazení býčího trendu, což může vést až ke změně trendu. U medvědího trendu se setkáváme s obdobnou situací. Změny v počtu obchodů je ale nutné interpretovat jako pravděpodobnost změny trendu, ne však jako jistý fakt. Právě z tohoto důvodu se na objemy obchodů díváme jako na doplňkový ukazatel, který nám společně s hodnotou kurzu pomáhá se správně rozhodnout o možném budoucím pohybu kurzu.

Dowova teorie je i v současnosti velmi využívanou metodou technické analýzy, a to i přes fakt, že její základy byly položeny již více než před 100 lety. Díky jejímu správnému pochopení se již několikrát podařilo predikovat správný vývoj trhu, a i přes výskyty falešných signálů se její úspěšnost prokazuje. Velmi silná je právě v identifikaci hlavního trendu pomocí sledování postupně rostoucích lokálních maxim či minim trendu, což z ní dělá ideální metodu pro dlouhodobé investice. Jedná se o metodu, která následuje trend a díky tomu určuje, o jaký trend se jedná. Nedokáže však určit samotnou sílu trendu a dobu jeho trvání [14].

## 2.4 Trendová analýza

Trendová analýza si klade za cíl na základě historického vývoje směru kurzu vhodně predikovat budoucí vývoj. Setkáváme se zde se dvěma pomocnými liniemi, které nám slouží k vymezení vhodných obchodních pásem. Jedná se o linie podpory a linie odporu. Linie podpory charakterizuje přeprodaný trh. Investoři v blízkosti linie zastavují své prodeje kvůli tomu, že převis nabídky nad poptávkou klesl příliš nízko a ceny kurzů jsou příliš nízko. Je jen otázkou času, než se začnou podhodnocené kurzy opět nakupovat a vytvoří se nová poptávka. U linie odporu je situace opačná. Ta nám indikuje, že pokud kurzy vystoupají až k této hranici, většina investorů přestane za dané kurzy dále kupovat. Důvodem je, že to považují již za příliš drahé, a tím utlumují svoji poptávku. Pokud se kurzy vyskytují poblíž této linie, bývá trh označován jako překoupený a je jen otázkou času, kdy dojde k jeho postupnému klesání [4].

Nalezení trendu je pro úspěšnou technickou analýzu základním předpokladem. Pro podrobnější studium a identifikaci daného trendu se převážně využívají trendové linie, a to za pomoci sestavení hranic podpory a odporu. Můžeme se setkat i se třemi druhy trendových linií, a to v podobě rostoucí, klesající a stagnující trendové linie. U hlavní rostoucí trendové linie dochází ke spojování dvou a více minim do jedné hlavní linie. Pro klesající trendovou linii je situace podobná, spojují se dva a více vrcholů ve vývoji kurzu. U stagnující dochází k situaci, kdy se trh nepohybuje žádným výrazným směrem do žádné strany a dochází pouze k udržování stávající hladiny.

Důležité parametry pro trendové linie jsou jak délka její časové periody, po kterou zůstává trendová linie neporušená, tak i počty maxim a minim, které spojila svým sklonem. Vznik příliš strmé trendové linie značí, že nebude mít příliš dlouhého trvání. Čím déle zůstává linie neporušená, tím déle nastolený trend na trhu trvá a investoři investují na jeho základu. V praxi se investorům nejvíce osvědčilo změnit pokyn k prodeji či nákupu, pokud došlo k porušení trendové linie a kurz se posunul novým směrem o výraznější hodnotu. Díky tomu dokážeme rozlišit emocionální část, kdy se investor domnívá, že je vhodný čas začít prodávat, od analytického chování, při kterém investor začíná prodávat z důvodu proražení trendové linie.

### 2.4.1 Technické indikátory

Technické indikátory slouží jako ukazatele pro praktickou aplikaci na různých trzích. Dělí se do několika skupin podle jejich dílčího zaměření. Nejvýznamnější z nich jsou trendové, cenové a objemové indikátory. Jejich úkolem je poskytnout obchodníkovi další možnosti, podle kterých může odhadnout, jakým směrem bude trh dále pokračovat.

## 2.4.2 Trendové indikátory

Trendové indikátory se snaží následovat trend a díky tomu určovat jakým směrem se bude v budoucnosti dále vyvíjet. Nejvýznamnějším zástupcem je metoda klouzavého průměru. Tato metoda se snaží ukázat jak druh nastoleného trendu, tak i identifikovat jeho případnou změnu. Znázorňuje nám průměrnou hodnotu daných kurzů. Existuje několik variant klouzavého průměru [6].

### SMA

Jednoduchý klouzavý průměr se vyjadřuje jako aritmetický průměr na přesný počet historických údajů. Jeho velké výhody jsou v jednoduchosti a rychlosti samotného výpočtu. Nevýhodou ale je, že zohledňuje pouze data v ohraničeném datovém úseku a přiřazuje stejnou váhu všem hodnotám jak aktuálním, tak těm minulým. Klouzavý průměr je popsán rovnicí 2.1 a graficky znázorněný na obrázku 2.11.

$$\text{Jednoduchý klouzavý průměr} = \frac{P_M + P_{M-1} + \dots + P_{M-(n-1)}}{n} \quad (2.1)$$

$M$  značí délku periody

$P$  značí jednotlivé hodnoty kurzu, kde dolní indexy ukazují pořadí kurzových hodnot uvažovaných v rámci dané časové periody od nejmladší  $P_M$  po nejstarší  $P_{M-(n-1)}$  kde  $n$  je rovno  $M$

### EMA

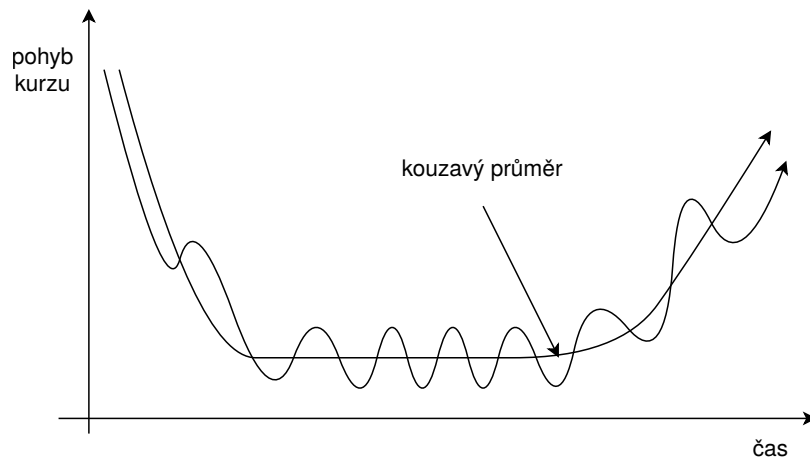
Jednou z nejoblíbenějších metod je exponenciální klouzavý průměr. Metoda spočívá v tom, že jednotlivým kurzům jsou při výpočtu přidělovány rozdílné váhy, které postupně exponenciálně rostou. Aktuální kurzy získávají nejvyšší váhy, zatímco nejstarší kurzy dostávají nejnižší váhy. Váha pro označení nejmladšího kurzu se nazývá vyhlazovací faktor. Ostatní váhy kurzů jsou přidělovány v závislosti na tomto vyhlazovacím faktoru. Jeden z nejpoužívanějších postupů pro výpočet exponenciálního klouzavého průměru lze zapsat tímto matematickým vztahem. Exponenciální klouzavý průměr je popsán rovnicí 2.3.

$$EMA_{\text{dnes}} = (EMA_{\text{včera}} \times (1 - ep)) + (P_{\text{dnes}} \times ep) \times n + (n - 1) + \dots + 2 + 1 \quad (2.3)$$

$EMA$  udává exponenciální klouzavý průměr

$ep$  udává vyhlazovací faktor

$P$  určuje samotný kurz instrumentu



Obrázek 2.11: Znázornění klouzavého průměru

### 2.4.3 Cenové indikátory

Jedná se o skupinu indikátorů, které jsou kalkulovány s využitím údajů o daných kurzech podobně jako u trendových. Při konstrukci indikátorů jsou jednotlivé objemy obchodů zanedbávány, avšak u některých metod mohou sloužit jako sekundární faktor. Dokáží také velice přesně zachytit intenzitu kolísání kurzu, aby bylo poté možné posoudit zesílení či zeslabení trendu, a tím následně zvažovat pravděpodobnost jeho změny.

Jeden z velmi oblíbených cenových indikátorů je index relativní síly (RSI) [4]. Jedná se o poměrně pomalejší oscilátor s přesně vymezenými a standardizovanými mezemi pro které jeho postupy platí. Metoda si klade za cíl přesně změřit vnitřní relativní sílu daného instrumentu [8]. Při výsledném výpočtu jsou zohledňovány jak záporné, tak i kladné změny kurzu v daném časovém úseku, díky čemuž dochází k vyjádření vnitřní síly daného instrumentu v čase. Hodnoty indikátoru kolísají v rozmezí od 0 do 100. Pokud hodnota vystoupí nad úroveň 70 signalizuje to překoupený trh, jakmile klesne pod 30 signalizuje nám to předprodaný trh. Index relativní síly je popsán rovnicí 2.4.

$$RSI = 100 - \frac{100}{1 + \frac{U}{D}} \quad (2.4)$$

$U$  je průměr kladných kurzových změn v daném časovém období

$D$  je průměr záporných kurzových změn v daném časovém období

### 2.4.4 Objemové indikátory

Zde je největší důraz kladen právě na sledování tendencí ve vývoji obchodů. Je-li na trhu nastolený silný trend, tak by měl být dále doprovázen rostoucím objemem obchodů. Pokud je trend slabý, měl by být doprovázen sníženým objemem obchodů. Za oslabení trendu je analyticky považován náhlý pokles obchodů, který může vyústit v samotnou změnu trendu. Platí pravidlo, že objem obchodů potvrzuje trend. Problém ale nastává v tom, že objemové indikátory bývají častým místem výskytu nejednoznačných signálů, stejně jako je tomu i u ostatních indikátorů. Právě proto se pro zvyšování spolehlivosti využívají kombinace s jinými typy indikátorů.

Jeden z nejznámějších indikátorů, se kterým se zde setkáváme, je právě indikátor peněžního toku (MFI) [4]. Tento indikátor funguje na podobném principu jako index relativní síly u cenových indikátorů. Jediný rozdíl nastává ve skutečnosti, že zde se pracuje jak s údaji o objemech provedených obchodů, tak i průměry kurzů za dané časové období. Vychází z předpokladu, že při rostoucích kurzech objem provedených transakcí výrazně roste, kdežto při klesajících kurzech výrazně klesá. Pro správné sestavení indikátoru je potřeba správně sestavit hodnotu toku peněz za dané časové období. Tok peněz se stanovuje jako součin průměrného denního kurzu a objemu obchodů.

## 2.5 Fundamentální analýza

Technická analýza ovšem není jediným možným stylem přístupu při analýze vývoje trhu. Fundamentální analýza si klade za cíl zjistit přesnou příčinu, proč došlo ke změně kurzu a na základě jakých faktorů. Například pokud je v nějaké zemi vyhlášena vysoká míra nezaměstnanosti, fundamentální analýza tyto údaje vezme v potaz a na jejich základě doporučí nákup či prodej. Díky tomuto přístupu je však mnohem náročnější na množství relevantních dat. Zde se snažíme najít odpověď, zda je správné daný finanční instrument kupovat [5].

