



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY



FAKULTA PODNIKATELSKÁ  
ÚSTAV EKONOMIKY

FACULTY OF BUSINESS AND MANAGEMENT  
INSTITUTE OF ECONOMICS

PREDIKCE VÝVOJE AKCIOVÉHO TRHU  
PROSTŘEDNICTVÍM  
TECHNICKÉ A PSYCHOLOGICKÉ ANALÝZY

STOCK MARKET PREDICTION VIA TECHNICAL AND PSYCHOLOGICAL ANALYSIS

DIPLOMOVÁ PRÁCE

MASTER'S THESIS

AUTOR PRÁCE

AUTHOR

Ing. PATRIK PETŘÍK

VEDOUCÍ PRÁCE

SUPERVISOR

prof. Ing. OLDŘICH REJNUŠ, CSc.

BRNO 2010

# ZADÁNÍ DIPLOMOVÉ PRÁCE

**Petrík Patrik, Ing.**

---

Podnikové finance a obchod (6208T090)

Ředitel ústavu Vám v souladu se zákonem č.111/1998 o vysokých školách, Studijním a zkušebním řádem VUT v Brně a Směrnicí děkana pro realizaci bakalářských a magisterských studijních programů zadává diplomovou práci s názvem:

**Predikce vývoje akciového trhu prostřednictvím  
technické a psychologické analýzy**

v anglickém jazyce:

**Stock Market Prediction via Technical and Psychological Analysis**

Pokyny pro vypracování:

Úvod

Vymezení problému a cíle práce

Teoretická východiska práce

Analýza problému a současné situace

Vlastní návrhy řešení, přínos návrhů řešení

Závěr

Seznam použité literatury

Přílohy

Seznam odborné literatury:

REJNUŠ, O.: Finanční trhy, Ostrava, Key Publishing, 2008. ISBN 978-80-87071-87-8

DUNIS, CH.: Forecasting Financial Markets - Exchange Rates, Interest Rates And Asset Management, John Wiley & Sons, 1996. ISBN 0471966533

FANTA, J.: Počítačové technologie na kapitálových trzích, Praha, Computer Press, 1998. ISBN 80-7226-073-1

Vedoucí diplomové práce: prof. Ing. Oldřich Rejnuš, CSc.

Termín odevzdání diplomové práce je stanoven časovým plánem akademického roku 2009/2010.

L.S.

---

Ing. Tomáš Meluzín, Ph.D.  
Ředitel ústavu

---

doc. RNDr. Anna Putnová, Ph.D., MBA

V Brně, dne 24.05.2010

## Abstrakt

Tato práce se zabývá predikcí akciového trhu. Nejprve jsou uvedena teoretická východiska práce a to především principy technické a psychologické analýzy. Dále jsou uvedeny metody z oblasti umělé inteligence, které je možno využívat při metodách zabývajících se predikcí. V rámci práce je navržen systém pro predikci a je implementována jedna jeho část. Tato část je testována a v závěru jsou diskutovány dosažené výsledky.

## Abstract

This work deals with stock market prediction via technical and psychological analysis. We introduce theoretical resources of technical and psychological analysis. We also introduce some methods of artificial intelligence, specially neural networks and genetic algorithms. We design a system for stock market prediction. We implement and test a part of system. In conclusion we discuss results.

## Klíčová slova

Predikce, technická analýza, psychologická analýza, klasifikace, naivní Bayesův klasifikátor, neuronová síť, evoluční algoritmy.

## Keywords

Stock market prediction, technical analysis, psychological analysis, classification, naive Bayes classifier, neural network, genetic algorithm.

## Citace

Patrik Petřík: Predikce vývoje akciového trhu prostřednictvím technické a psychologické analýzy, diplomová práce, Brno, FP VUT v Brně, 2010

# Predikce vývoje akciového trhu prostřednictvím technické a psychologické analýzy

## Prohlášení

Prohlašuji, že předložená diplomová práce je původní a zpracoval jsem ji samostatně. Prohlašuji, že citace použitých pramenů je úplná, že jsem ve své práci neporušil autorská práva (ve smyslu Zákona č. 121/2000 Sb., o právu autorském a o právech souvisejících s právem autorským).

.....  
Patrik Petřík  
25. května 2010

## Poděkování

Tímto bych rád poděkoval prof. Oldřichu Rejnušovi, vedoucímu diplomové práce, za ochotu a čas, který mi při tvorbě této práce věnoval, za cenné rady a za motivaci do dalšího studia.

© Patrik Petřík, 2010.

*Tato práce vznikla jako školní dílo na Vysokém učení technickém v Brně, Fakultě podnikatelské. Práce je chráněna autorským zákonem a její užití bez udělení oprávnění autorem je nezákonné, s výjimkou zákonem definovaných případů.*

# Obsah

<b>Obsah</b>	<b>2</b>
<b>1 Úvod</b>	<b>3</b>
1.1 Členění práce	4
<b>2 Vymezení problému a cíle práce</b>	<b>5</b>
2.1 Akciový trh	5
2.1.1 Významné akciové trhy	6
2.2 Predikce časových řad	6
2.3 Vymezení problému	6
2.4 Cíle práce	7
<b>3 Teoretická východiska práce</b>	<b>8</b>
3.1 Technická analýza	8
3.1.1 Dowova teorie	9
3.1.2 Teorie Elliottových vln	9
3.1.3 Grafické zobrazování	13
3.1.4 Technické indikátory	15
3.2 Psychologická analýza	16
3.2.1 Keynesova spekulativní rovnovážná hypotéza	17
3.2.2 Kostolanyho burzovní psychologie	17
3.2.3 Teorie spekulativních bublin	18
3.2.4 Drasnarova koncepce psychologické analýzy	20
3.3 Predikce časových řad	20
3.4 Neuronové sítě	21
3.4.1 Dopředná neuronová síť	22
3.4.2 Další typy neuronových sítí	23
3.4.3 Software neuronových sítí	23
3.4.4 Aplikace v ekonomii	24
3.5 Fuzzy množiny	24
3.6 Evoluční algoritmy	25
3.6.1 Princip	25
3.6.2 Aplikace v ekonomii	26
3.7 Teorie chaosu	26
3.8 Klasifikace	27
3.8.1 Bayesovská klasifikace	27
3.8.2 Klasifikace textů	27
3.9 Dolování dat	29

<b>4</b>	<b>Analýza problému a současné situace</b>	<b>31</b>
4.1	Specifika . . . . .	31
4.2	Technická analýza . . . . .	32
4.3	Fundamentální analýza . . . . .	33
4.4	Psychologická analýza . . . . .	33
4.5	Aplikace . . . . .	34
<b>5</b>	<b>Vlastní návrhy řešení, přínos návrhů řešení</b>	<b>36</b>
5.1	Návrh řešení . . . . .	36
5.1.1	Modul pro správu zpráv . . . . .	38
5.1.2	Modul pro psychologickou analýzu . . . . .	39
5.1.3	Modul pro fundamentální analýzu . . . . .	39
5.1.4	Modul pro technickou analýzu . . . . .	39
5.1.5	Modul pro rozhodovací část systému . . . . .	40
5.1.6	Návrh vybrané části systému . . . . .	41
5.1.7	Implementace vybrané části systému . . . . .	44
5.1.8	Testování . . . . .	45
5.1.9	Zhodnocení výsledků testů . . . . .	47
5.1.10	Přínos návrhů řešení . . . . .	49
<b>6</b>	<b>Závěr</b>	<b>50</b>
	<b>Použitá literatura</b>	<b>52</b>
	<b>Seznam použitých zkratk a symbolů</b>	<b>53</b>
	<b>Seznam příloh</b>	<b>54</b>
<b>A</b>	<b>Seznamy hledaných slov</b>	<b>55</b>
<b>B</b>	<b>Ukázkový výpis testovacího skriptu</b>	<b>57</b>

# Kapitola 1

## Úvod

Obchodování na burze se díky informačním technologiím a nejen jim dostává k stále širší veřejnosti a to i navzdory finanční krizi, která proběhla v nedávných letech. Obchodování na burze se tak stává jednou z možností, jak si lidé mohou vydělávat peníze, ale i o peníze přijít. Na rozdíl od jiných zaměstnání, tato činnost s sebou přináší jistá specifika, která u jiných nenacházíme. Jedná se především o znalosti, které dotýčný subjekt, jež se chce účastnit obchodování na burze, musí mít a dále o specifika, která patří k samotnému obchodování. Jedním z nich je riziko spojené s neurčitostí vývoje daného obchodovaného titulu.

Vznikly různé metody a teorie spojené s predikcí vývoje akciového trhu, které dělíme do dvou základních přístupů. Prvním z těchto přístupů je založen na technické analýze, kdy budoucí hodnoty vývoje vznikají na základě historických dat, konkrétně se uvažují většinou a především historické pohyby cen či objemy zobchodovaných titulů. Druhý z přístupů je založen na psychologii. Psychologická analýza vychází z předpokladu, že akciové trhy jsou silně ovlivněny chováním burzovního publika, které reaguje na různé impulzy. Každá z výše uvedených analýz bude podrobněji popsána v následujících kapitolách práce.

Rozvoj informačních technologií, ke kterému došlo v nedávné době, znamená i nové možnosti právě u výše zmíněných analýz, ať už technologické či psychologické. Zvláště u technologické analýzy se rozvoj informačních technologií projevil nejvíce, neboť umožnil efektivněji pracovat s daty, která jsou právě pro technickou analýzu tolik potřebná. Kromě možnosti efektivněji zpracovávat data informační technologie přinesli i nové přístupy v oblasti tzv. soft computingu. Jedná se především o možnost využití neuronových sítí, fuzzy logiky, evolučních algoritmů či nových metod klasifikace. V neposlední řadě je třeba zmínit rozvoj internetu, díky kterému je obchodování přístupné širší veřejnosti.

Tato práce si bere za cíl přiblížit problematiku využití informačních technologií právě v oblasti predikce vývoje akciového trhu. Přiblížit metody, které je možné využít v rámci různých analýz, ať už technické či psychologické. Popsat princip technické a psychologické analýzy, popsat principy jednotlivých přístupů z oblasti soft computingu.

Hlavní částí práce je studium využití principů z oblasti soft computingu v oblasti psychologické analýzy, neboť této není věnováno tolik pozornosti jako analýze technické. V rámci této práce se zaměříme především na možnost predikce vývoje akciového trhu pomocí klasifikace zpráv, které byly vydány v deníku The New York Times. Motivací k této práci je především autorovo studium na Fakultě podnikatelské VUT v Brně a na Fakultě informačních technologií VUT v Brně a jeho zájem právě o oblast využití informačních technologií při predikci vývoje akciového trhu.



## 1.1 Členění práce

Celá práce je členěna do kapitol s názvem podle tématického okruhu, které popisují.

### 1. Úvod

Kapitola **1**, kterou právě čtete, slouží k uvedení do širšího kontextu problematiky, a nastiňuje její členění.

### 2. Vymezení problému a cíle práce

V kapitole **2** přiblížíme nejprve základní pojmy z oblasti finančních trhů, problematiku spojenou s predikcí časových řad a základní pojmy z oblasti predikce, na jejichž základě poté vymezíme problém a cíl práce, kterému se budeme věnovat v dalších kapitolách.

### 3. Teoretická východiska práce

V kapitole **3** jsou přiblížena teoretická východiska práce. Kapitola je rozdělena do několika podkapitol. Nejprve je přiblížena problematika technické a psychologické analýzy. V dalších podkapitolách jsou dále uvedeny pojmy z oblasti soft computingu, které budeme v dalším textu práce využívat. Jedná se především o neuronové sítě, predikci časových řad a klasifikaci.

### 4. Analýza problému a současné situace

Kapitola **4** se zabývá analýzou problému predikce vývoje akciového trhu. V této kapitole se zabýváme problémem využití metod a přístupů právě v oblasti technické a psychologické analýzy. Zmiňujeme výsledky, kterých bylo v této oblasti dosaženo a to především v současné době.

### 5. Vlastní návrhy řešení, přínos návrhů řešení

Kapitola **5** je stěžejní částí práce. Na základě poznatků z předcházejících kapitol navrhujeme zjednodušený systém pro predikci vývoje akciového trhu, jehož zvolenou část naimplementujeme a otestujeme. Na základě provedených testů vyhodnotíme přínos provedené části.

### 6. Závěr

Kapitola **6** obsahuje shrnutí dosažených výsledků. Diskutuje budoucí možný rozvoj této práce a přínos práce pro autora.

### 7. Seznam použité literatury

V této části je uvedena literatura, ze které jsme čerpali při tvorbě této práce.

### 8. Přílohy

Přílohy představují především výstupy testovacích skriptů, které jsou podkladem pro závěry učiněné v práci.

## Kapitola 2

# Vymezení problému a cíle práce

Tato práce, která je zaměřena na predikci akciového trhu, přibližuje především využití metod z oblasti soft computingu při predikci vývoje akciového trhu. Je ovšem třeba uvést tuto problematiku do širšího kontextu finančních trhů. V této kapitole se nejprve zaměříme na akciový trh jako celek. Vycházíme především z [21]. Přiblížíme jeho charakteristiky a posléze se zaměříme na problematiku predikce vývoje tohoto trhu. Na základě uvedených informací definujeme jednotlivé cíle práce.

### 2.1 Akciový trh

Akciovým trhem rozumíme burzovní veřejný sekundární organizovaný trh [21], na kterém ve vyspělých zemích obchoduje i veřejnost (na rozdíl od burz devizových, na nichž obchodují především banky). Tyto trhy jsou považovány za nejvýznamnější. Obecně se dá říci, že čím významnější jednotlivé burzy jsou (a s nimi spojený trh), tím kvalitnější a světově známější obchodní tituly se na těchto trzích obchodují.

Veřejný trhem rozumíme trh, na kterém mohou obchodovat všichni potencionální zájemci a dané předmětné tituly jsou obchodovány za nejvyšší nabídnuté ceny. Pod pojmem sekundární trh rozumíme trh, na kterém se obchoduje s cennými papíry, které již byly v minulosti vydány, na rozdíl od trhu primárního, na kterém dochází právě k emisi nových cenných papírů. Pod pojmem organizovaný trh chápeme trh, který je organizován jistou institucí, konkrétně u burzovního organizovaného trhu je touto institucí burza, která má za úkol agregovat veřejnou nabídku a poptávku a párovat dané obchodní příkazy. V této práci se tedy zaměříme právě na tento typ trhu.

Každý akciový trh má jistá specifika a svou historii. Především jej charakterizuje fluktuace v úrovni jeho cen, vztah mezi cenami jednotlivých akcií a celku, jejich zisky a dividendami [25]. Investor, který se chce účastnit obchodování na daném akciovém trhu by měl znát jeho charakteristiky, vliv hospodářského cyklu na vývoj trhu atd.

Přestože platí, že celý finanční trh lze členit podle několika hledisek na různé dílčí trhy, přesto se tyto části vzájemně ovlivňují a existují mezi nimi vazby. Mezi významný faktor integrující jednotlivé dílčí segmenty patří i spekulace, jak je uvedeno v [21]. Právě spekulace, kterou představují spekulanti ovlivňující především tržní ceny obchodovaných titulů na daných trzích jsou z pohledu této práce prvkem trhu, který je obtížně predikovatelný a je vysoce rizikový.

### 2.1.1 Významné akciové trhy

Mezi nejvýznamnější burzy patří především New York Stock Exchange (NYSE), která je považována za největší, nejlikvidnější a nejvýznamnější burzu cenných papírů. Kromě významných amerických společností jsou na této burze obchodovány také významné zahraniční tituly. V roce 2006 se tato burza spojila s burzovní aliancí Euronext a vytvořili tak nejsilnější uskupení. Reprezentativním indexem NYSE je Dow Jones Industrial Average.

Další významnou burzou je Tokyo Stock Exchange (TSE), NASDAQ (další burza působící v Americe), Euronext (významné burzovní uskupení se sídlem v Paříži), London Stock Exchange (LSE), Shanghai Stock Exchange, Hong Kong Stock Exchange. Mezi významné indexy tak patří Dow Jones, S&P 500, NASDAQ a další [21].

## 2.2 Predikce časových řad

S problematikou predikce vývoje akciového trhu úzce souvisí predikce vývoje časových řad, neboť na ceny daného indexu či cenného papíru se rámci časového hlediska díváme jako na časovou řadu. Pojem „časová řada“ se využívá v mnoha odvětvích, jejichž společnou oblastí zájmu je právě predikce vývoje dané časové řady, porozumění mechanismu generování hodnot, zjištění různých charakteristik této časové řady a podobně. Více k časovým řadám je uvedeno v kapitole 3.

Predikce časových řad je nejvýznamnější částí analýzy časových řad. Existuje mnoho faktorů této predikce, především horizont předpovědi, forma předpovědi a přesnost předpovědi jsou klíčovými faktory. Přístupů k predikci existuje několik a jsou zmíněny v kapitole 3, kde se na predikci časové řady díváme především z pohledu technické analýzy, tedy bereme v úvahu právě historická data, pomocí kterých se snažíme predikovat budoucí hodnoty vývoje časové řady.

## 2.3 Vymezení problému

V této práci se tedy zaměříme na predikci vývoje akciového trhu. Na problematiku predikce vývoje akciového trhu lze pohlížet z několika hledisek. Právě rozdělení analýzy na technickou a psychologickou vymezuje jednotlivé oblasti problematiky predikce.

Technická analýza, blíže popsána v kapitole 3, při predikci vývoje využívá pouze historických dat. Jak je uvedeno v [16], tak technická analýza vychází především z historických cen a to proto, neboť tento přístup předpokládá, že ať už cenu daného titulu a její změnu může ovlivnit cokoli, počínaje od ekonomických, politických a dalších vlivů, jsou tyto vlivy zahrnuty v této ceně. Jak je dále uvedeno v [16] tento přístup se liší právě od fundamentální analýzy, která naopak studuje vlivy ekonomických sil, které mají za následek změnu dané ceny.

Ať už se jedná o technickou či fundamentální analýzu, obě jsou zaměřeny na studium buď ceny daného titulu, nebo ekonomických sil, která vychází z daných ekonomických dat, a ovlivňují tak cenu daného obchodovaného titulu. Psychologická analýza je na rozdíl od výše zmíněných analýz zaměřena na člověka, jeho chování a možnost ovlivnění jeho chování, které v konečném důsledku vede k ovlivnění budoucího vývoje obchodovaného titulu. Otázkou tedy je, co může ovlivnit chování člověka. Jaké informace jsou schopny

ovlivnit rozhodování? Jakou formu tyto informace mají a jak velký vliv na dané rozhodnutí mají v konečném důsledku?

## 2.4 Cíle práce

Cílem této práce je tedy analýza metod využívaných při predikci vývoje akciového trhu v rámci technické a psychologické analýzy. Především analýza metod využívaných v rámci psychologické analýzy je z pohledu této práce klíčová, neboť tato oblast není tolik studována jako oblast technické analýzy.

V rámci této práce navrhne zjednodušený model systému pro predikci vývoje akciového trhu a dále vybereme část tohoto systému, kterou naimplementujeme. Tato část bude, jak bylo zmíněno výše, zaměřena na psychologickou analýzu, konkrétně na klasifikaci zpráv, kdy třídy této klasifikace budou určovat právě pohyb daného akciového trhu, který bude reprezentován indexem.

Danou část systému budeme testovat z několika kritérií a stanovíme na základě těchto výsledků závěry týkající se použitých postupů a parametrů. V neposlední řadě budeme také diskutovat možný budoucí vývoj této práce.

Cíle práce:

1. Analýza metod využívaných při predikci vývoje akciového trhu
2. Navržení zjednodušeného systému pro predikci vývoje akciového trhu
3. Výběr části systému a její implementace
4. Provedení experimentů a otestování dané části
5. Zhodnocení výsledků experimentů
6. Diskuze nad budoucím možným vývojem práce

## Kapitola 3

# Teoretická východiska práce

V této kapitole přiblížíme teoretické pojmy, přístupy a metody, které jsou nutné pro další kapitoly a úzce souvisí s danou tematikou. Kapitola lze co do zaměření rozdělit na dvě části. V první části, která je prezentována prvními dvěma podkapitolami, jsou přiblíženy pojmy z oblasti predikce vývoje akciového trhu, především pojmy z oblasti technické analýzy a psychologické analýzy. Zdrojem informací jsou pro nás především knihy [21, 16].

