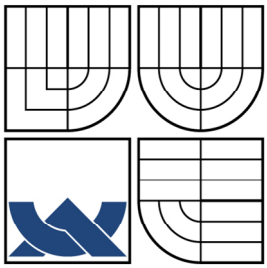
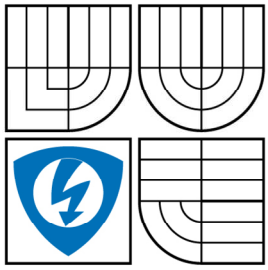


VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY



FAKULTA ELEKTROTECHNIKY A
KOMUNIKAČNÍCH TECHNOLOGIÍ
ÚSTAV AUTOMATIZACE A MĚŘÍCÍ TECHNIKY

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION
DEPARTMENT OF CONTROL AND INSTRUMENTATION

ROZEZNÁVÁNÍ VZORŮ V DYNAMICKÝCH DATECH
PATTERN RECOGNITION IN TEMPORAL DATA

DIPLOMOVÁ PRÁCE
MASTER´S THESIS

AUTOR PRÁCE — Bc. STANISLAV HOVANEC
AUTHOR

VEDOUCÍ PRÁCE — Ing. PETR HONZÍK, Ph.D.
SUPERVISOR

BRNO 2009

VYSOKÉ UČENÍ
TECHNICKÉ V BRNĚ

Fakulta elektrotechniky
a komunikačních technologií

Ústav automatizace a měřicí techniky

Diplomová práce

magisterský navazující studijní obor
Kybernetika, automatizace a měření

Student:	Bc. Stanislav Hovanec	***ID:***	89537
Ročník:	2	***Akademický rok:***	2008/2009

NÁZEV TÉMATU:

Rozeznávání vzorů v dynamických datech

POKYNY PRO VYPRACOVÁNÍ:

Cílem diplomové práce je provést rešerši v oblasti popisu a analýzy časově proměnných dat a vyhledávání předem vybraných vzorů v časových řadách. Na reálných datech ověřte a vyhodnoťte vybranou techniku a dosaženou přesnost predikce. Můžete využít data z oblasti burzy cenných papírů nebo komoditních trhů.

DOPORUČENÁ LITERATURA:

Dle vlastního literárního průzkumu a doporučení vedoucího práce.

Termín zadání:	9.2.2009	***Termín odevzdání:***	25.5.2009

Vedoucí práce: Ing. Petr Honzík, Ph.D.

prof. Ing. Pavel Jura, CSc.
Předseda oborové rady

UPOZORNĚNÍ:

Autor diplomové práce nesmí při vytváření diplomové práce porušit autorská práve třetích osob, zejména nesmí zasahovat nedovoleným způsobem do cizích autorských práv osobnostních a musí si být plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., včetně možných trestněprávních důsledků vyplývajících z ustanovení § 152 trestního zákona č. 140/1961 Sb.

Vysoké učení technické v Brně

Fakulta elektrotechniky a komunikačních technologií

Ústav automatizace a měřící techniky

Rozeznávání vzorů v dynamických datech

Diplomová práce

Specializace studia:	Kybernetika, automatizace a měření
Student:	Bc. Stanislav Hovanec
Vedoucí:	Ing. Petr Honzík, Ph.D.

Abstrakt :

V této diplomové práci je prvně provedena rešerše v oblasti popisu a analýzy časově proměnných dat. Práce následně uvádí problematiku technické analýzy burzovních grafů, cenových vzorů, indikátorů technické analýzy a metody *Pure price action*. Na praktickém příkladě dvou reálných obchodů je vysvětlena aplikace metody *Pure price action* v praxi za účelem vyhledávání a rozpoznávání cenových vzorů, analýzy a predikce dalšího možného vývoje ceny finančního instrumentu. Také je krátce vysvětlen samotný proces úspěšného obchodování. Následně je probráno téma *Rozeznávání Vzorů* a metoda *Učení založené na instancích*.
V praktické části je pomocí programu MATLAB realizován algoritmus rozeznávání cenového vzoru *Korekce* pro nákup a prodej v dynamických časových řadách, konkrétně v cenových burzovních grafech, jakými jsou komoditní nebo akciové průběhy cen v závislosti na čase. Pro analýzu časových řad je použita metoda *Pure price action.* Metoda *Učení založené na instancích* řeší algoritmus vyhledávání cenových vzorů. Vytvořený algoritmus je ověřen na reálných datech 5-ti minutového úsečkového grafu USA indexu Dow Jones za rok 2006, 2007 a 2008. Dosažená úspěšnost a přesnost predikce je vyhodnocena pomocí *Equity křivky*.

Klíčová slova : časové řady, analýza časových řad, technická analýza, cenové vzory, indikátory technické analýzy, cenové grafy, Pure Price Action, rozeznávání vzorů, učení založené na instancích, predikce, normalizace, Equity křivka.

Brno University of Technology

The Faculty of Electrical Engineering and Communication

Department of Control, Measurement and Instrumentation

Pattern recognition in temporal data

Master's Thesis

Specialisation of study: Cybernetics, Control and Measurement

Student: Bc. Stanislav Hovanec

Supervisor: Ing. Petr Honzík, Ph.D.

Abstract :

This diploma work initially conduct research in the area of descriptions and analysis of time series. The thesis then proceed to introduce the problems of technical analysis of price charts as well as indicators, price patterns and method of *Pure Price Action*. The method *Pure Price Action* is demonstrated in this work in two practical examples of its application to real businesses with a view to discovering and analyzing price patterns, as well as analysis and prediction of future price and financial evolution. This analysis is an introduction to the processes of successful business, following on from this we discuss the theme of *Pattern Recognition* and the *Instance Based Learning* method.

The practical aspect of this work is carried out with the aid of a MATLAB applied algorithm for the analysis of the price pattern *Correction* for sale and purchase in dynamic time segments, specifically in trading price graphs, like those used for commodities or stock trading. For the analysis of time series we use the *Pure Price Action* method. The *Instance Based Learning* method is used by the algorithm to recognize price patterns. The created algorithm is verified on real data of a 5 minute time series of the US Dow Jones price charts for the years 2006, 2007, 2008. The achieved accuracy is evaluated with the aid of *Equity Curves*.

Key words : time series, analysis of time series, technical analysis, price patterns, technical indicators, price charts, Pure Price Action, Pattern Recognition, Instance Based Learning, prediction, normalization, Equity curve.

Bibliografická citace

HOVANEC, S. *Rozeznávání vzorů v dynamických datech*. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, 2009. 70 stran. Vedoucí diplomové práce Ing. Petr Honzík, Ph.D.

P r o h l á š e n í

„Prohlašuji, že svou diplomovou práci na téma *Rozeznávání vzorů v dynamických datech* jsem vypracoval samostatně pod vedením vedoucího diplomové práce a s použitím odborné literatury a dalších informačních zdrojů, které jsou všechny citovány v práci a uvedeny v seznamu literatury na konci práce.

Jako autor uvedené diplomové práce dále prohlašuji, že v souvislosti s vytvořením této diplomové práce jsem úmyslně neporušil autorská práva třetích osob, zejména jsem úmyslně nezasáhl nedovoleným způsobem do cizích autorských práv osobnostních a jsem si plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., včetně možných trestněprávních důsledků vyplývajících z ustanovení § 152 trestního zákona č. 140/1961 Sb.“

V Brně dne : Podpis:

P o d ě k o v á n í

Děkuji svému vedoucímu diplomové práce Ing. Petru Honzíkovi, Ph.D. za účinnou metodickou, pedagogickou a odbornou pomoc, jeho velkorysou vstřícnost, a jeho další cenné rady při zpracování mé diplomové práce.

Na tomto místě také velmi děkuji celé své rodině za jejich maximální podporu, toleranci a jejich lásku, a to nejen po dobu mého studia na vysoké škole.

OBSAH

1. ÚVOD 9
2. ČASOVÉ ŘADY 12
2.1 Základní dělení časových řad 13
2.2 Analýza časových řad 14
2.3 Popisné charakteristiky časových řad 15
2.4 Úpravy časových řad 16
2.5 Specifické problémy časových řad 16
2.6 Metody analýzy časových řad 17
2.7 Modely časových řad 19
3. TECHNICKÁ ANALÝZA 21
3.1 Typy grafů 22
3.2 Indikátory technické analýzy 22
3.3 Pure price action 24
3.4 Cenový vzor *Korekce* 24
3.5 Aplikace Pure Price Action v praxi 27
4. ROZPOZNÁVÁNÍ VZORŮ 30
4.1 Terminologie rozpoznávání vzorů 31
4.2 Problém s rozhodovacími mezemi 32
4.3 Hlavní cíle algoritmu Pattern Recognition 34
4.4 Komponenty systému rozpoznávání vzorů 34
5. UČENÍ ZALOŽENÉ NA INSTANCÍCH 36
5.1 Princip IBL 36
5.2 Charakteristiky IBL 37
5.3 Výhody IBL 38
5.4 Nevýhody IBL 38
5.5 Metrika 39
5.6 Normalizace 39
5.7 Chybová funkce 40
6. PROGRAMOVÁ IMPLEMENTACE 41
6.1 Cenový vzor *Korekce* pro nákup 44

6.1.1 Definice vzoru 44
6.1.2 Nalezené vzory 45
6.1.3 Následující průběh u nalezených vzorů 47
6.1.4 Úspěšnost predikce vzoru 50
6.1.5 Equity křivka 51
6.2 Cenový vzor *Korekce* pro prodej 54
6.2.1 Definice vzoru 54
6.2.2 Nalezené vzory 54
6.2.3 Následující průběh u nalezených vzorů 57
6.2.4 Úspěšnost predikce vzoru 59
6.2.5 Equity křivka 60
7. ZÁVĚR 63
8. POUŽÍTÁ LITERATURA 65

SEZNAM OBRÁZKŮ

Obrázek 1: Dynamická časová řada – intradenní vývoj indexu Dow Jones 11
Obrázek 2: Dynamická časová řada – vývoj ceny zlata na denním grafu 11
Obrázek 3: Ukázka indikátorů technické analýzy (Stochastic a MACD)..... 23
Obrázek 4: Formace cenového vzoru *Korekce* 25
Obrázek 5: Předpokládaný průběh cenového vzoru *Korekce* 26
Obrázek 6: Cenový vzor *Korekce* pro prodej (1x) a pro nákup (3x) 26
Obrázek 7: Ukázka využití *Pure price action* v praxi..... 28
Obrázek 8: Ideální grafická reprezentace dat a rozhodovacích mezí [3]..... 32
Obrázek 9: Realističtější grafická reprezentace dat a rozhodovacích mezí [3] 33
Obrázek 10: Komponenty systému rozpoznávání vzorů [3]..... 35
Obrázek 11: Testovací data pro vyhledávání cenových vzorů, trh YM, 2006..... 41
Obrázek 12: Testovací data pro vyhledávání cenových vzorů, trh YM, 2007..... 42
Obrázek 13: Testovací data pro vyhledávání cenových vzorů, trh YM, 2008..... 42
Obrázek 14: Definice průběhu cenového vzoru *Korekce* 44
Obrázek 15: Nalezený vzor *Korekce* v testovacích datech 45
Obrázek 16: Nalezený vzor *Korekce* v testovacích datech 46
Obrázek 17: Nalezený vzor *Korekce* v testovacích datech 46
Obrázek 18: Nalezený vzor *Korekce* v testovacích datech 47
Obrázek 19: Následující průběh (červeně) u nalezeného vzoru..... 47
Obrázek 20: Následující průběh (červeně) u nalezeného vzoru..... 48
Obrázek 21: Následující průběh (červeně) u nalezeného vzoru..... 48
Obrázek 22: Následující průběh (červeně) u nalezeného vzoru..... 48
Obrázek 23: Následující průběh (červeně) u nalezeného vzoru..... 49
Obrázek 24: Následující průběh (červeně) u nalezeného vzoru..... 49
Obrázek 25: Četnosti nalezených ziskových a ztrátových vzorů *Korekce* - nákup ... 51
Obrázek 26: Equity křivka obchodů, pozice long, rok 2006..... 52
Obrázek 27: Equity křivka obchodů, pozice long, rok 2007..... 52
Obrázek 28: Equity křivka obchodů, pozice long, rok 2008..... 53
Obrázek 29: Definice průběhu cenového vzoru *Korekce* 54
Obrázek 30: Nalezený vzor *Korekce* v testovacích datech 55

Obrázek 31: Nalezený vzor *Korekce* v testovacích datech 56
Obrázek 32: Nalezený vzor *Korekce* v testovacích datech 56
Obrázek 33: Nalezený vzor *Korekce* v testovacích datech 57
Obrázek 34: Následující průběh (červeně) u nalezeného vzoru..... 57
Obrázek 35: Následující průběh (červeně) u nalezeného vzoru..... 58
Obrázek 36: Následující průběh (červeně) u nalezeného vzoru..... 58
Obrázek 37: Následující průběh (červeně) u nalezeného vzoru..... 59
Obrázek 38: Četnosti nalezených ziskových a ztrátových vzorů *Korekce* - prodej ... 60
Obrázek 39: Equity křivka obchodů, pozice short, rok 2006..... 61
Obrázek 40: Equity křivka obchodů, pozice short, rok 2007..... 61
Obrázek 41: Equity křivka obchodů, pozice short, rok 2008..... 62

SEZNAM TABULEK

Tabulka 1: Počet nalezených vzorů *Korekce* - nákup 45
Tabulka 2: Četnosti nalezených vzorů *Korekce* – pozice long 50
Tabulka 3: Ziskovost nalezených vzorů – pozice long 53
Tabulka 4: Počet nalezených vzorů *Korekce* - prodej..... 55
Tabulka 5: Četnosti nalezených vzorů *Korekce* – pozice short 59
Tabulka 6: Ziskovost nalezených vzorů – pozice short 62

1. ÚVOD

Vědecká disciplína *Rozpoznávání vzorů, ang. Pattern Recognition (PR),* se zabývá automatickým strojovým rozpoznáváním různých objektů, obrazů nebo signálů. Každý živý tvor na zemi má vrozenou schopnost identifikovat a klasifikovat vzory, jako například potravu, objekty, nepřátele, melodie, slova... Pro lidstvo je velkou výzvou naučit stroje nebo automatické systémy přesně rozpoznávat potřebné vzory.

Vyhledávání vzorů se řadí do kategorie strojového učení. To také definuje rozpoznávání vzorů jako zaznamenávání nezpracovaných dat a uskutečnění určité akce na základě kategorie těchto dat. Tedy rozpoznávání vzorů usiluje o klasifikaci dat na základě informací získaných ze vzorů nebo také na základě apriorní znalosti.

V této diplomové práci bude provedena rešerše v oblasti rozpoznávání vzorů: základní terminologie, hlavní cíle algoritmu Pattern Recognition, jednotlivé komponenty automatického systému pro rozpoznávaní vzorů. Také bude představena velmi efektivní metoda *Učení založené na instancích, ang. Instance Based Learning (IBL),* která se používá pro vyhledávání vzorů. Metoda IBL bude dále použita v programové implementaci.

Tato diplomová práce se kromě rozpoznávání vzorů zabývá i časovými řadami, což jsou sekvence datových bodů, které se měří zpravidla ve stejných časových intervalech. Studií časových řad se zabývá mnoho různých věd jako například statistika, zpracování digitálních signálů, ekonomie, fyzika, strojové učení, medicína... Časovou řadou může být například průběh venkovní teploty nebo tlaku v závislosti na čase nebo také průběh tepové frekvence při zátěži sportovce. I v této oblasti časových řad bude proveden základní teoretický průzkum, zejména pak popis a analýza dynamických časových řad. Dále bude uvedena problematika technické analýzy burzovních grafů, cenových vzorů, indikátorů technické analýzy a metody *Pure price action*. Na praktickém příkladě dvou reálných obchodů bude vysvětlena aplikace metody *Pure price action* v praxi za účelem vyhledávání a rozpoznávání

cenových vzorů, analýzy a predikce dalšího možného vývoje ceny finančního instrumentu. Také bude krátce vysvětlen samotný proces úspěšného obchodování.