Při fundamentální analýze jsou velmi podrobně zohledňovány finanční, politické, ekonomické a další faktory, které ovlivňují výsledný kurz. Na základě těchto dat se poté vhodně vytvoří predikce následujícího vývoje do budoucna. Získáváme tak i odpověď na otázku, proč se kurz zrovna změnil. Ale ani zde nejsou zohledňovány psychologické a iracionální faktory, které mají vliv na vývoj kurzu.

Základními cíli u obou druhů analýz je správné načasování nákupních a prodejních signálů. I přesto, že oba typy volí zcela odlišné přístupy, snaží se docílit stejného a správně stanovit prognózu pohybu kurzů. Podhodnocené kurzy, které se fundamentální analýza snaží nalézt mají výhodu v tom, že umožňují investorovi nakoupit za velmi nízkou cenu s vidinou vysokého zhodnocení vložené investice. K tomu, aby se daly správně nalézt podhodnocené kurzy, je potřeba provést podrobné zkoumání vnitřní hodnoty kurzu [16].

Rozdíly mezi oběma styly analýz spočívají již v samotném pohledu na data, se kterými pracují. Obě analýzy pracují nejen s historickými daty, ale i s daty aktuálními, avšak u fundamentální analýzy se aktuální data podílí mnohem výrazněji na změně kurzu. Trendy se mohou měnit v rámci minut, ale podhodnocené kurzy mohou na své zhodnocení čekat i dlouhé měsíce.

## Kapitola 3

# Klasifikace dat

Klasifikace dat představuje nástroj pro řešení problému, při kterém je potřeba správně vyhodnotit vzájemnou podobnost dvou skupin dat. Většinou se jedná minimálně o dvě rozdílné množiny dat. Samotná vzájemná klasifikace pak probíhá na principu porovnání jednotlivých prvků v obou množinách. Postupně se snažíme nalézt co nejpodobnější vzájemnou shodu pro jednotlivé úseky obou množin.

Algoritmy, které implementují pravidla pro samotnou klasifikaci se nazývají klasifikátory a řadíme je do několika kategorií. Například ty klasifikátory, jenž pracují pouze s celočíselnými hodnotami se využívají při klasifikaci, kde si můžeme dovolit jistou míru zkreslení výsledného porovnání, nebo když není potřeba pracovat i s čísly obsahujícími desetinnou část. Vyžaduje-li situace pracovat s reálnými čísly, existují klasifikátory, které se pro tuto práci dají využít.

Ve své práci vycházím z předpokladu, že již mám předem vytvořenou testovací množinu dat, která je využívána právě pro potřeby porovnávání se zkoumaným vzorkem dat, jenž představuje aktuálně hledaný trend na trhu. Každé porovnávání je vždy nezávislé na předchozím a určuje, do jaké míry se dané dvě skupiny dat vzájemně shodují.

### 3.1 Vyhledávání na základě měření vzdáleností

Tento druh metod představuje postupy, které slouží pro kvantifikaci blízkých objektů v daném zkoumaném metrickém prostoru. Samotné vyhledávání se nejčastěji využívá v případech, kdy množiny vytvořených dat obsahují velmi značné množství dat a je potřeba rychle a efektivně najít co nejpodobnější vzorec hodnot dat, která odpovídají vstupním datům. Samotné vyhledávání je přizpůsobené pro celou řadu úloh. Od prohledávání v metrickém prostoru za účelem nalezení nejpodobnějších hodnot, až po složitější druhy, kde je potřeba využití stromových modelů sloužících k porovnávání graficky reprezentovaných dat. Jednotlivé metody dále nevycházejí z předpokladu, že data jsou již v množinách seřazena [15].

Pro účely této práce byly vybrány algoritmy umožňující vyhledávání dat v již předem vytvořené databázi obsahující velké množství diskrétních hodnot. Z toho důvodu jsou použité algoritmy optimalizované právě na co nejefektivnější vyhledávání v daném metrickém prostoru. První vybranou metodou je Jaccardův koeficient. Tato metoda je dále využívána dvěma způsoby. První způsob představuje nezměněnou verzi původního algoritmu. Ve druhém způsobu byla metoda mírně upravena a rozšířena, díky čemuž v následném testování zastává roli další metody. Poslední vybranou metodou z této oblasti je porovnávání na základě rozdílu vzdáleností (Edit distance).

Druhá skupina metod, která byla vybrána, využívá pro vzájemnou kvantifikaci mezi množinami metody z oblasti K nejbližších sousedů. Na samotné prvky množin se pak nahlíží jako na vektory a provádí se výpočet vzdálenosti dvou vektorů. Cílem těchto metod je nalezení co nejmenší vzájemné vzdálenosti jednotlivých prvků uložených v daných množinách. První vybranou metodou z této oblasti je vyhledávání pomocí Manhattan vzdálenosti. Druhá metoda využívá Euclidean vzdálenost pro vzájemné porovnání. Třetí metodou byla zvolena Minkowskiho vzdálenost, která vychází z Euclidean, ale nabízí jiný pohled na její zpracování [12].

### 3.1.1 Jaccardův koeficient

Jaccardův koeficient popisuje způsob vzájemného porovnávání dvou sad dat na základě přesné shody číselných hodnot na odpovídajících pozicích. Při porovnávání vycházíme z toho, že data jsou uložena ve dvou množinách o stejné délce a při porovnání dochází k jejich vzájemnému posuzování na základě jednotlivých položek, které jsou uvnitř nich uloženy. Dochází k výpočtu vzájemné podobnosti, která je vyjadřována v procentech od 0% až do 100%. Čím je procento vyšší, tím je vzájemná podobnost dvou sad vyšší. Samotné porovnávání probíhá postupně pro jednotlivé hodnoty uložené v množinách. Mechanismus porovnání je popsán následující rovnicí 3.1.

$$d(A,B) = \frac{A \cap B}{A \cup B} * 100\% \quad (3.1)$$

**A** udává množinu dat, která chceme porovnávat s daty uloženými v databázi

**B** udává množinu dat, která jsou uložena v databázi

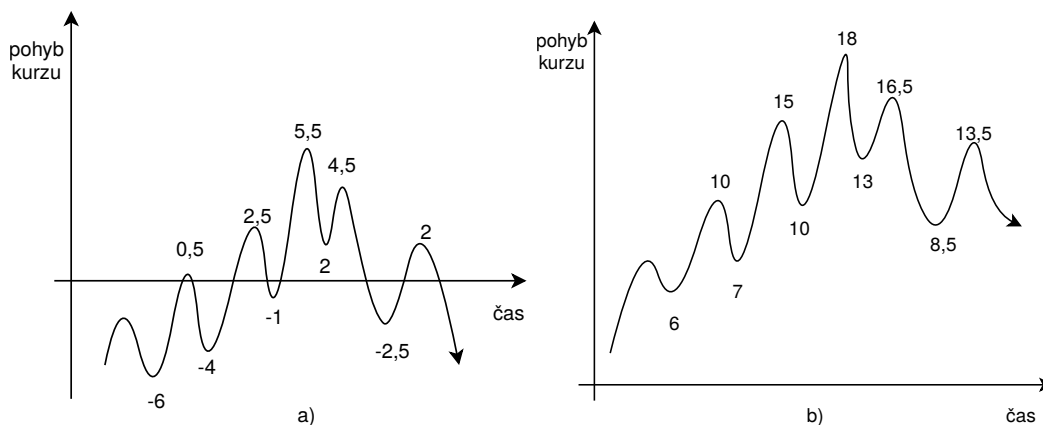
**d(A,B)** určuje vzájemnou podobnost zkoumaných množin

### Ukázka použití Jaccardova koeficientu

Následující ukázka demonstruje použití Jaccardova koeficientu na dvou množinách dat obsahující 10 různých položek, graficky znázorněných na obrázku 3.1. Nejprve je potřeba obě množiny dat patřičně upravit pomocí normalizačního koeficientu, který je popsán rovnicí 3.2. Normalizační koeficient slouží k určení vzájemné vzdálenosti obou množin, a na jeho základě jsou pak pro jednotlivé prvky zkoumané množiny vypočteny posunuté souřadnice. Výsledné dvě množiny se tímto vzájemně proloží. Díky tomu bude následně možné je vzájemně porovnávat, neboť budou reprezentovat data nacházející se na společné ose s jedním vzájemným bodem dotyku.

$$\text{normalizační koeficient} = \text{poslední prvek první množiny} - \text{poslední prvek druhé množiny} \quad (3.2)$$





Obrázek 3.1: Grafy znázorňující trendy na trhu

Pro názornost jsou zde hodnoty obou množin zaneseny do tabulky, kde jednotlivé řádky reprezentují danou množinu. Tabulka 3.1 zobrazuje množiny před provedením vzájemné normalizace. Tabulka 3.2 již popisuje množiny po provedení vzájemné normalizace.

první množina	-6	0,5	-4	2,5	-1	5,5	2	4,5	-2,5	2,5
druhá množina	6	10	7	15	10	18	13	16,5	8,5	13,5

Tabulka 3.1: Hodnoty množin před provedením normalizace

první množina	5	11,5	7	13,5	10	16,5	13	16	9,3	13,5
druhá množina	6	10	7	15	10	18	13	16,5	8,5	13,5

Tabulka 3.2: Hodnoty množin po provedení normalizace

Jakmile je zkoumaná množina znormalizována, přecházíme k výpočtu Jaccardova koeficientu. Zde již stačí pro obě množiny použít rovnici 3.2. Samotný výpočet je znázorněn rovnicí 3.1, kde vidíme, že vzájemná podobnost těchto dvou množin je 25%. Opakující se prvky jsou z toho důvodu, že je potřeba hlídat i pozici, na které se hodnota vyskytla.