Druhá část kapitoly je zaměřena na metody z oblasti soft computingu či umělé inteligence, které je možné a vhodné využívat při řešení daných problémů. Především přiblížíme neuronové sítě, fuzzy logiku, genetické algoritmy, dolování dat, klasifikaci a další. U jednotlivých samozřejmě zmíníme i oblast, kde jsou dané metody využívány nejvíce a přiblížíme i některé aktuální výsledky, kterých bylo dosaženo.

### 3.1 Technická analýza

*„Technickou analýzou rozumíme studium pohybů kurzů na trhu, primárně ve formě grafů, za účelem predikce budoucích cen“* [16]. Technickou analýzu využíváme pro předpověď budoucího vývoje kurzů jednotlivých akciových titulů či konkrétních akciových indexů, reprezentující daný akciový trh. Jak je uvedeno v [21]: *„Cílem technických analytiků je tedy jednak analyzovat vývoj kurzů akcií (resp. akciových indexů) a následně predikovat směry jejich budoucích kurzových změn (resp. vývoje celého akciového trhu), jednak určovat co nejvhodnější okamžiky k provádění obchodů (neboli provádět jejich tzv. časování /timing/).“*

Technická analýza je založena na třech základních předpokladech [16]:

1. V tržních akcích (změna ceny, kurzů, objemu apod.) je zahrnuto vše.
2. Pro pohyb ceny (kurzu) existuje trend.
3. Historie se opakuje.

Každý z těchto předpokladů je pro technickou analýzu klíčový. To, že v daných tržních akcích jsou zahrnuty všechny klíčové aspekty, které působí na pohyb cen či kurzů na daném trhu zjednodušuje pohled na problematiku predikce budoucího vývoje, neboť na rozdíl od fundamentální analýzy není třeba studovat různé ekonomické ukazatele týkající se daného obchodovaného titulu. Dalším klíčovým předpokladem je právě existence trendů v pohybu

daných cen (kurzů). Právě dále zmíněné teorie, jež vznikly v minulosti, jsou založeny na existenci těchto trendů a budoucí hodnoty jsou predikovány na základě těchto trendů. Posledním klíčovým předpokladem je opakující se historie. Právě pomocí časových řad se techničtí analytici pokoušejí hledat opakující se trendy ve vývoji daných kurzů a predikovat tak budoucí hodnoty.

Jak je uvedeno v [21]. Právě rozvoj informačních technologií a výpočetní techniky umožnil širší využití technické analýzy, neboť zpracovávání časových řad, tvorba a vyhodnocování grafů bylo značně pracné. Výpočetní technika tak umožnila nejenom rozšíření využití technické analýzy, ale i vývoj dalších druhů různých technických indikátorů využívaných při predikci pomocí technické analýzy. Jak uvidíme dále v práci, informační technologie nejenom, že usnadňují samotné zpracování historických dat a tvorbu daných indikátorů, ale jsou schopny při použití vhodných metod rozhodovat o samotném potenciaálním vývoji daného kurzu či ceny.

### 3.1.1 Dowova teorie

Dowova teorie patří mezi historicky první ucelenou teorii zaměřenou na problematiku predikce vývoje trendů akciových trhů [21]. Teorie je pojmenována po svém tvůrci, jímž byl Charles Dow a tvoří základní kámen technické analýzy. „*Dowova teorie vychází z předpokladu, že se vývoj kurzů většiny akcií pohybuje stejným směrem, jakým se vyvíjí celý akciový trh*“ [21]. Právě pro zkoumání trhu jako celku Charles Dow vytvořil historicky první akciové burzovní indexy, mezi které patří „*Dow Jones Industrial Average*“.

Trendy, které Charles Dow získal analýzou uzavíracích denních hodnot lze dělit podle hlediska délky jejich trvání nebo podle směru vývoje akciového trhu.

Z pohledu délky trvání rozlišoval tři druhy trendů. *Primární* trendy, jejichž délka je omezena od jednoho roku až do několika let, považoval Charles Dow za nejvýznamnější, neboť zahrnují nejvíce faktorů. *Sekundární* trendy, vyznačující se dobou trvání od několika týdnů do několika měsíců, které představují dočasnou korekci primárních trendů (reverzní pohyby). *Terciální* trendy, jejichž délka je zpravidla pouze několik dní, představující pouze krátkodobé výkyvy kurzů.

Na základě směru, který převládá v *primárním* trendu Charles Dow rozlišoval následující druhy trhů. Rostoucí (*býčí*) trh, vyznačující se růstem burzovního indexu, klesající (*medvědí*) trh, vyznačující se poklesem daného burzovního indexu a postranní trh, který nemá žádný z uvedených trendů.

Mezi další principy Dowovy teorie patří, že „*objem musí potvrzovat trend*“ [21], který hovoří o vztahu růstu obchodovaných objemů s růstem daných kurzů.

Mezi hlavní nedostatky této teorie patří především následující skutečnosti. Teorie je zaměřena pouze na primární trend, což omezuje potenciaálního uživatele, neboť nemůže teorii aplikovat i na krátkodobé trendy. Signály, které Dowova teorie přináší jsou indikovány se zpožděním a navíc mnohdy nejsou jednoznačné apod. I přes existenci nedostatků, sloužila Dowova teorie jako historicky první ucelená teorie, kterou bylo možné použít při predikci vývoje daného akciového trhu.

Širší popis Dowovy teorie lze nalézt v [21] a především v [16].

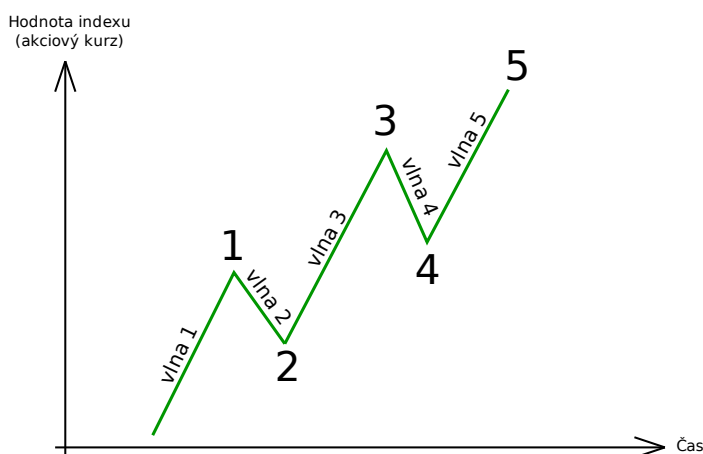
### 3.1.2 Teorie Elliottových vln

Další významnou teorií, která tvoří teoretický základ technické analýzy je *teorie Elliottových vln*. Tato teorie, jež byla publikována v roce 1938, je inspirována existencí period

v přírodních jevech. Tvůrce této teorie byl ovlivněn výše zmíněnou Dowovou teorií, která samotná má mnoho společného s teorií vln [16].

Teorie Elliottových vln podrobně popisuje chování trhu. Mnoho oblastí lidské činnosti lze popsat pomocí teorie vln, ovšem nejpobulárnější je aplikace této teorie právě v oblasti akciových trhů [9].

Teorie je založena na několika principech. Prvním principem zní, že trh roste v impulsní vlně, která se skládá z pěti podvln, které jsou označovány číselně 1, 2, 3, 4, 5 viz. obrázek 3.1. Vlny 1, 3, 5 jsou ve směru růstu trhu, zatímco vlny 2, 4 mají opačný trend. Elliott dále konstatuje, že vlna 2 nikdy neklesne pod počátek vlny 1, vlna 3 není nejkratší vlnou a vlna 4 vždy končí nad koncem vlny 1.



Obrázek 3.1: Impulsní vlna - Elliottovy vlny.

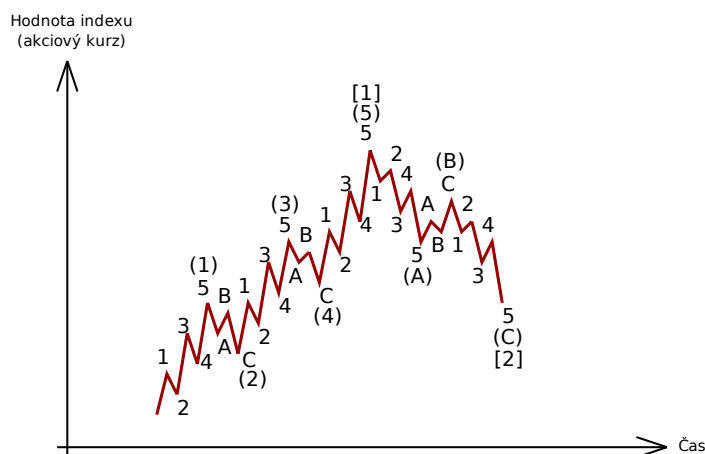
Další princip hovoří o dvou módech vln. Prvním jsou tzv. impulsní vlny a druhým jsou tzv. vlny korekční. Další princip hovoří o kompletním cyklu, který je složen z osmi vln, které jsou rozděleny do dvou fází. V první fázi je obsaženo pět vln (výše zmíněných) a v korekční fázi jsou obsaženy tři vlny označené písmeny A, B, C.

Princip hierarchické struktury dále hovoří o skládání Elliottových vln obdobně jako je tomu u fraktálů, kdy v tomto případě každá vlna je zároveň podvlnou vlny vyšší hierarchie a zároveň lze tuto vlnu rozložit na jednotlivé podvlny. Tento princip ilustruje obrázek 3.2.

Zajímavou souvislostí se jeví souvislost mezi možným počtem jednotlivých podvln v rámci hierarchického uspořádání a mezi Fibonacciho posloupností, jak je uvedeno v [9].

V rámci snazšího rozlišení jednotlivých hierarchických úrovní byla zavedena jednotná mezinárodní klasifikace. Úrovně jsou rozděleny do devíti skupin, které se vyznačují především různou délkou trvání dané vlny, každá úroveň má i speciální značení číslování [9, 21]:

1. Grand Supercycle - od několika desetiletí po staletí
2. Supercycle - od několika let po desetiletí
3. Cycle - od jednoho roku po několik let
4. Primary - od několika měsíců do cca dvou let
5. Intermediate - týdny až několik měsíců



Obrázek 3.2: Princip hierarchického uspořádání - Elliottovy vlny.

6. Minor - týdny
7. Minute - dny
8. Minuette - hodiny
9. Sub-minuette - minuty

Jak bylo uvedeno výše, vlny dělíme na impulzní a korekční. Impulzní jsou ty, které jsou orientovány ve směru hlavního trendu. Jak vidíme z obrázku 3.2 impulzní vlna není vždy v rostoucím směru, nýbrž může být i klesající. Při analyzování Elliottových vln je třeba nejprve zjistit, jaká je její nadřazená vlna, zdali je to vlna impulzní či korekční. Podle toho poté víme, kolik podvln daná vlna může mít a tedy o které typy podvln se může jednat [21].

Jak je uvedeno v [21], je třeba si uvědomit rozdíl mezi idealizovanou představou aplikace teorie Elliottových vln a reálnou aplikací tohoto principu. Z [21] vychází, že v rámci idealizované představy lze jednotlivé podvlny a impulzy, které stojí za vznikem těchto podvln charakterizovat následovně.

Pokud je nadřazenou vlnou impulzní vlna býčího trendu. Potom první impulzní vlna, která je zpravidla nejkratší, bývá označována jako „odraz ode dna“ a je základní vlnou nového cyklu. Korekční vlna, která jí následuje označujeme jako „test poklesu“, kdy na základě špatných fundamentálních podmínek dochází ke snížení úrovně indexu. Druhá impulzní vlna, často označována jako „mocná vlna“, je charakterizována růstem ekonomiky (růst prosperity, očekávání), bývá strmější než předchozí vlny a realizované obchody bývají nejvyšší. Druhá korekční vlna je výsledkem zklamání nad končící růstovou fází a poslední třetí impulzní vlna představující „závěrečný vzestup“ vzniká na základě zlepšení ekonomických podmínek, kdy ovšem v závěru převažuje psychologie nadhodnocení [21].

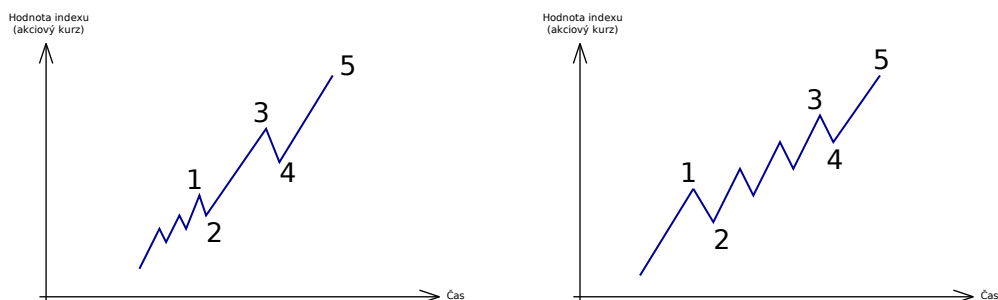
V případě, že nadřazenou vlnou je korekční vlna býčího trendu, potom první impulzní podvlna, představuje korekci daného akciového trhu. Korekční podvlna koriguje probíhající pokles a druhá impulzní vlna představuje pokračující pokles až k dosažení nového dna.

Jelikož reálné impulzní a korekční vlny nespĺňujú přesně výše popsané principy, především identifikaci jednotlivých podvln dané vlny, proto Elliottova teorie přináší mnoho rozšíření



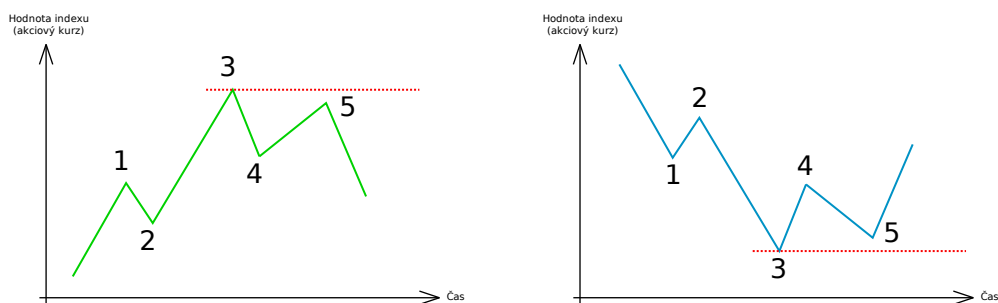
právě pro identifikaci a klasifikaci těchto podvln. V rámci této teorie tak například rozlišujeme několik druhů impulzních vln, jako jsou „rozšířené impulzní vlny“, „potlačená pátá vlna“ a „diagonální trojúhelníky“. Některé z těchto vln jsou znázorněny na následujících obrázcích.

Obrázek 3.3 znázorňuje právě případ, kdy je některá z podvln rozšířená, konkrétně v levé části obrázku se jedná o případ rozšíření vlny číslo 1 a v pravé části obrázku se jedná o rozšíření vlny číslo 3.



Obrázek 3.3: Rozšířená impulzní vlna - Elliottovy vlny.

Obrázek 3.4 znázorňuje případ potlačené páté vlny. K tomuto jevu dochází pokud vrchol páté vlny nepřesáhne vrchol vlny číslo 3. Na obrázku je uveden jak případ „býčího“, tak „medvědího“ trhu.



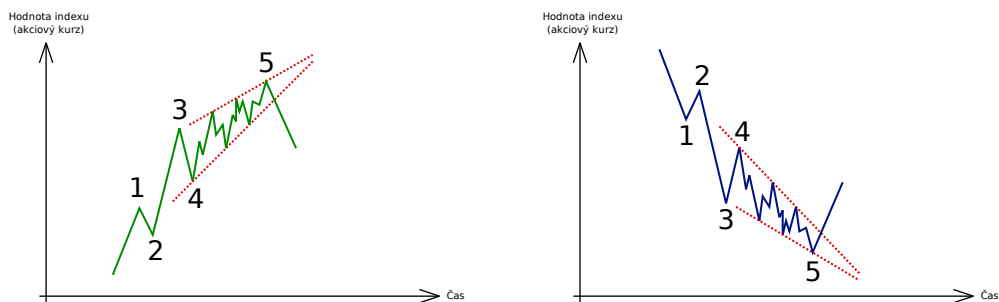
Obrázek 3.4: Potlačená pátá vlna - Elliottovy vlny.

„Třetí z variant jsou diagonální trojúhelníky. Lze je nalézt častěji v páté vlně a obecně jsou jevem, který je možné identifikovat v koncových bodech větších cenových pohybů, indikující vyčerpání daného trendu“ [6]. Graficky jsou tyto trojúhelníky znázorněny na obrázku 3.5.

Jak je uvedeno v [6], pokud se u páté vlny vyskytuje jeden z výše uvedených případů, indikuje to většinou dramatický obrat v dosavadním trendu.

U korekčních vln je rozlišováno několik skupin, jedná se o tzv. „zig-zag“ vlny, „ploché vlny“, trojúhelníky a další, které se dále liší počtem podvln, které obsahují. Podrobnější informace lze nalézt především v [9].

Jak vidíme, počet různých druhů vln je rozsáhlý a vezmeme-li v úvahu reálné případy, kdy může docházet k různým ať už menším či větším odchylkám od výše definovaných druhů, je zřejmé, že právě v této oblasti lze s výhodou využít možnosti výpočetní techniky.

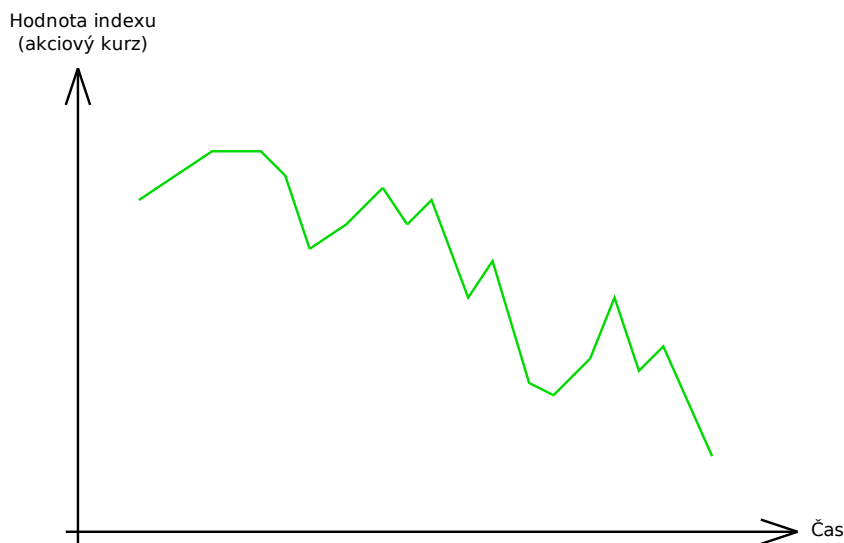


Obrázek 3.5: Diagonální trojúhelníky - Elliottovy vlny.

### 3.1.3 Grafické zobrazování

Jelikož technická analýza pracuje s historickými daty, jako jsou především kurzy a objemy obchodovaných titulů, využívá různých typů grafických zobrazovacích technik. Díky grafům, jež reprezentují požadované hodnoty časově vhodně uspořádané lze získat požadované odhady budoucích hodnot a ty vhodně vizualizovat [6]. V této podkapitole přiblížíme základní techniky, jež se využívají při vizualizaci dat potřebných pro technickou analýzu. V následujícím textu jsou popsány čárové grafy, čárkové a svíčkové grafy či technika „Kagi“.

Na obrázku 3.6 je znázorněn čárový graf. Tento graf je pro svou jednoduchost využíván především v analýze primárních trendů.

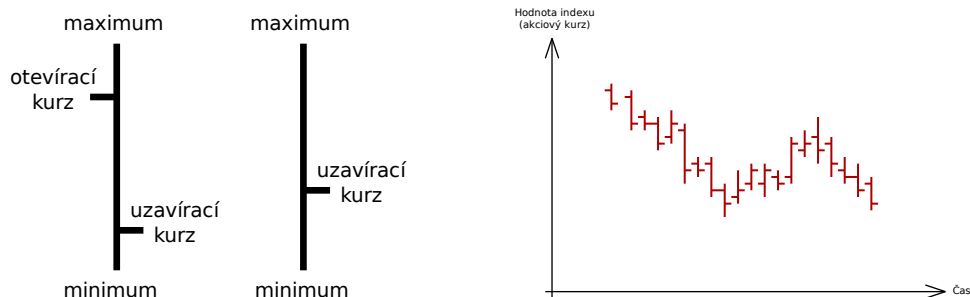


Obrázek 3.6: Čárový graf.