V praktické části bude realizován pomocí programu MATLAB algoritmus pro rozpoznávání vzorů v dynamických časových řadách, konkrétně v cenových burzovních grafech, jakými jsou komoditní nebo akciové průběhy cen v závislosti na čase. Na základě různých vzorů v cenových grafech je možné s určitou pravděpodobností predikovat další vývoj této dynamické časové řady. Konkrétně se bude jednat o vyhledávání cenového vzoru *Korekce* pro nákup a pro prodej daného obchodního titulu. Pro analýzu časových řad bude použita metoda *Pure price action.* Metoda *Učení založené na instancích* bude řešit algoritmus pro vyhledávání cenových vzorů. Vyhledávání vzorů proběhne v datech za rok 2006, 2007 a 2008, na 5-ti minutovém úsečkovém grafu akciového USA indexu *Dow Jones.* Závěrem bude vyhodnocena úspěšnost nalezených vzorů a jejich podobnost s ideálním průběhem vzoru *Korekce.* Úspěšnost predikce nalezených vzorů a hypotetických obchodů bude zobrazena v tzv. *Equity křivce.*

Obrázek 1 názorně ilustruje příklad dynamické časové řady, a to *intradenní* průběh ceny indexu *Dow Jones* na 3 minutovém úsečkovém grafu. *Obrázek 2* zobrazuje vývoj ceny zlata v závislosti na čase, konkrétně na denním úsečkovém grafu. Cílem této práce je vytvořit program, který umí sám vyhledat a klasifikovat vzory v podobných cenových grafech. Například zkušený expert-obchodník rozpozná během *intradenního* obchodování v průběhu ceny na *Obrázku 1* hned několik cenových vzorů, jako např. *Korekce* nebo *Otočení Trendu*. Dále na *Obrázku 2* jsou další zřetelné cenové vzory: tzv. *Oops, 8020, Gap, Gap Reversal* a další. Na základě aktuálních cenových vzorů a dle obchodního systému pak dělá expert rozhodnutí o obchodování.

Obrázek 1: Dynamická časová řada – intradenní vývoj indexu Dow Jones

Obrázek 2: Dynamická časová řada – vývoj ceny zlata na denním grafu

2. ČASOVÉ ŘADY

Studií časových řad se zabývá mnoho různých věd jako například statistika, fyzika, ekonomie, zpracování digitálních signálů, strojové učení, medicína... Dalo by se říci, že nejvýznamnější aplikací statistických metod v mnoha vědách je právě analýza časových řad.

Časové řady jsou sekvence hodnot ukazatelů, které se měří obvykle v konstantních časových intervalech. Časovou řadu s ukazatelem ***y*** a časovou proměnnou ***t*** s celkovým počtem pozorování ***n*** můžeme zapisovat například takto: $y_1, y_2,\ldots,y_n$, neboli y_t , $t = 1,\ldots,n$.

Další možný zápis časové řady je například pomocí vektoru:
$Y = \{ Y_1, Y_2, \ldots \}$, kde jednotlivá data Y_1, Y_2, představují postupnou sekvenci prvků časové řady. Zobrazovat časové řady můžeme například pomocí spojnicových grafů, kde vodorovná osa zaznamenává časovou proměnnou ***t***, svislá osa zobrazuje hodnoty ukazatele ***y*** analyzované časové řady.

Časovou řadou může být například průběh venkovní teploty nebo tlaku v závislosti na čase nebo také průběh tepové frekvence při zátěži sportovce, vývoj inflace, vývoj míry nezaměstnanosti, atd. Další příklady dynamických časových řad zobrazují *Obrázek 1* a *Obrázek 2* v úvodní kapitole na straně 11 a 12. Jedná se o průběh cen burzovních titulů na úsečkových grafech. V této diplomové práci budou vyhledávány cenové vzory právě v takovýchto časových řadách.

2.1 ZÁKLADNÍ DĚLENÍ ČASOVÝCH ŘAD

Časové řady dělíme na:

- **deterministické,**
- **stochastické,**
- **ekvidistantní,**
- **neekvidistantní.**

Deterministické časové řady můžeme bezchybně a dokonale predikovat při znalosti způsobu jejich generace. Tyto řady v sobě neobsahují žádný náhodný prvek. Příkladem deterministické časové řady je třeba posloupnost hodnot funkce sinus nebo cosinus.

Stochastické časové řady ovšem nemůžeme bezchybně predikovat, protože obsahují náhodné prvky. Většina ekonomických časových řad, jako například vývoj inflace nebo míry nezaměstnanosti, jsou stochastické časové řady.

Ekvidistantní časové řady mají konstantní časový interval mezi dvěma po sobě jdoucími údaji. Většina časových řad bývá ekvidistantní.

Neekvidistantní časové řady mají různé časové intervaly v datové sekvenci. Práce s neekvidistantními časovými řadami je mnohem náročnější.

Literatura [16] uvádí dělení časových řad dle charakteru ukazatele na tyto druhy:

- **okamžikové –** hodnota ukazatele je uvedena k určitému okamžiku času *t*, jako například celkový počet zaměstnanců v určitém čase,
- **intervalové –** velikost sledovaného ukazatele závisí na délce intervalu, za který je sledován, např. měsíční náklady na provoz.

Dále podle druhu ukazatelů dělíme časové řady na tyto typy:

- **s absolutním ukazatelem (očištěné),**
- **s odvozeným ukazatelem (poměrové).**

Časová řada s absolutním ukazatelem je prvotní řada bez úprav. Jednotlivé prvky v řadě jsou získané například měřením nebo pozorováním.

Časová řada s odvozeným ukazatelem je upravená nebo transformovaná původní časová řada s absolutním ukazatelem. Obvyklý způsob transformace je výpočet průměrů a indexů. Některé úpravy mění vlastnosti původní časové řady. Pokud například časovou řadu náhodných nezávislých veličin transformujeme pomocí klouzavého průměru, budou jednotlivé prvky závislé mezi sebou.

2.2 ANALÝZA ČASOVÝCH ŘAD

Podle literatury [16] je základním cílem analýzy časových řad ***snaha porozumět principům, na základě kterých se generují hodnoty řady***. Tedy pokud se podaří tento mechanismus odhalit, je možné předpovídat budoucí vývoj systému a v některých případech dokonce samotný systém řídit a optimalizovat jeho vývoj.

Pomocí přesného ***modelu*** časové řady můžeme pochopit mechanismus vzniku dané časové řady. Model zobrazuje hypotézu o vztahu mezi vysvětlovanými a vysvětlujícími veličinami. Ve většině případech má model podobu několika stochastických rovnic. Tedy je důležité vytvořit přesný model analyzované časové řady. Pomocí tohoto získaného modelu je možné studovat podmínky, které mají různé vlivy na vznik a průběh hodnot analyzované časové řady. Pokud můžeme řídit některé důležité parametry, které mají vlivy na naší časovou řadu, je možné v některých případech dokonce ovlivnit další vývoj časové řady.

Při analýze časových řad bývá následující postup:

- **porozumění charakteru časové řady,**
- **úprava časové řady,**
- **vytvoření vhodného modelu analyzované časové řady,**
- **identifikace modelu (odhadnout velikosti jeho parametrů),**
- **test modelu,**
- **verifikace modelu na základě empirických dat.**

2.3 POPISNÉ CHARAKTERISTIKY ČASOVÝCH ŘAD

Autoři [17] uvádějí několik způsobů popisných charakteristik.

- **Charakteristiky polohy (průměry) –** velmi často je nutné zjistit průměrné hodnoty analyzované časové řady, jako například prostý aritmetický průměr, vážený aritmetický průměr nebo vážený chronologický průměr.
- **Charakteristiky variability –** ve statistice se nejvíce používají rozptyl a směrodatná odchylka pro určení míry variability.
- **Míry dynamiky –** základní vlastnosti chování časových řad můžeme charakterizovat pomocí první diference (absolutní přírůstek), relativního přírůstku, koeficientu růstu, meziročního koeficientu růstu.
- **Korelace –** vyjadřuje relativní míru závislosti ve vzájemném vývoji dvou časových řad.
- **Stacionární řada –** změny v průměru nebo variabilitě časové řady neovlivňují její chování. Tedy nemůžeme odlišit jeden úsek řady od druhého.
- **Nestacionární řada –** změny v průměru nebo variabilitě časové řady ovlivňují její chování.

2.4 ÚPRAVY ČASOVÝCH ŘAD

V mnoha případech je nutné původní časovou řadu upravit nebo transformovat. Nejčastější úpravy jsou tyto:

- **doplnění chybějících hodnot,**
- **vyhlazování časových řad,**
- **transformace měřítka,**
- **kombinace časových řad,**
- **časový posun,**
- **sezónní diference,**
- **kumulativní součet.**

2.5 SPECIFICKÉ PROBLÉMY ČASOVÝCH ŘAD

Časové řady mají hned několik specifických rysů, které musíme brát v úvahu při jejich zpracování.

- **Problémy s kalendářem –** různá délka měsíců, různý počet pracovních dnů a víkendů v měsíci, pohyblivé svátky.
- **Problémy s délkou časových řad –** je nutné volit optimální počet pozorování, záleží na použitém modelu analýzy. Např. pro Box-Jenkinsovskou metodologii je minimální počet 50 pozorování. Naopak při velkém počtu pozorování může být průběžně měněná charakteristika modelu, což komplikuje analýzu a modelování.
- **Problémy s volbou časových bodů pozorování –** většina reálných časových řad bývá spojitá, ovšem analýzu musíme provádět s diskrétní časovou řadou. Tedy musíme volit správný způsob diskretizace (prostá diskretizace, akumulace, průměrování hodnot).
- **Problémy s nesrovnalostí jednotlivých měření –** např. růst cenové hladiny, růst kvality, změny struktury vzorku.

2.6 METODY ANALÝZY ČASOVÝCH ŘAD

Existuje mnoho metod analýzy časových řad. Jak je uvedeno v [17], výběr metody analýzy závisí na řadě faktorů:

- **účel analýzy,**
- **typ časové řady,**
- **dostupná databáze,**
- **softwarové a hardwarové vybavení,**
- **zkušenosti experta, statistika.**

Mezi základní metody analýzy časových řad patří [16]:

- **expertní (kvalitativní) metody,**
- **grafická a psychologická analýza,**
- **dekompozice časových řad,**
- **Box-Jenkinsovská metodologie,**
- **spektrální analýza,**
- **lineární dynamické (ekonometrické) modely**

Expertní (kvalitativní) metody se používají v případech, kde není možnost optimální kvantifikace analyzované veličiny. Pro dlouhodobé predikce vykazují velmi kvalitní výsledky. Nevýhodou kvalitativní metody je vysoká náročnost, také je nákladná. Nejčastěji používané expertní metody jsou *Delphi metoda, dotazování zákazníků, dotazování prodejců.*

Grafická a psychologická analýza časových řad patří mezi vysoce efektivní a nejjednodušší metody analýzy. Vychází se přímo z časové řady, kterou reprezentuje graf. V daném grafu je možné jednoduše analyzovat časovou řadu. Zkušený expert rozpozná krátkodobé a dlouhodobé trendy, různé vzory chování („patterny“) časové řady, které se opakují. Tato metoda se velice často aplikuje při

technické analýze burzovních grafů, kde mívá vysokou úspěšnost při predikci jak krátkodobých, tak i střednědobých trendů. Grafická metoda bude detailněji probrána v navazující kapitole „Technická analýza“, konkrétně v podkapitolách 3.3, 3.4, 3.5.

Dekompozice časových řad předpokládá, že náhodný proces, který generuje časovou řadu, je závislý jen na čase. Také vychází z toho, že časovou řadu můžeme rozdělit na několik nezávislých složek, jako je trend, cykličnost, sezónnost a náhodnost. Poté je snazší postupně identifikovat chování jednotlivých složek. Časovou řadu tedy tvoří trend a složka sezónní, cyklická a náhodná. Systematickou složkou se souhrnně nazývá trend, sezónní a cyklická složka. Na tuto systematickou složku se zaměřuje metoda dekompozice časových řad a náhodná složka je zcela opomíjená. Jednotlivá pozorování se považují za zcela nezávislá.

Box-Jenkinsovská metodogie při tvorbě modelu časové řady bere v úvahu analýzu náhodné složky, která může být tvořena korelovanými (závislými) veličinami. Jádro této metody spočívá hlavně v této korelační analýze. Hlavním cílem je vyšetření vzájemné závislosti jednotlivých prvků řady s různým zpožděním a závislosti na náhodném vstupu. Výhodou této metody je vysoká flexibilita, adaptují se rychle na změněný charakter časové řady.

V lineárních dynamických a regresních modelech se jedná o příčinné (kauzální) modely, kde je analyzována proměnná vysvětlována pomocí jedné nebo více vysvětlujících proměnných. Jde tedy o odhalení příčinné vazby mezi veličinami. Ve většině případech se předpokládají lineární nebo linearizované závislosti mezi proměnnými. Literatura [16] také uvádí příklad, kde při modelování inflace může být vysvětlovanou proměnnou cenová hladina, vysvětlujícími proměnnými reálný hrubý domácí produkt a množství peněz v oběhu. Modely se konstruují pomocí teoretických předpokladů, tato metoda analýzy se velmi užívá v ekonometrii.

Spektrální analýza se také nazývá analýza ve spektrální doméně nebo tzv. fourierová analýza, kde se zkoumaná časová řada považuje za nekonečnou směs sinusových a kosinusových křivek s různými frekvencemi a amplitudami. Předchozí metody prováděly analýzu v časové doméně. Spektrální analýza vyšetřuje spektrum řady - zjistí intenzitu zastoupení jednotlivých frekvencí a jejich parametrů v naší časové řadě. Pomocí spektrální analýzy se také může posuzovat zpoždění ve vývoji mezi dvěma veličinami – a to porovnáním jejich spekter.

2.7 MODELY ČASOVÝCH ŘAD

Analýza časových řad má za cíl vytvořit a identifikovat model analyzované řady, tedy nalézt matematický popis modelu, a odhadnout jeho parametry. Nejčastěji se používají tyto modely časových řad:

- **Modely náhodných procházek**
- **Modely klouzavých průměrů (MA)**
- **Autoregresní modely (AR)**
- **Autoregresní modely klouzavých průměrů (ARMA)**
- **Autoregresní integrované modely klouzavých průměrů (ARIMA)**

U modelů náhodných procházek [18] modelujeme jednoduchý **stochastický proces**. Tato řada je **nestacionární**, protože průměr a rozptyl se mění, není konstantní. Použijeme-li se model náhodných procházek k popisu dynamického chování, pak půjde o nestacionární model časových řad. Avšak časové řady cenových změn jsou již generovány stacionárním ryze náhodným procesem, nazývaným také jako bílý šum.

Pomocí **modelů klouzavých průměrů (MA)** můžeme jednoduše a efektivně modelovat dynamiku **stacionárních** časových řad. Klouzavé součty označují proces, ve kterém je hodnota vysvětlované veličiny tvořená lineární kombinací současné hodnoty a minulých hodnot náhodné veličiny. U náhodné veličiny nemůže určit její hodnotu ani průběh. Existují ovšem techniky, které umožňují odhadnout náhodnou veličinu přímo z dat.

Autoregresní modely (AR) představují další přístup k modelování časové struktury **stacionárních** časových řad. Vyjádření vysvětlované veličiny je funkce několika předcházejících pozorování – lineární kombinací minulých hodnot této řady. Dle [18] se v praxi při modelování časových řad setkáváme s případy, kdy stacionární náhodný proces, generující jednotlivá pozorování, nevyhovuje zcela předpokladům MA, resp. AR modelů. V této situaci je adekvátní taková specifikace modelu časové řady, jejíž složky vycházejí z principu kombinace AR a MA procesů – vzniká tak model ARMA.

Autoregresní model klouzavých průměrů ARMA (p,q) spojuje oba základní modely MA a AR. Jedná se tedy o smíšený model časové řady, kde *p* odpovídá výrazů autoregresního typu AR a *q* reprezentuje zpožděné klouzavé průměry MA.

Autoregresní integrovaný model klouzavých průměrů ARIMA se používá pro práci s nestacionárními časovými řadami. ARIMA se aplikuje na modelování integrovaných procesů, které můžeme převést na stacionární procesy pomocí diferenciací určitého řádu. Model ARIMA (n_a, d, n_c) pak odpovídá ARMA(n_a, n_c) modelu d-té diference integrované časové řady.

3. TECHNICKÁ ANALÝZA

Technická analýza je aplikována na všechny finanční produkty za účelem predikce budoucích cenových pohybů. Predikce probíhá na základě analyzování a vyhodnocování minulých a aktuálních burzovních dat, které se skládají z ceny (Price), objemu proběhlých obchodů (Volume), počtu otevřených pozic (Open Interest), volatility (Volatility), počtu umístěných příkazů na trhu, nebo také i z korelace daného finančního instrumentu s jiným (například korelace mezi indexy YM, SP, NQ, nebo korelace ceny zlata s dluhopisy).

Predikce cenových pohybů může být také prováděna pomocí fundamentální analýzy, která používá například ekonomické informace, policikou situaci, úrokové sazby, zásoby komodit… Tedy technická analýza vychází jen z informací, které generuje samotný pohyb ceny analyzovaného finančního instrumentu.

Zkušený expert dokáže s vysokou pravděpodobností konzistentně předvídat krátkodobé (v horizontu minut až jednoho týdne) i střednědobé (v horizontu přibližně až čtyři týdny) cenové pohyby, a to s pomocí technické analýzy, která určuje, jaký pohyb je v daný okamžik nejpravděpodobnější. Tedy ceny se nepohybují náhodně, ale pohybují se v trendech krátkodobých, střednědobých a také i dlouhodobých.