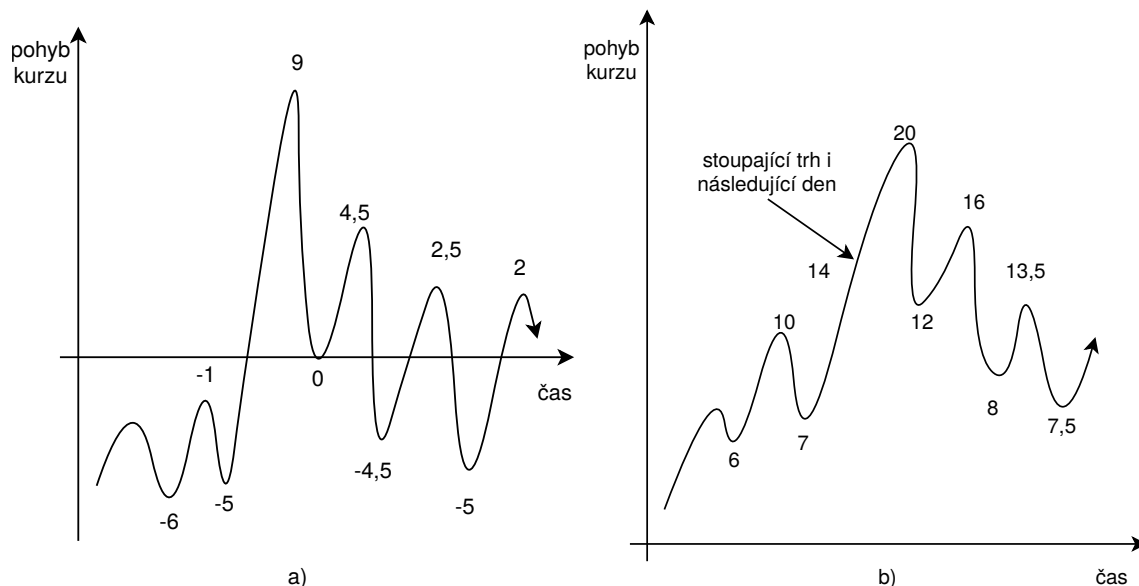
$$\begin{aligned}
 d(A,B) &= \frac{\{7, 10, 13, 13,5\}}{\{5, 6, 7, 8,5, 9,3, 10, 10, 11,5, 13, 13,5, 13,5, 15, 16, 16,5, 16,5, 18\}} * 100\% = \\
 &= \frac{4}{16} * 100\% = 25\%
 \end{aligned}
 \tag{3.1}$$

### 3.1.2 Rozšíření Jaccardova koeficientu

Pro potřeby této práce byl Jaccardův koeficient rozšířen, aby se z jeho pomoci daly porovnávat i dvě množiny, které jsou na první pohled značně rozdílné. Toto rozšíření bylo vytvořeno aby v případě, že se obě množiny nebudou přímo shodovat stále bylo možné detekovat správný směr vývoje trendu. Rozšíření zde spočívá v tom, že nedochází pouze k přímému porovnání jednotlivých vzájemně odpovídajících položek, ale i k porovnání položek nacházejících se o jednu hodnotu vlevo či vpravo.

## Ukázka použití rozšířeného Jaccardova koeficientu

Následující ukázka demonstruje použití rozšířeného Jaccardova koeficientu na dvou množinách dat obsahující 10 různých položek, graficky znázorněných na obrázku 3.2. Rozdíl v porovnání s klasickou Jaccardovou metodou je zde zvýrazněn v grafu.



Obrázek 3.2: Grafy znázorňují nastolený stejný trend na trhu s rozdílnou dobou trvání

Jednotlivé hodnoty z obou množin jsou zde pro zvýšení názornosti zaneseny do tabulky 3.3, kde řádky reprezentují dané množiny, před provedením vzájemné normalizace. Stejně jako v předchozím případě se i zde provede normalizace podle rovnice 3.1. Výsledná množiny po provedení normalizace však budou v tomto případě obsahovat dvě různě znormalizované skupiny dat. Pro názornost jsou data zanesena do tabulky 3.4. Dvě skupiny dat zde vznikly z důvodu toho, že vzájemná normalizace se provádí i pro hodnoty, které se nachází o jednu pozici vlevo. Takto upravená data nám umožňují například pro hodnotu 20 ve znormalizované tabulce použít i hodnotu 21 jako podobnou, a to i přes fakt, že je o jednu pozici vedle.

první množina	-6	-1	-5	9	0	4,5	-4,5	2,5	-5	2
druhá množina	6	10	7	14	20	12	16	8	13,5	7,5

Tabulka 3.3: Hodnoty zanesené z grafů výše před provedením normalizace

první množina	1	5	1	15	6	10,5	2	8,5	1	6
první množina	6	11	7	21	12	16,5	8	14,5	7	12
druhá množina	6	10	7	14	20	12	16	8	13,5	7,5

Tabulka 3.4: Hodnoty po provedení normalizace pro dvě různé odchylky

Jakmile je zkoumaná množina znormalizována, přecházíme k samotnému výpočtu Jaccardova rozšířeného koeficientu. Zde je nutné pro každý prvek provést dva až tři dílčí vý-

počty průniků jednotlivých prvků množin. Například pro první znormalizovanou množinu z tabulky 3.4 to bude 0% a pro druhou hodnotu 20%. Následující výpočet je potřeba provést v každém kroku Jaccardova koeficientu. První znormalizovaná množina je znázorněna rovnicí 3.4 a druhá 3.5. Výsledky se na konci vzájemně porovnají a ten s větším parametrem shody je vybrán. Opakující se prvky jsou z toho důvodu, že je potřeba hlídat i pozici, na které se hodnota vyskytla.

$$d(A,B) = \frac{\emptyset}{\{0, 1, 1, 2, 5, 6, 6, 6, 7, 7.5, 8, 8.5, 10, 10.5, 12, 13.5, 14, 15, 16, 20\}} * 100\% = 0\% \quad (3.4)$$

$$d(A,B) = \frac{\{6, 7, 8, 12\}}{\{6, 7, 7, 7.5, 8, 10, 11, 12, 12, 13.5, 14, 14.5, 16, 16.5, 20, 21\}} * 100\% = \frac{6}{16} * 100\% = 20\% \quad (3.5)$$

### 3.1.3 Rozdíl vzdáleností

Rozdíl vzdáleností (anglicky Edit distance) umožňuje porovnávat dvě množiny a určit jaké nezbytné množství kroků je potřeba učinit, aby se tyto dvě množiny kompletně rovnaly. Porovnávání probíhá postupně pro jednotlivé položky obou množin. Výsledná hodnota celého porovnání nakonec stanoví přesný počet potřebných kroků ke změně. Cílem této metody je dosáhnout co nejmenšího počtu kroků, které by vedly k určení vzájemné podobnosti [15].

Názornost této metody je zobrazena na uvedeném příkladu, který využívá data z tabulky 3.3. Jednotlivé prvky z tabulky jsou zaneseny do tabulky 3.5, kde je znázorněno, jak samotné porovnávání probíhá. Například pro první dva prvky množin znázorněných v tabulce bude porovnávání probíhat následujícím způsobem. V kroku 1 se provede porovnání kolik změn je potřeba učinit, aby se modifikovala množina obsahující prvek {0} na množinu obsahující prvek {0}. Počet změn je roven 0. V druhém kroku je situace obdobná a zkoumáme kolik změn je potřeba učinit, aby se množina {0,6.5} rovnala {0,4}. Počet změn je roven jedné. Kolik kroků je potřebných pro změnu posloupnosti {0;6.5} na {0;4} -> 1 krok. Výsledný rozdíl při porovnání obou množin je nakonec roven sedmi.

		0	6.5	2	8.5	5	11.5	8	10.5	2.5	8
	0	1	2	3	4	5	6	7	8	9	10
0	1	0	1	2	3	4	5	6	7	8	9
4	2	1	1	2	3	4	5	6	7	8	9
1	3	2	2	2	3	4	5	6	7	8	9
9	4	3	3	3	3	4	5	6	7	8	9
4	5	4	4	4	4	4	5	6	7	8	9
12	6	5	5	5	5	5	5	6	7	8	9
7	7	6	6	6	6	6	6	6	7	8	9
10.5	8	7	7	7	7	7	7	7	6	7	8
2.5	9	8	8	8	8	8	8	8	7	6	7
7.5	10	9	9	9	9	9	9	9	8	7	7

Tabulka 3.5: znázornění použití rozdílu vzdáleností

### 3.1.4 Manhattanova vzdálenost

Manhattanova vzdálenost představuje prvního zástupce odlišné skupiny metod, které pro výpočet své vzájemné vzdálenosti využívají  $K$ -nejbližších sousedů [10]. Samotná metoda vyjadřuje poměr vzdáleností dvou vektorů. Název metody je odvozen od uspořádání ulic do tvaru připomínající mříž, jenž se nacházejí na Manhattanu. Metoda má reprezentovat potřebu nalezení nejkratší cesty k cíli pro auta, která se na Manhattanu nacházejí [19].

Metoda má své využití v případech, kdy je potřeba prohledávání metrického prostoru. Zde pracuje na principu měření podle os  $x$  a  $y$ , které spolu svírají pravý úhel. Samotný výpočet spočívá v tom, že se pro jednotlivé dva vektory vypočte jejich vzájemný rozdíl v absolutní hodnotě. Pro potřeby této práce se osa  $x$  zanedbává a vzájemné porovnání se provádí jen na ose  $y$ . Tento vztah je reprezentován rovnicí 3.6. Velkou výhodou této metody je její vysoká rychlost, avšak daní za to je menší přesnost v porovnání s ostatními metodami.

$$d(A,B) = |N_1 - M_1| + |N_2 - M_2| + \dots + |N_n - M_n| \quad (3.6)$$

$N$  značí prvek z množiny dat, který chceme porovnávat s daty uloženými v databázi

$M$  značí prvek z množiny dat, který slouží k porovnávání

$n$  určuje index daného prvku v množině

$d(A,B)$  určuje vzájemnou podobnost zkoumaných množin

### 3.1.5 Euclideaná vzdálenost

Euclideaná vzdálenost vyjadřuje vzájemnou vzdálenost dvou vektorů v metrickém prostoru. Ke svému výpočtu metoda využívá Pythagorovu větu v nezměněné podobě. Ta udává, že vzájemná vzdálenost mezi dvěma body v metrickém prostoru je dána rozdílem jejich souřadnic umocněných na druhou a to celé odmocněno dvěma. Pro potřeby této práce budou opět souřadnice na ose  $x$  zanedbány a výpočet se bude provádět pouze pro hodnoty na ose  $y$  [1].

Výsledek této metody je hodnota v rozmezí nula až maximální vypočítaná hodnota. Naprostou shodu mezi dvěma vektory nám značí hodnota nula. V případě neshodnosti je velikost dána maximální hodnotou dosaženou ve výpočtu. Pro porovnání dvou množin bude výpočet odpovídat rovnici 3.7.

$$d(A,B) = \sqrt{(N_1 - M_1)^2 + (N_2 - M_2)^2 + \dots + (N_n - M_n)^2} \quad (3.7)$$

$N$  značí prvek z množiny dat, který chceme porovnávat s daty uloženými v databázi

$M$  značí prvek z množiny dat, který slouží k porovnávání

$n$  určuje index daného prvku v množině

$d(A,B)$  určuje vzájemnou podobnost zkoumaných množin

### 3.1.6 Minkowskiho vzdálenost

Minkowskiho vzdálenost představuje obecnou metodu pro popis vzájemné vzdálenosti v metrickém prostoru. Ta je reprezentována jako poměr vzdálenosti dvou vektorů. Díky svému obecnému popisu umožňuje, aby z ní jiné metody mohly vycházet a stávat se tak

obecnějšími ukázkami této metody. Názorně je to vidět na příkladu Euclidean vzdálenosti, která z ní vyplývá a zobecňuje ji. Minkowskiho metoda však nabízí mnohem vyšší možnosti využití pro hledání největší shody dat. Problém zde však v porovnání s předchozí metodou nastává kvůli zvýšení časové náročnosti na samotný výpočet [11]. Pro porovnání dvou množin bude výpočet odpovídat rovnici 3.8.

$$d(A,B) = \sqrt[k]{(N_1 - M_1)^k + (N_2 - M_2)^k + \dots + (N_n - M_n)^k} \quad (3.8)$$

**N** značí prvek z množiny dat, který chceme porovnávat s daty uloženými v databázi

**M** značí prvek z množiny dat, který slouží k porovnávání

**n** určuje index daného prvku v množině

**k** umožňuje nám zvolit hodnotu exponentu, se kterou bude metoda pracovat

**d(A,B)** určuje vzájemnou podobnost zkoumaných množin

Nejčastějšími parametry hodnoty **k** jsou 1, 2, 3 a 4. Při parametru jedna se metoda nazývá Hammingova vzdálenost a při parametru dva Euclidean vzdálenost. Pro účely této práce bylo vybráno použití hodnoty čtyři. Za cenu zvýšení výpočetní zátěže je očekáváno zlepšení přesnosti nalezení správné vzdálenosti.