Složitějšími typy grafů jsou grafy čárkové a svíčkové. Tyto grafy zobrazují až čtyři druhy hodnot denních kurzů. Především se jedná o kurz otevírací, maximální, minimální a uzavírací. V případě čárkových grafů je minimum a maximum znázorněno rozpětím mezi nejvyšším a nejnižším bodem a otevírací a uzavírací kurz je určen vodorovnými čárkami po

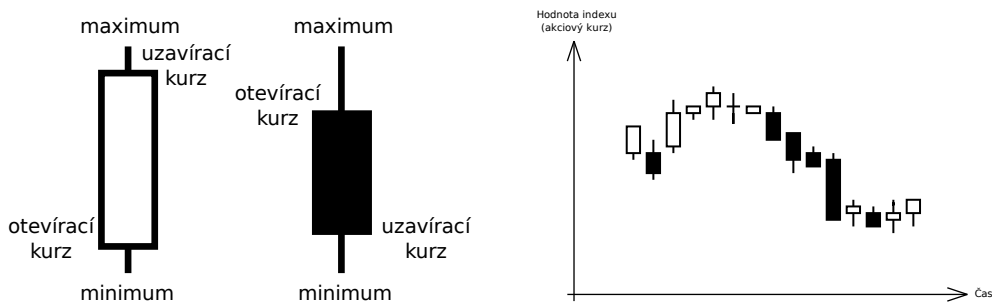
stranách, jak je vidět na obrázku 3.7.

U svíčkového grafu, jež pochází z Japonska je tomu obdobně, zde ovšem navíc lze na první pohled vidět výsledek daného časového okamžiku díky zbarvení těla „svíčky“. Pokud je otevírací kurz nižší než uzavírací používá se bílé barvy, pokud je tomu naopak používá se tmavé - černé barvy nebo se může využívat více barevných kombinací - zelená a červená apod. Schématickou ukázkou grafu lze vidět na obrázku 3.8.



Obrázek 3.7: Čárkový graf.

V levé části obrázku 3.7 je vysvětlen význam jednotlivých značek, v pravé je schématická ukáзка podoby čárkového grafu.



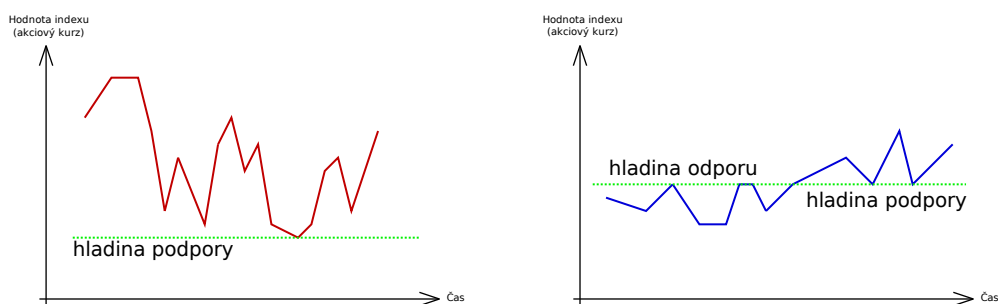
Obrázek 3.8: Svíčkový graf.

Kombinací japonské grafiky a grafiky užívané v západních zemích je možné získat další možnosti zobrazení historických dat a trendů v nich obsažených. „Kagi“ grafy jsou založeny na denních nebo týdenních uzavíracích cenách [6]. Při této technice se využívá různé tučnosti čar, která tak znamená buď pokračování v přechozím trendu či naopak. Více o této technice zobrazování lze nalézt v [6].

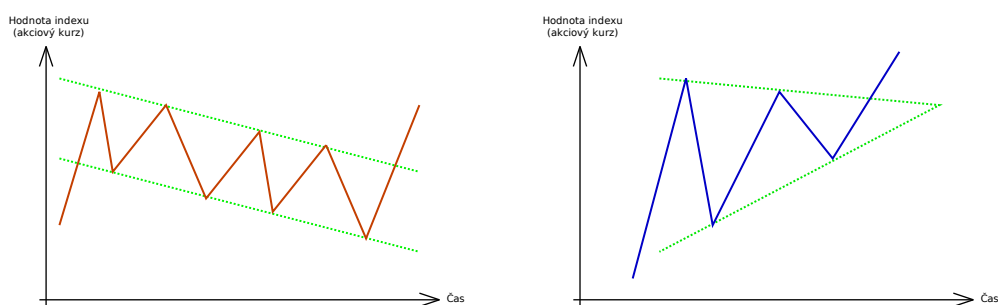
Grafické zobrazování slouží posléze jako podklad pro grafickou analýzu, kdy na základě zobrazených informací hledáme určité specifické vzory, jež mohou indikovat změny v probíhajících trendech a mohou tak sloužit jako indikátory nákupu či prodeje daného titulu.

Analyzujeme tak především vzestupné a sestupné trendy a dále různé formace, jež se mohou v grafickém zobrazení vyskytovat [21]. V rámci analýzy vzestupných a sestupných trendů je využíváno například cenových hladin, jež se označují jako hladiny odporu či podpory. „Techničtí analytici se snaží především rozpoznávat cenové hladiny (úrovně), na nichž se dají očekávat změny ve vývoji akciových kurzů (resp. změny v růstu či poklesu akciových trhů)“ [21]. Tyto hladiny schématicky zobrazuje obrázek 3.9.

Mezi formace, jež lze identifikovat v rámci grafické analýzy patří reverzní formace, konsolidační formace a mezery. Reverzní formace slouží i identifikace změn v probíhajících trendech. Rozlišují se různé typy „vrcholů“, od jednoduchých zaoblených vrcholů po dvojitě vrcholy, formace hlava a ramena apod. Konsolidační formace slouží k identifikaci pokračování původního trendu. Lze rozlišovat několik formací, mezi základní patří vlajka, praporek, konsolidační hlava a ramena či trojúhelníkové formace. Ukázky některých formací představuje obrázek 3.10. Dalším typem formací jsou mezery, jež jsou znázorněny prázdnými místy v grafech, které tak signalizují časový úsek, ve kterém se neuskutečnil s daným titulem žádný obchod. Kromě výše zmiňovaných formací existuje ještě mnoho jiných, jako například „AB-CD“ formace, „motýlí“ formace a podobně [17].



Obrázek 3.9: Hladiny podpory a odporu.



Obrázek 3.10: Konsolidační formace.

Na obrázku 3.10 je v levé části zobrazena formace „vlajka“ a v pravé části formace „praporek“. Obě formace představují pauzu uprostřed trendu rychlého růstu či poklesu akciových kurzů.

Na závěr této podkapitoly jen zmiňme, že v rámci grafické analýzy neuvažujeme pouze kurz (cenu) daného akciového indexu, nýbrž i další charakteristiky. Jednou z nich je i objem a jeho změna, která má významnou souvislost s pokračováním daného trendu apod.

### 3.1.4 Technické indikátory

Tyto indikátory, které jsou představovány funkcemi, jejichž parametry je jistý specifický počet historických dat, slouží ke stejnému účelu jako výše zmíněné formace a linie u grafické analýzy. Indikátory dělíme na cenové a objemové.

Mezi nepoužívanější indikátory patří klouzavé průměry, pásma, oscilátory a cenově objemové a objemové indikátory. Právě klouzavé průměry jsou nepoužívanějším nástrojem technické analýzy. Vyhlažují prudké výkyvy kurzů a identifikují tak probíhající trend.

Existuje několik druhů klouzavých průměrů, od jednoduchých klouzavých průměrů, vážených klouzavých průměrů až po exponenciální klouzavé průměry apod. Více o technických indikátorech lze nalézt v [21].

## 3.2 Psychologická analýza

V této podkapitole se zaměříme na psychologickou analýzu, která se zaměřuje na chování člověka, jako subjektu, jež realizuje obchody na akciovém trhu, neboť kromě sledování vývoje kurzů či objemu obchodovaných titulů je třeba nezapomenout, že obchod je také závislý na aktuálním stavu psychiky daného obchodníka. V této kapitole tak přiblížíme nejprve různé druhy investorů, s nimiž se lze na trhu setkat a poté také hypotézy a teorie, jež jsou s psychologickou analýzou spjaty. Vycházíme především z literatury [21, 8].

Význam psychologie resp. sociální psychologie, která se zabývá skupinovým myšlením a jednáním, je všeobecně znám. Každý obchodník je jedinečná osobnost, kterou charakterizují její vlastnosti, zkušenosti, aktuální zdravotní stav apod. Tyto charakteristiky mají samozřejmě vliv i na samotná rozhodnutí, jež souvisí s obchodováním na akciovém trhu.

Účastníky trhu tak můžeme intuitivně rozdělit na několik typů. Jedni mají pozitivní vztah k riskování, jiní jsou naopak opatrní. Někteří preferují využívání rozsáhlých možností informačních technologií a analýz s nimi spojenými zatímco jiní se spoléhají na svou intuici [8]. Na základě provedených empirických výzkumů byly klasifikovány jisté skupiny investorů.

**investor s vnitřním konfliktem** – tento typ investora se vyznačuje svou nerozhodností, úzkostí a závistí, které v konečném důsledku mnohdy vedou k tzv. hře s nulovým součtem, kdy výsledek investování se blíží nule a samotné investice tak nemají význam;

**investor prožívající depresi** – investor nedokáže vnímat skutečnou hodnotu rizika vzhledem k depresivním stavům, jež prožívá, což opět vede k nutným chybám, kterých se dopouští v rámci své investiční činnosti;

**investor s paranoidními rysy** – typickým rysem u tohoto typu investora je malá sebedůvěra, nevěří svým obchodním rozhodnutím a spíše spoléhá na jiné, kterým posléze dává za vinu malé zisky či neúspěchy, mění často strategii obchodování, neboť si není jist její správností;

**neklidný investor** – investor je příliš opatrný a precizně pracující člověk, jehož smysl pro detail a perfektní výsledek jde za únosnou hranici, vede k neustálému prohlížení grafů, práci s velkým počtem indikátorů a neustálému návratu zpět ke starým materiálům, rozhodování tohoto typu investora je pro „vnitřní neklid“ takto ztíženo;

**investor mstitel** – tento typ investora je představován silnou riskující osobností, která se nespokojí pouze s malým ziskem a která si trh personifikuje, pokud tak utrpí ztrátu, je schopna učinit vše, aby se pomstila, tato osoba má touhu dobýt trh riskantní operací;

**investor herec** – specifická je psychika náchylná k vnějškovému řízení, investor hraje roli, jež je v konkrétní situaci žádoucí, tato role ovšem může vyvolat konflikt, pokud je v rozporu s charakterem osobnosti investora, což může vést s chybným rozhodnutím.

Výše zmíněna klasifikace je uvedena v [8] a není konečným výčtem. Obvykle ani není možné klasifikovat daného jedince přesně do jedné dané třídy. Přínosem je především možnost (sebe)reflexe motivů, postojů či emocí investora. Jak vidíme z uvedených tříd může být jednání investorů pro třetí osoby mnohdy těžko pochopitelné. S tím souvisejí například i tzv. spekulativní bubliny zmíněné dále.

Psychologická analýza především zahrnuje teoretické koncepce jako jsou:

1. Keynesova spekulativní rovnovážná hypotéza,
2. Kostolany burzovní psychologie,
3. Teorie spekulativních bublin,
4. Drasnarova koncepce psychologické analýzy,

jejichž popis lze nalézt v [21, 23]. Tyto teoretické koncepce se zabývají především vlivem masové psychologie na pohyb kurzů daných obchodovaných titulů. V následujícím textu zmíníme některé tyto koncepce podrobněji.

### 3.2.1 Keynesova spekulativní rovnovážná hypotéza

Keynes v této teorii poprvé hovoří o chování investorů jako o hybné síle změny kurzů. Zavádí pojmy jako spekulace a podnikavost. Podnikavostí rozumí činnost vyplývající z predikce výnosu daného instrumentu na základě fundamentální analýzy. Spekulací rozumí rozhodování založené na prognóze kolektivní psychologie. Za faktory, jež ovlivňují chování akciových trhů považuje především:

- narůstající podíl vlastnictví akcií v rukou nezkušených investorů, kteří tak nejsou schopni je kvalifikovaně spravovat,
- nadměrné reakce akciových trhů na různé události, které mnohdy nejsou dostatečně významné a mají jen pomíjivý charakter, což má za následek přílišnou volatilitu kurzů,
- chování investičního publika je ovlivňováno kolektivní psychologií velkého počtu neinformovaných jednotlivců,
- investiční rozhodování se tak zaměřuje na predikci budoucího chování investičního publika, které svými rozhodnutími významně ovlivňuje pohyb kurzů.

Důležitým poznatkem této teorie je, že psychologická analýza je analýzou krátkodobou [21].

### 3.2.2 Kostolanyho burzovní psychologie

Kostolanyho burzovní psychologie předpokládá, že v krátkém období je chování burzovního publika ovlivněno více různými událostmi, zatímco v dlouhém období (více jak jeden rok) jsou hlavními faktory fundamentální ukazatele [21].

Kostolanyho burzovní psychologie uvažuje dva typy účastníku burzovního obchodování:

**hráči** – kteří chtějí dosáhnout co nejrychleji zisků, proto nejednají na základě fundamentálních ukazatelů, nýbrž na základě nových informací a událostí a chovají se tak spíše emotivně, vzhledem k tomu, že se zaměřují na krátké časové úseky jsou i jejich zisky menší, představují podle Kostolanyho až 90% účastníků;

**spekulanti** – orientují se na dlouhodobější transakce, jejich chování je podloženo konkrétními argumenty vycházející z fundamentálních ukazatelů, tito účastníci obchodů jsou podle Kostolanyho jednoznačně úspěšnější než „hráči“.

Kostolanyho teorie rozlišuje situaci na trhu podle toho, která z uvedených skupin drží v dané době většinu akcií ve svých rukou. Podle toho pak predikuje možný budoucí vývoj kurzů. Konkrétně rozlišuje následující situace:

1. kurzy rostou a rostou objemy obchodů – přesun akcií od spekulantů ke hráčům, vytváření tzv. „překoupeného“ trhu, který je vysoce rizikový,
2. kurzy klesají, ale objemy obchodů rostou – panika v řadách hráčů přesouvá akcie do rukou spekulantů, vzniká tak „předprodaný“ trh,
3. kurzy klesají a klesají objemy obchodů – tato situace znamená očekávání dalšího hlubšího poklesu
4. kurzy rostou, ale nerostou objemy obchodů – buď hráči kvůli přetrvávajícímu pesimismu nechtějí nakupovat nebo nemohou, neboť nemají dostatek finančních prostředků.

Toky akcií mezi hráči a spekulanty tak vytvářejí nekonečný koloběh, jež se skládá z fází: vzestupná korektura (nízké objemy, nízké kurzy), vzestupná doprovodná fáze (zvyšující se kurzy), přehánění vzestupného trendu (euforie, velký počet obchodů), sestupná korektura (nízký obrat), sestupná doprovodná fáze (pokles kurzů akcií) a přehánění sestupného trendu (kurzy klesají) [21].

Kostolanyho teorie tak říká, že bychom měli kupovat ve fázích vzestupné korektury a přehnaného sestupného trendu, vyčkávat ve fázích vzestupné a sestupné doprovodné fáze a prodávat ve fázích přehnaného vzestupného trendu a sestupné korektury.

### 3.2.3 Teorie spekulativních bublin

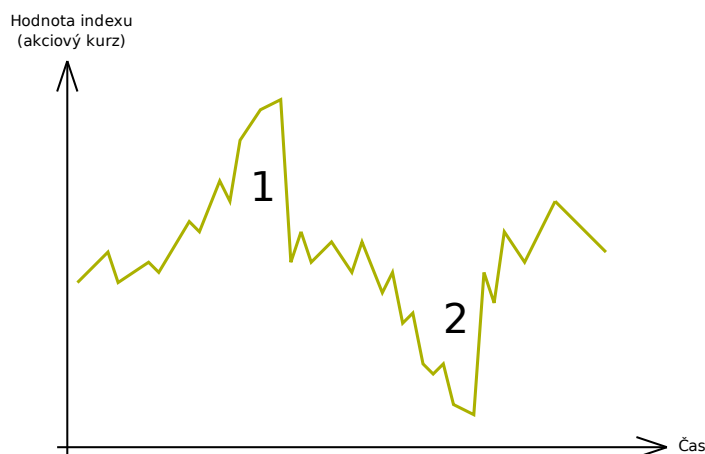
Spekulativními bublinami rozumíme takový vývoj na akciových trzích, kdy po určité době růstu (poklesu) kurzů, dochází k prudkému opačnému trendu a obojí je spojeno s velkým objemem obchodů. Významným znakem je rozdíl mezi kurzovou hodnotou a vnitřní hodnotou akcie daného titulu definovanou fundamentální analýzou [21].

Existují dva typy spekulativních bublin. U prvního dochází nejprve k růstu kurzů nastane na trhu neočekávaný prudký propad nesoucí ztráty. U druhého dochází k poklesu kurzů, který je posléze následován prudkým vzestupem, jak znázorňuje obrázek 3.11.

Vznik spekulativní bublin bývá většinou spojen s nadměrnou reakcí na nějakou událost, která ovlivňuje chování burzovního publika, které tak působí na změnu kurzů. Přesto názory na vznik spekulativních bublin nejsou jednotné. Existují teorie, které spojují vznik spekulativních bublin s chováním monetárních autorit, jež je spojeno s expanzivní monetární politikou a nadměrnou peněžní likviditou. Spekulativní bubliny tak vedou k neefektivní alokaci zdrojů, zesilování hospodářských cyklů a snížení důvěry ve finanční systém [11].

V rámci historie patří mezi nejvýznamnější projevy spekulativních bublin především:

**Tulipánová mánie** – poptávka po tulipánových cibulkách byla tak velká, že způsobila neustálý růst ceny tohoto předmětu obchodování, že se vyplatilo spekulovat o jeho dalším možném růstu a následném výhodném prodeji až nakonec nastal prudký cenový pokles spojen s obchodním vystřízlivěním [8];



Obrázek 3.11: Spekulativní bubliny.

**Tichomořská bublina** – anglická společnost South Sea Company byla založena za účelem získání financí za zaplacení úroků z válečného dluhu, společnost získala monopol na obchodování se zeměmi Jižní Ameriky, jelikož se předmět obchodování a počáteční výsledky zdáli být slibné, rostla poptávka po akciích této společnosti, což přinášelo zisky prvním investorům, jak rostl kurz akcií, uvědomili si tito, že situace je neúnosná, prodej akcií vyvolal prudký propad kurzu [11];

**Velký krach na New Yorkské burze 1929** – růst kurzu většiny akcií vyvolal vidinu velkých zisků pro nové příchozí nezkušené investory, tito využívali úvěrů bank, které ručily nakoupenými akciemi, optimismus na trzích byl náhle vystřídán pesimismem, který zapříčinil prudký propad kurzů, krach na burze posléze znamenal vznik ekonomické krize [21, 11];

**Krach na americkém akciovém trhu 1987** – „... v době prasknutí bubliny na trh nepřišli žádné informace, které by mohly způsobit takový pád, přičemž stejně tak neexistovaly ani žádné fundamentální důvody pro rychlý růst kurzů ...“ [21];

**Krach japonského akciového trhu** – japonské banky vlastnily akcie průmyslových podniků, jež naopak vlastnily akcie bank, pokles kurzů akcií vyvolal pokles možností bank úvěrovat, což ovlivnilo hospodářské výsledky firem, což opět negativně ovlivnilo situaci bank, důsledky tohoto krachu byly dlouhodobé [21];

**Internetová horečka** – rozvoj IT firem, jež se soustředili na tržní podíl a zisky v budoucnu a nízké úrokové sazby způsobily, že kurzy akcií těchto firem rostly a staly se silně nadhodnocenými, když v roce 2000 bublina praskla a firmy, jež byly orientovány na zisky v budoucnu přišli o možnosti financování, mnoho firem zkrachovalo a přežily pouze nejsilnější [21, 11].