Samozřejmě ze samotné podstaty trhů pohyby cen bývají často iracionální a neexistuje maximální jistota v predikci jejich dalšího vývoje. Cenu udávají obchodníci, kteří také můžou jednat emotivně a iracionálně. Je proto důležité počítat i s tím, že daný predikovaný pohyb naší dynamické časové řady bude zcela odlišný od reality. Obchodování na burze s sebou nese značné riziko ztráty kapitálu. Ovšem toto není úkol pro technickou analýzu, ale zabývá se tím oblast *Risk Management* a *Money Management.* Vše dohromady pak tvoří tzv. *Obchodní systém*, který definuje, kdy vstupovat, kolik riskovat, a kdy vystupovat z obchodní pozice. Velice důležitým faktorem při úspěšné predikci finančního instrumentu je používání citu a intuice, které se nedají naprogramovat. Proto žádné automatické obchodní systémy zdaleka nedosahují výkonnosti zkušeného experta při samotném procesu obchodování.

3.1 TYPY GRAFŮ

Pro předpověď pohybů cen se prvotně vychází ze samotných grafů, kde na vodorovné ose je zobrazen čas, na svislé ose zpravidla cena analyzovaného instrumentu. Nejvíce se používají tyto tři různé typy grafů:

- **spojnicový graf (Liniový graf, Line chart),**
- **úsečkový graf (OHLC, Schodový graf, Bar Chart),**
- **svíčkový graf (Candlestick).**

Existují také další alternativní grafy, jako například *Volume graf, Tick graf,* nebo *Range bar.* V této práci bude použit pouze úsečkový graf, který má největší vypovídací schopnost, podobně jako svíčkový graf. Na *Obrázku 1* a *Obrázku 2* v úvodní kapitole byly zobrazeny úsečkové grafy, kde každá jednotlivá úsečka udává detailní průběh ceny v daném časovém intervalu. Každý „schod“ je tvořen ze třech částí, kde svislá přímka zobrazuje maximum (High) a minimum (Low), kratší vodorovná část z levé strany představuje otevírací cenu (Open), stejně jako uzavírací cena (Close), která je ze strany pravé. Pomocí schodového grafu je možné nacházet jednotlivé cenové vzory a rozpoznat krátkodobé nebo i dlouhodobé trendy a predikovat jejich otočení.

Předvídat nejpravděpodobnější pohyb na naší dynamické časové řadě můžeme pomocí indikátorů technické analýzy, nebo pomocí *Pure price action („Čisté chování ceny“).*

3.2 INDIKÁTORY TECHNICKÉ ANALÝZY

Indikátory technické analýzy si můžeme zjednodušeně představit jako různé křivky, které vznikají matematickými výpočty z několika posledních cen zkoumaného finančního instrumentu. Protože indikátory vycházejí zejména ze

samotné ceny (OHLC), která se následně přepočítá dle matematických vztahů, indikátory tedy představují opožděný derivát ceny. Tento fakt může poněkud negativně ovlivnit predikci zkoumané časové řady. Existuje celá řada indikátorů technické analýzy:

- **Indikátory trendu:** klouzavé průměry MA, exponenciální klouzavé průměry EMA, indikátor konvergence a divergence MACD, Williams Percent %R, …
- **Indikátory volatility:** Standardní odchylka SD, Bollingerova pásma, Index výběru komodit, …
- **Indikátory hybnosti:** Swingový index akumulace ASI, Index komoditních kanálů CCI, Dynamický index momenta DMI, …
- **Indikátory cyklů:** Cyklické křivky CL, Fibonacciho časové zóny FT, …
- **Indikátory tržní síly**

No *Obrázku 3* je zobrazen 2-denní průběh ceny a technické indikátory Stochastic a MACD. Jedná se o úsečkový graf s časovou periodou (Time Frame) 5min. Nastavení parametrů použitých technických indikátorů záleží na optimální volbě experta, který vychází z dlouhodobého studia daného indikátoru a trhu. Například někteří experti používají pro nastavení indikátoru MACD hodnoty 9, 13 a 26 z Fibonacciho řady.

Obrázek 3: Ukázka indikátorů technické analýzy (Stochastic a MACD)

3.3 PURE PRICE ACTION

Price Action je výraz pro slovní spojení „chování ceny.“ Pokud při analýze cenového průběhu používáme jen indikátory technické analýzy a tedy ignorujeme samotný průběh ceny zkoumaného finančního instrumentu, bude predikce málo efektivní. Je proto výhodnější používat informace ze samotného průběhu naší dynamické řady, tedy grafu ceny bez jakékoli úpravy. Vůbec nejefektivnější predikcí dosáhneme s použitím tzv. metody *Pure Price Action,* při které se experti rozhodují pouze na základě průběhu samotné ceny. Tedy pro analýzu nepoužívají indikátory technické analýzy, které mají tendenci být opožděné. U této metody se jedná o grafickou a psychologickou metodu analýzy časových řad. Pro kvalitní predikci pomocí *Pure Price Action* je nutné umět vyhledávat cenové vzory přímo na grafech. Také je velice důležité mít vyvinuté cit a intuici pro chování trhů, nejedná se totiž o exaktní vědu.

Cenové vzory, nebo také cenové formace, patterny, reprezentují určité chování aktérů v daném trhu. Cenových vzorů existuje v podstatě nekonečně mnoho, v cenových grafech se vyskytují s různou distribucí a s odlišnou úspěšností. Jmenovat můžeme například technické formace *Dvojitý Vrchol, Dvojité Dno, Obchodní Pásmo, Korekce, Trojúhelník.*

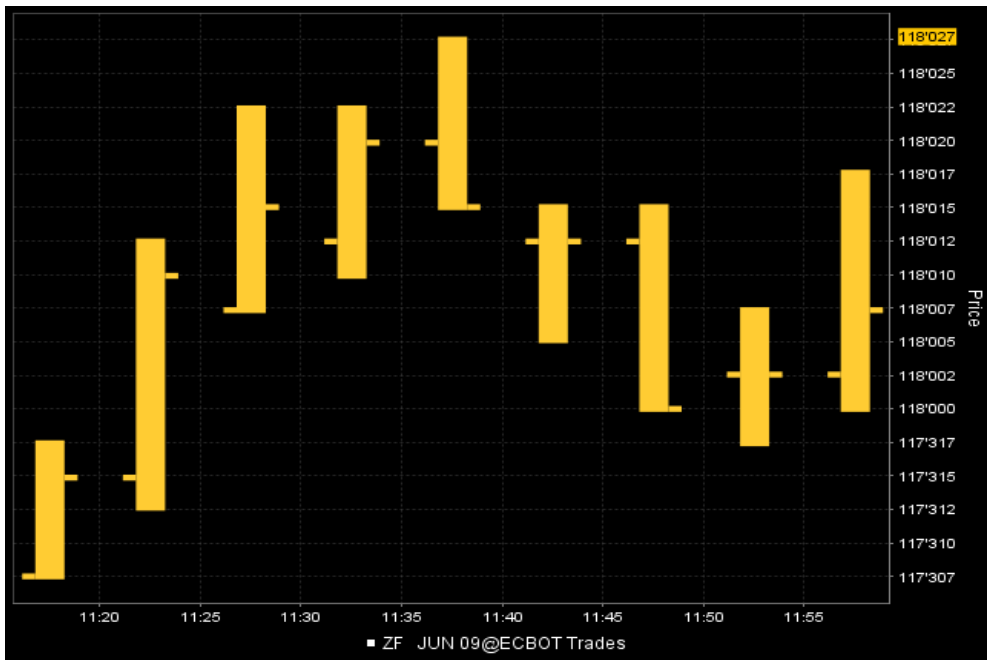
V této práci se budeme zabývat vyhledáváním cenových vzorů *Korekce* pro nákup a pro prodej jen pomocí metody *Pure Price Action.*

3.4 CENOVÝ VZOR *KOREKCE*

Jedná se o vysoce úspěšný a konzervativní cenový vzor s velmi častým výskytem. Existuje celá řada typu korekcí, jejich kombinací a různých nuancí. V této práci je vybraný jediný typ korekce, který bude vyhledáván v naší dynamické řadě.

Na *Obrázku 4* je zobrazen cenový vzor pro nákup. Jedná se o specifickou korekci, kde prvně cena postupně roste a vytvoří lokální maximum, je tedy předpoklad „býčího“ trendu. Následně dochází k malému lokálnímu propadu ceny, k malé korekci ceny. Zastavení tohoto klesání je na předposlední úsečce. Poté, na poslední úsečce, dochází k prudkému otočení korekce ceny a náznak pokračování

předchozího trendu. V tomto momentě, tedy na poslední úsečce, došlo k vytvoření jednoduchého cenového vzoru. Pro experta, který pro predikci používá jen metodu *Pure Price Action*, se jedná o validní signál pro nákup analyzovaného instrumentu. Tento vzor vychází z předpokladu dalšího pokračování trendu a útoku ceny na lokální maximum. Vstup do pozice (nákup) probíhá těsně před pokračováním původního trendu. Ochranný příkaz (prodej), který definuje míru rizika, je umístěn v okolí lokálního minima vytvořeného při korekci.

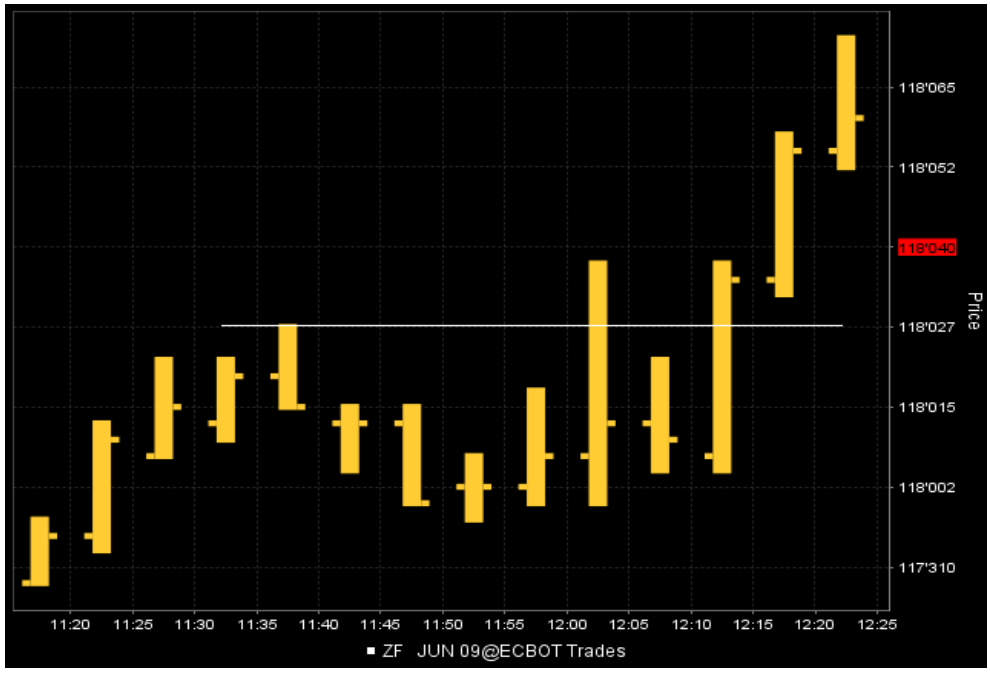


Obrázek 4: Formace cenového vzoru *Korekce*

Na *Obrázku 5* je vyobrazen další vývoj našeho popisovaného vzoru se zvýrazněným lokálním maximem. Jde vidět, že cena útočila na lokálním maximum a později ho překonala a vytvořila nové maximum. Metod výstupů se ziskem (prodej v případě „long“ pozice) existuje celá řada, záleží na samotném stylu obchodníka. Nejjednodušším a nejrychlejším výstupem může být například umístění obchodních příkazů těsně pod lokální maximum, které se budou exekuovat při testování ceny lokálního maxima.

Tedy pro cenový vzor *Korekce* je vždy predikován podobný průběh jako na *Obrázku 5*. Tento následný vývoj má větší pravděpodobnost než vývoj ceny v opačném směru. Na *Obrázku 6* je zobrazen jeden cenový vzor pro prodej (pozice

„short“) a tři cenové vzory pro nákup (pozice „long“.) Bílou čárou jsou vždy zvýrazněny lokální minima, resp. maxima.



Obrázek 5: Předpokládaný průběh cenového vzoru *Korekce*

Obrázek 6: Cenový vzor *Korekce* pro prodej (1x) a pro nákup (3x)

3.5 APLIKACE PURE PRICE ACTION V PRAXI

Obrázek 7 ilustruje úspěšnou aplikaci metody *Pure Price Action*, kterou používá autor této práce během reálného obchodování. V osmi krocích na *Obrázku 7* bude představen samotný proces analýzy časové řady a její predikce, rozpoznávání a vyhledávání cenových vzorů, různé metody výstupů a řízení rizika, používání citu a intuice. Tedy na praktickém příkladě bude krátce vysvětlen princip úspěšného obchodování.

Na *Obrázku 7* je 5-ti minutový úsečkový graf zobrazující vývoj ceny USA dluhopisů během dvou dnů. Během prvního dne obchodník prvně čeká a vyhledává cenové vzory, které sám obchoduje a uzná za vhodné. V bodě č. 1 se vyskytl cenový vzor signalizující zastavení předchozího „medvědího“ trendu (propad ceny) a jeho možné otočení. Předpokládaná predikce pohybu ceny byla dle levé šipky – útok ceny na dnešní maximum a krátkodobé pokračování „uptrendu“ dle denního grafu. Tedy v bodě č. 1 byly otevřeny nákupní pozice s ochranným příkazem (prodej) pod dnešním minimem. Jedná se jen o pravděpodobnost pohybu, proto je nutné mít risk pod kontrolou. Vhodné je riskovat cca 1-3% kapitálu na jeden obchod. Podle toho se volí množství otevřených kontraktů. Velikost risku na jeden otevřený kontrakt v této konkrétní situaci odpovídá cca 70 USD (započteny komise i případný skluz v plnění.) V tomto likvidním trhu je možné bez výrazných problému otevírat pozice až o velikosti 1000 kontraktů. Pro větší pozice je nutné počítat s většími skluzy v plnění, případně používat jinou techniku vstupů a výstupů.

Následný pohyb naší časové řady potvrdil předpokládané otočení trendu. Během následujících cca 50 min bylo vytvořeno lokální maximum a byla určitá pravděpodobnost korekce před dalším růstem. Tedy bylo nutné setrvat a přečkat podobnou korekci ceny, jaká byla popisována v předchozí kapitole. Ochranný příkaz byl na stále bezpečném a vhodném místě. Dále došlo ke krátkému testu dnešního minima a velmi rychlému odrazu a pokračování v novém uptrendu.

Autor vždy používá tři sady obchodních pozic – tedy počet otevřených kontraktů je násobkem tří a používá tak až tři různé výstupy. V bodě č. 2 byly všechny tři sady otevřených pozic najednou posunuty z ochranného příkazu až na

úroveň bodu č. 2 a došlo tak k zálohování akumulovaného zisku. První sada otevřených kontraktů vždy vystupuje se ziskem během dne vstupu na nějaké signifikantní cenové úrovni, kterou opět odhadneme pomocí technické analýzy. V tomto případě tomu odpovídá dnešní maximum v bodě č. 3. Opět pomocí analýzy cenového grafu byl predikován další potenciál pokračování uptrendu a možný ranní *gap* (otevírací mezera), proto byly zbylé dvě sady pozic otevřeny až do dalšího dne.

Obrázek 7: Ukázka využití *Pure price action* v praxi.

Následující den (bod č. 4) došlo k rannímu *gapu* a k prudkému otočení ceny. Vznikl tak cenový vzor *Oops*, který má vysokou pravděpodobnost úspěšnosti predikce. Tedy v tomto případě byl vyhodnocen konec včerejšího uptrendu a predikován útok ceny na včerejší minimum dne, jak ukazuje pravá šipka. Proto v bodě č. 4 došlo k uzavření zbylých dvou sad (realizace zisků prodejním

příkazem) a zároveň otevření nových třech sad pozic do pozice *short* (prodej). Ochranný příkaz (nákup) byl umístněn na vhodném místě v bodě č. 5. Risk tohoto obchodu byl opět do 3% celkového kapitálu. Pokud cena dosáhne úrovně ochranného příkazu, cenový vzor *Oops* nebude úspěšný pro tento den. V bodě č. 6 byl opět rezervován akumulovaný zisk se všemi sadami. Výstup se ziskem pro 1. sadu byl očekáván při testu signifikantní cenové úrovni včerejšího minima. K jeho realizaci došlo v bodě č. 7 (nákup). Jelikož byl následně vyčerpaný pohyb, slabá aktivita obchodníků, nebyl dostatek prodávajících, kteří by cenu tlačili níže, byly zbylé dvě sady uzavřeny v bodě č. 8. V některých případech bývá 2. i 3. sada otevřených pozic držena na denní bázi, to ovšem záleží na aktuální situaci.