## 3.2 Shrnutí jednotlivých metod

Rozdíly mezi jednotlivými metodami primárně spočívají v tom, jakým způsobem je provedeno vzájemné porovnání vzdálenosti, a jak se to projeví do výsledné náročnosti výpočtu. U některých metod, jako je například rozšířený Jaccardův koeficient, bude zvýšení náročnosti výpočtu znát nejvíce, kvůli porovnávání nejen samotné hodnoty, ale i její sousední hodnoty ve vzdálenosti jedna. Naopak u metody Manhattan, kde dochází pouze k výpočtu absolutní hodnoty vzájemné vzdálenosti dvou vektorů se očekávají nároky na výpočetní výkon řádově nižší.

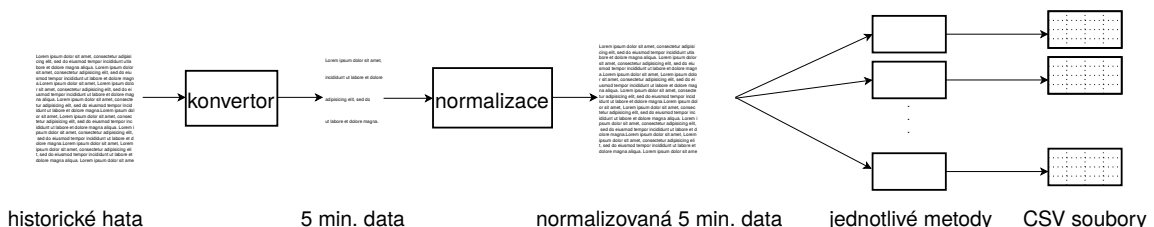
Vzhledem k tomu, že všechny zde zmíněné metody dokáží určit vzájemný poměr vzdáleností pouze pro dané parametry dvou množin, může nastávat problém, že vzájemná podobnost dvou množin dat bude velmi nízká, a to i za předpokladu, že jsou si tyto množiny velmi podobné. Ideálním řešením tohoto problému je využití vzájemné odchylky. Odchylka popisuje, jaká maximální vzájemná vzdálenost mezi dvěma prvky daných dvou skupin dat je ještě přijatelná a jde je pak považovat za shodné. Podrobněji je tato část rozepsána v sekci věnující se samotnému experimentování s danými metodami v kapitole 5.1.

## Kapitola 4

# Návrh a implementace metod pro analyzování trhu

Obsahem této kapitoly je popis návrhu a implementace jednotlivých metod pro potřeby analyzování právě sledovaného trendu na trhu. Jako testovací prostředí byl zvolen OS Linux a pro samotnou implementaci byl vybrán skriptovací jazyk Python. Ten byl zvolen z důvodu toho, že při testovací části daných metod není potřeba docílit vysoké optimalizace na výkon, který by nám představovalo použití například jazyka C, ale vystačíme si s dostatečnou silou právě jazyka Python.

V jednotlivých podkapitolách je popsán způsob, jakým byly jednotlivé skripty konstruovány, na základě jakých parametrů pracují a jak vypadá jejich výstupní soubor. Tato sada skriptů má za cíl rozložit práci s analyzováním daného trendu na jednotlivé dílčí kroky, které pak následně mají společně vhodně predikovat vývoj trendu. Na obrázku 4.1 je prezentováno schéma systému pro predikci vývoje trhu.



Obrázek 4.1: Schéma architektury

### 4.1 Příprava dat z trhu

Samotná analýza začíná výběrem vhodného měnového trhu, pro který se bude provádět predikce budoucího vývoje trendu. Tento výběr je velmi důležitý, neboť pro některé aktivnější měnové páry, jako je například EUR/USD, budou získaná data více dynamičtější a pro účely následného vyhodnocování a predikování trendu budou mít mnohem vyšší vypovídající hodnotu než trhy, které po většinu svého vývoje pouze stagnují.

Data, která reprezentují historický vývoj směru daného kurzu, byla získána z volně dostupného zdroje na adrese [www.histdata.com](http://www.histdata.com). Tato stránka nabízí k dispozici značné množství dat pro jednotlivé měnové páry. Pro ty lze dohledat i záznamy historických hodnot,

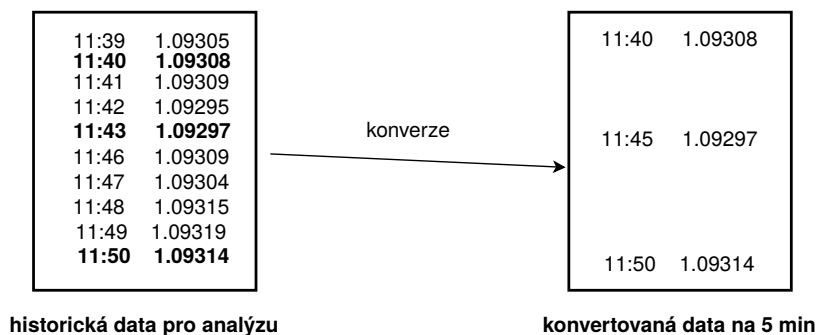
kteře sahají až patnáct let do minulosti. Pro účely této práce byla využita data pouze po roce 2008 a to z důvodu doporučeného v knize [4], kde autor nabádá využívat data pro technickou analýzu pouze po tomto roce kvůli finanční krizi, která ovlivňovala hodnoty kurzů před tímto rokem.

Získaná data reprezentují vývoj dané měny, kde je pro každou minutu zaznamenána patřičná hodnota vývoje kurzu. V této práci zde byl vytvořen skript, který nejprve všechna tato data z minulých let spojí do jednoho souboru. Ten bude reprezentovat sumu historických dat a bude sloužit jako vstupní soubor pro následné zpracování. Dále umožňuje na základě vstupního parametru upravit data na patřičné časové rámce, která již nebudou obsahovat obchody provedené pro každou minutu. To znamená, že pokud bychom chtěli zkoumat trh pouze po úsecích dlouhých pět minut, tento skript dostane na vstup právě parametr pět a celou historickou skupinu dat naseká tak, aby obsahovala pouze data s rozstupem každých pět minut. Množina přípustných hodnot, pro které skript dokáže data upravit, je od jedné minuty až po šedesát minut, což reprezentuje případ, kdybychom chtěli zkoumat vývoj trhu na základě hodinových dat.

Důvod, proč je vhodné provést úpravu časových údajů, které jsou v historické databázi, je především kvůli rychlosti následného zpracování. Vzhledem k tomu, že data se upraví na časové rámce například po pěti minutách, umožní metodám provést analýzu historického vývoje a vytvořit predikci budoucího. Pokud by data reprezentovala minutové údaje, proces porovnávání by trval velmi dlouho.

Problém zde však představují mezery v zaznamenaných datech. Tyto mezery se mohou projevat například chybějícími daty po dobu od jedné minuty až v řádu desítek minut. Proto je tohle potřeba brát v úvahu. Skript, který provádí danou úpravu dat, tohle v úvahu bere a provádí nalezení nejbližší hodnoty, která bude odpovídat zadanému koeficientu pro nasekání.

Například může nastat situace, kdyby skript byl spouštěn s parametrem pro úpravu na obchody provedené každých pět minut. Údaje, které by v něm byly uloženy, by už číslo pět neobsahovaly a jako adekvátní náhrada by byla zvolena nejbližší hodnota směrem k přecházejícímu číslu, jenž reprezentuje pětiminutový údaj. V krajním případě se může stát, že kvůli vzniklé mezeře v zaznamenaných datech bude výsledný soubor obsahovat obchod provedený po pěti minutách a další provedený například po deseti. Příklad konverze dat na pětiminutové úseky je znázorněn na obrázku 4.2. Tento problém je v následných skriptech řešen tak, že se nesnažíme nalézt jen jednu nejpodobnější kombinaci dat. Při vyšším počtu nalezených podobných datech je tento problém zanedbán. Tento proces je popsán v následujících kapitolách.



Obrázek 4.2: Konverze dat na 5-ti minutové úseky

## 4.2 Normalizace dat z trhu

Jakmile jsou data pomocí předcházejícího skriptu upravena tak, aby jejich vzdálenosti pro jednotlivé časové záznamy odpovídaly požadovaným vstupním parametrům, je celý soubor historických dat předán dalšímu skriptu, jenž má za cíl tato data patřičně znormalizovat vůči aktuálnímu sledovanému trendu. To v praxi znamená, že pro celý soubor obsahující historická data bude interně v těle skriptu provedena normalizace na úseky dat, které budou svou délkou odpovídat aktuálnímu zkoumanému trendu. Tato délka je dána vstupním parametrem skriptu. Jednotlivé úseky jsou v těle skriptu ukládány do seznamu.

Soubor historických dat obsahuje celkem čtyři dílčí informace o každém dílčím kurzu. Tyto informace jsou hodnoty uzavírajícího, otevírajícího, maximálního a minimálního kurzu, se kterým se v daném úseku obchodovalo [4]. V této práci byly všechny skripty nastaveny tak, aby pracovaly pouze s jednou hodnotou, kterou zde reprezentuje uzavírající kurz. Díky tomu dochází ke značné odchylce, způsobené tím, že se bude dále pracovat pouze s jednou hodnotou ze čtyř.

V průběhu samotného ukládání jednotlivých úseků do seznamu je pro ně vždy vypočítán normalizační koeficient, který je popsán rovnicí 3.2 z třetí kapitoly. Na základě získaného normalizačního koeficientu je následně provedena samotná normalizace. Ta zajistí, aby jednotlivé úseky byly patřičně posunuty na společnou osu. Na této ose budou mít oba úseky jeden společný bod dotyku. Ostatní prvky se díky normalizačnímu koeficientu taktéž proloží na osu, díky čemuž jsou oba dva grafy vyvážené a ve stejném výchozím stavu.

Jako jednotlivé parametry skriptu slouží v tomto případě délka, pro kterou se provádí vytvoření úseků dat. Tato délka odpovídá velikosti aktuálně zkoumaného úseku dat, aby obě množiny byly stejně dlouhé a byla zvolena v rozmezí od třiceti minut až po dvě a půl hodiny. Pro takto vytvořené úseky dat je následně provedena jejich normalizace. Jako vstupní soubor zde slouží soubor historických dat, které jsou výstupním souborem předcházejícího skriptu.