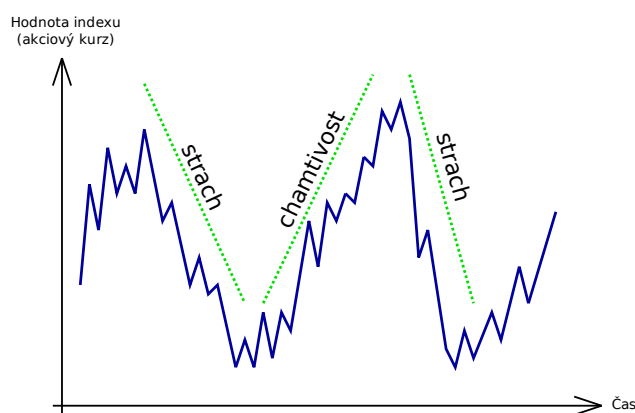
Z uvedených historických údajů tedy vyplývá, že na vzestup a pád akciových trhů z pohledu spekulativních bublin má vliv především monetární politika centrálních bank (především včasnost použití vhodných monetárních nástrojů může zmírnit negativní projevy), propojení bankovní a podnikové sféry (což nejlépe ukazuje krach japonského akciového trhu), nakupování akcií na úvěr a souběžný vývoj na ostatních trzích (devizový



trh, trh nemovitostí apod.), jelikož ekonomický systém je provázaný a jednotlivé součásti se vzájemně ovlivňují. A v neposlední řadě významným faktorem je dodržování zákonů a předpisů, které souvisí s obchodováním na finančních trzích [21].

### 3.2.4 Drasnarova koncepce psychologické analýzy

Autor této teorie vysvětluje růst či pokles kurzů dvěma opačnými lidskými vlastnostmi. Jsou jimi chamtivost a strach. Podle toho, která převládá, dochází buď k růstu či poklesu akciových kurzů [21]. Chamtivost je spojena s růstem kurzů. Snaha dosahovat vysokých zisků podporuje poptávku, která má za vznik růst kurzu. Tento trend je posléze vystřídán strachem o ztrátu nabytého bohatství, dochází k poklesu poptávky a tím i k poklesu kurzů. Tyto dvě klíčové vlastnosti se střídají a podle toho dochází i ke střídání trendů v rámci kurzů akciových titulů, což znázorňuje i následující obrázek 3.12.



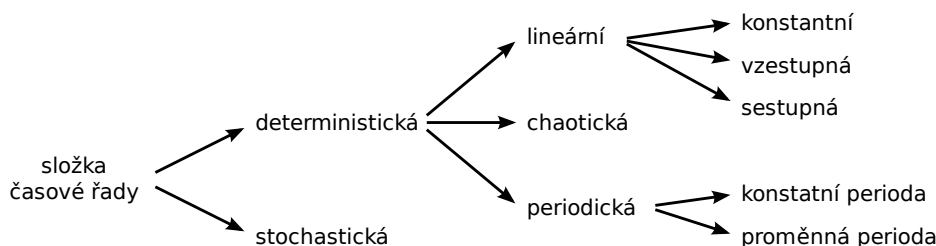
Obrázek 3.12: Drasnarova koncepce psychologické analýzy.

## 3.3 Predikce časových řad

Predikce časových řad je problematika, která se dotýká mnoha oborů, od statistiky po ekonomické obory apod. V rámci predikce časových řad se snažíme předpovědět budoucí hodnoty, nalézt mechanismy, na nichž jsou závislé hodnoty časové řady a faktory, které je ovlivňují [4].

Pod pojmem časová řada si můžeme představit sekvenci hodnot, jež je závislá na čase. Příkladem jsou například počty narozených dětí v daných dnech, množství vyrobených produktů v podniku či cena určitého titulu na daném akciovém trhu. Jak vidíme, oblast predikce časových řad rozsáhlá a každý obor, do kterého zasahuje přináší své vlastní specifické nástroje pro predikci.

Časová řada se může skládat ze dvou složek 3.13. Těmi jsou složka deterministická a stochastická. Deterministickou složku lze dělit dále na složky trendové, složky s konstantní nebo nekonstantní periodou apod. Samotná časová řada lze tak rozložit na několik složek a každá může mít z pohledu predikce svou „váhu“.



Obrázek 3.13: Složky časové řady. Schéma převzato z [4].

Pro predikci časových řad se využívá mnoho metod, jak z oblasti statistiky, tak především z oblasti umělé inteligence či soft computingu. Jedná se především o neuronové sítě, fuzzy logiku. V rámci různých optimalizačních problémů spojených s predikcí časové řady je také možno použít evolučních algoritmů apod.

Při predikci časové řady hrají klíčový význam metody využití při predikci a znalost složek časové řady. Při predikci je vhodné vyhodnocovat chyby předpovědi a zlepšovat podle těchto chyb výsledný model [4].

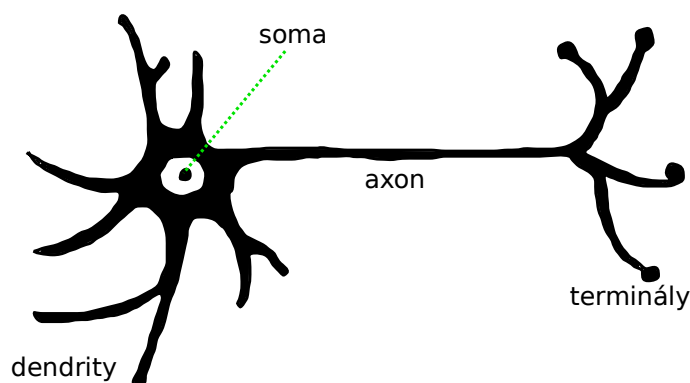
V následujících podkapitolách přiblížíme právě důležité nástroje, jež lze využít při predikci časových řad, především neuronové sítě, evoluční algoritmy, fuzzy logiku. Jelikož se v další části práce věnujeme psychologické analýze připomeneme také oblast dolování dat a klasifikace, na kterou navážeme v praktické části práce.

### 3.4 Neuronové sítě

V této podkapitole přiblížíme problematiku neuronových sítí. Jedná se o systém složený z jednoduchých prvků (neuronů), jež jsou inspirovány biologickými neurony 3.14.

Tyto jednotlivé neurony jsou vzájemně propojeny a jsou rozděleny do vrstev. Systém má vstupy, výstupy a dále je charakterizován váhami a prahovými hodnotami. Jsou to právě neuronové sítě, které jsou využívány v oblasti klasifikace a predikce časových řad, které je možno využít při predikci vývoje akciového trhu a mohou nahradit statistické odhady založené například na klouzavých průměrech [8]. Existuje několik druhů neuronových sítí, některé jsou uvedeny v následujících podkapitolách.

Neuronové sítě přináší při řešení jistého problému jinou koncepci, než je koncepce John von Neumanna. Na rozdíl od této koncepce, neuronové sítě neznají přesný algoritmus řešení daného problému, ale naopak tento algoritmus tzv. procesem učení hledají. U neuronových sítí je tak kladen především důraz na strukturu dat, která je představována nastavením vah hran, jež propojují jednotlivé neurony dané sítě. Hledáním vhodné struktury a tedy vhodných vah se zabývá učení neuronové sítě, které užívá určitého učícího algoritmu, který hledá optimální řešení pro danou úlohu [8]. Samotná problematika učení je rozsáhlá a je spojena s několika problémy jako je na příklad „přetrénování“, které znamená, že síť ztratila schopnost generalizace tj. použitelnost na obecné či podobné problémy. Problematika neuronových sítí je obsáhlá a tato práce si nebere za cíl uvést všechny parametry, které identifikují neuronovou síť, její různé typy či proces učení. Připomeneme pouze nejjzákladnější

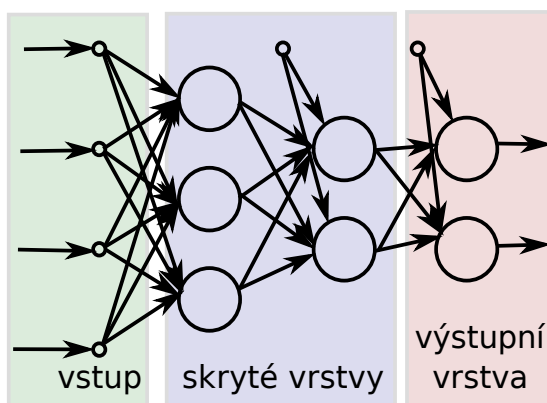


Obrázek 3.14: Schéma biologického neuronu.

problematiku a typy sítí, především ve spojitosti s predikcí časových řad.

### 3.4.1 Dopředná neuronová síť

Tento typ neuronové sítě patří mezi nejrozšířenější, je tomu tak proto, neboť její struktura je relativně jednoduchá a osvědčila se právě při různých problémech spojených s klasifikací či predikcí. Jak vidíme na obrázku 3.15, má tato síť několik vstupů, neurony jsou rozděleny do několika vrstev a platí, že hrany nejsou orientovány do vrstvy předcházejí, odtud vychází název sítě „dopředná“.



Obrázek 3.15: Dopředná neuronová síť.

Z obrázku 3.15 je patrné rozložení sítě, kdy nejprve vstupní vrstva distribuuje hodnoty ze vstupu skrytým vrstvám, jež jsou následovány výstupní vrstvou. Počet skrytých vrstev je závislý na povaze úlohy, kterou síť „řeší“.

Sít' je složena z neuronů. Neuron představuje matematický model, jehož činnost lze popsat vzorcem:

$$y = f(u) = f(u(\vec{x})), \quad (3.1)$$

kde  $\vec{x}$  je vektor vstupních hodnot,  $u$  je bázová funkce a  $f$  je aktivační funkce. Speciálním případem neuronové sítě je tzv. perceptron, tedy síť obsahující pouze jeden neuron.

Neuron je tedy charakterizován bázovou funkcí, která převádí vstupní vektor  $\vec{x}$  na hodnotu  $u(\vec{x})$ , která je posléze vstupem aktivační funkce. U tohoto typu sítě je použita lineární bázová funkce:

$$u = \omega_0 + \sum_{i=1}^n \omega_i x_i, \quad (3.2)$$

kde  $\vec{x}$  je vektor vstupních hodnot daného neuronu a  $\vec{\omega}$  je vektor vah příslušných k daným vstupům.

Jako aktivační funkce se většinou využívá buď skoková aktivační funkce nebo pro predikci časových řad různé typy spojitých funkcí, především „sigmoida“, hyperbolický tangens či lineární funkce [12].

Tento typ neuronové sítě je také nazýván „Neuronová síť Backpropagation“ a to podle metody učení, který se u ní využívá. Algoritmus „Backpropagation“ je založen na zpětném šíření chyby v rámci učení sítě spojeným s úpravou vah umělých neuronů. V rámci učení jsou důležité parametry jako koeficient učení, jenž ovlivňuje rychlost učení a kvalitu výsledku, dále momentum koeficient, který napomáhá učicímu algoritmu vyhnout se lokálním extrémům. Pro každou úlohu, na kterou se síť učí jsou pak specifické další parametry učení jako je například počet „epoch“, počáteční nastavení vah, počet skrytých vrstev a jejich neuronů. Je viditelné, že problematika neuronových sítí je obsáhlá, proto podrobnější informace lze nalézt především v [12].

### 3.4.2 Další typy neuronových sítí

Především v rámci predikce časových řad je možno využít dalších typů neuronových sítí, konkrétně neuronová síť s radiální bázovou funkcí. Tato síť má vstupní a skrytou vrstvu s radiálními jednotkami a obvykle lineární výstupní vrstvu. Učení této sítě je obvykle rychlejší, ovšem výsledná síť je obvykle pomalejší.

Dalšími typy sítí jsou neuronové sítě vyšších řádů, Elmanova síť (rekurentní), Jordanova síť atd. Síť a experimenty s nimi na časových řadách jsou například zmíněny v [18].

### 3.4.3 Software neuronových sítí

Software, jenž je zaměřen na neuronové sítě, existuje celá řada, ať už od profesionálních placených řešení po volně šiřitelné nástroje či knihovny, které je možné využít v rámci daného programovacího jazyka.

Z profesionálních řešení jsou to především Statistica Neural Networks, Neural Connection, NeuroShell a podobně [8]. V rámci volně šiřitelného software jsou to JavaNNS, Emergent, Matlab Neural Network Toolbox apod. V ekonomii lze využít především Neural Connection [8].

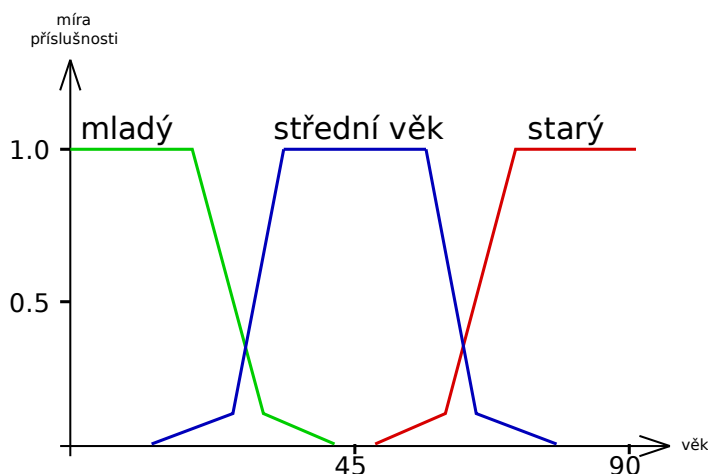
### 3.4.4 Aplikace v ekonomii

Z [8, 18], plyne možnost využití neuronových sítí v oblasti predikce časových řad. Tedy právě v oblasti technické analýzy. Jak vidíme v [8], konkrétně porovnání empirických a vypočítaných dat pomocí neuronové sítě, vyšla neuronová síť s lepším výsledkem. Neuronové sítě lze však s úspěchem využívat i při jiných „problémech“ v rámci ekonomie jako například klasifikace apod.

## 3.5 Fuzzy množiny

Fuzzy množiny spojujeme se zobecněním obyčejných množin, kdy každému prvku přísluší tzv. stupeň příslušnosti prvku do dané množiny. Na rozdíl od klasické teorie množin prvek do dané fuzzy množiny náleží s určitou mírou. Fuzzy množiny a fuzzy logika, jež je na nich založena, se tak více přibližuje klasickému lidskému uvažování, jež neprobíhá v přesných pojmech, nýbrž zahrnuje jistou míru nejistoty, nejednoznačnosti či neurčitosti [24].

V rámci fuzzy logiky jsou využívány tzv. lingvistické proměnné, jejichž hodnotami jsou lingvistické termy 3.16. Těmito termy jsou slova z nějaké množiny slov. V rámci teorie fuzzy množin jsou využívány pojmy jako nosič, singleton, jádro, normální fuzzy množina, řezy fuzzy množiny atd. Jsou také definovány operace nad těmito množinami jako rovnost, podmnožina, průnik, sjednocení, doplněk a podobně [24].

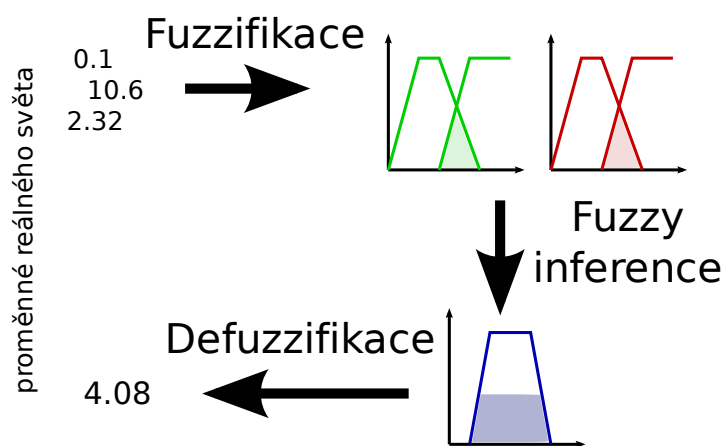


Obrázek 3.16: Lingvistické proměnné a jejich hodnoty - Fuzzy množiny.

Využití fuzzy logiky je již v dnešní době široké od různých počítačových systémů, expertních systémů, přístrojové techniky, spotřební elektroniky atd. Využití fuzzy logiky tak lze najít i v oblasti kapitálového trhu [8].

Fuzzy množiny jsou využívány v oblasti expertních systémů, které jsou založeny na pravidlech. Tyto systémy jsou využívány například v oblasti podpory rozhodování či predikce vývoje budoucích kurzů. Schéma obecného fuzzy systému je znázorněno na obrázku 3.17.

V rámci fuzzifikace jsou reálné proměnné převedeny na lingvistické proměnné reprezentující dané fuzzy množiny a stupeň členství v nich. V následujícím kroku „fuzzy inference“ jsou aplikována pravidla, jejichž výstupem jsou lingvistické proměnné vyjadřující



Obrázek 3.17: Schéma fuzzy systému - Fuzzy množiny.

míru příslušnosti do dané fuzzy množiny. V posledním kroku „defuzzifikace“ jsou tyto lingvistické proměnné převedeny na reálné proměnné.

## 3.6 Evoluční algoritmy

Dalším konceptem, který se inspiruje biologií, jsou evoluční algoritmy. Oblast evolučních algoritmů je rozsáhlá a je představována především obory jako jsou: genetické algoritmy, evoluční strategie, evoluční programování, genetické programování, klasifikační systémy atd. V této podkapitole uvedeme základní principy, na nichž jsou založeny evoluční algoritmy a jejich uplatnění při predikci vývoje akciového trhu. Vycházíme především z [22, 8].

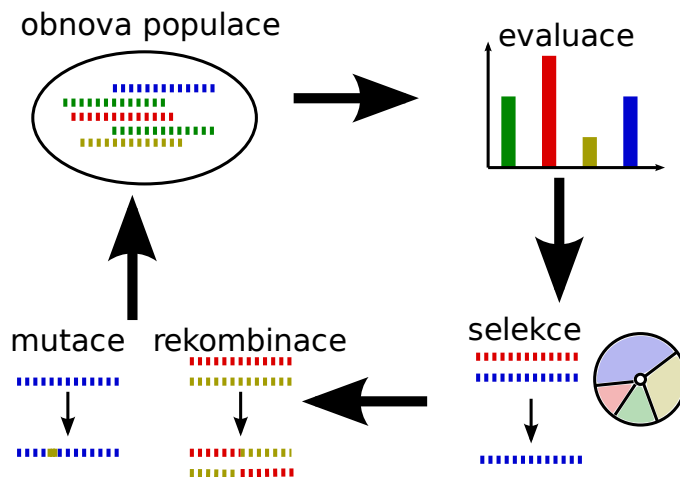
Evolučními algoritmy můžeme chápat proces, který „hledá“ co neoptimálnější řešení daného problému. Při tomto „hledání“ využívá nahodilosti, různých postupů selekce, rekombinace a mutace. Kritériem úspěšnosti daného jedince, který představuje daný vzorek hledaného řešení, je tzv. „fitness“. Abychom mohli použít evolučních algoritmů je v první řadě nutné mít možnost daný problém vhodně zakódovat do chromozomu.

Tato technika se využívá především při hledání optimálních řešení, která jsou závislá na přijatých kritériích. V oblasti kapitálových trhů tak lze evoluční algoritmy využít například při hledání optimálního portfolia aktiv [8].

### 3.6.1 Princip

Princip evolučního algoritmu je schematicky znázorněn na obrázku 3.18.

Na počátku algoritmu je vytvořena počáteční populace jedinců (ti představují potenciální řešení daného problému). Následně je vyhodnocena fitness každého jedince v populaci. Z takto ohodnocených jedinců jsou vybráni ti, na které bude aplikována mutace či rekombinace. Pro výběr jedinců se využívá několik metod, nejvýznamnější je tzv. „ruleta“, při níž mají jedinci s vyšší fitness větší šanci výběru. Využívá se také tzv. elitářství, jež zabezpečuje, že jedinci s nejlepší fitness jsou v další populaci obsaženi automaticky.



Obrázek 3.18: Schéma principu evoluční algoritmu.

Operace mutace a křížení (rekombinace) vytvářejí nové jedince na podobném principu jako je tomu v biologii. Nově vzniklí jedinci jsou poté zařazeni do nové populace. Výše popsany proces probíhá tak dlouho, dokud není nalezeno řešení s odpovídající fitness a zároveň dochází ke zlepšování fitness jedinců v populaci (konvergence k optimálnímu řešení).

Následuje stručný popis některých technik z oblasti evolučních algoritmů [22]:

**genetické algoritmy** – chromozom je pevné délky a obsahuje celočíselné (binární) hodnoty, využívá operátory mutace a křížení, jedná se o nejvyžívanější metodu;

**evoluční strategie** – chromozom obsahuje kromě zakódovaného řešení také tzv. strategické parametry, jež ovlivňují evoluci a mohou se tak vyvíjet spolu s daným řešením, hlavní důraz je kladen na mutaci;

**evoluční programování** – chromozom kóduje graf, je využíváno především mutace;

**genetické programování** – cílem evoluce je nalezení algoritmu (programu), který řeší danou problematiku, řešení je reprezentováno grafem (stromem), využívá jak křížení tak mutaci.