Na tomto příkladě je zřetelně vidět, že predikovat vývoj ceny finančních titulů je možné s velmi vysokou úspěšností. Je nutné ovládat technickou analýzu, která je vysoce důležitá. Ovšem pro úspěšné obchodování na burze je technická analýza v podstatě zanedbatelná vůči ostatním dovednostem. Zásadním problémem během samotného procesu obchodování je umět vnímat a ovládat sám sebe – jedná se o psychologii obchodování. Tento problém dokonce nevyřeší ani automatické obchodní systémy, které postrádají schopnost používat cit a intuici. Proto mají automatické obchodní systémy mnohem menší výkonnost a efektivitu než zkušení experti.

4. ROZPOZNÁVÁNÍ VZORŮ

Rozpoznávání vzorů, ang. Pattern Recognition (PR), je vysoce aktivní oblast výzkumů, ve které je stále mnoho nevyřešených problémů, a také stále více a více aplikací, které mohou velmi benefitovat z této vědní disciplíny.

Rozeznávání vzorů usiluje o klasifikaci dat (*vzorů, patternů*) na základě apriorní znalosti nebo na základě informací získaných ze vzorů [5]. Vzory jsou většinou soubory měření nebo pozorování, definující body ve vhodném vícedimenzionálním prostoru

Kompletní automatický systém pro rozpoznávání vzorů se skládá ze snímače, mechanismu extrakce znaků a dále z klasifikátoru. Snímač sbírá data, která mají být klasifikována nebo charakterizována. Mechanismus výběru znaků počítá číselné nebo symbolické informace z pozorovaných dat. Konečně klasifikátor poté třídí nebo charakterizuje pozorované data na základě vybraných znaků.

Klasifikační schéma je obvykle založeno na dostupnosti řady vzorů, které už byly zatříděny nebo popsány. Tato množina patternů se nazývá trénovací množina a výsledná učicí strategie je charakterizována jako „učení s učitelem.“ Učení může být také bez učitele, pokud není systému dána apriorní znalost patternu. V takovém případě systém sám stanovuje třídy na základě statistických pravidelných opakování vzorů.

Třídicí schéma obvykle používá statistický nebo strukturální přístup. Statistické rozpoznávání vzorů vychází z charakterizace patternů, kde se předpokládá, že vzory jsou generovány nějakým pravděpodobnostním systémem. Naproti tomu strukturální rozpoznávání patternů vychází ze strukturálních vzájemných vztahů mezi znaky patternů. Pro rozpoznávání vzorů může být použita

velmi široká škála algoritmů, od velmi jednoduchého Bayesova klasifikátoru až po velmi výkonné neuronové sítě.

Neobvyklým problémem v PR je vztah mezi daty, které mají být tříděna a mezi výkony různých klasifikátorů.

4.1 TERMINOLOGIE ROZPOZNÁVÁNÍ VZORŮ

Autor Robi Polikar [3] uvádí toto obvyklé používání terminologie pro PR:

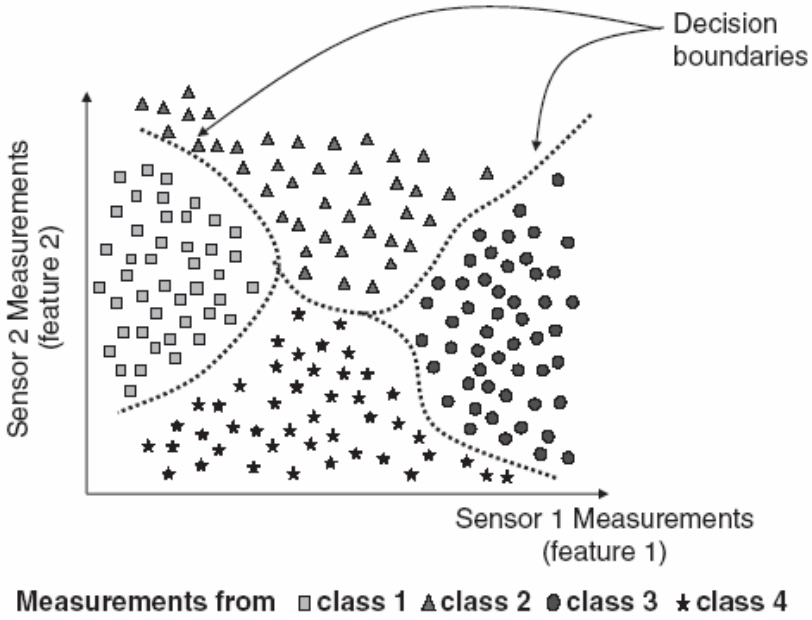
- **Znak (Feature)** – množina proměnných charakterizujících informaci identifikovaného objektu.
- **Vektor znaků x (Feature Vector)** – soubor d znaků uspořádaných do d-dimensionálního sloupcového vektoru.
- **Třída ω (Class)** – kategorie, do které patří daný objekt.
- **Vzor (Pattern)** – soubor znaků identifikovaného objektu společně se správnou třídou pro tento objekt.
- **Exemplář (Instance)** – jakákoli daná ukázka patternu.
- **Trénování (Training)** – procedura, při které se systém rozpoznávání vzorů naučí přiřadit vztahy mezi vektory znaků a odpovídajícími třídami.
- **Rozhodovací meze (Decision Boundaries)** – nelineární funkce, oddělující patterny odlišných tříd jeden od druhého.
- **Tréninková data (Training Data)** – data, na kterých je algoritmus rozpoznávání vzorů trénovaný.
- **Validační data (Validation Data)** – data, na kterých je vyhodnocen výkon algoritmu rozpoznávání vzorů
- **Reálné data (Field Data)** – neznámé data, na kterých se bude aplikovat algoritmus rozpoznávání patternů.
- **Nákladová funkce (Cost Function)** – kvantitativní úroveň reprezentující cenu chyby při třídění. Algoritmus rozpoznávání patternu je speciálně trénovaný pro minimalizaci nákladové funkce.

- **Model, Klasifikátor (Model, Classifier)** – systém rozpoznávání patternů, který nastavuje své parametry pro nalezení správných rozhodovacích mezí pomocí učícího algoritmu a tréninkových dat tak, aby nákladová funkce byla minimální.

Pro vyhodnocení výkonnosti trénovaného klasifikátoru se používají dva parametry:

- **Tréninková výkonnost (Training Performance)** – schopnost klasifikátoru správně identifikovat třídy tréninkových dat. Používá se pro určení délky trvání procedury trénování nebo taky pro určení správnosti naučení se na tréninkových dat.
- **Zobecněná výkonnost (Generalized Performance)** – schopnost klasifikátoru správně identifikovat třídy na dříve neviděných vzorech.

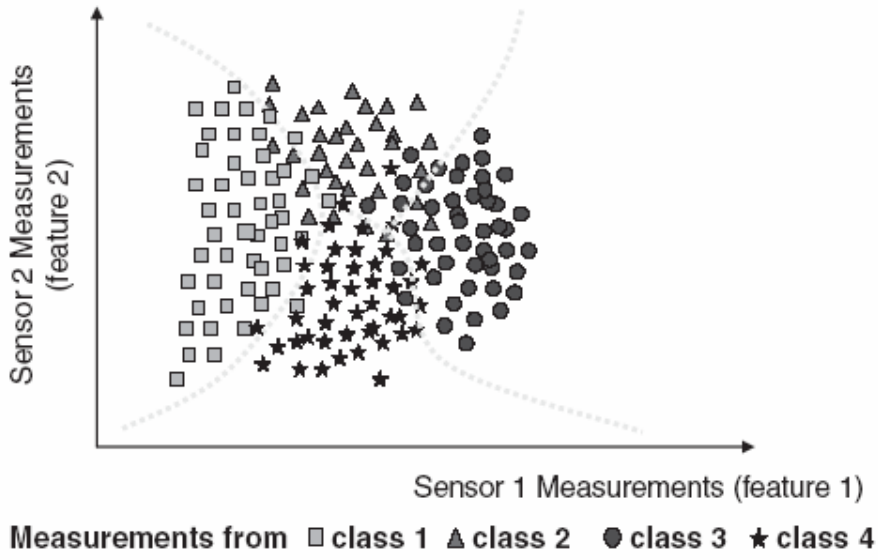
4.2 PROBLÉM S ROZHODOVACÍMI MEZEMI



Obrázek 8: Ideální grafická reprezentace dat a rozhodovacích mezí [3]

Na *Obrázku 8* je ilustrace hypotetického 2D příkladu, kde *Znak 1 (Feature)* může být například systolický krevní tlak a *Znak 2* může být váha pacienta. *Rozhodovací meze (Decision boundaries)* oddělují vzory do celkem čtyř různých kategorií dat. Čtyři různé třídy mohou indikovat počet srdečních infarktů u pacienta během posledních pěti let: žádný, jeden, dva, více než dva…

Obrázek 8 reprezentuje velmi optimistický až ideální scénář rozpoznávání patternů – vzory z jakékoliv třídy jsou perfektně oddělitelné od ostatních tříd pomocí přesně daných rozhodovacích mezí. Takové úlohy jsou považovány za velmi jednoduché pro většinu PR aplikací. Bohužel ve většině aplikací praktického charakteru data nespolupracují na takové ideální úrovni. *Obrázek 9* znázorňuje více realističtější scénář, kde se vzory z různých kategorií překrývají v prostoru. Je mnohem náročnější určit rozhodovací meze, které by přesně rozdělily instance jednotlivých tříd.



Obrázek 9: Realističtější grafická reprezentace dat a rozhodovacích mezí [3]

Cílem algoritmu PR je určit rozhodovací meze, které mají nejlepší zobecněnou výkonnost na neznámých datech. Perfektní tréninková výkonnost tedy vůbec nestačí.

Ovšem ve skutečnosti ani není moc zapotřebí hledat přesné hranice mezi vzory, protože převážná většina překrývání patternů je způsobená „šumem“ v datech. Dokonce nalezení přesných rozhodovacích mezí v tréninkových datech by znamenalo naučení se algoritmu na šum v datech. Permanentní učení se šumu zapříčiní horší zobecněnou výkonnost na testovaných datech. Tomuto velmi častému fenoménu v PR se říká „**přetrénování**“ a dá se mu předcházet několika způsoby.

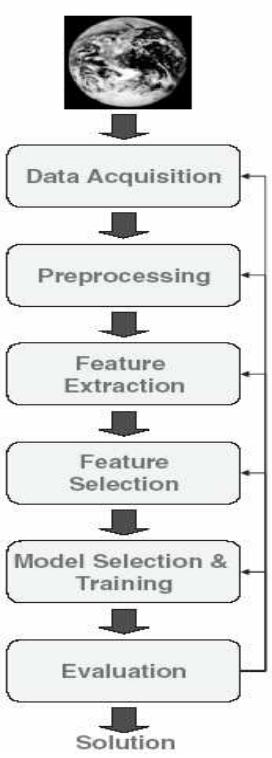
4.3 HLAVNÍ CÍLE ALGORITMU PATTERN RECOGNITION

Správně sestavený algoritmus pro rozpoznávání patternů má tedy za úkol:

- **Odhadnout přesně třídu odpovídající danému znakovému vektoru na základě nějaké předchozí znalosti získané pomocí procedury trénováni.**
- **Nalézt a určit správné rozhodovací meze, které jsou obecně nelineárními funkcemi.**
- **Minimalizovat nákladové funkce pomocí upravování parametrů systému.**

4.4 KOMPONENTY SYSTÉMU ROZPOZNÁVÁNÍ VZORŮ

Kompletní systém pro rozpoznávání vzorů se skládá z několika komponent, viz *Obrázek 10.*



Obrázek 10: Komponenty systému rozpoznávání vzorů [3]

- **Reálný Problém Rozpoznávání Vzorů** – Hledáme vzor z reálného světa. Jak ho klasifikovat?
- **Získáni Dat (Data Acquisition)** – Jak získat data? Kolik dat je třeba?
- **Předzpracování (Preprocessing)** – Proces odstranění šumu, filtrování, normalizace.
- **Extrakce Znaků (Feature Extraction)** – Extrakce důležitých rysů z dostupných dat.
- **Selekce Znaků (Feature Selection)** – Výběr množiny nejvíce relevantních znaků.
- **Výběr Modelu (Model Selection) –** Výběr správného typu modelu.
- **Trénink Modelu (Model Training)** – Trénink vybraného typu modelu pomocí vhodného učícího algoritmu.
- **Zhodnocení (Evaluation)** – Odhadnutí pravé zobecněné výkonnosti klasifikátoru v reálném světě. Zjištění míry jistoty tohoto závěru.
- **Řešení (Solution)** – Vyřešení klasifikačního problému z reálného prostředí a následné automatické rozhodnutí.

5. UČENÍ ZALOŽENÉ NA INSTANCÍCH

Pro rozpoznávání cenových vzorů na burzovních grafech byla použita velmi efektivní a jednoduchá metoda *Učení založené na instancích, ang. Instance Based Learning (IBL).* Dle literatury [13] metody typu IBL fungují na základě podobnosti nového případu s případy již známými. Jedná se o jednoduché a zároveň velmi účinné nástroje rozpoznávání patternů. Přestože metoda IBL je principiálně triviální, umožňuje velmi efektivně aproximovat i značně složité funkce.

5.1 PRINCIP IBL

Proces predikce výstupní hodnoty nového prvku probíhá jednoduše. Máme **nový prvek *Z*,** který je popsán **vstupními veličinami x_z.** Naším úkolem je určit **výstupní hodnotu y_z.** K dispozici máme **trénovací data,** která jsou také popsána stejnými **atributy *x*,** známe u nich však i správné **výstupní hodnoty *y*.**

Definujeme operaci **vzdálenost**, jejímž vstupem jsou dva prvky popsané vstupními atributy *x* a výstupem je kvantitativní vyjádření jejich vzájemné nepodobnosti, tedy vzdálenosti. Na základě této operace je v databázi uložených trénovacích záznamů nalezen předem daný počet prvků, které se svým popisem *x* nejvíce podobají prvku *Z*, který je určen vektorem x_z. Z této nově vytvořené množiny nejpodobnějších případů se bude následně určovat výstupní hodnota .

Literatura [13] dále uvádí, že algoritmus tohoto postupu může principielně probíhat dvěma odlišnými způsoby:

- Výstupní hodnota y_z je určena na základě **výstupních hodnot *y* nalezených sousedů** (popřípadě i jejich vzdáleností od prvku *Z*).
- Na základě **atributů *x* a *y* popisujících nalezené sousedy** je vytvořen lokálně platný model (např. lineární, rozhodovací strom, neuronová síť, …), který generalizuje znalost v okolí neznámého prvku a bude sloužit k určení výstupní hodnoty y_z. Při nastavování parametrů může být součástí chybové funkce vzdálenost sousedů od klasifikovaného prvku.

Společným rysem obou přístupů je **nalezení určitého počtu z uložených prvků, které jsou nové instanci nejpodobnější.**

5.2 CHARAKTERISTIKY IBL

Metody IBL se řadí mezi tzv. líné metody. Je pro ně specifické, že nevytvářejí žádný model. Výpočet sloužící k predikci výstupu provádějí až ve chvíli, kdy je předložen nový prvek k analyzování. Proto jsou metody IBL schopny dobře aproximovat závislosti uvnitř prostoru vymezeném trénovacími daty. Ovšem nejsou schopny dobré extrapolace, tedy predikce mimo oblast pokrytou uloženými záznamy.

Algoritmy IBL představují určitý přechod mezi učením s učitelem a bez učitele. Protože je známá požadovaná výstupní hodnota u každého trénovacího prvku, je IBL algoritmus s učitelem. Ovšem k žádnému „učení“ nedochází, díváme-li se na základní algoritmus k-NN (*k* nearest neighbor). Trénovací data jsou uložena a predikce nového prvku je počítána „průměrem“ nejbližších sousedů.

Metody IBL lze aplikovat na libovolné typy vstupních a výstupních dat (kvantitativní nebo kvalitativní). V případě klasifikace bývají IBL citlivé na počet prvků spadajících do jednotlivých tříd. Při použití většího počtu sousedů ke klasifikaci může početněji zastoupená třída přehlasovat třídu početně méně zastoupenou.

V této práci bude použita lokální metoda algoritmu IBL, kde se odvozuje predikce neznámého prvku na základě *k* nejbližších sousedů, přičemž *k* je menší než počet všech trénovacích prvků.

5.3 VÝHODY IBL

Autor [13] uvádí tyto výhody:

- Možnost **lokálně ovlivňovat přesnost** metody (např. uložení více záznamů nebo jejich znásobení).
- **Proces učení** je omezen pouze na ukládání záznamů – je tedy **velice rychlý**. Nejsou požadovány žádné významné apriorní znalosti o vztazích mezi proměnnými.
- **Inkrementální charakter** – protože nedochází ke generalizaci, je možné při získání nových dat tato jednoduše přidat.
- Předností IBL je také skutečnost, že místo odhadu **jedné funkce** popisující celý prostor řešení je vytvářen **nový jednoduchý model** pro každou nově klasifikovanou instanci. To umožňuje řešit i velice složité a komplexní problémy pomocí jednoduchých aproximací s lokální platností.