Výstupním souborem skriptu je speciálně upravená skupina dat, která jsou již znormalizována a svou délkou odpovídají velikosti hledaného aktuálního trendu. Data jsou připravena v takovém formátu, aby je dále jednotlivé skripty reprezentující samotné metody mohly snadno využívat a pracovat s nimi.

## 4.3 Jednotlivé metody

Jakmile jsou historická data znormalizována podle aktuálního sledovaného úseku dat, který reprezentuje aktuální sledovaný trend, přecházíme k samotnému využívání jednotlivých klasifikačních metod. Cílem těchto metod je na základě již znormalizovaného souboru provést nalezení co nejpodobnějšího úseku historických dat, jenž bude simulovat pokračování aktuálního trendu.

Samotná implementace metod byla rozdělena do dvou skupin. První skupina představuje metody, které výsledný poměr podobnosti dvou úseků dat vyjadřují v procentech od nuly až do sta, kde hodnota nula reprezentuje situaci, pro kterou se dané dvě skupiny dat vůbec navzájem nepodobají. Jejich podobnost postupně roste až k hodnotě sto, která reprezentuje maximální možnou podobnost. Mezi tyto metody patří Jaccardův koeficient, rozšířený Jaccardův koeficient a rozdíl vzdáleností. Druhou skupinu dat představují metody z kategorie K-nejbližších sousedů. Tyto metody výslednou podobnost představují číslem od nuly do maximálního vypočteného čísla. Čím blíže je číslo bodu nula, tím jsou si metody vzájemně podobnější.



Vstupními parametry jednotlivých metod jsou právě znormalizované soubory, které slouží pro hledání co nejpodobnějších úseků dat. Dále je pomocí vstupních parametrů nutné zadat, jakou délku má úsek trendu, který zrovna sledujeme. Mimo to je zde nutné zadat parametr, který udává jaké množství nejpodobnějších paternů se z danými parametry má v normalizovaném souboru nalézt.

Problém, který zde nastává je v tom, že při použití jednotlivých metod na samotná historická data, není šance na nalezení podobného výsledku velmi vysoká. To je způsobeno tím, že dva trendy se v historii úplně přesně neopakují. Z toho důvodu je potřeba vzít v potaz vzájemnou odchylku mezi aktuálními daty reprezentujícími hledaný patern a historickými daty, pro které je samotné vyhledávání prováděno. Tato odchylka je vyjádřena desetinným číslem, a udává, jak moc velká je přípustná rozdílnost mezi dvěma skupinami dat.

Například pokud bychom měli čísla 1.01511 a 1.01522, tak jejich vzájemná podobnost bude nízká, avšak při použití odchylky vzájemné vzálenosti například 0,00006 se obě dvě hodnoty začnou překrývat a jejich vzájemná shodnost bude sto procent. V praxi to znamená, že číslo 1.01511 bude pro následující odchylku zahrnovat hodnoty v rozmezí 1.01505 až 1.01517 a druhé číslo 1.01522 bude zahrnovat hodnoty od 1.01516 až po hodnotu 1.01528.

Samotný problém s nalezením vhodné vzájemné odchylky pro celou skupinu znormalizovaných dat není snadný, neboť každé zkoumání dat je odlišné a není možné předem odhadnout, jaká celková odchylka je pro výpočet potřebná. Z toho důvodu je proces řešení automatickou detekcí. To znamená, že skript provede první nalezení co nejpodobnější skupiny dat v historickém úseku dat při co nejnižší možné odchylce. Pokud bylo požadováno nalezení například 30 nejpodobnějších metod, tak skript nalezne 30 metod, které jsou podle něj co nejpodobnější. Jakmile jsou tyto výsledky získány, přecházíme do části automatické detekce odchylky. Tento proces se v případě nouze opakuje do té doby, dokud nebude nejvhodnější odchylka nalezena.

### 4.3.1 Postupy pro nalezení automatické odchylky

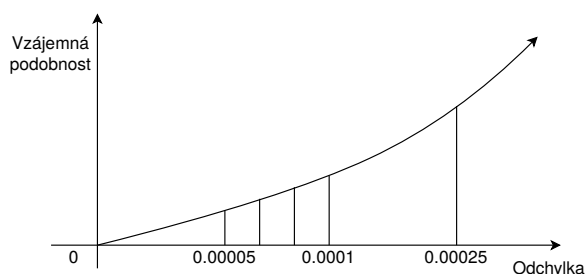
Pro první skupinu metod, kterou reprezentuje Jaccardův koeficient, rozšířený Jaccardův koeficient a rozdíl vzdáleností, probíhá výpočet automatické detekce odchylky následujícím způsobem. V prvním kroku skripty provedou nalezení co nejpodobnějších skupin dat s minimální vzájemnou odchylkou. Minimální odchylka se liší v závislosti na typu trendu, zda-li se jedná o krátkodobý nebo dlouhodobý. Samotná hodnota byla odvozena od experimentálního pozorování chování trhu.

Jakmile při výpočtu vzájemné podobnosti nastane situace, kdy shoda nalezených metod vůči aktuálnímu trendu bude například pouze 60%, dojde k procesu výpočtu automatické odchylky na základě právě dosažených procent, a to za účelem zvýšení procentuální vzájemné podobnosti pro následující opakování. Ve skriptu je nastavena minimální povolená podobnost dvou skupin dat na 80%. Pokud tato podmínka není splněna, provede se znovu výpočet vhodné vzájemné odchylky a následně opět výpočet podobnosti. Tento proces se opakuje, dokud se pro všechny nejpodobnější úseky dat, jejichž množství je dáno vstupním parametrem skriptu, nedosáhne minimální míry podobnosti 80%. Každé nalézání nejpodobnější skupiny dat trvá v rozmezí tří až pěti minut pro všechny jednotlivé metody.

U druhé skupiny dat je výpočet automatické odchylky odlišný. To je způsobeno tím, že vzájemná podobnost dvou skupin dat není reprezentována procentuálně, ale číselnou hodnotou v rozmezí 0 až maximální vypočtená hodnota. Proto u těchto metod výpočet automatické odchylky je prováděn rozdílným stylem. V prvním kroku je vypočtena vzájemná vzdálenost s minimální vzájemnou odchylkou, která je dána typem trendu který pozorujeme.

To znamená, že pro krátkodobý trend, bude velikost minimální odchylky různá v porovnání s trendem dlouhodobým. Velikost je dána množstvím možných podobných trendů. Při krátkodobém trendu jsou šance na nalezení vyššího počtu podobných trendů větší již při nižší odchylce v porovnání s trendem dlouhodobým.

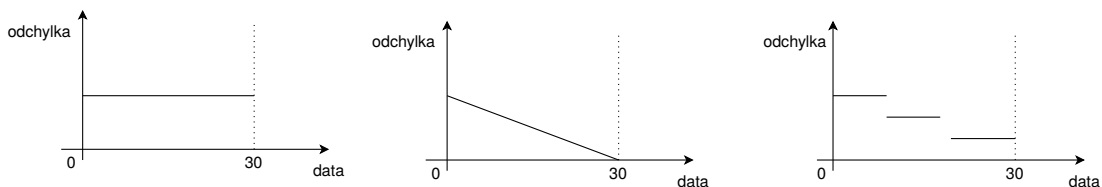
V následujícím průchodu je odchylka zvednuta a pozoruje se vzájemný rozdíl mezi hodnotami. Na základě daného rozdílu je vypočtena odchylka a pomocí níž se provede opětovný výpočet pro nalezení vzájemné podobnosti. Odchylka se vypočítává vzájemným poměrem aktuální procentuální shody a minimální procentuální shody, která je potřebná pro správné nalezené podobné množiny dat. Pokud se vzájemná hodnota nalezených podobných hodnot po dvou průchodech liší a nově vypočtená hodnota s odchylkou je vyšší, znamená to, že odchylka byla příliš vysoká, a dané dvě hodnoty se již od sebe více vzdalují. Pokud by hodnota vyšla po druhém provedení s odchylkou nižší, znamená to, že jsme úpravou odchylky dosáhli zvýšení přesnosti a našli podobnější skupiny dat. Na obrázku 4.3 je znázorněna závislost odchylky na podobnosti daných dvou množin.



Obrázek 4.3: Závislost odchylky na podobnosti

Avšak kvůli tomu, že hodnoty obchodů provedených aktuálně či krátce do minulosti mají vyšší vypovídající hodnotu, než hodnoty kurzů provedených více v minulosti, byly pro jednotlivé metody připraveny i různé varianty samotných metod sloužící pro práci s odchylkou. Jednotlivé způsoby jsou graficky znázorněny na obrázku 4.4. Pro každou metodu vznikají tři dílčí způsoby, jakými pracuje s danou odchylkou:

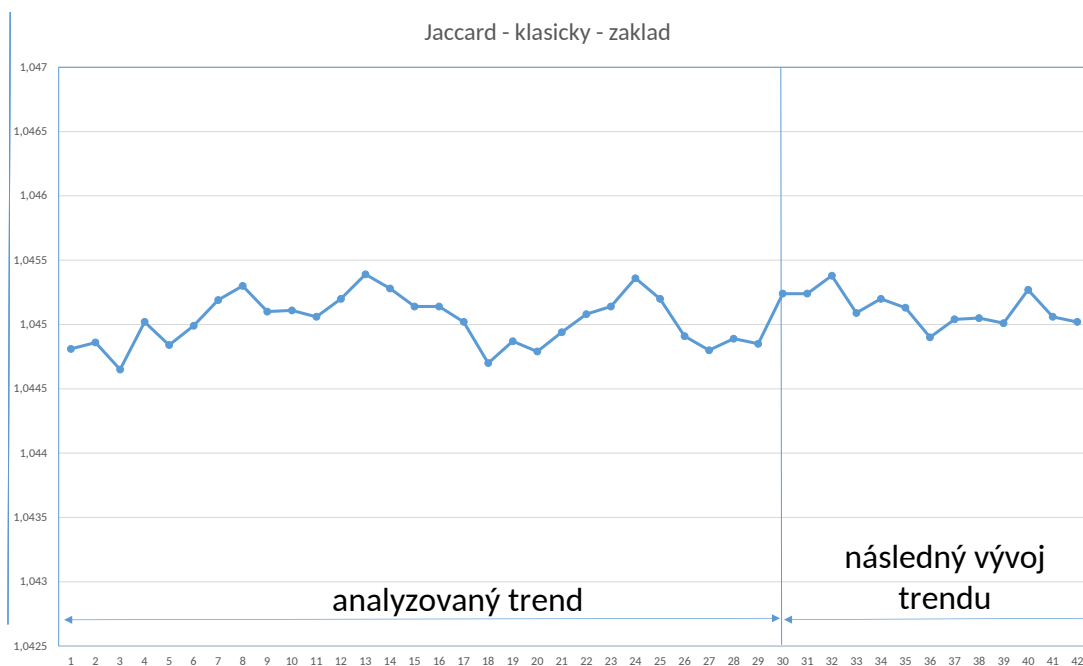
- První způsob je pomocí základního tvaru odchylky, která je stejná pro všechny jednotlivé položky ve zkoumaném úseku dat.
- Druhý způsob ukazuje hodnotu odchylky, která bude mírně růst směrem do minulosti, což znamená, že pro data, která jsou aktuální, bude tato odchylka nejnižší, ale pro data směrem do minulosti bude postupně růst až do délky zkoumaného vzorku dat, například do hodnoty 25, při zkoumání vzorku dat o délce 25.
- Třetí metodou je systematické rozdělení odchylky na jednotlivé úseky dat, pro které bude odchylka mírně rozdílná. Například pokud délka zkoumaného paternu bude 25, tak se tato délka rozdělí na čtvrtiny a pro jednotlivé úseky se bude provádět výpočet s rozdílnou odchylkou odpovídající tomu, ve které čtvrtině se bude nacházet. Opět platí, že nejnižší odchylka bude u co nejaktuálnějších dat. U ostatních bude postupně narůstat.