### 3.6.2 Aplikace v ekonomii

Evoluční algoritmy se využívají především při optimalizace struktury portfolia investora. Jedinci v populaci kódují složení portfolia a fitness funkce hodnotí jedince podle rizikovosti dané struktury [5].

## 3.7 Teorie chaosu

Tato teorie je zaměřena na chování nelineárních systémů, které ač se jeví jako náhodné až „chaotické“ přesto mají skrytý řád, jímž se řídí. Tato teorie je známá především v oblasti matematiky a fyziky, kdy pod takovým nelineárním systémem můžeme například chápat zemskou atmosféru [6, 4].

Důležitým pojmem v teorii chaosu je tzv. „Atraktor“, což je stavová veličina definující rovnováhu systému. Atraktor je stav, do kterého systém směřuje, ale samotný systém v něm nesetrvává. Atraktor může být bodový, cyklický či chaotický [4]. V případě chaotického atraktoru hovoříme o tzv. dynamické rovnováze.

Teorii chaosu zde zmiňujeme především v souvislosti s predikcí časových řad, kdy můžeme zkoumat „chaotičnost“ těchto řad, například v souvislosti s Hurstovým exponentem.

## 3.8 Klasifikace

Klasifikací rozumíme proces, kdy na základě vytvořeného modelu rozdělujeme data do daných klasifikačních tříd. Samotný proces klasifikace má dvě fáze. V první fázi pomocí trénovacích dat vytváříme model klasifikátoru pomocí metod „učení“. V druhé fázi použijeme daný model pro zařazení dat do daných tříd [14].

Data, která jsou využívány při klasifikaci musí být nejdříve upraveny do vhodné formy. Data jsou tak čištěna (odstranění šumu a doplnění chybějících hodnot), dále jsou odstraněna irelevantní či redundantní data a následně jsou data transformována za účelem přizpůsobení se klasifikačnímu modelu, což znamená jejich generalizaci a normalizaci.

V rámci klasifikace je možno využít několik technik či metod. Především se jedná o rozhodovací stromy, systémy rozhodovacích pravidel, statistické modely, naivní Bayesovská klasifikace, neuronové sítě atd.

### 3.8.1 Bayesovská klasifikace

Bayesovská klasifikace je založena na statistice. Konkrétně predikuje pravděpodobnost příslušnosti daného vzorku do definovaných tříd. Bayesovská klasifikace je založena na Bayesově větě, která hovoří o podmíněné pravděpodobnosti [14]. Právě Bayesova klasifikace, konkrétně „naivní Bayesův klasifikátor“ je co do výkonu srovnatelný s rozhodovacími stromy či neuronovými sítěmi a navíc je vhodný pro databáze s velkým objemem dat.

Naivní Bayesův klasifikátor získal jméno díky své jednoduchosti, která je založena především předpokladu nezávislosti atributů, které reprezentují data, jež chceme klasifikovat. Dalším nástrojem jsou Bayesovské důvěrné sítě představující grafický model vztahů [14].

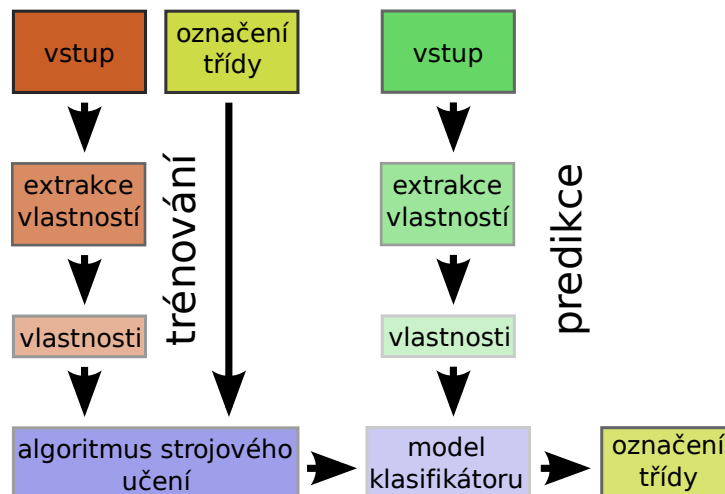
### 3.8.2 Klasifikace textů

Klasifikace textů stejně jako „obecně“ klasifikace představuje proces, která má dvě fáze. V první fázi je vytvořen model klasifikátoru, který je vytvořen „učením“ či „trénováním“ na dané množině testovacích hodnot. V druhé fázi je tento model využíván pro samotnou klasifikaci dat. Pod pojmem data v oblasti klasifikace textů rozumíme text, většinou ve formě dokumentů [1].

Cílem klasifikace textů je tedy danému textu (dokumentu) přiřadit vhodnou třídu. Základním problémem při této klasifikaci je identifikace takových vlastností textu, které jsou relevantní vzhledem k významu třídy, do které je klasifikován. Při identifikaci těchto vlastností lze využít některé optimalizační techniky zmíněné výše 3.6.

K trénování či učení klasifikátoru lze využít několika metod. Jednou z nich je tzv. „učení s učitelem“, kdy je množina dat rozdělena na trénovací množinu a testovací. V trénovací





Obrázek 3.19: Učení s učitelem, schéma převzato z [1]

množině každá entita obsahuje jak data určená pro klasifikaci, tak označení třídy, do které daná data patří. Model klasifikátoru se na této množině „trénuje“ a následně je na testovací množině ověřena jeho správnost. Schéma metody učení s učitelem je znázorněno na obrázku 3.19.

Na obrázku 3.19 tak vidíme, že klasifikace má dvě části a to trénování a samotnou predikci. Při trénování vzniká model klasifikátoru, který je následně využit při predikci.

Mezi známé klasifikační úlohy patří například rozhodování o tom, zdali daný email je spam či nikoli, klasifikace zpráv do oblastí jako sport, politika apod. Klasifikačních úloh existuje několik a to od klasifikace do více tříd po klasifikaci do tříd, které nejsou všechny definovány na počátku samotné klasifikace [1].

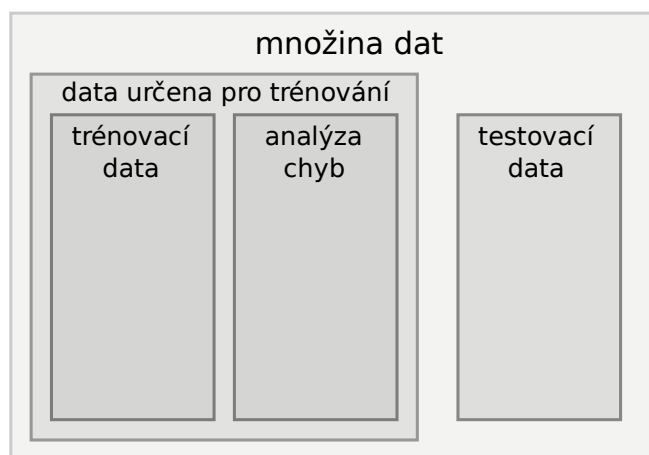
Velký vliv na úspěšnost klasifikace má výběr vhodných vlastností textu a jejich vhodné zakódování. Tato problematika je jednou ze „zajímavých“ oblastí samotné klasifikace textů. Výběr vhodných vlastností a jejich zakódování je závislé především na správném pochopení problematiky témat tříd, do nichž klasifikujeme.

Typickým postupem hledání vhodných vlastností textu je metoda „pokus-omyl“, kdy na základě provedených testů a jejich vyhodnocení provádíme úpravu extrahovaných vlastností. Z počátku lze extrahovat všechny možné vlastnosti, které jsou v průběhu testování odstraňovány až je nalezeno optimální řešení. Je třeba ovšem dát pozor na obdobný problém jako u neuronových sítí a to na tzv. „přetrénování“, kdy klasifikátor ztrácí schopnost generalizace a je tedy nepoužitelný pro „obecné“ problémy [1].

Při postupu „pokus-omyl“ se využívá rozdělení dat do dvou množin. V první množině jsou data určena pro „vývoj“ a v druhé data určena pro testování úspěšnosti. Data pro vývoj jsou dále rozdělena na trénovací data a testovací data, jež slouží pro hledání chyb v rámci testovacích dat. Rozdělení množiny dat je znázorněno na obrázku 3.20.

Před samotným trénováním a predikcí je třeba data, která chceme klasifikovat, upravit do vhodné formy, tak jak je to obecně u klasifikace běžné. U klasifikace textů se tak jedná především o [1, 15]:

**tokenizace** – jsou odstraněna interpunkční znaménka, řetězce znaků bez oddělovačů jsou považována za „tokeny“;



Obrázek 3.20: Rozdělení dat, schéma převzato z [1]

**normalizace** – písmena tokenů (slov) jsou převedena do jednotné podoby (buď na velká či malá písmena), slova jsou zkrácena do základního tvaru (kořen slova);

**odstranění irelevantních dat** – jsou odstraněna slova, jež nemají vliv na klasifikaci, jedná se především o tzv. „stop slova“ jako například členy, předložky a často užívaná slova.

Z takto upravených dat jsou posléze extrahovány žádané vlastnosti, které jsou využity při klasifikaci.

Co se týče extrakce vlastností z dat (v tomto případě z předzpracovaného textu), patří mezi nejjednodušší metodu zachycení informace, zdali se dané slovo v textu vyskytuje či nikoli. Zde ovšem vyvstává problém, na jaká slova, jejichž výskyt chceme zachytit, se zaměřit. Tato jednoduchá metoda dále nezachycuje kontext či význam dané věty zcela přesně, neboť chybí zachycení vazeb mezi více slovy, která teprve dávají větě správný význam. Proto se využívá složitějších technik, jako je tvorba „bigramů“ (dvojic) či „trigramů“, které jsou schopny lépe zachytit kontext dané věty [1].

Při klasifikaci textů jsou tak hlavními problémy:

1. identifikace klíčových vlastností textu
2. správné zakódování vlastností pro potřeby vytvoření modelu.

### 3.9 Dolování dat

Dolováním dat či získáváním znalostí z databází rozumíme obor informačních technologií, který je zaměřen na extrakci znalostí z velkého množství dat, analýzu dat apod. Dolování dat v této práci uvádíme především proto, že mnohé postupy a metody, které dolování dat využívá souvisí s výše uvedenými koncepty z oblasti soft computingu a umělé inteligence a to především v souvislosti predikcí či klasifikací [14].

Dolování dat je relativně nová oblast, která zažívá rozmach především s rozvojem databázových systémů a informačních technologií obecně. V neposlední řadě je významným faktorem rozvoj internetu.

Z pohledu predikce vývoje akciových trhů lze uvažovat o využití metod z oblasti dolování dat především v souvislosti s psychologickou analýzou. Například klasifikace zpráv, které jsou představovány daty v databázi a jejich vliv na pohyb burzy a jeho další vývoj je jedna ze zajímavých oblastí užití těchto technik.

Kromě výše uvedeného se dolování dat využívá také například v biomedicíně, analýze DNA, analýze finančních dat (klasifikace zákazníků, detekce finančních úniků apod.), telekomunikačních dat atd.

V této kapitole jsme uvedli teoretická východiska predikce vývoje akciového trhu, tedy především metody technické analýzy a psychologické analýzy. Dále jsme uvedli techniky či metody z oblasti umělé inteligence či soft computingu, které lze využít právě při technické či psychologické analýze.

V dalším textu se budeme zabývat především využitím uvedených metod či nástrojů v oblasti predikce vývoje akciového trhu. Především se zaměříme na psychologickou analýzu, které není věnováno tolik pozornosti jako analýze technické.

## Kapitola 4

# Analýza problému a současné situace

V této kapitole se zaměříme na analýzu problému predikce vývoje akciového trhu a to především ve spojitosti využití metod a postupů z oblasti umělé inteligence či soft computingu v rámci technické či psychologické analýzy.

Po teoretické stránce budeme navazovat na předchozí kapitolu, v níž jsou uvedeny základní teoretické poznatky z oblasti technické analýzy, psychologické analýzy, predikce časových řad a z oblasti umělé inteligence.

Uvedeme také současnou situaci v oblasti využití výše zmíněných metod a to především na základě prací či článků, které se zabývají touto tematikou. V dalších textu práce se již zaměříme především na užití metod v rámci psychologické analýzy.

### 4.1 Specifika

V této podkapitole si shrneme specifika či vlastnosti typické pro úspěšnou predikci vývoje akciového trhu. Především si připomeňme, že účastníky trhu jsou lidé, kteří na daném trhu provádí obchody a tak ovlivňují kurzy obchodovaných titulů. Tito lidé využívají při predikci metody, které jsou inspirovány výše zmíněnými teoriemi. Z výše uvedeného je tedy zřejmé, že:

- teorie využívané při predikci vývoje akciového trhu jsou mnohdy inspirovány procesy probíhajícími v přírodě (perioda),
- lidé „nepřemýšlí“ v přesných pojmech,
- pro třetí osoby se může zdát pohyb kurzů na trhu chaotický,
- pro úspěšné obchodování je třeba znát problematiku trhu,
- lidské rozhodování je ovlivněno aktuálním psychickým stavem a dostupnými informacemi,

samozřejmě bychom mohli najít více faktů, jež jsou specifické pro obchodování na burze. Tyto uvedené ale představují základ, z kterého budeme v dalším vycházet.

Nyní si připomeňme jednotlivé metody z oblasti umělé inteligence, které jsme uvedli v kapitole 3, a jejich stručnou charakteristiku:

**neuronové sítě** – inspirace z biologie, konkrétně inspirace lidským mozkem, učení pomocí příkladů (zkušenosti), klasifikace;

**fuzzy množiny a fuzzy logika** – stupeň příslušnosti do dané množiny, expertní systémy, pravidla, nejistota;

**evoluční algoritmy** – inspirace z biologie, význam „nahodilost“, optimalizační úlohy, hledání řešení v rámci daného stavového prostoru;

**dolování dat** – získávání znalostí z databází;

**teorie chaosu** – nelineární systémy se skrytým řádem, rovnováha systému.

Z výše uvedených výčtů vidíme, že mnohé metody z oblasti umělé inteligence mají „podobné“ charakteristiky (východiska, inspirace), jako metody z oblasti problematiky predikce vývoje akciového trhu. Především se jedná o „inspiraci z biologie“, neurčitost, chaotičnost a získávání informací. V navazujících podkapitolách přiblížíme propojení jednotlivých metod z oblasti umělé inteligence s metodami z oblasti predikce vývoje akciového trhu.

## 4.2 Technická analýza

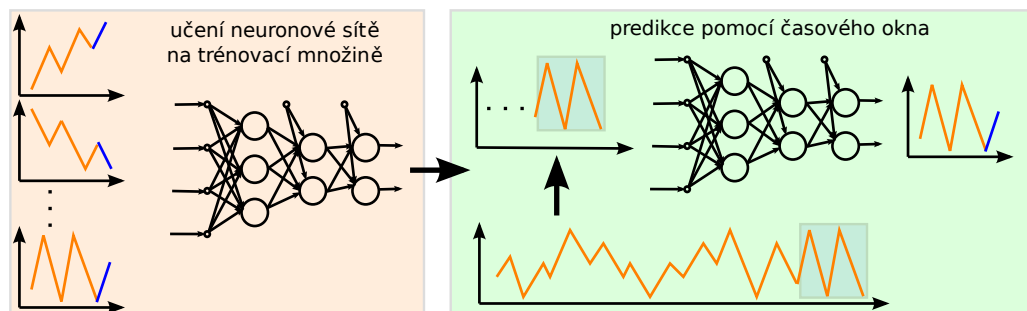
Princip technické analýzy je popsán v podkapitole 3.1. Ze znalostí o technické analýze vyplývá, že mezi obecně užívané nástroje technické analýzy patří především [6] trendové linie, klouzavé průměry, „Bollinger Bands“, „Directional Movement Index“, „Price and Volume Trend“, momentum, „Money Flow Index“ apod. Tyto metody se neustále vyvíjí a zdokonalují a to i přesto (nebo právě i proto), že existují argumenty proti technické analýze, jako je například správnost základních předpokladů technické analýzy atd.

Tyto metody (a další) úzce souvisejí s grafickou analýzou, neboť technická analýza studuje časovou řadu daného akciového indexu (titulu) a tu lze znázornit grafem, ve kterém se lze snáze orientovat. Metody grafické analýzy „hledají“ v daném grafu vzory, podle kterých lze usuzovat budoucí vývoj časové řady. Jinými slovy se tyto metody snaží o „klasifikaci“ daných vzorků do známých tříd vzorů a podle výsledku této klasifikace pak určit budoucí vývoj. Právě slovo klasifikace v této problematice hraje významnou roli.

Pro problematiku rozpoznávání (klasifikace) obrazců, písma či podpisů se využívají neuronové sítě. Pro predikci vývoje akciového trhu se využívají neuronové sítě také. Schéma principu využití neuronové sítě je znázorněno na obrázku 4.1. Využívá tzv. časového okna, kdy na vstupu neuronové sítě je časově uspořádaná posloupnost hodnot. Učení sítě probíhá tak, že se okno posunuje od počátku časové osy a síť se na daných vstupních hodnotách učí.

Vhodnost použití neuronových sítí v oblasti technické analýzy demonstruje například [5] nebo [13], je třeba ovšem připomenout, že úspěšnost použití neuronových sítí je závislá na kvalitě učícího procesu, na povaze problému a na faktorech ovlivňující nastavení neuronové sítě.

Další metodou, kterou lze s úspěchem využít při technické analýze je fuzzy logika. Jak je uvedeno v [7], techničtí analytici využívající nástroje jako trendové linie či cenové rámce



Obrázek 4.1: Predikce pomocí neuronové sítě - schéma.

nevědomky využívají metody fuzzy logiky tím, že v určitých pásmech nedefinují jednoznačně dané hodnoty. Především u nástroje jako jsou „Kagi“ linie je toto patrné. Využívanými nástroji z oblasti fuzzy logiky jsou tak například Fuzzy Candlesticks.

U technické analýzy je dále možné využít evolučních algoritmů a to především v souvislosti s různými optimalizačními problémy (nastavení vah neuronů dané neuronové sítě) apod. Z toho vyplývá, že v rámci technické analýzy je vhodné kombinovat více metod z oblasti umělé inteligence, neboť každá z nich má své specifické zaměření.

### 4.3 Fundamentální analýza

V rámci fundamentální analýzy, lze opět využít fuzzy logiku, neboť výstupem této analýzy je množství ukazatelů, které charakterizují daný podnik či odvětví, a investor (obchodník) se rozhoduje podle daných ukazatelů. Opět, pokud jsme schopni sestavit pravidla nutná pro fuzzy systém, můžeme využít fuzzy množin a fuzzy logiky.

Mezi nástroje využívané při technické analýze či fundamentální analýze patří [6] jak specializované nástroje jako například MetaStock, Computrac, Stable, Reuters Money Network, Windows On Wall Street, Money Marker, tak nástroje obecné nástroje či knihovny pro běžné programovací jazyky či jiné nástroje. Například různé moduly pro nástroje jako je Matlab, Statistica, Excel atd.

### 4.4 Psychologická analýza

V rámci psychologické analýzy lze opět využít metod jako jsou neuronové sítě či fuzzy logika, neboť i psychologická analýza a některé její teorie vychází z hodnot představující růst či pokles trhu či změny v objemech obchodů - tedy z časových řad reprezentující tyto hodnoty. Zde máme konkrétně na mysli Kostolanyho burzovní psychologii zmiňující závislost změn kurzů a obchodovaných objemů na tom, která skupina (hráči či spekulanti) drží většinu akcií na daném trhu.

V rámci teorie Spekulativních bublin lze využít přístupy jak z oblasti technické tak fundamentální analýzy a to především pro samotnou povahu (základ) této teorie. Tímto máme na mysli rozdíl mezi kurzovou hodnotou a vnitřní hodnotou akcie. Pro predikci kurzové hodnoty využíváme technickou analýzu a pro predikci vnitřní hodnoty využíváme

analýzu fundamentální. Proto v rámci této teorie můžeme využít především neuronové sítě, fuzzy logiku a evoluční algoritmy.