5.4 NEVÝHODY IBL

Autor [13] také uvádí i tyto nevýhody:

- Predikce výstupní hodnoty nového prvku může být **časově náročná**. Čím více trénovacích prvků je uloženo, tím náročnější je nalezení k nejbližších sousedů. Vzniká tak otázka, jak data co nejlépe uložit, aby bylo nalezení sousedů co nejrychlejší.
- Metody IBL jsou velice **citlivé na irelevantní atributy**. Pokud budou mezi 10 atributy pouze 3 relevantní a metodě bude předloženo všech 10 atributů, metoda bude predikovat velice nepřesně.
- V případě použití metody k nejbližších sousedů nedochází ke generalizaci (k té dochází až během klasifikace nové instance). Nevzniká obecný model, **znalost uložená v datech není pro člověka srozumitelná** a jednoduše interpretovatelná, i když metoda předikuje úspěšně.
- Zcela zásadní význam má **správná definice vzdálenosti** mezi dvěma prvky, což může být v případě kombinace více atributů různého typu obtížné.

5.5 METRIKA

Máme nový prvek *q* definovaný svými atributy a hledáme jeho výstupní hodnotu *y*. Podstatou IBL metody je nalezení určitého počtu prvků z trénovací databáze, které se nejvíce podobají prvku *q*. Na základě výstupních hodnot nalezených nejbližších sousedů dále usuzujeme, jaký bude nejpravděpodobnější výstup prvku nového. Vzniká otázka, jak určit, kteří sousedé jsou ti nejbližší?

Za účelem výběru nejbližších sousedů je nutné definovat vhodnou **metriku,** která vyjadřuje vzájemnou vzdálenost dvou libovolných prvků.

Dle literatury [13] je metrikou rozuměna funkce $d: \boldsymbol{X} \times \boldsymbol{X} \rightarrow R$ taková, že:

1. $x_1, x_2 \in \boldsymbol{X} : d(x_1, x_2) \geq 0$
2. $d(x_1, x_1) = 0$
3. $d(x_1, x_2) = d(x_2, x_1)$
4. $x_1, x_2, x_3 \in \boldsymbol{X} : d(x_1, x_2) + d(x_2, x_3) \geq d(x_1, x_3)$

5.6 NORMALIZACE

V případě používání mnoha vstupních atributů, které popisují různé veličiny v různých měřítkách, vzniká klasický problém, kdy jedna veličina je např. řádově ve stovkách, druhá v desetinách. Pro srovnávání vzdálenosti v rámci takových rozsahů veličin je nutné jednotlivé atributy normalizovat tak, aby byly hodnoty jejich metrik vzájemně srovnatelné.

V této práci je použita **normalizace rozsahem** (min-max normalizace), kde výsledný rozsah veličiny je v intervalu <0; 1>. Pro transformovanou veličinu *x* bude normalizovaná *i*-tá hodnota dána vztahem:

$$x'_i = \frac{x - \min(x_1 \ldots x_i)}{\max(x_1 \ldots x_i) - \min(x_1 \ldots x_i)} \qquad (1)$$

Funkce *min()* a *max()* vracejí hodnotu nejmenšího a největšího prvku transformované množiny.

5.7 CHYBOVÁ FUNKCE

Máme-li všechny trénovací data v množině D, pak chybová funkce E založená na výpočtu metodou nejmenších čtverců vyjadřuje celkovou chybu jako:

$$E = \frac{1}{2} \cdot \sum_{x \in D} (f(x) - \hat{f}(x))^2 \qquad (2)$$

Pro potřeby IBL je nutné omezit platnost chyby jen na nejbližších *k* sousedů, proto platí následující vztah:

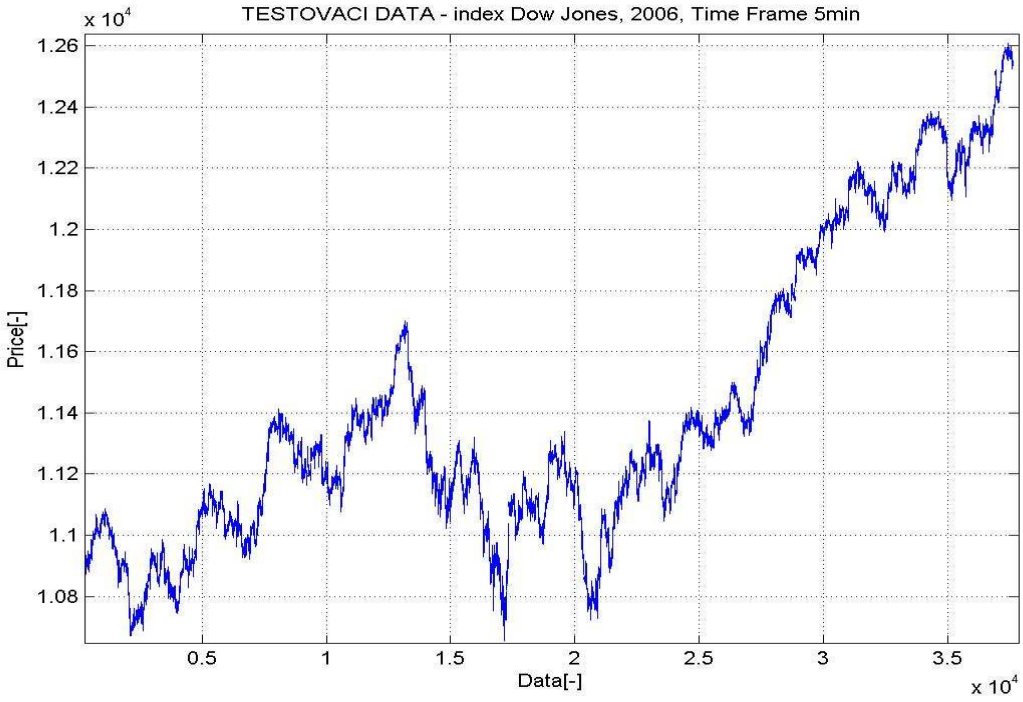
$$E(x_q) = \frac{1}{2} \cdot \sum_{i=1}^{k} (f(x) - \hat{f}(x))^2 \qquad (3)$$

6. PROGRAMOVÁ IMPLEMENTACE

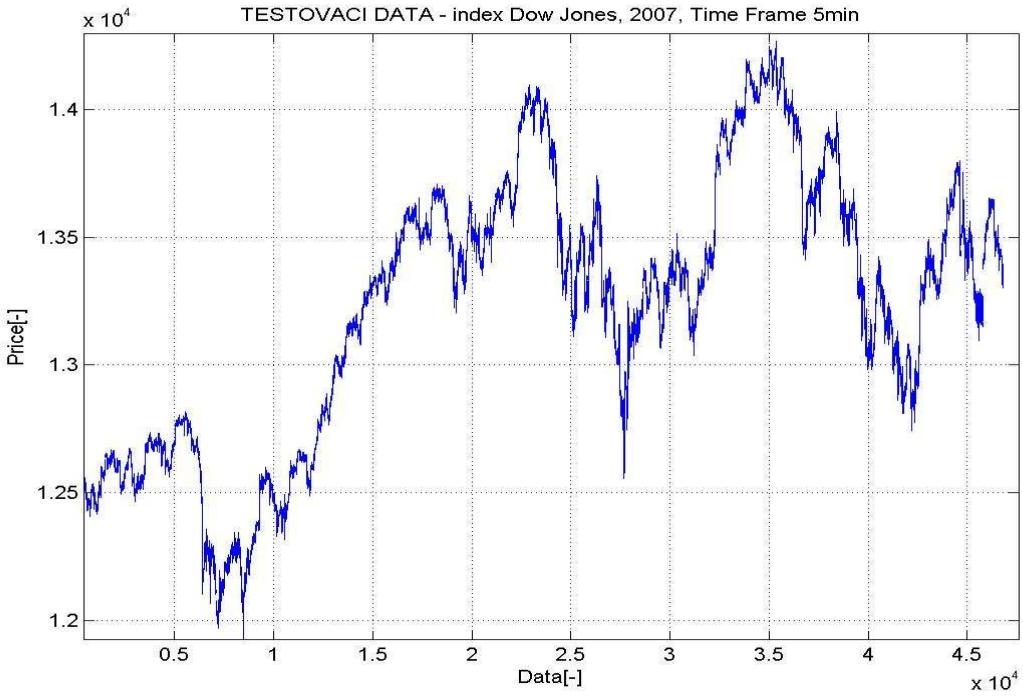
V prostředí programu MATLAB bylo úspěšně realizováno vyhledávání cenových vzorů *Korekce* pro nákup a prodej v dynamických datech. Na základě grafické a psychologické metody byla provedena analýza časové řady a následně pomoci metody učení založené na instancích byly vyhledávány cenové vzory. Jako testovací dynamická data byla použita 5-ti minutová reálná data z burzovního USA indexu Dow Jones za rok 2006, 2007 a 2008. Průběh této časové řady zobrazují *Obrázek 11*, *Obrázek 12*, a *Obrázek 13*. Jedná se o dynamickou časovou řadu obsahující dohromady přesně 142 423 řádků. Každý řádek je ve formátu:

Date,Time,Open,High,Low,Close,Volume. Ukázka hodnot naší časové řady ze dne 7. února 2008 při otevírání obchodní seance v 15:30 našeho času:

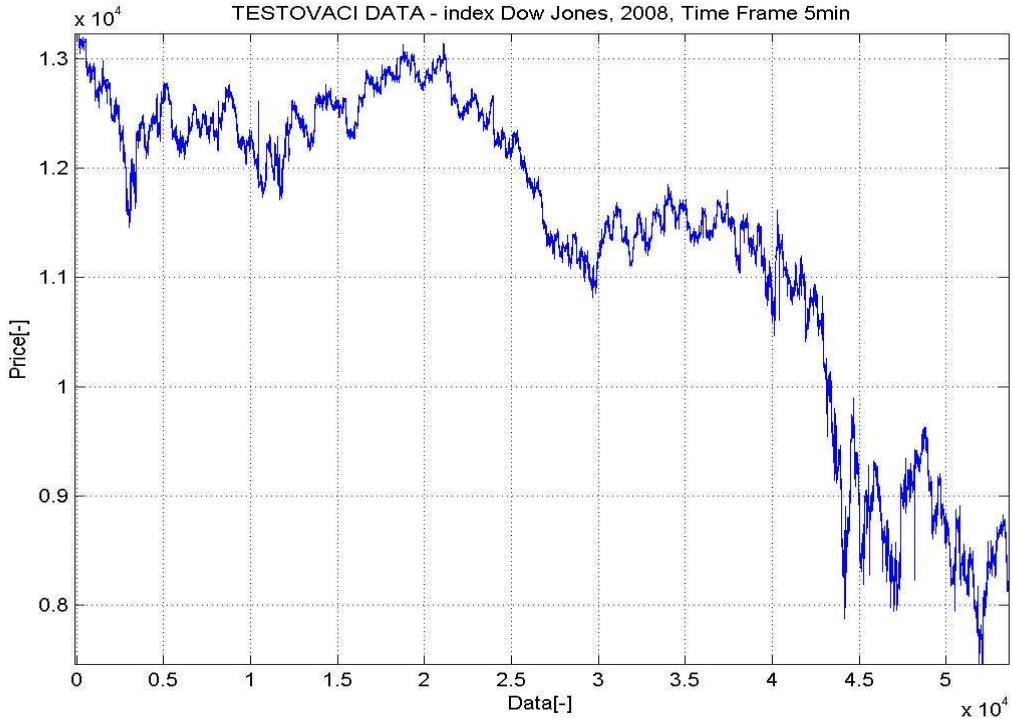
20080207,1530,12148.00,12160.00,12145.00,12160.00,1352
20080207,1535,12159.00,12184.00,12154.00,12181.00,8064
20080207,1540,12180.00,12209.00,12179.00,12193.00,9752



Obrázek 11: Testovací data pro vyhledávání cenových vzorů, trh YM, 2006



Obrázek 12: Testovací data pro vyhledávání cenových vzorů, trh YM, 2007



Obrázek 13: Testovací data pro vyhledávání cenových vzorů, trh YM, 2008

V programu MATLAB byl vytvořen algoritmus pro vyhledávání vybraných cenových vzorů. Pomocí metody IBL byl v datech vybrán vždy pattern, který se co nejvíce podobal ideálnímu průběhu. Jednotlivé kroky programu řeší následující problematiku:

1. Definice vyhledávaného cenového vzoru.
2. Načtení všech hodnot celé časové řady.
3. Postupné načítání jednotlivých bloků časové řady (o stejné délce jako vyhledávaný vzor).
4. Normalizace a další úpravy načteného bloku.
5. Analýza načteného bloku.
6. Aplikace metody IBL.
7. Porovnání podobnosti načteného bloku s definovaným vzorem.
8. Výpočet chybové funkce.
9. Selekce nalezených vzorů.
10. Vykreslení nalezeného vzoru a definovaného vzoru.
11. Načtení následujícího průběhu (pro nalezený vzor) časové řady.
12. Určení úspěšnosti nalezeného vzoru (zisk nebo ztráta).
13. Vykreslení následujícího průběhu u nalezeného vzoru (pro zisk i ztrátu).

Vyhledávání vzorů probíhalo pouze během regulérní obchodní seance, tedy vždy od 15:30 do 22:00 našeho času. Během této doby bývá největší likvidita a aktivita. Ačkoli většina elektronických trhů má nonstop obchodování, nemá smysl vyhledávat vzory během noční obchodní seance, kdy je téměř nulové obchodování a bývají špatné exekuce obchodních příkazů (tzv. „skluzy“ v plnění většího objemu obchodních příkazů.) Tedy v algoritmu byl zaveden časový filtr pro ignoraci noční obchodní seance.

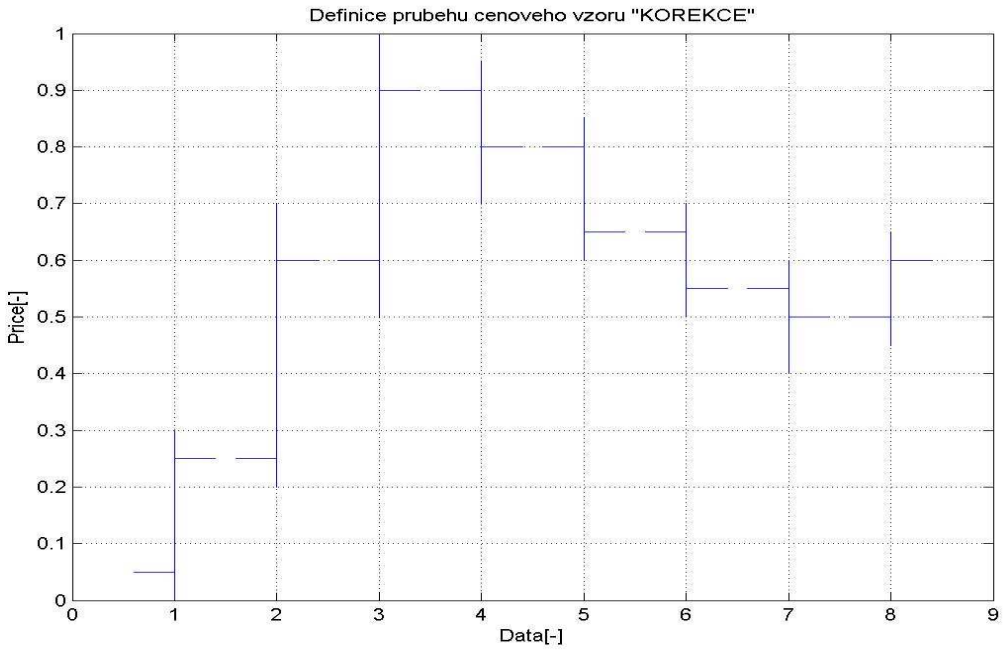
6.1 CENOVÝ VZOR *KOREKCE* PRO NÁKUP

6.1.1 Definice vzoru

Definice našeho cenového vzoru *Korekce* v programu MATLAB musí být zvlášť pro hodnoty Open, Hight, Low a Close ve formě normalizovaných vektorů:

```
K_O = [ 0.05 0.25 0.6 0.9  0.8  0.65 0.55 0.5];
K_H = [ 0.3  0.7  1   0.95 0.85 0.7  0.6  0.65];
K_L = [ 0    0.2  0.5 0.7  0.6  0.5  0.4  0.45];
K_C = [ 0.25 0.6  0.9 0.8  0.65 0.55 0.5  0.6];
```

Grafické zobrazení definovaného vzoru je na *Obrázku 14*, kde se jedná o velice ideální průběh ceny. Protože v trzích každý okamžik je naprosto unikátní, v reálném grafu přesně takový průběh objevíme zřídka. Zde nám jde zejména o nalezení velmi podobných průběhů, kde hodnoty High a Low budou přibližně odpovídat definovanému průběhu. Hodnoty Open a Close jsou v tomto případě méně podstatné. Samozřejmě existuje mnoho typů průběhů cenových korekcí a mnoho jejich různých nuancí. V této práci se zaměříme jen na jeden typ korekce, jaký vidíme právě na *Obrázku 14.*



Obrázek 14: Definice průběhu cenového vzoru *Korekce*

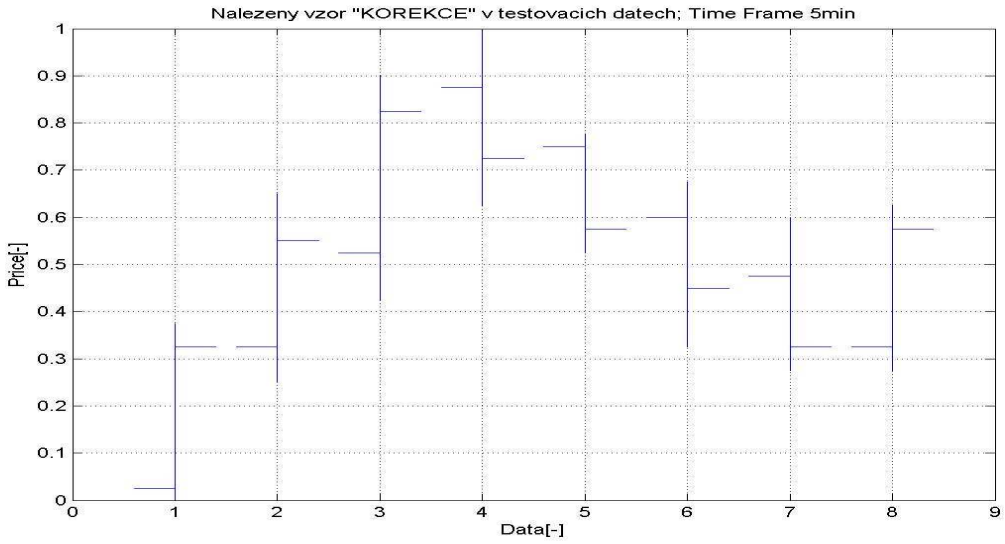
6.1.2 Nalezené vzory

Celkový počet programem nalezených cenových vzorů *Korekce* pro nákup v testovacích datech pro jednotlivé roky ukazuje *Tabulka 1*.