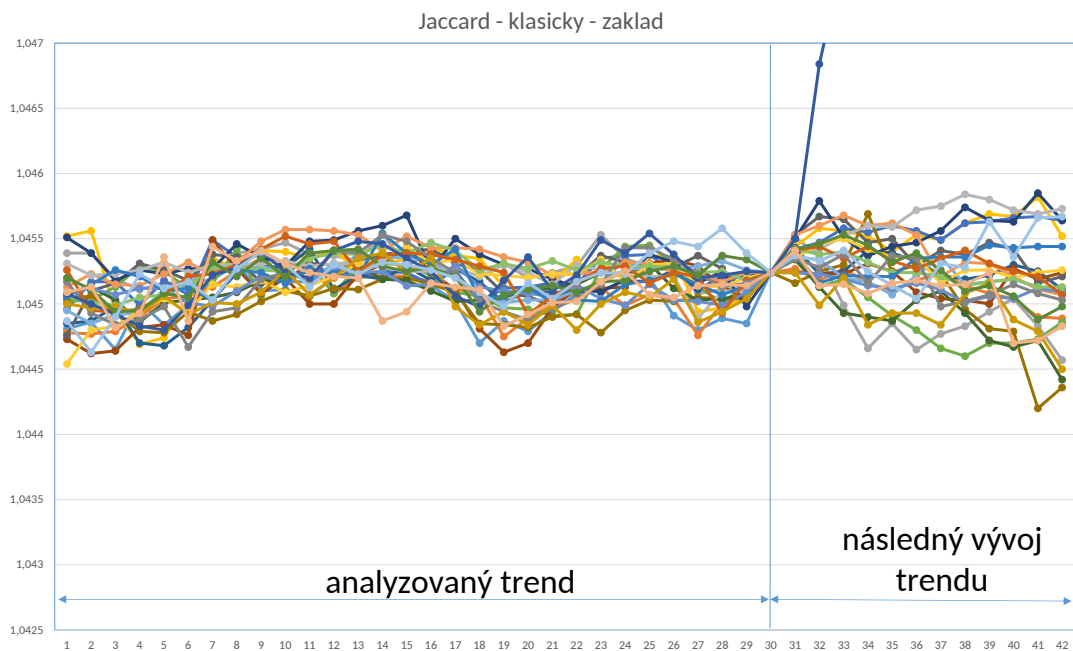
Obrázek 4.4: Jednotlivé způsoby pro práci s odchylkou

Na základě toho bude celkový počet jednotlivých skriptů vyšší. To znamená, že pro každou metodu budou potřeba tři dílčí verze jedné metody, kde každý z nich bude odchylku počítat zcela jiným způsobem. Tyto úpravy se projeví na celkovém zpomalení dané metody, ale získané pozitivum spočívá v tom, že nalezené skupiny dat, pro které podobnost platí, jsou přesnější.

Výstupními soubory pro jednotlivé skripty byly zvoleny csv soubory [17]. Ty obsahují seznam jednotlivých nejpodobnějších úseků dat, které odpovídají zkoumanému paternu. Díky tomu se tento soubor dá jednoduše vykreslit v tabulkovém editoru do podoby grafu a je názorně vidět, jak moc jsou si jednotlivé skupiny dat podobné. Na následující straně je ukázka výstupního csv souboru, jenž byl převeden do grafické podoby.



Obrázek 4.5: Graf reprezentující aktuální trend na trhu



Obrázek 4.6: 25 nejpodobnějších historických trendů

První graf 4.5 reprezentuje aktuální trend jenž na trhu právě probíhá. Následující obrázek 4.6 zobecňuje sadu nejpodobnějších dat, nalezených metodou pro rozdíl vzdáleností. Testování v tomto případě probíhalo pro nalezení 25 nejpodobnějších dat. V úseku a je zobrazen minulý vývoj a zároveň jakým směrem se jednotlivé metody dále vyvíjely. U jedné sady dat je patrné, že v minulosti velmi strmě začala stoupat. Z toho důvodu je výhodné mít větší sadu nejpodobnějších dat, aby se tyto vlivy patričně zanedbaly a zprůměrovaly s ostatními metodami. Je patrné, že trend na trhu po jedné hodině obchodování klesl, což nám i tyto metody predikovali.

Kromě csv souboru, který obsahuje skupinu co nejpodobnějších nalezených vzorků dat, generuje skript i další csv soubor, jenž data z předchozího csv souboru zobrazuje v podobě, která reprezentuje jakým směrem se trend vydal. Tuto informaci podává ve formě, která popisuje, kolikrát v podobné situaci šel vzhůru, kolikrát klesal a kolikrát stagnoval. Pro předešlý případ by predikce vypadala následujícím způsobem. Jedna nalezená metoda v minulosti klesla, osm stoupla a šestnáct klesalo. Z toho vyplývá, že metoda by doporučila investorovi, aby se zaměřil na to, že trh bude v následující hodině klesat. Názorně je výstup jednotlivých metod znázorněn na obrázku 4.7.

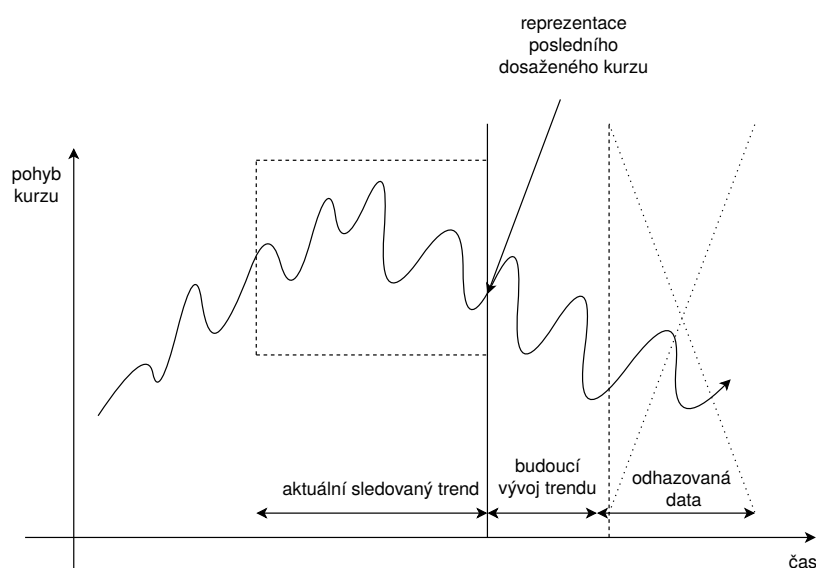
Rovno	1
Mensi	16
Vetsi	8

Obrázek 4.7: Vyhodnocení 25 nejpodobnějších trendů

## Kapitola 5

# Testování a vyhodnocení jednotlivých metod

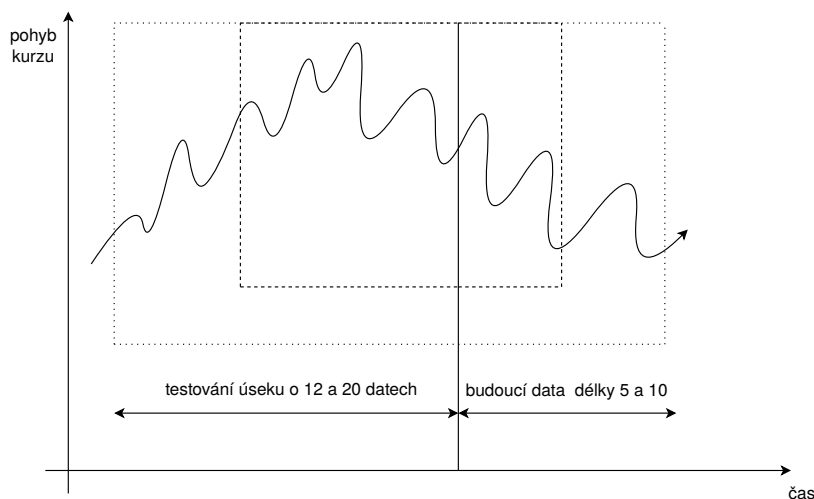
Po dokončení implementace samotných metod jednotlivých skriptů se přešlo k jejich postupnému testování. Každá jednotlivá metoda byla otestována na vstupních testovacích souborech, které simulovaly aktuální situaci na trhu. Celé testování probíhalo na datech, které byly patřičně upraveny, aby obsahovaly hodnoty kurzů po pětiminutových úsecích. Data potřebná pro simulaci se získávala tak, že se od konce testovacího souboru odřezávala část dat, aby na historických datech simuloval aktuální stav trhu. Způsob, jakým byla testovací data získávána, umožňoval nejen mít informaci o podobě daného trendu v minulosti, ale také informaci o tom, jak se daný trend dále vyvíjel. Díky tomu bylo umožněno porovnávat, zda tento typ přístupu k datům na principu technické analýzy funguje, či nikoliv. Ukázka, jakým způsobem probíhá příprava dat pro samotné testování je znázorněna na obrázku 5.1.



Obrázek 5.1: Ukázka způsobu přípravy dat pro následné simulace

Samotné testování probíhalo nejprve na jednom počítači, na kterém byl nainstalován OS linux. Avšak kvůli vysokým počtům simulací, které byly ve finální části testování pro-

váděny, bylo potřeba mnohonásobně zvýšit výpočetní výkon. K tomuto účelu byla využita počítačová laboratoř na VUT FIT, kde bylo testování prováděno na řadě počítačů, jejich počet postupně narůstal a v největší části testování dosahoval čísla osmnácti strojů. Problém, proč bylo potřeba tak vysokého počtu počítačů pro samotné výpočty spočíval v tom, že postupný průběh výpočtu i s analyzováním minulých trendů trval pro jednu simulaci od tří až do deseti minut. Vzhledem k tomu, že počty simulací byly v řádech tisíců, bylo potřeba většího množství výpočetních strojů, aby samotné testování mohlo být považováno za velmi objektivní a podrobné. Grafické znázornění jakým způsobem probíhalo samotné testování na datech je znázorněno na obrázku 5.2.