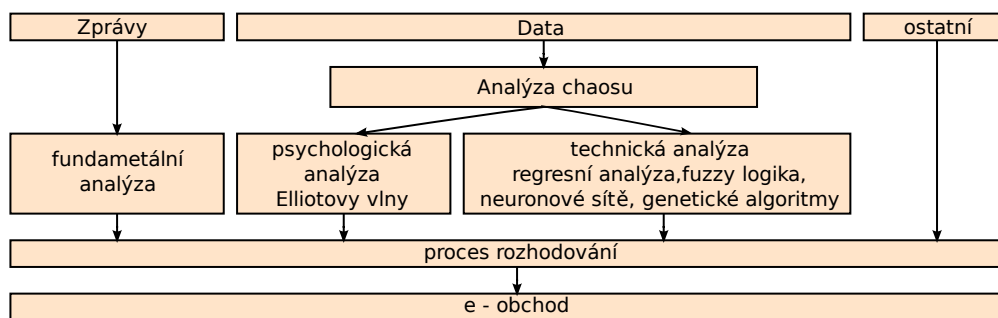
Další teorií je Keynesova spekulativní teorie. Keynes v rámci své spekulativní teorie považuje za jeden z hlavních faktorů ovlivňující chování akciových kurzů nadměrné reakce na události, které mnohdy nejsou dostatečně významné a mají jen pomíjivý charakter. V rámci této problematiky můžeme uvažovat o využití metod z oblasti dolování dat a klasifikace. Dolování dat především proto, neboť tak můžeme hledat právě ty události (reprezentované zprávami), které ač v podstatě nevýznamné měli podstatný vliv na chování burzovního publika. Klasifikaci v tomto případě můžeme využít právě pro rozhodování, zdali daná událost měla či neměla vliv na burzovní publikum. V rámci klasifikace tak lze opět využít například neuronové sítě či jiné klasifikátory (naivní Bayesův klasifikátor).

Práce, které by se zabývaly podobnou problematikou, jako je vliv zpráv na pohyb kurzů již existují, jsou jimi například [2] či [10]. Konkrétně [10] je pro následující text práce klíčový neboť se zabývá podobnou tematikou, které bude věnována praktická část této práce. Tento článek se zabývá možností predikce akciového trhu pomocí finančních zpráv. Konkrétně ukazuje, že predikce je možná v intervalu 20 minut před a po tom, co se daná zpráva stane veřejnou. Další prací, která hodnotí vliv zpráv na pohyb akciového trhu je například [3], která se soustředí především na rok 1987 a v úvahu nebere pouze finanční zprávy, ale i zprávy politické apod.

Především z [10] tedy vyplývá, že použití metod data miningu a klasifikace lze s úspěchem využít právě v oblasti psychologické analýzy. Je ovšem otázkou, za jakých podmínek lze využít predikci pomocí psychologické analýzy. Teorie říkají, že psychologická analýza je analýzou krátkodobou, což [10] potvrzuje. Kde ovšem leží hranice mezi krátkodobou analýzou a dlouhodobou? V této práci využijeme obdobných postupů, jako ve výše uvedených článcích, ale zaměříme se na delší časové úseky.

## 4.5 Aplikace

Jak tedy vidíme, je využití metod z oblasti umělé inteligence v oblasti predikce vývoje akciového trhu (ale i jiných trhů) vhodné. Příklad komplexního využití různých metod je uveden v [4]. Na obrázku 4.2 je schéma daného systému.



Obrázek 4.2: Schéma systému - převzato z [4].

Ze schématu je patrné, že systém je rozdělen jak horizontálně, tak vertikálně. Konkrétně horizontální rovina je členěna podle analýz využívaných v oblasti predikce vývoje akciového trhu, tedy analýza fundamentální, psychologická a technická. Pro každou z těchto metod,

jak bylo uvedeno výše jsou vhodné konkrétní nástroje z oblasti umělé inteligence. V rovině vertikální je pak členění možné pozorovat především u technické analýzy, kdy je nejprve provedena analýza chaotičnosti časové řady, jejíž výsledek je dále využit při technické analýze, a výsledek technické analýzy poté, spolu s dalšími výsledky, vstupuje do rozhodovacího procesu, jež je většinou představován fuzzy logikou [4].

V další kapitole navrhne obdoby systém, jehož části detailněji popíšeme a zvolenou část naimplementujeme a otestujeme.



## Kapitola 5

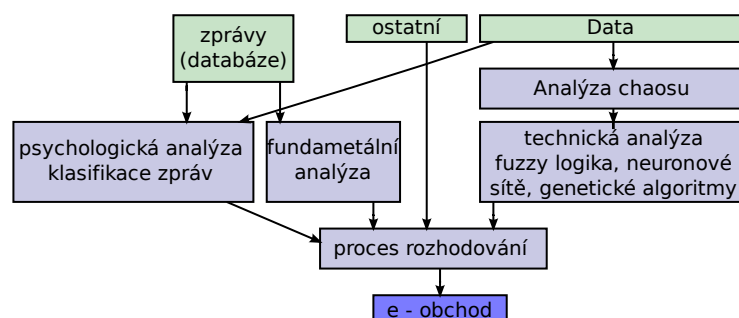
# Vlastní návrhy řešení, přínos návrhů řešení

Tato kapitola se zabývá návrhem vlastního systému pro predikci vývoje akciového trhu pomocí technické a psychologické analýzy, který využívá metody a nástroje z oblasti umělé inteligence či soft computingu. Návrh systému vychází z poznatků předchozích kapitol.

Nejprve tedy navrhne systém, jehož jednotlivé části dále podrobně popíšeme a zdůvodníme použití konkrétních metod z oblasti umělé inteligence. V další části práce se budeme zabývat konkrétním popisem zvolené části systému, její implementací a testováním. V závěru kapitoly vyhodnotíme přínos námi navrženého systému, jeho výhody a nevýhody.

### 5.1 Návrh řešení

Na obrázku 5.1 je znázorněno schéma navrženého systému. Oproti schématu 4.2 vidíme, že hlavní rozdíl je především v části psychologické analýzy, kdy schéma na obrázku 5.1 využívá k psychologické analýze především zprávy, které klasifikuje podle toho, jak jsou schopny ovlivnit pohyb kurzu na daném akciovém trhu.



Obrázek 5.1: Schéma navrženého systému.

Stručný popis systému a jeho jednotlivých částí je následující:

**zprávy (databáze)** – zprávy z různých zdrojů, které slouží pro extrakci informací, zprávy jsou využívány především v rámci psychologické analýzy, ale také v rámci fundamentální analýzy, zprávy jsou v „databázi“ uloženy v původním i v normalizovaném tvaru, nemusí se jednat přímo o databázové řešení, v rámci této práce jsme například využili uložení zpráv v podobně XML souborů, zprávy jsou u různých oblastí jako například finanční, politické apod., právě extrakce vhodných vlastností teprve určuje význam jednotlivých zpráv pro danou analýzu;

**psychologická analýza** – tato část systému predikuje pohyb kurzů na základě zpráv, které klasifikuje podle jejich vlastností, využívá klasifikátor (naivní Bayesův klasifikátor, neuronová síť apod.) který vytváří trénováním na základě historických zpráv, v rámci této části tak lze využít metody klasifikace, neuronových sítí, evoluční algoritmy apod., kritickým faktorem pro úspěšnost klasifikátoru je vhodná extrakce vlastností z textu (zpráv);

**fundamentální analýza** – obdobně jako u návrhu systému v [4] slouží tato část pro predikci vývoje z pohledu fundamentální analýzy, která je v tomto případě založena na finančních zprávách, které obsahují informace o finančních ukazatelích daných firem či oborů;

**data** – představuje, obdobně jako zprávy, část systému, která slouží pro správu historických dat, jež představují historické pohyby kurzů na daném akciovém trhu, kromě změn kurzů zachycuje také další hodnoty, jako pohyby objemů obchodů atd., umožňuje také různé vzorkování dat podle požadavků ostatních částí systému;

**analýza chaosu** – tato část systému slouží stejně jako u [4] k identifikace míry chaotičnosti časové řady, pokud řada nevykazuje příliš velkou míru chaotičnosti, je možné pokračovat v predikci pomocí technické analýzy, v jiném případě nemá význam pokračovat, neboť řada obsahuje významně stochastickou složku, která znesnadňuje predikci;

**technická analýza** – tato část představuje stěžejní část systému, využívá především různých typů neuronových sítí pomocí nichž predikuje budoucí vývoj akciového trhu, kromě neuronových sítí využívá také evoluční algoritmy, především pro optimalizaci nastavení vah v rámci neuronové sítě;

**ostatní** – představuje různé parametry, které jsou vloženy na základě zkušeností uživatele a slouží pro optimalizaci celkové predikce systému;

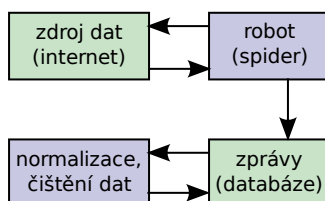
**proces rozhodování** – na základě výstupů jednotlivých modulů je v rámci získání „globálního“ výsledku využito fuzzy logiky, jednotlivé části mohou mít různé výstupy a v rámci této práce tak vzniká jedno konkrétní doporučení pro uživatele, kromě fuzzy logiky využívá také evolučních algoritmů pro hledání funkcí příslušnosti jednotlivých fuzzy množin;

**e-obchod** – poslední část systému, která na základě predikovaných hodnot generuje nákupní či prodejní signály.

Tak jako je trh dynamický systém, tak i námi navržený systém, jehož cílem je predikce vývoje akciového trhu, je také dynamický. Tato „dynamičnost“ je představována především využitím evolučních algoritmů, kdy jednotlivé části systému se „neustále“ vyvíjí, což je zobrazeno u popisu jednotlivých částí.

### 5.1.1 Modul pro správu zpráv

Tato část systému slouží pro ostatní navazující části. Umožňuje přístup k datům, uchovává data, aktualizuje databázi a provádí nad daty operace jako je normalizace, čištění dat, odstranění nevhodných dat atd. Schéma této části je znázorněno na obrázku 5.2.

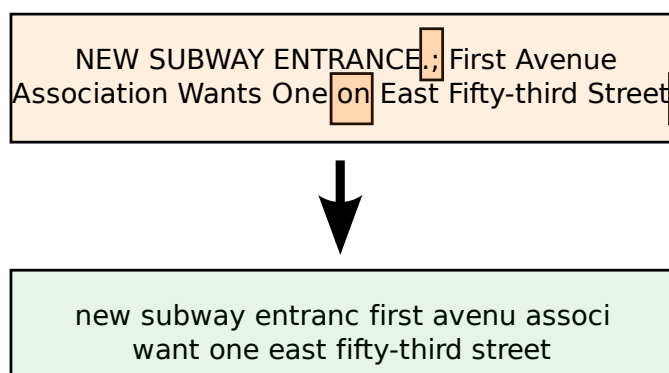


Obrázek 5.2: Schéma části pro správu databáze zpráv.

Hlavními akcemi jsou tedy aktualizace databáze pomocí robota (spidera) a úprava dat (normalizace atd.), princip je následující:

**aktualizace** – je prováděna pomocí robota (tzv. spider), který prochází dané zdroje na internetu a „stahuje“ nové zprávy do databáze, zdroje sami mohou být aktualizovány (přidávány či odstraňovány), jednou z forem zdrojů můžou být i RSS kanály apod., cílem je aktuální databáze, důraz je tedy kladen i na rychlost aktualizace;

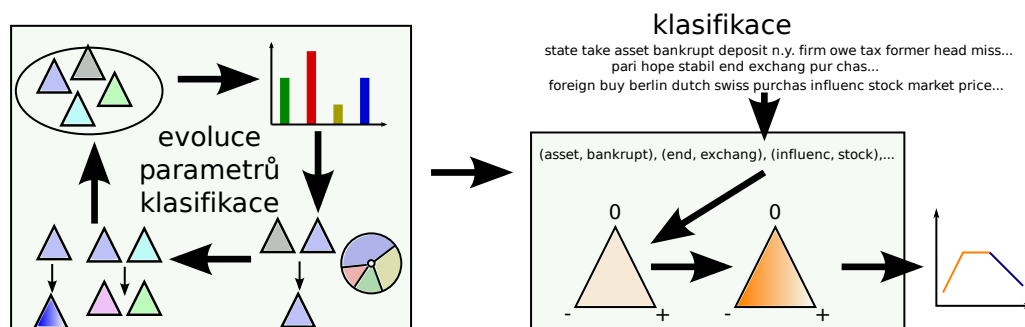
**úprava dat** – je také jednou za hlavních činností (metod) tohoto modulu, především jsou odstraněna nerelevantní data (například zprávy, které neobsahují v těle žádný obsah), texty jsou normalizovány (převedeny na malá písmena, odstraněna slova, která nemají vliv na význam - spojky, předložky, slova jsou převedena do základního tvaru), ukázka úpravy dat je na obrázku 5.3.



Obrázek 5.3: Úprava dat.

### 5.1.2 Modul pro psychologickou analýzu

Tento modul slouží pro klasifikaci zpráv do tříd, které prezentují vliv zpráv na pohyb kurzů na akciovém trhu. Ke klasifikaci je možné využívat několika nástrojů či metod. Jednou z nich jsou neuronové sítě nebo různé klasifikátory založené na pravděpodobnosti. V této práci uvažujeme především tzv. naivní Bayesův klasifikátor. Schéma činnosti modulu je znázorněno na obrázku 5.4.



Obrázek 5.4: Schéma části pro psychologickou analýzu.

Na obrázku 5.4 jsou znázorněny dvě části. Samotná klasifikace v pravé části obrázku představuje samotnou funkčnost klasifikátoru. Nejprve jsou z daných zpráv extrahovány vlastnosti nutné pro klasifikaci a poté jsou tyto zprávy (reprezentované extrahovanými vlastnostmi) klasifikovány pomocí klasifikátoru. V levé části obrázku je znázorněno využití evolučních algoritmů při optimalizaci klasifikátoru. Klasifikátor je v první řadě charakterizován různými parametry, jejichž vhodné nastavení je právě cílem využití evolučních algoritmů, v rámci evoluce se také vyvíjí množina konkrétních vlastností, které jsou extrahovány z textu a jsou poté využity při samotné klasifikaci. Díky tomuto principu je tento podsystém dynamický - reaguje tedy na změny ve skladbě zpráv a přizpůsobuje se tak konkrétním podmínkám.

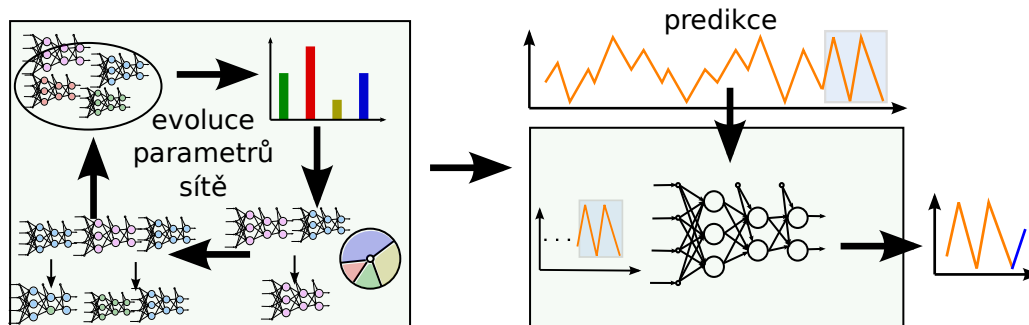
### 5.1.3 Modul pro fundamentální analýzu

Účelem tohoto modulu je extrakce informací z finančních zpráv. Jedná se především o zprávy zachycující finanční situaci konkrétních podniků působících v daném oboru. Podobně jako u modulu pro psychologickou analýzu má i tento dvě části. Jedna část představuje samotnou funkčnost modulu, kdy jsou z daných zpráv extrahována data a jsou odhadovány finanční ukazatele. Druhá část slouží pro evoluci různých parametrů, které slouží pro vhodné nastavení jak extrakce vlastností z textu tak samotných výpočtů.

### 5.1.4 Modul pro technickou analýzu

Tento modul, který představuje jednu ze stěžejních částí systému slouží pro predikci vývoje časové řady pomocí neuronových sítí. Konkrétně na základě historických dat pohybů kurzů či objemu prodeje jsou predikována data budoucí. V rámci nastavení (či výstupu) je možné data různým způsobem vzorkovat a podle toho se zaměřit na různá časová období. Schéma činnosti modulu je znázorněno na obrázku 5.5.

Tento modul se opět skládá ze dvou částí. V jedné se díky použití evolučních algoritmu v kombinaci s metodou učení „zpětného šíření chyby“ vyvíjí lepší neuronová síť. Druhá část slouží k samotné predikci pomocí časového okna.

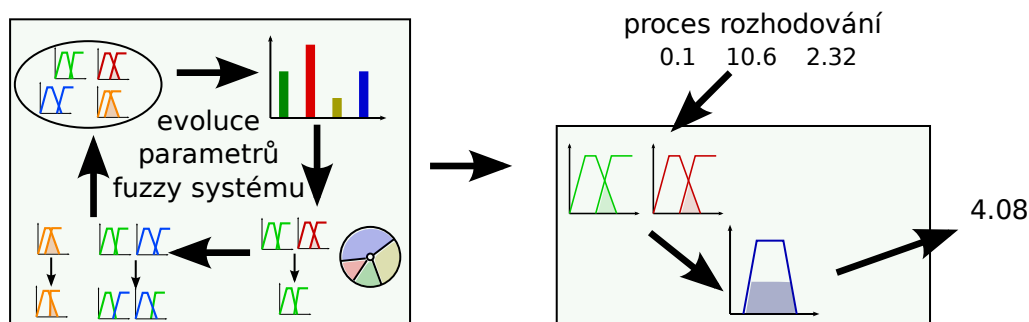


Obrázek 5.5: Schéma části pro technickou analýzu.

Princip evoluce v rámci neuronových sítí spočívá v kombinaci náhodného nastavení vah u jednotlivých neuronů a s následným „doučením“ pomocí klasické metody učení. V rámci evoluce se také vyvíjí různé parametry jako je například velikost časového okna a jeho přesná podoba, počet časových oken apod. Nejsilnější jedinec z populace pak ve vhodném časovém okamžiku nahrazuje instanci neuronové sítě využívanou při predikci.

### 5.1.5 Modul pro rozhodovací část systému

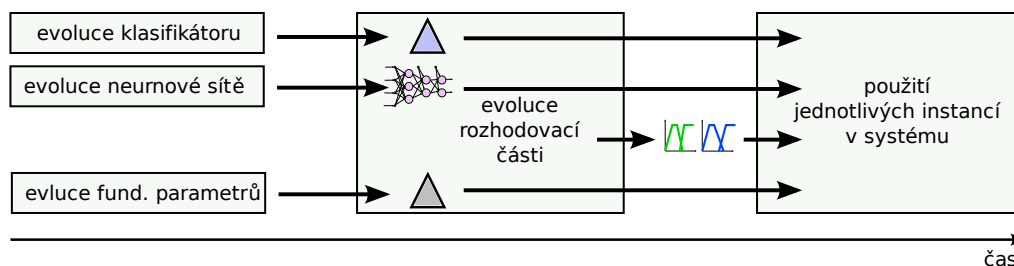
Tento modul slouží pro získání „globálního“ výsledku na základě výsledků jiných modulů systému. Je založen na fuzzy logice a opět se skládá ze dvou částí. První část slouží pro samotný proces rozhodování v rámci systému, kdy na základě vstupních dat a procesu fuzzyfikace jsou získány míry příslušnosti do jednotlivých fuzzy množin. V rámci aplikace pravidel je získána výsledná fuzzy množina spolu s mírou příslušnosti, ze které je procesem defuzzyfikace získána konkrétní hodnota. Druhá část slouží pro evoluci fuzzy systému - jsou hledány optimální funkce příslušnosti za základě „nejčerstvějších“ dat. Opět nejsilnější jedinec z populace fuzzy systémů ve vhodném okamžiku nahrazuje instanci právě využívaného fuzzy systému.



Obrázek 5.6: Schéma části pro rozhodování.

Na obrázku je zachyceno schéma rozhodovací části systému spolu s evolucí vhodných funkcí příslušnosti a dalších parametrů.