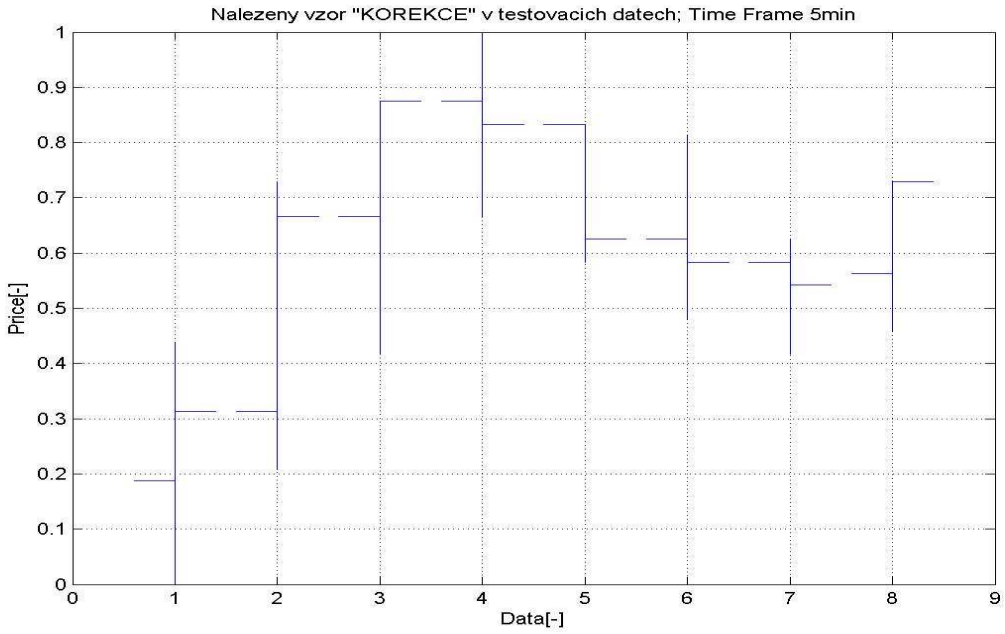
Rok	Celkem [Počet]
2006	126
2007	97
2008	75

Tabulka 1: Počet nalezených vzorů *Korekce* - nákup

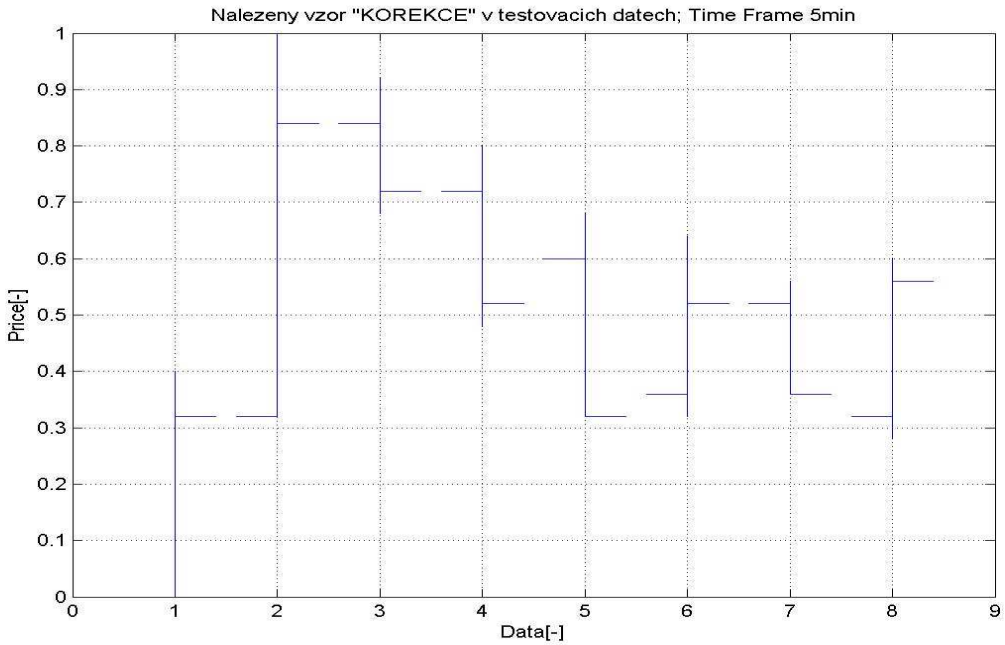
Vytvořený algoritmus v programu MATLAB také jednotlivě vykreslí každý nalezený cenový vzor, a to v normalizovaném úsečkovém grafu. Příklady nalezených vzorů zobrazují *Obrázek 15* až *Obrázek 18*. Je možné vidět různé nuance definovaného cenového vzoru.



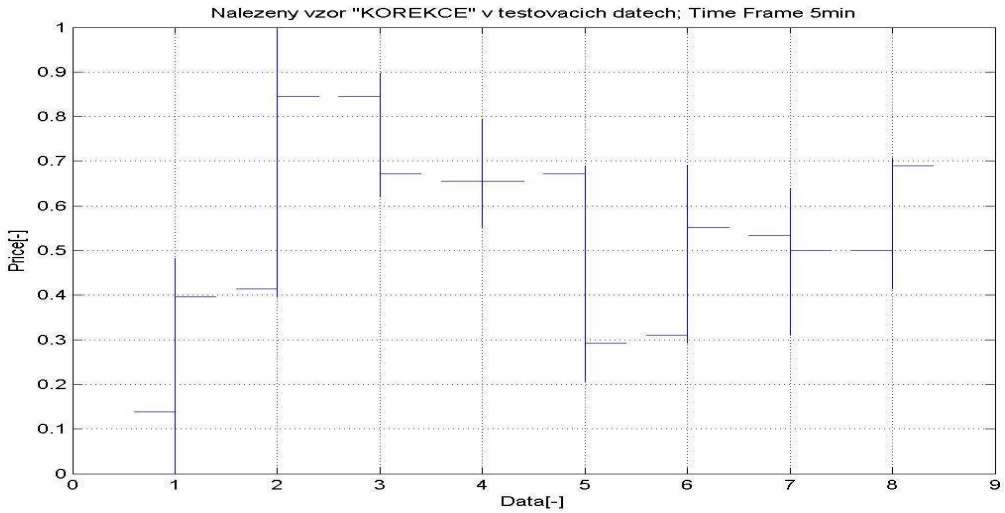
Obrázek 15: Nalezený vzor *Korekce* v testovacích datech



Obrázek 16: Nalezený vzor *Korekce* v testovacích datech



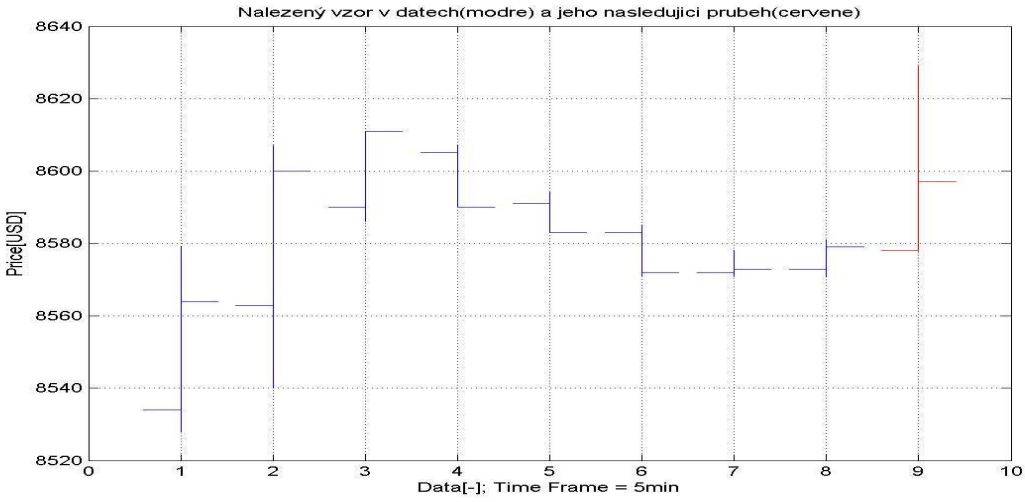
Obrázek 17: Nalezený vzor *Korekce* v testovacích datech



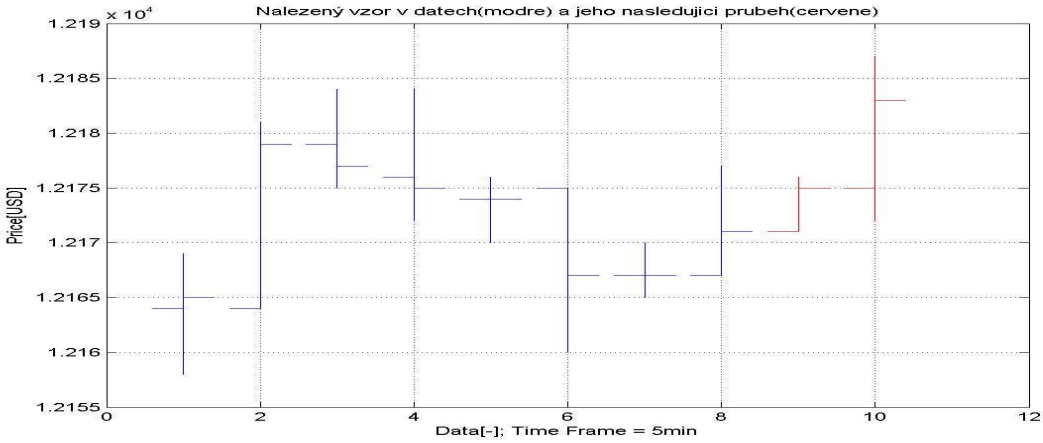
Obrázek 18: Nalezený vzor *Korekce* v testovacích datech

6.1.3 Následující průběh u nalezených vzorů

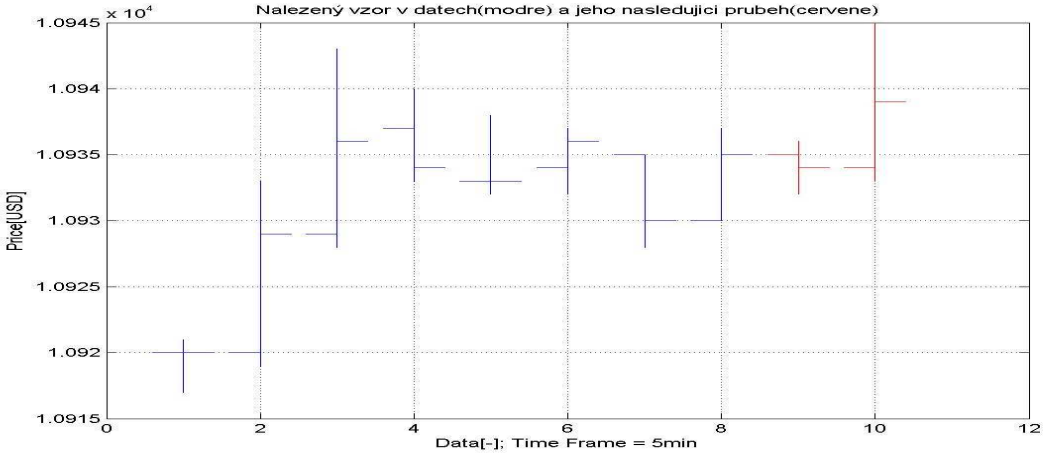
Dalším výstupem programu je vykreslení následujícího průběhu ceny u každého nalezeného vzoru. Vykreslení probíhá na nenormalizovaném úsečkovém grafu, viz *Obrázek 19* až *Obrázek 24*, kde průběh ceny nalezeného vzoru zobrazuje modrá barva, a následující průběh ceny znázorňuje barva červená.



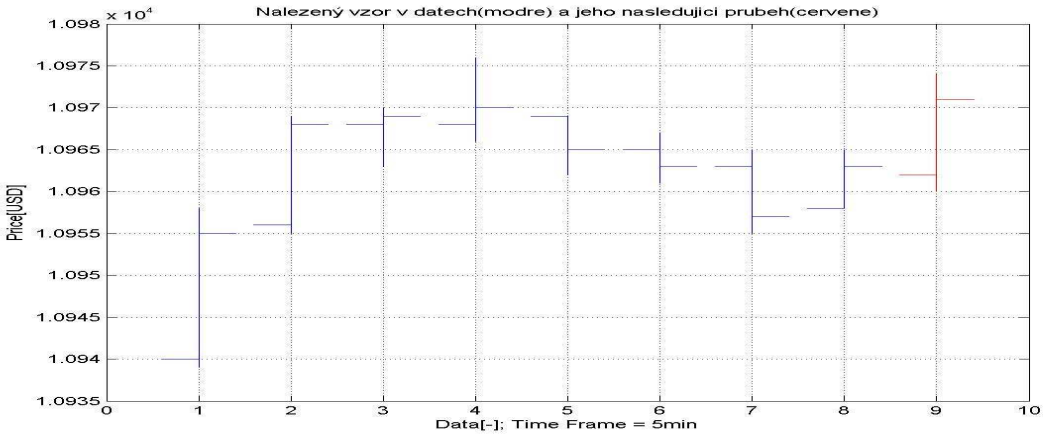
Obrázek 19: Následující průběh (červeně) u nalezeného vzoru



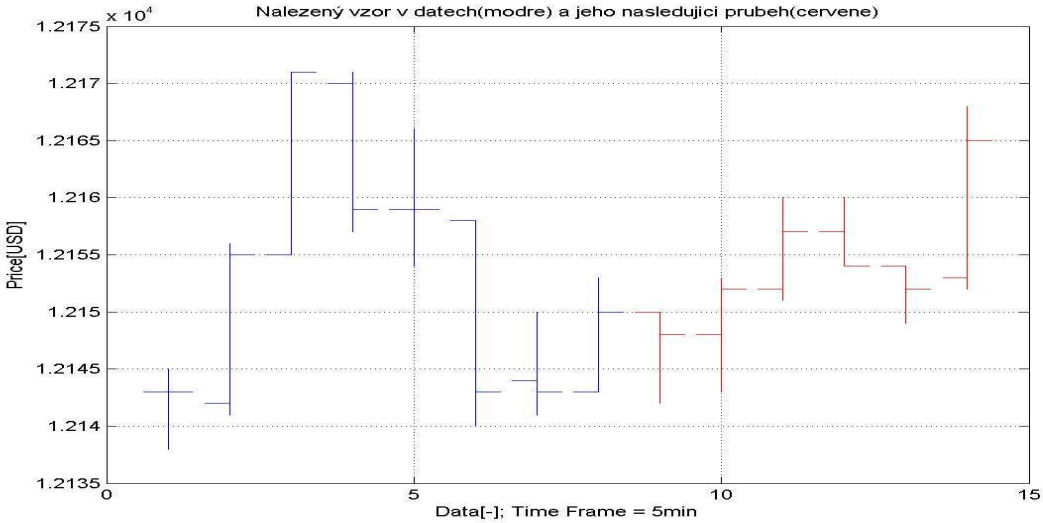
Obrázek 20: Následující průběh (červeně) u nalezeného vzoru