Obrázek 5.2: Reprezentace způsobu testování vzorku dat

## 5.1 Etapy testování

Testování bylo rozděleno do několika etap, kde každá z nich reprezentovala různý přístup k nalezení co nejpodobnějších trendů v minulosti, které by odpovídaly aktuálnímu stavu trhu. Samotné rozdíly mezi jednotlivými postupy spočívaly v tom, na jak velkých úsecích dat a s jakými parametry se samotné testování provádělo. To znamená, že v některých případech se při testování porovnávaly úseky dat o délce odpovídající obchodování za dvě a půl hodiny, v jiných případech se testování projevovalo na úsecích délky odpovídající pouze hodině. Další parametr, na který se testování zaměřovalo, byl pohled do budoucnosti, aneb jak moc velkou predikci a s jakou přesností bude tento druh analýzy trhu nabízet. Kromě toho bylo testování prováděno i na různých trzích. Kromě EUR/USD se testování provádělo na USD/GBP.

### 5.1.1 První etapa testování

První etapa představovala testování v podobě pouhých deseti vzorků. Každý ze vzorků simuloval pohled zpět o dvě a půl hodiny s výhledem do budoucnosti na jednu hodinu. Následující sada simulací byla spuštěna pro všech šest metod a jejich dílčích variant, které se vzájemně lišily různým pohledem na výpočet přijatelné vzájemné odchylky. To v praxi znamenalo, že bylo spuštěno celkem deset simulací pro osmnáct metod, tedy sto osmdesát

simulací. Doba provádění jednotlivé simulace se pohybovala v rozmezí od tří do deseti minut v závislosti na tom, jak dlouho skriptu trvalo, než našel nejvhodnější odchylku. Nakonec se celková doba provádění simulací v této etapě pohybovala kolem patnácti hodin.

Výsledky z první etapy byly velmi pozitivní, avšak jejich vypovídající hodnota nebyla příliš vysoká, kvůli malému vzorku dat, jenž představoval pouze deset simulací. Problém však byl i v tom, že pokud by došlo ke zvýšení počtu simulací, znamenalo by to zpomalení získávání výsledků ze simulací. Tento problém se podařilo vyřešit pomocí sady počítačů na VUT FIT, jejichž výpočetní výkon byl propůjčen pro účely těchto simulací. Díky tomu mohly být v druhé etapě samotné skupiny testovaných dat mnohonásobně rozšířeny, a experimentování s různými vstupními parametry skriptů bylo již možné delegovat na jednotlivé stroje. Nicméně tato etapa testování prokázala, že zvolené metody jsou použitelné a tento druh technické analýzy je vhodný.

### 5.1.2 Druhá etapa testování

Druhá etapa testování spočívala v testování sady dat o celkovém počtu sto vzorků reprezentujících jednotlivé trendy na trhu, což je desetinásobný nárůst v porovnání s předcházející etapou. Testovací sada byla sestavena z pětiminutových úseků, a jejím cílem bylo nalézt metody, které naleznou nejvhodnější sadu dat. Jednotlivé simulace se vzájemně lišily primárně v různých pohledech na jak velké úseky do minulosti se budou zaměřovat. Zda-li to bude jen sledování krátkodobého trendu do vzdálenosti okolo jedné hodiny zpět, nebo dlouhodobého, kde sledování bylo až dvě a půl hodiny zpátky. Mimo toho se provádělo testování i na rozdílně dlouhé predikce vývoje směrem do budoucnosti. Cílem bylo, zjistit zda bude větší úspěšnost u predikce spíše dlouhodobého trendu, nebo naopak u trendu krátkodobého. V následujících odstavcích bude popsáno několik vybraných simulací i s výsledky, jakých se v jejich průběhu dosahovalo.

#### Simulace s krokem 2,5h do minulosti a predikcí 1h do budoucnosti

V této části testování probíhalo celkem tisíc osm set testů s celkovou délkou běhu sto padesát hodin. Tato zátěž byla rozdělena na celkem šest počítačů, kterým zabralo dvacet pět hodin vypočítat výsledky. Samotné spuštění jednotlivých simulací probíhalo pomocí automatického skriptu napsaného v jazyce Bash, který měl za úkol postupně ve správném pořadí zapínat jednotlivé skripty, které si mezi sebou pak adekvátně předávaly parametry. Jediným parametrem skriptu bylo číslo udávající celkový požadovaný počet simulací, což v tomto případě bylo číslo sto.

Z výsledků simulací vyplynulo, že metoda využívající rozdílů vzdáleností byla pro tuto sadu dat nejvhodnější a dokázala vhodným způsobem predikovat budoucí vývoj kurzu. Samotné vyhodnocení jedné varianty vždy spočívalo v tom, že bylo nalezeno  $N$  nejpodobnějších úseků dat, které predikovaly budoucí vývoj směru trendu na základě toho, jak se trend vyvíjel u nich. Výsledek byl prezentován jako poměr trendů, které stoupaly, trendů které klesaly, a trendů které stagnovaly. Výsledná úspěšnost jednotlivých metod byla vypočítána na základě těchto hodnot a porovnána, jak skutečně se trend pohyboval. V tomto případě byl poměr správně detekovaných, ve srovnání se špatně detekovanými, ve výši okolo deseti procent u metody využívající rozdílů vzdáleností. Nutno dodat, že pro tuto úspěšnost bylo nezbytné, aby vzájemný poměr trendů, které stoupaly a klesaly, byl v rozdílu minimálně osm, až pak se dalo dosahovat takto vysokých čísel pro správně detekované trendy. Pro názornost jsou dosažené výsledky z metody rozdílů vzdáleností reprezentovány tabulkou 5.1, kde jednotlivé hodnoty určují vzájemný poměr úspěšné predikce v porovnání s neúspěšnou.

vzájemný rozdíl	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
základní odchylka	-12	-8	-5	-6	-5	-4	-1	4	2	2	2	2	2	0	0
stoupající odchylka	-7	-7	-7	-6	-5	-2	0	2	1	-3	-4	-4	-3	-2	-2
rovnoměrná odchylka	-2	-6	-6	-9	-9	-2	-1	2	1	3	3	1	1	-2	-1

Tabulka 5.1: reprezentace dosažených výsledků pro metodu rozdílu vzdáleností

### Simulace s krokem 2,5h do minulosti a predikcí 15 min do budoucnosti

Následující simulace představuje možnost predikce trendu pouze pro krátkodobý budoucí trend ve srovnání s předcházející metodou, kde jsme se zaměřovali na trend dlouhodobý. Celkový počet simulací byl opět zvolen pro hodnotu sto, což pro všechny metody dalo tisíc osm set simulací. Pro samotný výpočet bylo opět využito laboratoři na VUT FIT, kde samotný výpočet za použití šesti stojů trval dvacet pět hodin.

Dosažené výsledky jsou reprezentovány tabulkou 5.2, kde jsou reprezentovány tři vybrané metody. Číslo u každé metody reprezentuje typ výpočtu vzájemné odchylky. Hodnota jedna reprezentuje základní odchylky pro všechny data stejnou. Hodnota dva reprezentuje případ, kdy hodnota odchylky postupně klesá a hodnota tři popisuje případ kdy hodnota je pro určité úseky zkoumaných dat rozdílná.

vzájemný rozdíl	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Jaccardova vzdál. 1	6	6	2	-1	0	-2	-1	0	-1	0	0	-1	0	0	0
Jaccardova vzdál. 2	7	1	-1	3	-1	-1	0	3	4	4	3	1	1	1	1
Jaccardova vzdál. 3	14	13	10	3	3	2	0	0	-2	3	2	2	2	1	1
Rozdíl vzdál. 3	-20	-8	-7	-2	-3	0	1	2	2	6	6	2	2	2	2
Rozdíl vzdál. 1	9	7	5	2	1	4	8	6	5	5	4	2	3	2	2
Rozdíl vzdál. 2	2	2	2	-1	-1	4	5	4	3	3	3	3	3	0	1
Euklidovská vzdál. 3	7	3	2	2	3	-1	-4	-5	-5	-5	-4	-3	-2	-1	-1
Euklidovská vzdál. 1	3	3	2	-1	0	-7	-10	-4	-3	0	0	-3	-3	-1	-1
Euklidovská vzdál. 2	-3	-1	-1	-4	-3	-6	-7	-5	-5	0	0	-1	-1	-1	-1

Tabulka 5.2: reprezentace dosažených výsledků pro vybrané metody

### Simulace s krokem 1h do minulosti a predikcí 15 min do budoucnosti

Poslední simulace, která zde bude zmíněna, je metoda, u které byl pohled do minulosti zkrácen na délku pouhé hodiny, a pro predikci budoucího trendu byla zvolena délka patnácti minut. Tato simulace byla zaměřena na to, aby z velké míry zanedbávala historický vývoj kurzu a sledovala pouze velmi krátký úsek trhu, pro který se snažila najít velké množství podobných úseků dat. Díky tomu, že se tyto dvě skupiny dat takřka rovnaly, představovala tato metoda velmi nadějný přístup pro predikci vývoje kurzu. Problém, který to přinášelo a se kterým se muselo počítat byl, že metoda velmi zanedbávala okolní souvislosti, což se při dlouhodobějším pozorování mohlo projevit.

Výsledky této simulace však byly velmi slibné. Všechny metody pro porovnávání a hledání nejbližších vzorků dat se osvědčily, a výsledný poměr doporučení, kam se trend na trhu bude posouvat, byl slibný. Nejlépe se v této simulaci osvědčila metoda pro Euklideánskou vzdálenost, která měla hodnotu vzájemného rozdílu nejvyšší ze všech testů provedených v této etapě testování. Hned za ní se opět umístila metoda pro měření rozdílu vzdáleností.



Tato metoda se ve všech testovaných simulacích osvědčila, a proto ji považuji za jednu z nejvýkonnějších. V tabulce 5.3, jsou reprezentovány dvě vybrané metody, které měly v simulaci nejlepší výsledky.

vzájemný rozdíl	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Rozdíl vzdál. 1	7	8	11	6	2	3	1	8	6	4	5	3	2	6	7
Rozdíl vzdál. 2	9	8	8	5	4	4	6	4	5	5	3	1	1	-1	0
Rozdíl vzdál. 3	6	4	7	6	5	7	9	6	6	4	2	3	3	2	3
Euklidovská vzdál. 1	4	7	8	4	5	6	7	4	3	1	-2	2	4	5	3
Euklidovská vzdál. 2	19	14	12	11	12	12	12	11	10	10	9	5	3	2	1
Euklidovská vzdál. 3	13	14	13	13	15	12	11	8	7	4	5	5	2	1	0

Tabulka 5.3: Reprezentace dosažených výsledků pro vybrané metody

### 5.1.3 Třetí etapa testování

Poslední etapa testování byla značně odlišná. Tato odlišnost spočívala v tom, že skript, který měl za úkol provádět automatické zapínání jednotlivých metod i s výpočty normalizací, byl nastaven na hodnotu deset tisíc. Problém tady ale spočíval v tom, že celková doba provádění simulací by trvala velmi dlouho. Z toho důvodu bylo simulování nastaveno na nižší hodnoty. To znamená testování pouze krátkodobého trendu s výhledem do budoucnosti pouze na patnáct minut. I přes tyto úpravy testování probíhalo po dobu šesti dnů a na konci bylo zastaveno, kvůli tomu, že laboratoř již bylo potřeba využívat i někým jiným. Avšak po dobu kdy simulace běžela, stihla vyhodnotit tisíce dílčích případů, které byly dále podrobeny analýze. Bohužel nebylo testování provedeno pro úplně všechny metody, ale pouze pro polovinu, takže získaná data nelze považovat za kompletní.