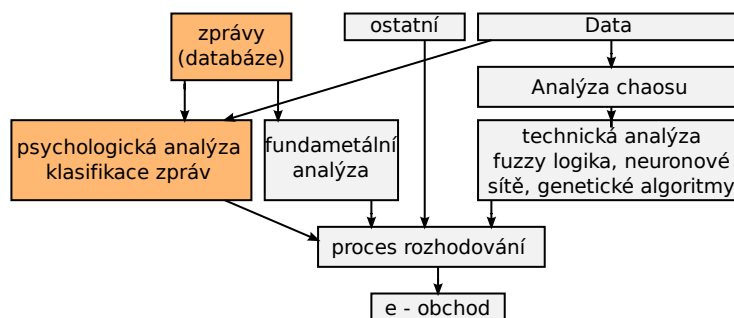
Na „globální“ úrovni probíhá evoluce následovně. V rámci nižších vrstev systému, jako je například modul psychologické, fundamentální a technické analýzy probíhá evoluce, jejímž výsledkem jsou nejsilnější jedinci, kteří jsou adeпти na nahrazení stávajících instancí, přesto k nahrazení nedochází okamžitě. Vyšší úrovně systému se musí těmto novým instancím přizpůsobit, proto je nejprve provedena evoluce na vyšší úrovni, při níž jsou využívána data a výstupy z budoucích instancí. Teprve poté, co jsou na všech úrovních optimalizovány parametry jednotlivých instancí a jsou vytvořeni díky evoluci „lepší“ jedinci, jsou v rámci systému nahrazeny všechny stávající instance novými. Toto je znázorněno na obrázku 5.7.



Obrázek 5.7: Schéma hierarchie evolucí.

### 5.1.6 Návrh vybrané části systému

Vybranou částí, která je implementována v rámci práce je část pro správu zpráv a modul pro psychologickou analýzu. Části jsou zvýrazněny na obrázku 5.8.



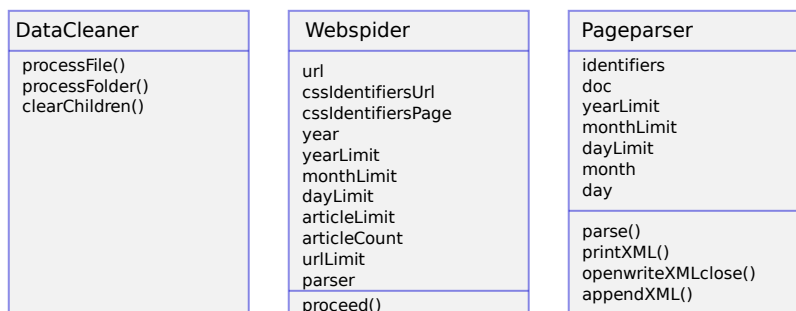
Obrázek 5.8: Znázornění vybrané části systému pro implementaci a testování.

Konkrétně se v rámci jednotlivých částí jedná o:

**zprávy** – je realizována část pro získávání zpráv z internetu (třída *WebSpider*, pomocí struktury stránky je tato procházena a jsou ukládány vhodné zprávy do databáze, v rámci tohoto modulu je také realizována část pro úpravu dat jako je normalizace atd. (třída *DataCleaner*, realizace je zaměřena pouze na části související s modelem představující psychologickou analýzu, tj. není realizována část pro správu zpráv určených pro fundamentální analýzu;

**psychologická analýza** – je realizován modul pro klasifikaci zpráv do tříd podle pohybu kurzu daného titulu (indexu), využívá naivního Bayesovského klasifikátoru, jedná se o třídu `NBClassifier`, která obsahuje metody pro trénování (učení) klasifikátoru, pro samotnou klasifikaci a další.

Schémata tříd `DataCleaner`, `Webspider` a `Pageparser` jsou znázorněna na obrázku 5.9.



Obrázek 5.9: Třída `DataCleaner`, `Webspider` a `Pageparser`.

V rámci třídy `DataCleaner` jsou významné především následující metody:

`processFile()` – tato metoda vytvoří z daného souboru XML strom a na každý textový uzel stromu zavolá metodu `clearChildren()`;

`processFolder()` – metoda „projde“ obsah dané složky na každý soubor ve složce zavolá metodu `processFile()`;

`clearChildren()` – metoda znormalizuje textový obsah daného XML elementu, nejprve převede všechna písmena na mála písmena, poté převede slova do základního tvaru a odstraní tzv. „stop“ slova a interpunkci, na závěr uloží XML strom do nového souboru.

Třída `Webspider` je využívána pro stahování nových zpráv z internetu, konkrétně obsahuje jednu metodu:

`proceed()` – tato metoda na základě předaných parametrů, jako je URL či hloubka zanoření, prochází daný „web“ a pokud se dostane na nejnižší úroveň (stránka představuje danou zprávu) zavolá objekt pro extrakci informací ze stránky a jejich uložení do souboru XML,

kromě toho tato třída obsahuje řadu proměnných, které určují adresu webu, různá limitní omezení jak pro datum či adresu dané stránky.

Třída `Pageparser` představuje nástroj pro extrakci informací ze stránky je a jejich uložení do XML souboru. Především se jedná o metody:

`parse()` – metoda na základě kódu internetové stránky a CSS selektorů vytvoří XML strom reprezentující obsah vybraných částí stránky;

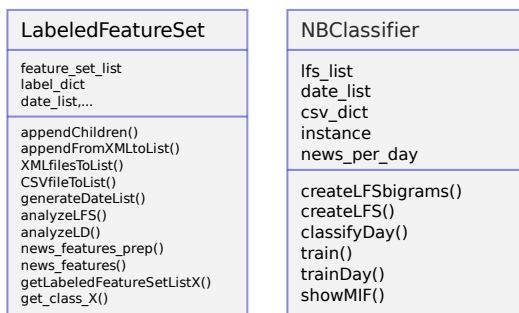
`printXML()` – metoda vytiskne na standardní výstup reprezentaci XML stromu;

`openwriteXMLclose()` – metoda otevře daný soubor, uloží do něj reprezentaci XML stromu a soubor uzavře;

`appendXML()` – metoda volá metodu `parse()` pro vytvoření XML reprezentace dané stránky, která je poté připojena k danému souboru,

kromě výše uvedených metod třída obsahuje proměnné pro uložení data daných zpráv či jejich omezení.

Funkčnost modulu psychologické analýzy je reprezentována třídami `NBClassifier` a `LabeledFeatureSet`, které jsou znázorněny na obrázku 5.10. Připomeňme, že tyto třídy představují funkčnost naivního Bayesovského klasifikátoru. V rámci vybrané části systému není řešena problematika evoluce klasifikátoru, jak bylo zmíněno výše.



Obrázek 5.10: Třída `NBClassifier` a `LabeledFeatureSet`.

Třída `LabeledFeatureSet` slouží pro přípravu dat pro samotnou instanci klasifikátoru. Popis metod přiblíží funkčnost této třídy:

`XMLfilesToList()` – metoda slouží pro převedení daných XML souborů do proměnné typu seznam (`list`), volá při tom metodu `appendFromXMLtoList()`, proměnnými seznamu jsou slovníky (heterogenní datový typ);

`appendFromXMLtoList()` – vytvoří XML strom z daného souboru a na každý uzel stromu zavolá metodu `appendChildren()`;

`appendChildren()` – daný uzel stromu je převeden na položku předaného slovníku;

`CSVfileToList()` – převede daný CSV soubor obsahující data o historických pohybech kurzů daného indexu na proměnnou datového typu výčet (slovník);

`generateDateList()` – vygeneruje seznam dat (datum) v daném časovém intervalu;

`analyzeLFS()`, `analyzeLD()` – metody slouží pro získání statistik nad danými seznamy;

`news_features_prep()`, `news_features()` – tato skupina metod slouží pro extrakci vlastností z textu, metod je více právě podle daného způsobu reprezentace vlastností;

`getLabeledFeatureSetListX()` – metoda vrací seznam proměnných reprezentující jak extrahované vlastnosti z textů tak jim přiřazené třídy;

`get_class_X()` – metody na základě parametrů vrací třídy, které jsou poté přiřazeny daným zprávám,



kromě výše uvedených zpráv tato třída obsahuje proměnné pro reprezentaci výše uvedených seznamů či slovníků, které slouží pro uložení extrahovaných vlastností ze zpráv a jim přiřazených tříd.

Třída `NBClassifier` je stěžejní třídou navrženého systému, využívá objektu výše uvedené třídy `LabeledFeatureSet`. Třída implementuje následující metody:

`createLFSbigrams()` – metoda vytvoří seznamy hodnot reprezentující zprávy, pohyby kurz atd., v rámci extrakce vlastností ze zpráv jsou využívány metody, jejichž výsledky jsou reprezentovány bigramy; metoda slouží pro „inicializaci“ většiny potřebných proměnných před samotným trénováním klasifikátoru;

`createLFS()` – obdobně jako předchozí metoda, tato také slouží pro inicializaci většiny potřebných dat (reprezentovaných seznamy), které poté slouží pro trénování a samotnou klasifikaci;

`classifyDay()` – metoda slouží pro klasifikaci zpráv z daného dne, na základě klasifikace jednotlivých zpráv vrací globální výsledek za celý den;

`train()` – metoda slouží pro trénování klasifikátoru na základě vytvořených seznamů dat (trénovací a testovací);

`trainDay()` – metoda slouží pro trénování klasifikátoru, je kladen důraz na trénování podle dnů.

### 5.1.7 Implementace vybrané části systému

Implementačním jazykem je jazyk Python [19]. Jedná se o skriptovací, dynamický, objektový, silně typovaný jazyk. Předností toho programovacího jazyka jsou vysokoúrovňové datové typy jako seznam (list), slovník (heterogenní datový typ). V rámci implementace byly využity vlastnosti jazyka umožňující objektově orientovaný přístup a vysokoúrovňové datové typy a operace s nimi spojené.

Zdrojové soubory byly rozděleny do modulů, kde každá třída byla umístěna v samostatném modulu. Konkrétně se jedná o:

`classifier.py` – tento modul představuje implementaci třídy `NBClassifier` v jazyce Python;

`labeledFeatureSet.py` – tento modul je implementací třídy `LabeledFeatureSet`;

`dataCleaner.py` – modul implementující třídu `DataCleaner`, v jazyce Python;

`webspider.py` – modul implementující třídu `Webspider`;

`pageparser.py` – modul implementující třídu `Pageparser`.

Pro samotné testování slouží skript `test.py`, v rámci kterého lze nastavit parametry pro jednotlivé objekty výše uvedených tříd. Při implementaci byl využit modul `nltk`, který umožňuje zpracování řetězců a textů právě pro „analýzu“ přirozeného jazyka v jazyce Python [1].

Zdrojové soubory jsou uloženy na příloženém CD. Implementace spolu s testováním byly provedeny na operačním systému Ubuntu Linux (jádro 2.6.31-21).

### 5.1.8 Testování

Tato podkapitola se zabývá problematikou testování implementovaného řešení. V rámci testování byly měněny různé parametry a byla měřena úspěšnost predikce zvoleného modelu. Je třeba připomenout, že právě optimální nastavení parametrů je otázkou různých optimalizačních technik, mezi které patří zmíněné evoluční algoritmy. Tato část systému, zabývající se optimalizací parametrů klasifikátoru ovšem implementována nebyla a proto dále zmíněné výsledky nelze považovat za „optimální“.

V rámci testování jsme se zaměřili na následující parametry:

**počet zpráv za den** – je zřejmé, že predikovaná hodnota, resp. její kvalita, je závislá na počtu zpráv za daný den, těžko lze například predikovat pohyb kurzu, pokud jej predikujeme na základě jedné zprávy;

**extrahované vlastnosti z textu** – nejsložitější stránka testů, extrahované vlastnosti jsou základním a nejdůležitějším parametrem při klasifikaci textů, obtížnost extrakce vhodných vlastností lze považovat za problematiku přesahující úroveň diplomové práce a proto jsme se v rámci této práce zaměřili na „jednodušší řešení“ které je představováno:

- intuitivním definováním základních ekonomických či finančních slov, která budeme v rámci klasifikovaného textu vyhledávat,
- intuitivním definováním citově zabarvených slov - ať už pozitivních či negativních,
- omezování těchto množin za účelem získání informací o vlivu na predikci,

ukázka těchto množin je v příloze A,

**velikost dat** – zde máme na mysli především časové omezení testovací a trénovací množiny (napří. interval 2-5 let);

**další parametry** – například počet klasifikovaných tříd atd., těmto se ovšem vzhledem k výsledkům nebylo vhodné více věnovat.

Testování bylo provedeno na operačním systému Ubuntu Linux (jádro 2.6.31-21). V rámci testování jsme měřili úspěšnost predikce na množině testovacích dat. Testovací data byla získána z následujících zdrojů:

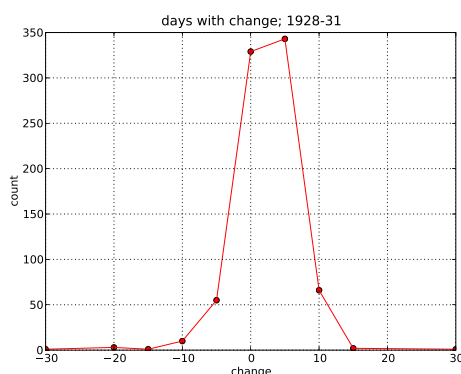
- <http://spiderbites.nytimes.com/> - Main Index - Site Map - NYTimes.com - zprávy pro klasifikaci,
- <http://finance.yahoo.com> - Yahoo! Finance - Business Finance, Stock Market, Quotes, News - soubor s historickými hodnotami pohybu indexu Dow Jones,

jelikož data uvedená na těchto adresách podléhají smluvním podmínkám, je potřeba si je pro testování skriptů stáhnout „individuálně“. U zpráv postačí využít robota pro stažení zpráv s vhodnou adresou.

Mezi další parametry patří především „limit kurzu“, tím máme na mysli minimální změnu kurzu za den (rozdíl mezi otevírací a uzavírací hodnotou). Jak vidíme na obrázku 5.11, je drtivá většina denních změn blízka nule.

Následuje několik grafů, které popisují testování změn různých parametrů.

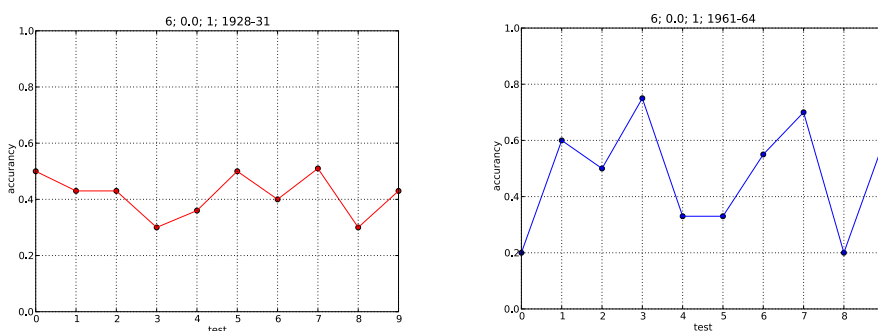
Tento jev, kdy většina hodnot je blízka nule není vhodný pro trénování klasifikátoru, neboť příliš mnoho dat je obsaženo blízko nuly a tak nejsou vhodné pro trénování. Proto



Obrázek 5.11: Počet dnů s danou denní změnou kurzu, 1928-1929.

byl zaveden parametr „limit kurzu“ který určuje minimální absolutní hodnotu změny kurzu za den, aby tato byla využita při klasifikaci.

Obrázek 5.12 znázorňuje výsledky testů při „limit kurzu“ rovnu nule. Tj. v rámci trénování jsou využity všechny hodnoty - i ty blízké nule.

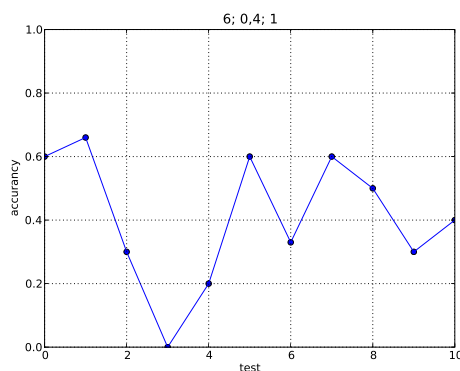


Obrázek 5.12: Test - minimálně 6 zpráv za den, limit kurzu při učení 0.0, minimální počet bigramů 1, 1928-1931 (vlevo) 1961-64 (vpravo).

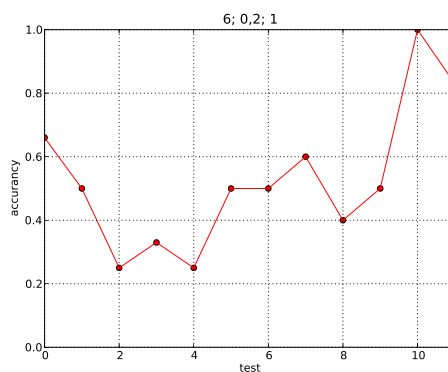
Obrázek 5.13 znázorňuje testování při následujících parametrech - minimální počet zpráv za den: 6, limit změny kurzu při učení: 0.4, minimální počet bigramů: 1, v rozmezí let 1928-1929. Bylo provedeno pouze 10 testů, neboť již z výsledků těchto testů je patrné, že parametry nejsou nastaveny správně, neboť úspěšnost nedosahuje průměrně ani 50%.

Obrázek 5.14 znázorňuje testování při následujících parametrech - minimální počet zpráv za den: 6, limit změny kurzu při učení: 0.2, minimální počet bigramů: 1, v rozmezí let 1928-1929. Opět vidíme, že úspěšnost systému s těmito parametry je nízká a výstup je prakticky náhodný (hodnoty kolísají od 0% po 100%). Více testů s těmito parametry nebylo prováděno právě kvůli nízké úspěšnosti.

V dalším testu jsme rozšířili časový interval na roky 1928-1931. Jak vidíme na obrázků 5.15 a 5.16 ani zde jsme nedosáhli průměrně vyšší úspěšnosti jak 50%.



Obrázek 5.13: Test - minimálně 6 zpráv za den, limit kurzu při učení 0.4, minimální počet bigramů 1, 1928-1929.

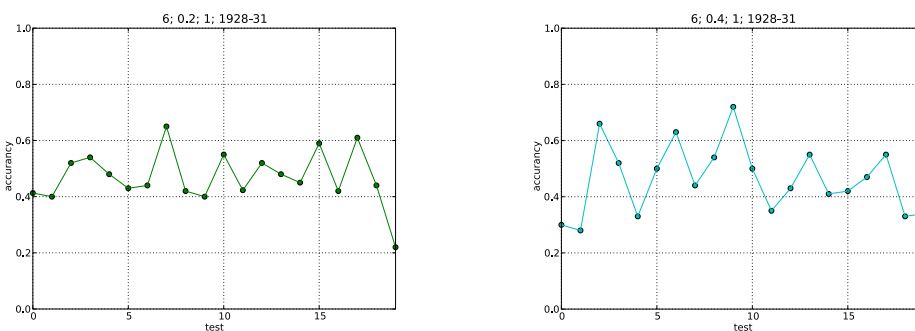


Obrázek 5.14: Test - minimálně 6 zpráv za den, limit kurzu při učení 0.2, minimální počet bigramů 1, 1928-1929.

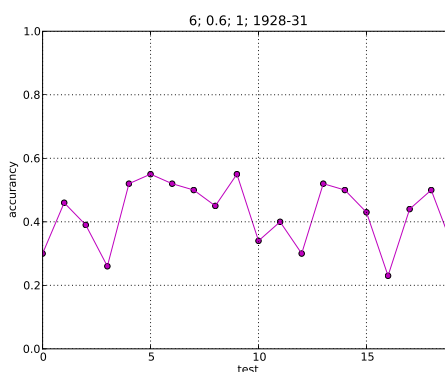
### 5.1.9 Zhodnocení výsledků testů

Jak vidíme z výše uvedených grafů modul pro klasifikaci zpráv má nízkou úspěšnost. To je zapříčiněno následujícími důvody:

- v rámci testování byly měněny parametry klasifikátoru, přesto jsme nenalezli takové nastavení, při kterém by úspěšnost dosahovala alespoň průměrných 60%, je otázkou, zdali by použití pokročilých optimalizačních technik nepřineslo lepší výsledky,
- při testování jsme různě měnili množiny extrahovaných vlastností z textů, přesto jsme nenalezli optimální složení těchto množin, pokud byly množiny příliš malé, potom zpráv pro klasifikaci bylo také málo, což ovlivnilo i výsledky testování, pokud naopak množiny byly příliš velké zapříčinilo to výskyt nerelevantních zpráv, které tak ovlivnili trénování klasifikátoru,
- samotné použití naivního Bayesovského klasifikátoru, jako jednoho z nejjednodušších (ale přitom do výkonu srovnatelných s neuronovými sítěmi), nebyla patrně správná volba, čemuž odpovídají i výsledky provedených experimentů,



Obrázek 5.15: Test - minimálně 6 zpráv za den, limit kurzu při učení 0.2 (vlevo) 0.4 (vpravo), minimální počet bigramů 1, 1928-1931.