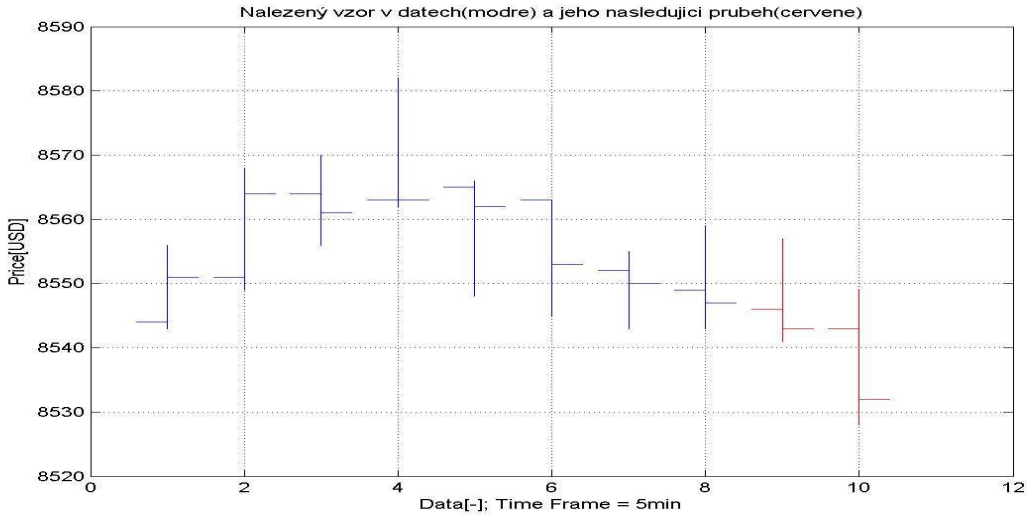
Obrázek 21: Následující průběh (červeně) u nalezeného vzoru



Obrázek 22: Následující průběh (červeně) u nalezeného vzoru



Obrázek 23: Následující průběh (červeně) u nalezeného vzoru



Obrázek 24: Následující průběh (červeně) u nalezeného vzoru

6.1.4 Úspěšnost predikce vzoru

Další část programu spočívá ve vyhodnocení úspěšnosti nalezeného cenového vzoru. Pro zjednodušení byly zvoleny dvě proměnné:

- ***PT* pro výstup z pozice se ziskem na úrovni 0.9 v normalizovaném grafu,**
- ***SL* pro výstup z pozice se ztrátou na úrovni menší než minimum korekce.**

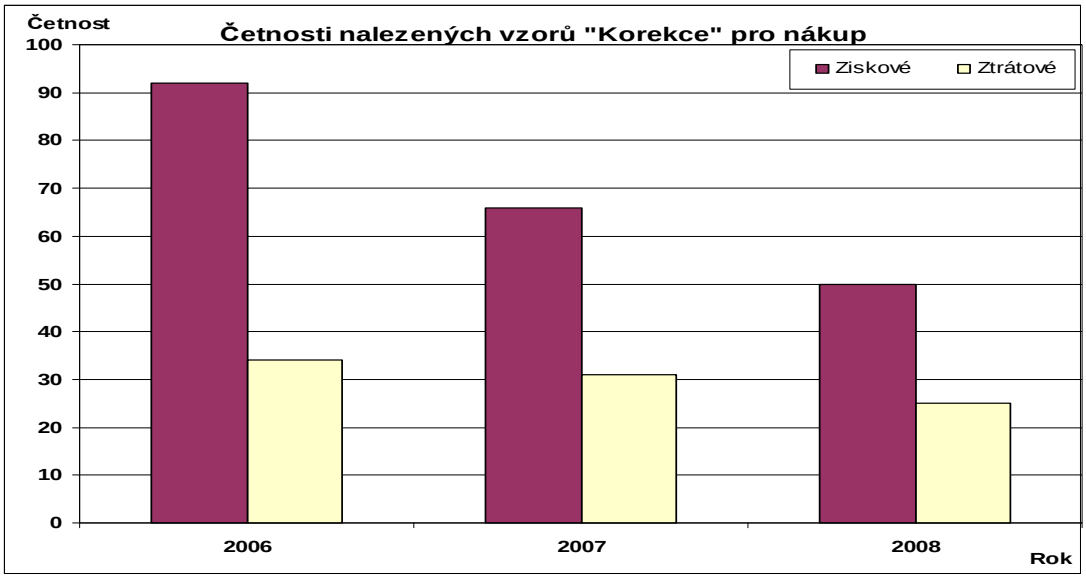
Tedy za ziskový obchod byl považován ten, ve kterém následující průběh ceny protkl hranici volitelné proměnné *PT (Profit Target)*, která symbolizuje výstup se ziskem z otevřené pozice. V tomto velmi zjednodušeném případě se jedná o výstup s malým ziskem, protože proměnná *PT* je nastavená na hodnotu 0.9. Ovšem pokud cena útočí na lokální maximum (hodnota 1 v normalizovaném grafu), pak je vysoká pravděpodobnost vytvoření nového lokálního maxima. Samozřejmě existují dokonalejší metody výstupu se ziskem, které ovládá s pomocí vycvičeného citu samotný obchodník. Jelikož naprogramovat cit není možné, proto je volen tento zjednodušený výstup.

Naopak ztrátový obchod je takový, kdy nedošlo k testu lokálního maxima a cena pokračovala dále v korekci, nebo také došlo k otočení trendu. Výstup se ztrátou je v okolí lokálního minima. Zkušený obchodník opět s pomocí citu a objektivity umí ukončit ještě dříve ztrátový obchod, nebo dokonce v některých případech vystoupit s malým ziskem na pokrytí nákladů.

Tabulka 2 a *Obrázek 25* udávají četnosti ziskových a ztrátových hypotetických obchodů, které byly uskutečněny při nalezení vzoru.

Rok	Celkem	Ziskové		Ztrátové	
	[Počet]	[Počet]	[%]	[Počet]	[%]
2006	126	92	73	34	27
2007	97	66	68	31	32
2008	75	50	67	25	33

Tabulka 2: Četnosti nalezených vzorů *Korekce* – pozice long



Obrázek 25: Četnosti nalezených ziskových a ztrátových vzorů *Korekce* - nákup

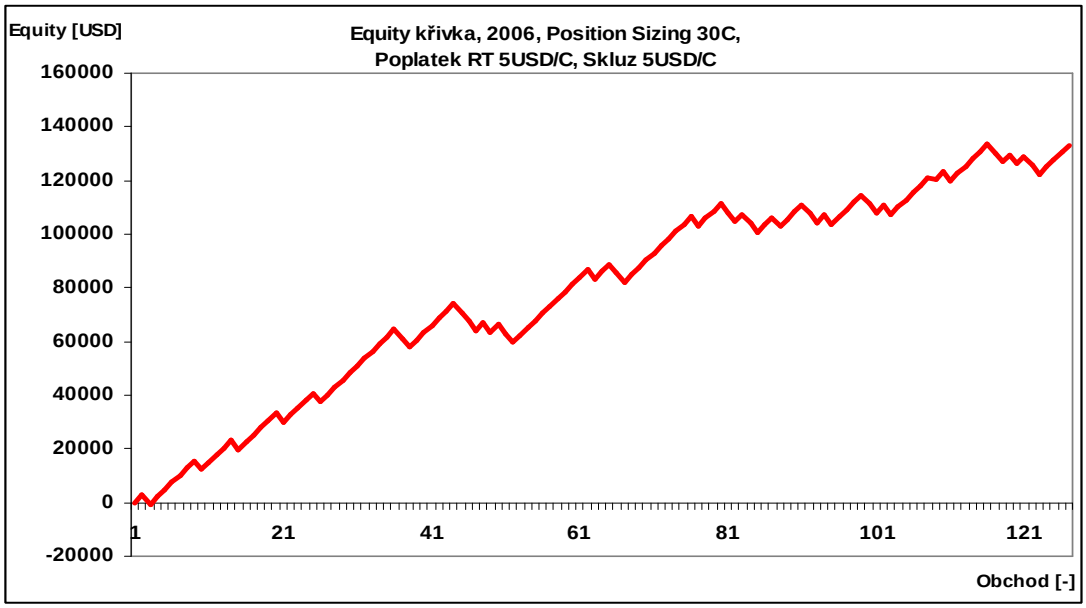
6.1.5 Equity křivka

Equity křivka graficky vyjadřuje úspěšnost obchodního systému. Byl vytvořen zjednodušený automatický obchodní systém, který definuje vstupy a výstupy. Jednotlivé obchody generují zisky a ztráty a dle jejich distribuce se vyvíjí equity křivka.

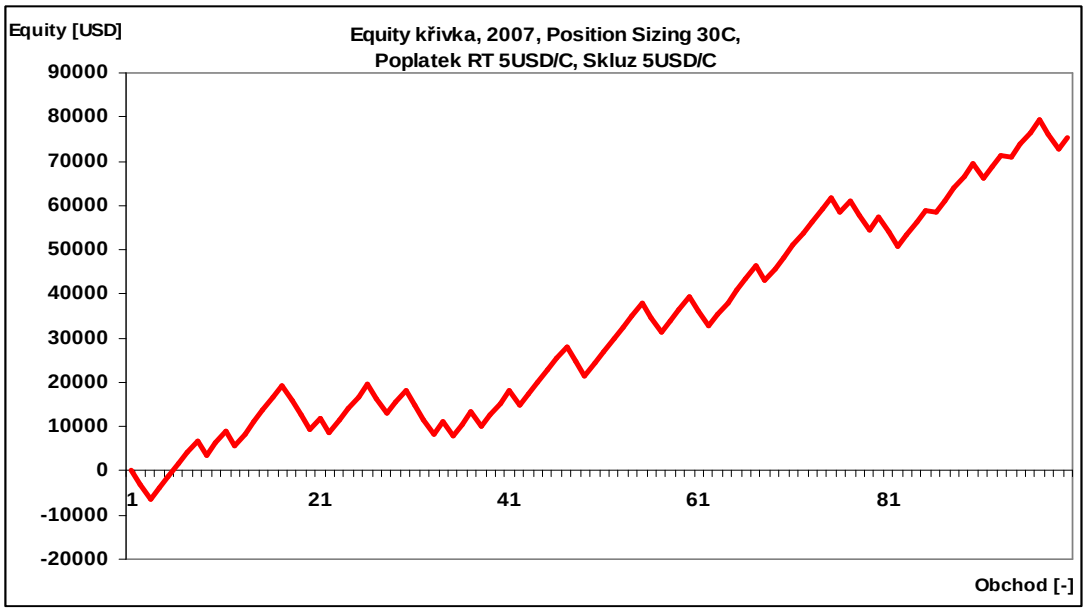
Jako nejvhodnější byly zvoleny tyto parametry (myšleno pro řešení programu, tedy ne jako nejefektivnější parametry):

1. Výstup s maximální ztrátou na 1 kontrakt = 110 USD (100 USD + poplatky 5 USD + skluz 5 USD).
2. Výstup se ziskem na 1 kontrakt = 95 USD (100 USD – poplatky 5 USD).
3. Risk Management = risk maximálně 3% kapitálu na 1 obchod.
4. Money Management = 4 000 USD na 1 kontrakt.
5. Position Sizing = konst = 30 kontraktů (likvidita trhu YM toto množství kontraktů uspokojuje).
6. Počáteční obchodní kapitál = 150 000 USD.

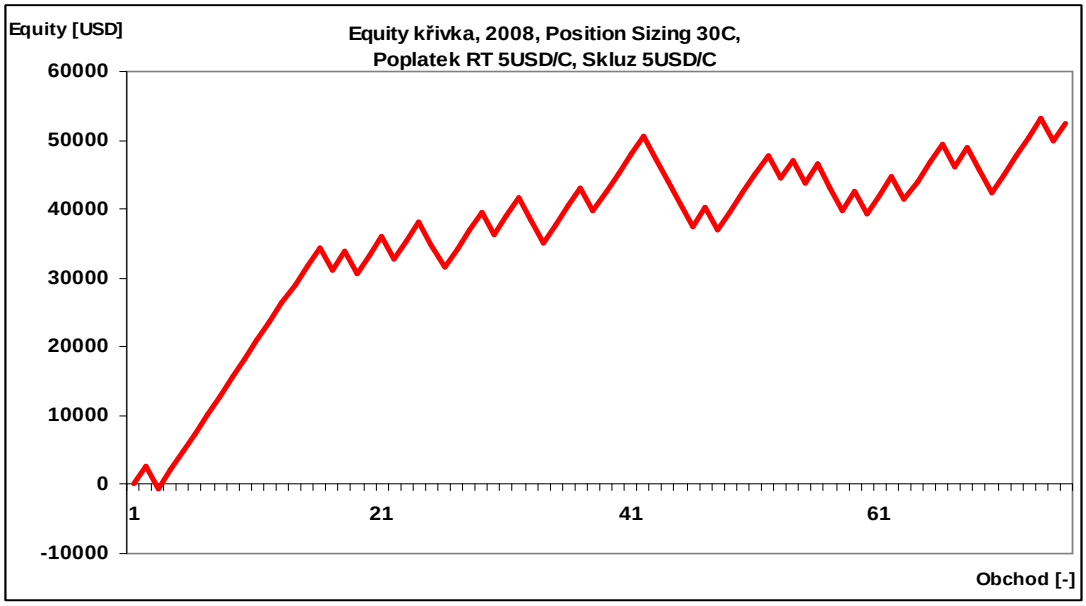
Pro tyto parametry bude výsledná equity křivka vypadat dle *Obrázek 26, Obrázek 27* a *Obrázek 28.*



Obrázek 26: Equity křivka obchodů, pozice long, rok 2006



Obrázek 27: Equity křivka obchodů, pozice long, rok 2007



Obrázek 28: Equity křivka obchodů, pozice long, rok 2008

Tabulka 3 udává ziskovost nalezených vzorů během jednotlivých roků.

Rok	Zisk	
	[USD]	[%]
2006	133200	88,8
2007	75300	50,2
2008	52500	35

Tabulka 3: Ziskovost nalezených vzorů – pozice long

6.2 CENOVÝ VZOR *KOREKCE* PRO PRODEJ

6.2.1 Definice vzoru

Definice našeho cenového vzoru *Korekce* v programu MATLAB musí být zvlášť pro hodnoty Open, Hight, Low a Close ve formě normalizovaných vektorů:

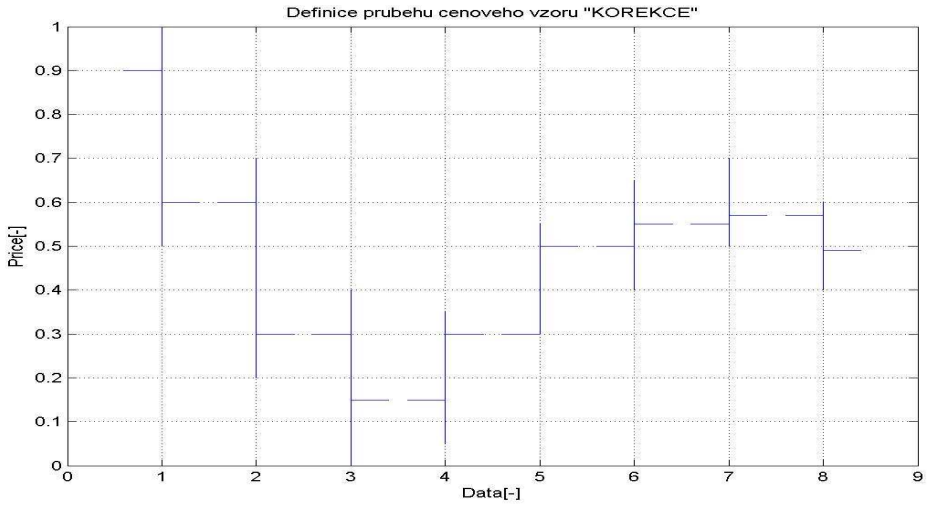
K_O = [0.9 0.6 0.3 0.15 0.3 0.5 0.55 0.57];

K_H = [1 0.7 0.4 0.35 0.55 0.65 0.7 0.6];

K_L = [0.5 0.2 0 0.05 0.3 0.4 0.5 0.4];

K_C = [0.6 0.3 0.15 0.3 0.5 0.55 0.57 0.49];