Avšak z výsledků dlouhodobější analýzy trhu vyplývá, že jednotlivé metody jsou úspěšné i při dlouhodobějším pozorování, bohužel jejich procentuální úspěšnost není již tak vysoká. Tento problém bude nejpravděpodobněji způsoben jednak nečekaným vývojem trhu, který technická analýza nemá šanci predikovat, ale také tím, že analýza je prováděna pouze na základě konečné hodnoty kurzu v daném čase. Možná řešení, která by zvýšila vzájemnou shodu jsou popsána v následující kapitole.

## 5.2 Možnosti dalšího testování

Pro následné rozšíření a testování za účelem nalezení vyšší přesnosti v predikci vývoje budoucího trendu trhu navrhuji následující řadu úprav. Ty by mohly potencionálně zvýšit celkový počet správně predikovaných hodnot budoucího vývoje trhu. Návrhy pro jednotlivá rozšíření byly postupně nacházeny v průběhu provádění simulací, kdy se částečně začalo ukazovat, jaký potenciál samotné úpravy nabízejí. Rozdílné druhy rozšíření jsou nastíněny dále.

### 5.2.1 Rozšíření pro analýzu historického trendu na základě výkyvů trendu

Následující rozšíření reprezentuje variantu, při které by se hledání vzájemné podobnosti nezaměřovalo pouze na samotné hodnoty uzavírajících kurzů v jednotlivém čase provedených v minulosti. Pro samotné skupiny dat, které by byly vyhodnoceny jako nejpodobnější vůči hledanému trendu, by bylo puštěno ještě jedno porovnávání, které by reprezentoval

skriptem pro účely nalezení naprosté totožnosti tvaru minulého trendu. To znamená, že by bylo zkoumáno, jestli v každém kroku vývoje aktuálního trendu, který stoupá či klesá, byla tato skutečnost dodržena i v nalezeném podobném kurzu z historických. Pokud by nalezená data odpovídala nejen velikostí, ale i skutečným vývojem historickému kurzu, lze očekávat výrazné zvýšení přesnosti samotného porovnávání.

Například pokud bychom sledovali krátkodobý trend na trhu pro jednu hodinu zpátky do minulosti a zde by celou dobu probíhalo stoupání bez jediného výkyvu, tak bychom se snažili nalézt v minulosti takový úsek dat, který by po znormalizování nabízel hodnoty, které budou co nejvíce odpovídat aktuálnímu trendu. Zároveň by však musel splňovat podmínku, která říká, že po celou dobu svého vývoje data musí reprezentovat stoupavou tendenci a růst bez výkyvů.

Popsané rozšíření bylo zčásti již implementováno, ne však pro celý úsek historických dat, nýbrž pouze pro data pět kroků do minulosti. Již při testování na dvou tisících hodnotách tato metoda nabízela slibné predikce, avšak nelze ji brát jako plně fungující, protože nebyla plně otestována a ověřena. Proto je zde zmíněna jako návrh na budoucí plné implementování za účelem zvýšení úspěšnosti budoucí predikce.

### 5.2.2 Rozšíření pro analýzu nejen konečné uzavírající hodnoty kurzu

Rozšíření doporučuje využívat při technické analýze trhu nejen uzavírající hodnotu daného kurzu, ale i ostatní informace, které informují o otevírající, maximální a minimální hodnotě kurzu. Díky tomu by se samotné nalezené úseky dat daly považovat za velmi podobné až totožné, co se hodnot získaných z historického obchodování týká. Pokud bychom dané rozšíření zkombinovali i s rozšířením předcházejícím, nalezená data na trhu by představovala potencionálně co nejpřesnější podobnost hledaného trendu vůči historickému, jaký lze ze samotných dat získat. Problém, který by u daného rozšíření vzniknul by spočíval ve čtyřnásobném zvýšení náročnosti samotného výpočtu, neboť by bylo nezbytné provádět dílčí kroky pro všechny jednotlivé údaje o historickém kurzu. Proto by se následující varianta dala používat pouze pro predikci dlouhodobého kurzu.

## 5.3 Shrnutí dosažených výsledků z testování

Jednotlivé simulace, které byly prováděny pro dílčí metody prokázaly, že tento typ technické analýzy je možný a data, která z něj získáme se dají považovat za přesná. Problém však představuje nalezení nejvhodnější kombinace dat parametrů, pro které budou mít tyto metody nejvyšší úspěšnost.

Při vyhodnocování výsledků získaných z jednotlivých simulací bylo zjištěno následující. Nejlépe si s nalezením co nejpodobnějšího vzorku dat poradila metoda pro rozdíl vzdáleností. Tato metoda skoro ve všech simulacích nabízela kladné predikce. V některých simulacích se to pohybovalo kolem hranice nula až jedna, což při dlouhodobějším sledování trhu nepřináší patřičný užitek. Avšak byly zde i metody, u kterých se hodnota, kterou metoda v simulacích predikovala pohybovala kolem úspěšné predikce okolo 2-3 % čísel, což lze považovat za úspěšný výsledek.

Nasazení této metody technické analýzy na finanční platformu a její následné uvedení do provozu na reálném trhu bych ještě nedoporučoval, a raději bych se zaměřil na podrobnější simulace a implementaci rozšíření. Až na základě dosažených výsledků by bylo možné uvažovat o nasazení na reálný trh.

# Kapitola 6

## Závěr

Cílem této práce bylo ověření teorie a naprogramování softwaru pro detekci budoucího směru vývoje trendu na trhu na principu technické analýzy. K tomu bylo potřeba vytvořit řadu skriptů, pomocí nichž pak následně probíhala analýza daného trendu za využití algoritmů z oblasti strojového učení, přesněji klasifikace dat. Bakalářská práce je zaměřená právě na sadu vybraných algoritmů, které postupně implementuje a podrobuje rozsáhlému testování.

V rámci bakalářské práce byly nejprve nastudovány principy, na jakých samotné obchodování na trhu funguje. Největší důraz při studiu byl kladen na analýzu Forexového trhu, který byl dále v práci využíván. Součástí tohoto popisu je i část, která se zabývá samotnou technickou analýzou, na jejímž základě je tato práce postavena. Je zde i zmínka o fundamentální analýze, která obsahuje zcela odlišný způsob pohledu na samotný trh, avšak ta nebyla dále v práci nijak využívána.

Každá vybraná metoda byla podrobně popsána a analyzována pro účely hledání nejpodobnějších skupin dat, které byly obsahem testovacích souborů. Cílem této analýzy bylo, zjistit, jak přesně budou jednotlivé metody detekovat daný trend a zda-li budou vhodně předikovat i jeho budoucí vývoj. Na základě této analýzy byly implementované jednotlivé skripty. Implementace skriptů byla realizována v jazyce Python na systému Linux. Jejich implementace pro účely této práce byla pouze ve fázi testovací. Samotné napojení na obchodní platformu a jeho následné testování zde neproběhlo a je připravené až pro další rozšíření práce.

Testování výsledného softwaru probíhalo na laboratorních počítačích, které se nacházejí na fakultě VUT FIT. Důvodem pro toto testování byla potřeba vyššího výpočetního výkonu pro velmi rozsáhlé testovací skupiny dat. Na základě poznatků získaných ze samotného testování bylo provedeno následné vyhodnocení dosažených výsledků pro účely analýzy, která metoda dosahovala nejlepších výsledků.

Z naměřených výsledků vyplynulo, že metoda rozdílů vzdáleností nabízí nejvhodnější přístup pro nalezení vhodné predikce dat na trhu. Mezi další metody, které prokazovaly kvalitní výsledky patří Euclideánská vzdálenost a Jaccardův koeficient. Při testování se prokázalo, že je vhodnější se více zaměřovat na krátkodobé trendy v porovnání s dlouhodobými, a to z důvodu toho, že množství dat pro vzájemné porovnání je mnohem vyšší.

Na tuto bakalářskou práci se dá dále navazovat, a to způsoby popsány v kapitole 5, kde je řada dalších metod, které se v průběhu testování objevovaly a nabízely možnosti zvyšování predikce budoucího trendu.

# Literatura

- [1] Barrett, P.: *Euclidean Distance raw, normalized, and double-scaled coefficients*. 2005, [Online; navštíveno 20.03.2017].  
URL <http://www.pbarrett.net/>
- [2] D. Štýbr, P. O., P. Klepetko: *Začínáme investovat a obchodovat na kapitálových trzích*. Grada, 2011, ISBN 978-80-247-3648-8.
- [3] Hartman, O.: *Jak se stát forexovým obchodníkem*. FXstreet.cz spol. s.r.o., 2014, ISBN 978-80-904418-3-5.
- [4] Jitka Veselá, M. O.: *Technická analýza na akciových, měnových a komoditních trzích*. Ekopress, 2015, ISBN 978-80-87865-22-4.
- [5] K. Lien, P. O., P. Klepetko: *Forex: ziskové intradenní a swingové obchodní strategie : jak na technickou a fundamentální analýzu pro úspěch na finančních trzích*. FXstreet.cz, 2013, ISBN 978-809-0441-828.
- [6] Kahn, M.: *A beginner's guide to charting financial markets: a practical introduction to technical analysis for investors*. Harriman House, 2007, ISBN 19-056-4121-4.
- [7] Kaufman, P.: *A short course in technical trading*. Wiley, 2003, ISBN 04-712-6848-8.
- [8] Larson, M.: *12 simple technical indicators that really work*. Marketplace Books, 2007, ISBN 15-928-0290-7.
- [9] Ludvík Turek, O. H.: *První kroky na FOREXu*. Computer Press, 2009, ISBN 978-80-251-2006-4.
- [10] M. Meloun, J. M.: *Statistická analýza experimentálních dat*. Academia, 2004, ISBN 80-200-1254-0.
- [11] M. Meloun, J. M.: *Kompedium statistického zpracování dat: Metody a řešené úlohy*. Academia, 2006, ISBN 80-200-1396-2.
- [12] Mitchell, T.: *M. Machine Learning*. McGraw Hill, 1997, ISBN 0070428077.
- [13] Murphy, J.: *Technical analysis of the financial markets*. Prentice Hall Press, 1986, ISBN 978-0138980085.
- [14] Oldřich, R.: *Finanční trhy*. Grada, 2014, ISBN 978-80-247-3671-6.
- [15] Pavel Zezula, G. A. M. B., Vlastislav Dohnal: *Similarity Search: The Metric Space Approach*. Springer, 2005, ISBN 0-387-29146-6.

- [16] Radomír Čížek, R. H.: *Prognóza finančních trhů*. BizBooks, 2014, ISBN 978-80-265-0063-6.
- [17] Shafranovich, Y.: *Common Format and MIME Type for Comma-Separated Values (CSV) Files*. RFC 4180, 2005, [Online; navštíveno 20.04.2017].  
URL <https://www.ietf.org/rfc/rfc4180.txt>
- [18] Turner, T.: *A beginner's guide to short-term trading*. Adams Media, 2002, ISBN 15-806-2570-3.
- [19] Zacharski, R.: *A Programmer's Guide to Data Mining*. 2015, [Online; navštíveno 25.04.2017].  
URL <http://guidetodatamining.com/>