Obrázek 5.16: Test - minimálně 6 zpráv za den, limit kurzu při učení 0.6, minimální počet bigramů 1, 1928-1931.

- úzké rozpětí testovaných dat - testy byly provedeny na datech z let 1928-1931 a 1960-1964, je otázkou, zdali by v jiných časových intervalech nebylo dosaženo lepších výsledků, zde ovšem musíme uvést, že testování je náročné na čas a proto byly použity „omezené zdroje“,
- vzorkování historický dat - na základě [10] a také vzhledem k našim dosaženým výsledkům můžeme říci, že vzorkování historických dat nebylo vhodné, je otázkou, jakých výsledků by bylo dosaženo pokud bychom pohyby kurzů vzorkovaly „intradenně“ - tedy s kratší časový interval.

Přestože jsme tedy nedosáhli uspokojivých výsledků, lze říci, že jsme ukázali, že použití naivního Bayesovského klasifikátoru v problematice klasifikace zpráv za účelem predikce vývoje akciového trhu není vhodné. Dalším významným faktorem bylo již zmíněné vzorkování historických dat. Tyto znalosti, kterých jsme dosáhli díky provedeným testům tak mohou posloužit pro budoucí vývoj v této oblasti - tedy využití jiných metod a jiných nastavení vzorkování historických dat.

### 5.1.10 Přínos návrhů řešení

Hlavním přínosem výše uvedeného systému je jeho schopnost adaptace na měnící se podmínky akciového trhu. Příkladem takto měnících se podmínek může například v rámci klasifikace textu být změna používaných pojmů v oblasti finanční problematiky. Například v 30. letech minulého století se ve finančních zprávách vyskytovaly jiná sousloví nežli dnes. Kromě toho se vyvíjí nové technologie (a tím pádem i jejich pojmenování) a klasifikační systémy musí být aktualizovány na tyto nové názvy těchto technologií, pokud tedy mají spojitost s predikcí.

Je diskutabilní použití evoluce v tolika částech systému. Jak víme, evoluce, především ohodnocení všech jedinců v populaci je náročné na strojový čas. Přesto, vzhledem k výsledkům v současné době [20], lze říci, že rozvoj informačních technologií, především možnost paralelizace výpočtů přináší nové možnosti právě v evolučních algoritmech a tedy i v dalších oblastech jako je například predikce vývoje akciového trhu.

Velkým přínosem v rámci implementace a testování zvolené části bylo zjištění vhodnosti využití použitých metod, především naivního Bayesovského klasifikátoru, který na základě výsledků není pro tuto problematiku vhodný. Dále bylo zjištěno, že použité vzorkování historických dat není pro analýzu vlivu zpráv na pohyby kurzů vhodné, neboť metody dosahovali nízkých úrovní úspěšnosti.

## Kapitola 6

### Závěr

S rostoucím využitím informačních technologií a s vývojem nových metod v oblasti umělé inteligence se naskytá možnost využití těchto prostředků v oblasti predikce vývoje akciového trhu. V této práci jsme se zaměřili právě na využití těchto nástrojů a technik v této oblasti a to především v rámci psychologické analýzy.

Nejprve jsme přiblížily technickou a psychologickou analýzu, které jsou využívány při predikci vývoje akciového trhu, především jsme uvedli známé teorie jako je „Teorie Elliotových vln“, „teorii Spekulativních bublin“ apod. V rámci metod využívaných v oblasti umělé inteligence jsme uvedli neuronové sítě, evoluční algoritmy a metody z oblasti klasifikace.

V hlavní části této práce jsme navrhli systém využívající metody z oblasti umělé inteligence a to tak, aby systém predikoval co nejpřesnější výsledky. Proto jsme v rámci jednotlivých částí systému hojně využili evolučních algoritmů, abychom tak navržený systém přizpůsobili neustále se měnícím podmínkám v oblasti akciového trhu.

Pro implementaci jsme vybrali část systému, která je zaměřena na klasifikaci zpráv podle jejich vlivu na vývoj akciového trhu. Konkrétně jsme zprávy klasifikovali do dvou skupin podle toho, zdali kurz roste či klesá, možnost stagnace jsme zahrnuli do jedné z výše uvedených skupin). V rámci tohoto modulu jsme využili naivní Bayesův klasifikátor a znalosti z oblasti klasifikace textů. Při implementaci jsme využili modul `nltk`, představující nástroje pro analýzu textů v jazyce „Python“.

Výsledky dosažené v této práci napovídají, že využití naivního Bayesova klasifikátoru v této oblasti není vhodné. Je otázkou, jakých výsledků by bylo dosaženo, pokud by byl systém implementován jako celek a byly tak využity optimalizační techniky pro parametry tohoto klasifikátoru.

V rámci budoucího rozvoje lze implementovat jiné části systému, jako je například modul technické analýzy či fundamentální analýzy. V rámci psychologické analýzy by bylo vhodné otestovat funkčnost s použitím jiného klasifikátoru či jiného vzorkování časové řady.

Přínosem pro autora v rámci této práce bylo detailní seznámení se s metodami analýz v oblasti predikce vývoje akciového trhu a jejich spojení s metodami z oblasti umělé inteligence a využít tak znalosti jak z oblasti finančních trhů, tak z oblasti informačních technologií.

# Literatura

- [1] Bird, S.; Klein, E.; Loper, E.: *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit*. O'Reilly, 2009, ISBN 978-0-596-51649-9, [Online; navštíveno 17.5.2010].  
URL <http://www.nltk.org/book>
- [2] Chan, W. S.: Stock Price Reaction to News and No-News: Drift and Reversal After Headlines. In *MIT Sloan School of Management, Working Paper*, 2002, str. 223260.
- [3] Cutler, D. M.; Poterba, J. M.; Summers, L. H.: What Moves Stock Prices? NBER Working Papers 2538, National Bureau of Economic Research, Inc, Červenec 1989, [Online; navštíveno 17.5.2010].  
URL <http://ideas.repec.org/p/nbr/nberwo/2538.html>
- [4] Dostál, P.: *Pokročilé metody analýz a modelování v podnikatelství a veřejné správě*. CERM, 2009, ISBN 9788072046058, 342 s.
- [5] Dunis, C. L.: *Forecasting Financial Markets: Exchange Rates, Interest Rates and Asset Management*. Wiley, 1996, ISBN 0471966533, 324 s.
- [6] Fanta, J.: *Počítačové technologie na kapitálových trzích*. Computer Press, 1998, ISBN 80-7226-073-1, 240 s.
- [7] Fanta, J.: *Technologie umělé inteligence na kapitálových trzích*. Karolinum, 1999, ISBN 80-7184-866-2, 89 s.
- [8] Fanta, J.: *Psychologie, algoritmy a umělá inteligence na kapitálových trzích*. Grada, 2006, ISBN 80-247-0024-7, 168 s.
- [9] Frost, A. J.; Prechter, R. R.: *Elliott wave principle: key to market behavior*. Elliott Wave International Inc., 1998, ISBN 0932750435, 244 s.
- [10] Gidófalvi, G.: Using News Articles to Predict Stock Price Movements. Technická zpráva, Department of Computer Science and Engineering, University of California, San Diego, 2001, [Online; navštíveno 17.5.2010].  
URL <http://130.203.133.150/viewdoc/summary?doi=10.1.1.161.887>
- [11] Gregorová, Z.: Spekulativní bubliny na finančních trzích. 2009, [Online; navštíveno 8.5.2010].  
URL [http://is.muni.cz/th/206901/esf\\_b/](http://is.muni.cz/th/206901/esf_b/)
- [12] Hu, Y. H.; Hwang, J.-N.: *Handbook of neural network signal processing*. CRC Press, 2002, ISBN 0849323592, 408 s.



- [13] Hutchinson, J. M.; Lo, A. W.; Poggio, T.: A Nonparametric Approach to Pricing and Hedging Derivative Securities Via Learning Networks. NBER Working Papers 4718, National Bureau of Economic Research, Inc, Únor 1995, [Online; navštíveno 17.5.2010].  
URL <http://ideas.repec.org/p/nbr/nberwo/4718.html>
- [14] Kamber, J. H. . M.: *Data mining: concepts and techniques*. Morgan Kaufmann, 2006, ISBN 1558609016, 770 s.
- [15] Markov, Z.; Larose, D. T.: *Data Mining the Web*. Wiley, 2006, ISBN 9780471666554.
- [16] Murphy, J. J.: *Technical Analysis of the Financial Markets: A Comprehensive Guide to Trading Methods and Applications*. New York Institute of Finance, 1999, ISBN 978-0735200661, 576 s.
- [17] Pesavento, L.; Joufflas, L.: *Trade What You See: How to Profit from Pattern Recognition*. John Wiley and Sons, 2007, ISBN 9780470106761, 201 s.
- [18] Petr, L.: Predikce pomocí neuronové sítě. 2005, [Online; navštíveno 8.5.2010].  
URL <http://www.fit.vutbr.cz/study/DP/DP.php?id=3297&y=2004&st=Lomovciv>
- [19] Pilgrim, M.: *Dive Into Python*. APress, 2004, ISBN 1590593561.
- [20] Pospíchal, P.: Akcelerace genetického algoritmu s využitím GPU. 2009, [Online; navštíveno 18.5.2010].  
URL <http://www.fit.vutbr.cz/study/DP/rpfile.php?id=6540>
- [21] Rejnuš, O.: *Finanční trhy*. KEY Publishing, 2008, ISBN 978-80-87071-87-8, 559 s.
- [22] Smith, A. E. E. . J. E.: *Introduction to evolutionary computing*. Springer, 2003, ISBN 3540401849, 299 s.
- [23] Sojka, Z.; Dostál, P.: *Eliotovy vlny*. Librix.eu, 2008, ISBN 9788073996307, 272 s.
- [24] Zimmermann, H.-J.: *Fuzzy set theory—and its applications*. Springer, 2001, ISBN 0792374355, 514 s.
- [25] Zweig, J.: *Inteligentní investor*. Grada Publishing a.s., 2007, ISBN 8024717921, 9788024717920, 503 s.

# Seznam použitých zkratek a symbolů

**CD** – Compact Disc

**CSS** – Cascading Style Sheets

**CSV** – Comma-separated values

**LSE** – London Stock Exchange

**NASDAQ** – National Association of Securities Dealers Automated Quotations

**NYSE** – New York Stock Exchange

**RSS** – Really Simple Syndication

**S&P 500** – Standard & Poor's 500

**TSE** – Tokyo Stock Exchange

**XML** – Extensible Markup Language

# Seznam příloh

- A Seznamy hledaných slov
- B Ukázkový výpis testovacího skriptu
- C Datový nosič CD s kompletní implementací a zdrojovými texty práce

## Příloha A

### Seznamy hledaných slov

```
word_features = ['consum', 'dollar', 'risk', 're-elect', 'bill', 'rate',
'public', 'broker', 'survey', 'chang', 'depart', 'campaign', 'select',
'asset', 'strike', 'compani', 'hold', 'tariff', 'account', 'work',
'executor', 'social', 'control', 'claim', 'household', 'agenc', 'rye',
'stock', 'trade', 'report', 'return', 'tribut', 'investor', 'bank',
'possess', 'oblig', 'debentur', 'profit', 'syndic', 'exchang', 'rank',
'vote', 'system', 'expenditur', 'need', 'by-elect', 'price', 'react',
'plan', 'bomb', 'gold', 'portfolio', 'equiti', 'manufactur', 'busi',
'factor', 'bureau', 'earn', 'commod', 'result', 'expert', 'reform',
'import', 'dividend', 'contract', 'subsidi', 'interest', 'firm', 'demand',
'budget', 'macroeconomi', 'econom', 'check', 'oil', 'inflat', 'reward',
'money', 'wage', 'corpor', 'export', 'short-term', 'deficit', 'wheat',
'market', 'promot', 'monetari', 'banker', 'produc', 'war', 'analyst',
'invest', 'gold', 'emiss', 'rate', 'reserv', 'return', 'insur', 'liabil',
'fed', 'currenc', 'transact', 'econom', 'asset', 'interest', 'capit',
'regul', 'risk', 'financi', 'repo', 'fund', 'valu', 'liquid', 'discount',
'invest', 'credit']

self.word_features_positive = ['relea', 'excess', 'strong',
'great', 'honor', 'extra', 'wit', 'promot', 'extravag', 'desir', 'award',
'high', 'attract', 'pure', 'stabil', 'better', 'good', 'upward', 'regul',
'safeti', 'believ', 'sympathi', 'optimist', 'posit', 'modern', 'strength',
'reach', 'clear', 'cheapen', 'increas', 'fair', 'best', 'favor', 'succeed',
'improv', 'success', 'popular', 'right', 'trust', 'wealth', 'happi',
'enlarg', 'right', 'old', 'prefer', 'prosper', 'apt', 'remun', 'adequ',
'opportun', 'proper', 'desir', 'skill', 'superior', 'fit', 'profit',
'pay', 'opul', 'happi', 'better', 'felicit', 'suffici', 'suitabl',
'advis', 'suit', 'forward', 'conveni', 'perspect', 'match', 'applic',
'product', 'good', 'vers', 'amelior', 'season', 'propiti', 'fruit',
'gain', 'go-ahead', 'becom', 'grow', 'present', 'practic', 'experienc',
'appropri', 'success', 'lucr', 'melior', 'elig', 'expedi', 'train',
'apposit', 'improv']

self.word_features_negative = ['executor', 'strike', 'risk',
'broke', 'lower', 'failur', 'shock', 'neg', 'critic', 'end', 'reject',
```

'crumbl', 'deficit', 'fall', 'break', 'heavi', 'depress', 'confus',  
'restrict', 'attack', 'lowest', 'fear', 'bad', 'weak', 'blame', 'cancel',  
'curb', 'worri', 'unwarr', 'afraid', 'dearer', 'loser', 'stagger', 'down',  
'fought', 'troubl', 'deadlock', 'powerless', 'frighten', 'strongli',  
'retrogress', 'harm', 'drop', 'diminut', 'inauspici', 'turning-point',  
'sag', 'menac', 'poorli', 'impair', 'aggrav', 'recess', 'decreas',  
'bust', 'shake', 'amiss', 'subsid', 'faulti', 'retrovers', 'incorrect',  
'depress', 'worthless', 'hapless', 'fate', 'decay', 'defect', 'labefact',  
'laps', 'declin', 'sinist', 'bad', 'invalid', 'black', 'low',  
'sick', 'threaten', 'decad', 'calamit', 'inferior', 'collaps', 'poor',  
'unlucki', 'diseas', 'comedown', 'ill-star', 'fals', 'weak', 'bale',  
'slump', 'evil', 'improp', 'badli', 'wrong', 'ill-fat', 'wrongli', 'dire',  
'fall', 'bankruptci', 'wick', 'failur', 'smash', 'revers', 'deterior',  
'ill', 'erron', 'unfortun', 'omin', 'inaccur', 'problem', 'degrad', 'dip',  
'crisi', 'viciou']

## Příloha B

# Ukázkový výpis testovacího skriptu

Ukázkový výpis testovacího skriptu

cet bigramu: 1

border limit: 0.2

Minimum zprav za den: 6

testovací data splnuji limit borderu

```
set(['1931-10-28', '1931-5-15', '1929-8-15', '1931-2-4', '1931-12-30',  
'1931-1-27', '1929-10-22', '1929-11-11', '1930-5-15', '1929-8-7',  
'1931-7-7', '1929-2-25', '1929-7-30', '1930-12-4', '1931-11-24',  
'1930-1-21', '1931-5-25', '1931-4-28', '1931-1-15', '1928-12-31',  
'1930-9-25', '1930-9-24', '1931-10-1', '1930-5-7', '1930-2-6',  
'1930-11-19', '1929-4-2', '1929-7-23', '1930-9-9'])
```

```
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
```

sample: ((, 'max'), '1928-11-1')

sample: (("contains(('heavi', 'demand'))": True, 'max'), '1928-11-1')

length (train): 595

length (test): 195

```
xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
```

'max': 311, '0.0': 284

minimum: 284

train set len:568

```
xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
```

accuracy: 0.45641025641

Most Informative Features

contains(('depress', 'stock')) = True 0.0 : max = 2.3 : 1.0

contains(('weak', 'stock')) = True 0.0 : max = 2.2 : 1.0

contains(('lowest', 'price')) = True max : 0.0 = 2.1 : 1.0

contains(('low', 'price')) = True max : 0.0 = 1.9 : 1.0

contains(('better', 'busi')) = True max : 0.0 = 1.7 : 1.0

contains(('good', 'busi')) = True max : 0.0 = 1.7 : 1.0

contains(('good', 'demand')) = True 0.0 : max = 1.7 : 1.0

contains(('heavi', 'trade')) = True 0.0 : max = 1.7 : 1.0

contains(('restrict', 'export')) = True max : 0.0 = 1.7 : 1.0

contains(('improv', 'stock')) = True max : 0.0 = 1.7 : 1.0

```

contains(('depress', 'result')) = True 0.0 : max = 1.7 : 1.0
contains(('improv', 'demand')) = True max : 0.0 = 1.7 : 1.0
contains(('improv', 'corpor')) = True max : 0.0 = 1.7 : 1.0
contains(('grow', 'import')) = True max : 0.0 = 1.7 : 1.0
contains(('regul', 'compani')) = True 0.0 : max = 1.7 : 1.0
contains(('break', 'stock')) = True 0.0 : max = 1.7 : 1.0
contains(('fall', 'price')) = True 0.0 : max = 1.4 : 1.0
contains(('high', 'rate')) = True 0.0 : max = 1.4 : 1.0
contains(('popular', 'price')) = True max : 0.0 = 1.4 : 1.0
contains(('high', 'price')) = True max : 0.0 = 1.3 : 1.0
contains(('improv', 'busi')) = True max : 0.0 = 1.3 : 1.0
contains(('extra', 'dividend')) = True max : 0.0 = 1.2 : 1.0
contains(('low', 'price')) = None 0.0 : max = 1.0 : 1.0
contains(('lowest', 'price')) = None 0.0 : max = 1.0 : 1.0
contains(('break', 'stock')) = None max : 0.0 = 1.0 : 1.0
contains(('weak', 'stock')) = None max : 0.0 = 1.0 : 1.0
contains(('lowest', 'rate')) = None 0.0 : max = 1.0 : 1.0
contains(('improv', 'econom')) = None max : 0.0 = 1.0 : 1.0
contains(('depress', 'stock')) = None max : 0.0 = 1.0 : 1.0
contains(('restrict', 'import')) = None max : 0.0 = 1.0 : 1.0
test of classifier
('1929-4-2', 1, 0.75, 'max', 8)
('1930-2-6', 0, 0.33333333333333331, '0.0', 6)
('1929-8-15', 0, 0.0, '0.0', 6)
('1930-9-24', 0, 0.0, '0.0', 6)
('1930-9-25', 0, 0.0, '0.0', 9)
('1931-5-15', 0, 0.16666666666666666, '0.0', 6)
('1931-7-7', 0, 0.25, '0.0', 8)
('1931-12-30', 1, 0.66666666666666663, 'max', 6)
('1931-11-24', 1, 1.0, 'max', 6)
('1929-8-7', 0, 0.33333333333333331, '0.0', 6)
('1931-1-15', 0, 0.0, '0.0', 7)
('1929-2-25', 1, 1.0, 'max', 6)
('1930-5-7', 0, 0.2857142857142857, '0.0', 7)
('1929-7-30', 1, 0.7142857142857143, 'max', 7)
('1931-5-25', 0, 0.44444444444444442, '0.0', 9)
('1931-2-4', 1, 1.0, 'max', 6)
('1929-7-23', 1, 1.0, 'max', 6)
('1930-11-19', 1, 0.8571428571428571, 'max', 7)
('1929-11-11', 0, 0.16666666666666666, '0.0', 6)
('1930-9-9', 1, 1.0, 'max', 6)
('1928-12-31', 1, 0.8571428571428571, 'max', 7)
('1931-10-28', 0, 0.16666666666666666, '0.0', 6)
('1931-4-28', 0, 0.5, '0.0', 6)
('1930-5-15', 0, 0.0, '0.0', 7)
('1930-12-4', 0, 0.0, '0.0', 7)
('1930-1-21', 1, 1.0, 'max', 7)
('1931-1-27', 0, 0.0, '0.0', 8)

```

```
('1931-10-1', 0, 0.42857142857142855, '0.0', 7)
('1929-10-22', 1, 0.6666666666666663, 'max', 6)
uspesnost
0.413793103448
```