Grafické zobrazení definovaného vzoru je na *Obrázku 29.*



Obrázek 29: Definice průběhu cenového vzoru *Korekce*

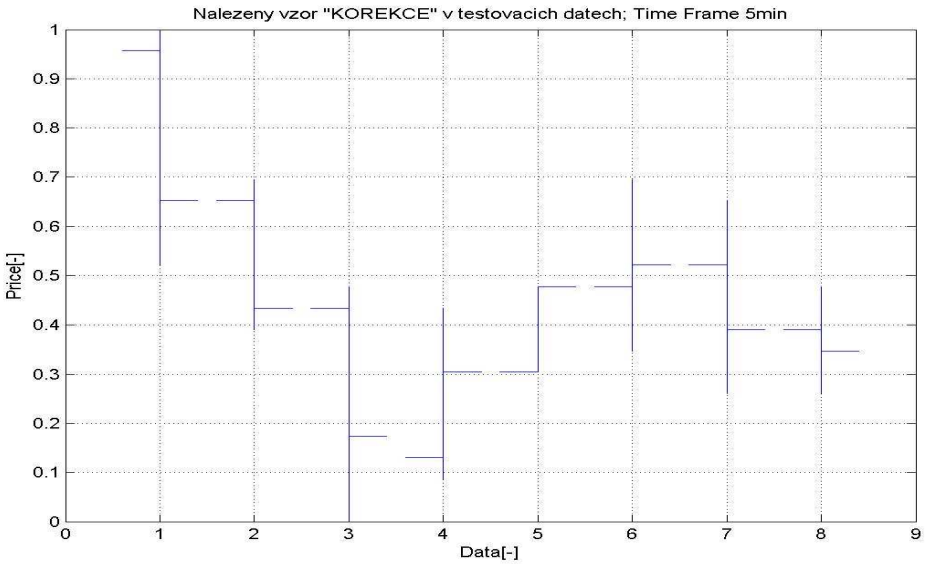
6.2.2 Nalezené vzory

Celkový počet programem nalezených cenových vzorů *Korekce* pro prodej (otevření „short“ pozice) v testovacích datech pro jednotlivé roky ukazuje *Tabulka 4.*

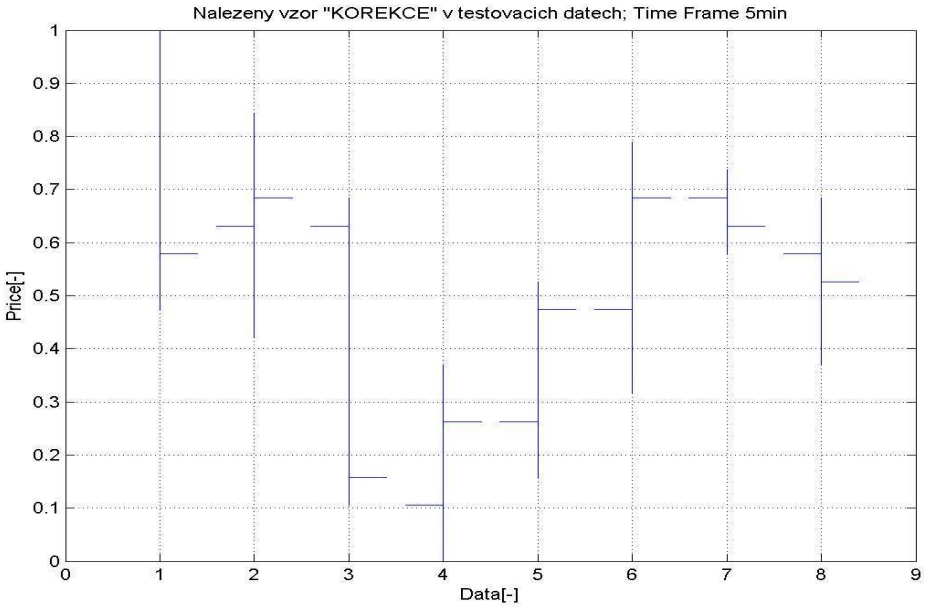
Rok	Celkem [Počet]
2006	216
2007	249
2008	197

Tabulka 4: Počet nalezených vzorů *Korekce* - prodej

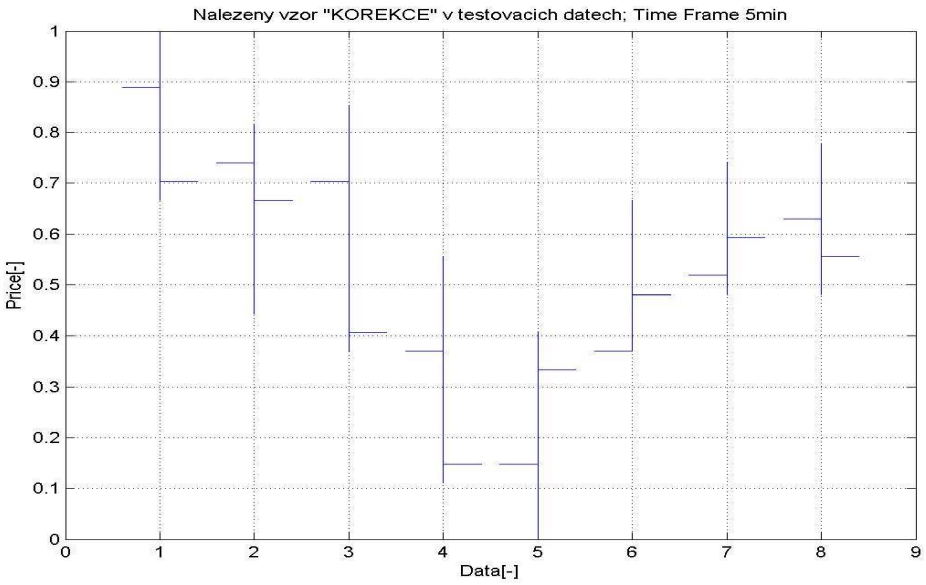
Vytvořený algoritmus v programu MATLAB také jednotlivě vykreslí každý nalezený cenový vzor, a to v normalizovaném úsečkovém grafu. Příklady nalezených vzorů zobrazují *Obrázek 30* až *Obrázek 33*. Je možné vidět různé nuance definovaného cenového vzoru.



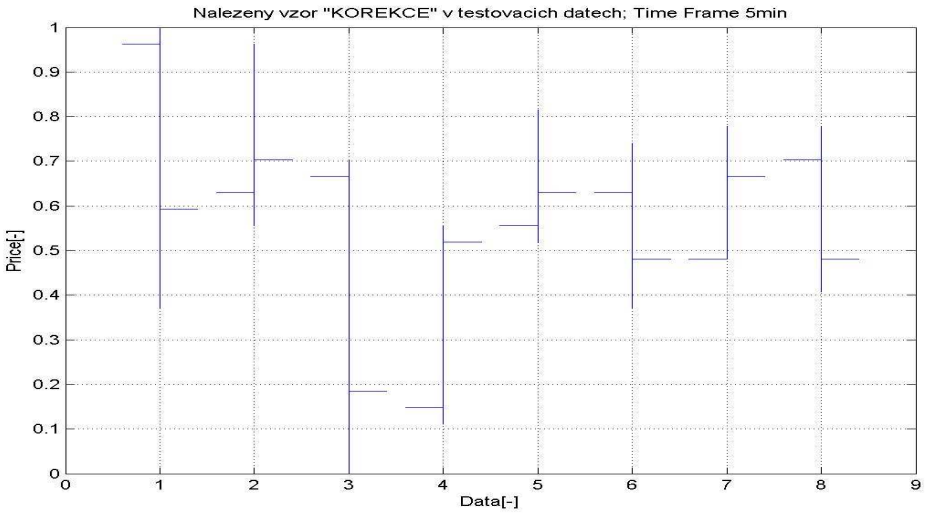
Obrázek 30: Nalezený vzor *Korekce* v testovacích datech



Obrázek 31: Nalezený vzor *Korekce* v testovacích datech



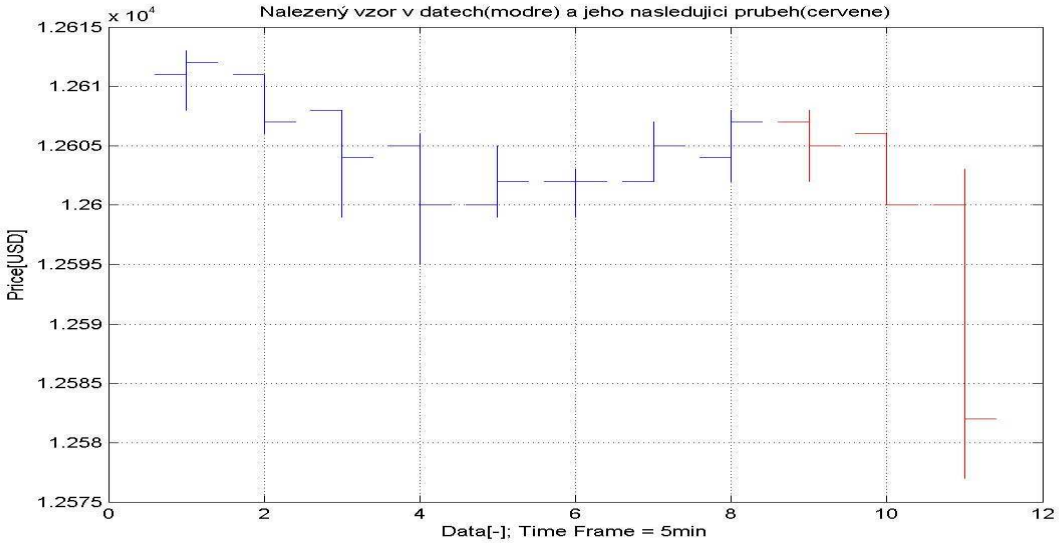
Obrázek 32: Nalezený vzor *Korekce* v testovacích datech



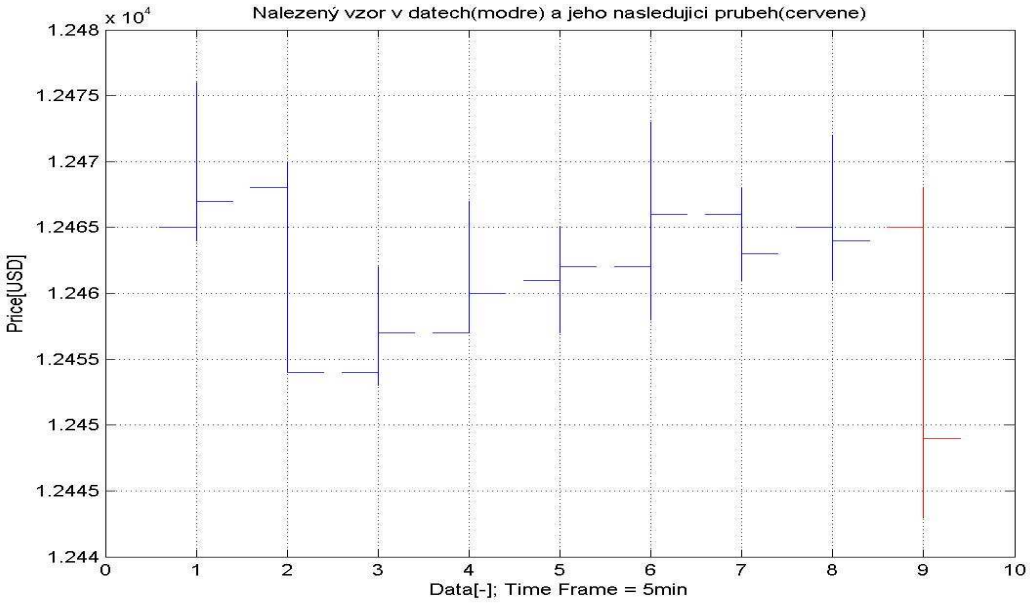
Obrázek 33: Nalezený vzor *Korekce* v testovacích datech

6.2.3 Následující průběh u nalezených vzorů

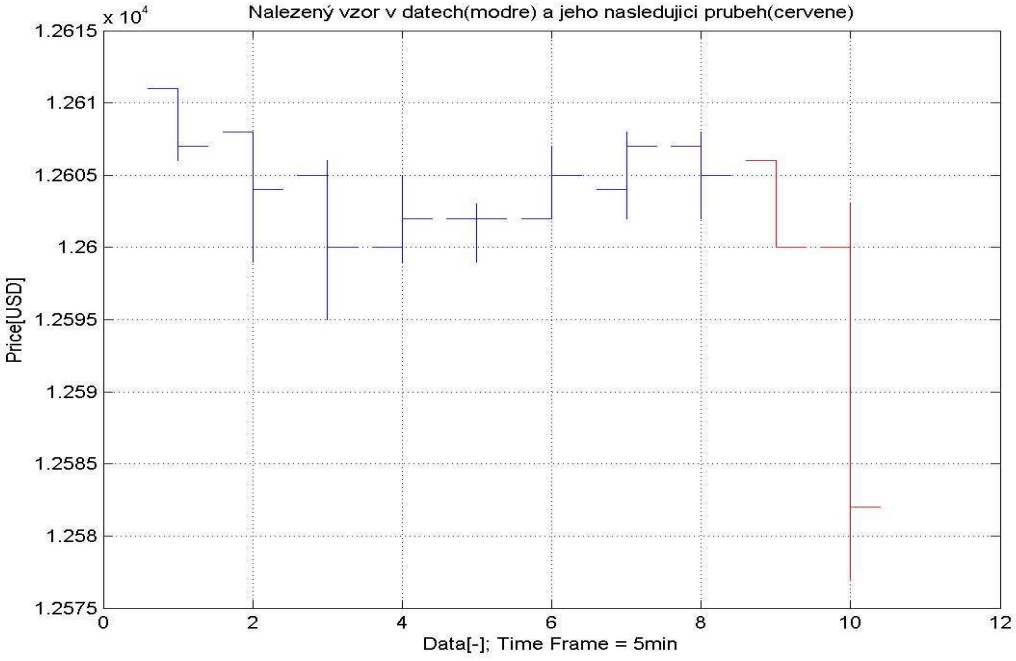
Dalším výstupem programu je vykreslení následujícího průběhu ceny u každého nalezeného vzoru. Vykreslení probíhá na nenormalizovaném úsečkovém grafu, viz *Obrázek 34* až *Obrázek 37*, kde průběh ceny nalezeného vzoru zobrazuje modrá barva, a následující průběh ceny znázorňuje barva červená.



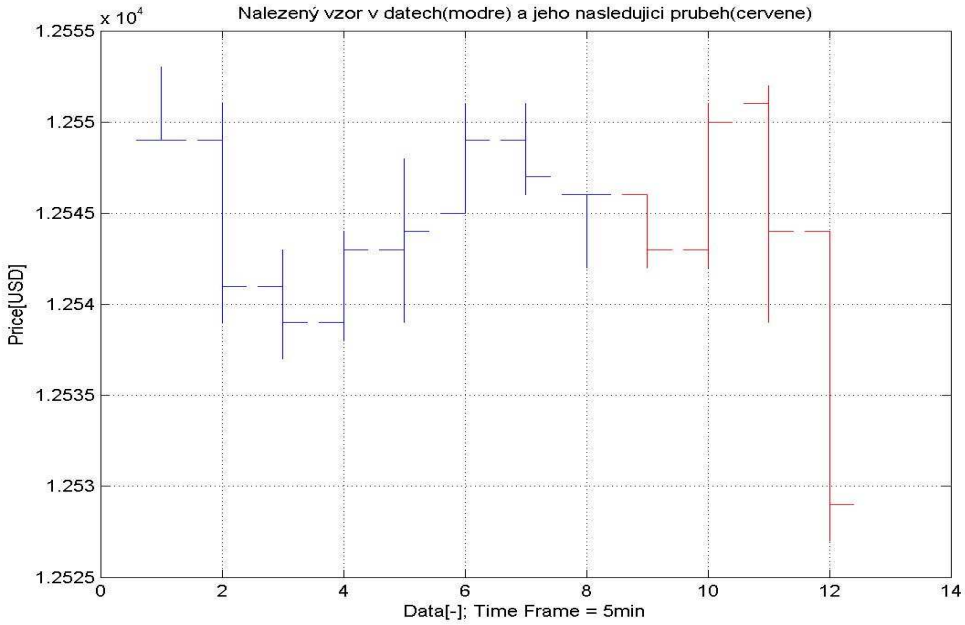
Obrázek 34: Následující průběh (červeně) u nalezeného vzoru



Obrázek 35: Následující průběh (červeně) u nalezeného vzoru



Obrázek 36: Následující průběh (červeně) u nalezeného vzoru



Obrázek 37: Následující průběh (červeně) u nalezeného vzoru

6.2.4 Úspěšnost predikce vzoru

Další část programu spočívá ve vyhodnocení úspěšnosti nalezeného cenového vzoru. Pro zjednodušení byly zvoleny dvě proměnné:

- ***PT* pro výstup z pozice se ziskem na úrovni 0.1 v normalizovaném grafu,**
- ***SL* pro výstup z pozice se ztrátou na úrovni větší než maximum korekce.**

Tabulka 5 a *Obrázek 38* udávají četnosti ziskových a ztrátových hypotetických obchodů, které byly uskutečněny při nalezení vzoru.

Rok	Celkem	Ziskové		Ztrátové	
	[Počet]	[Počet]	[%]	[Počet]	[%]
2006	216	133	62	83	38
2007	249	152	61	97	39
2008	197	107	54	90	46

Tabulka 5: Četnosti nalezených vzorů *Korekce* – pozice short

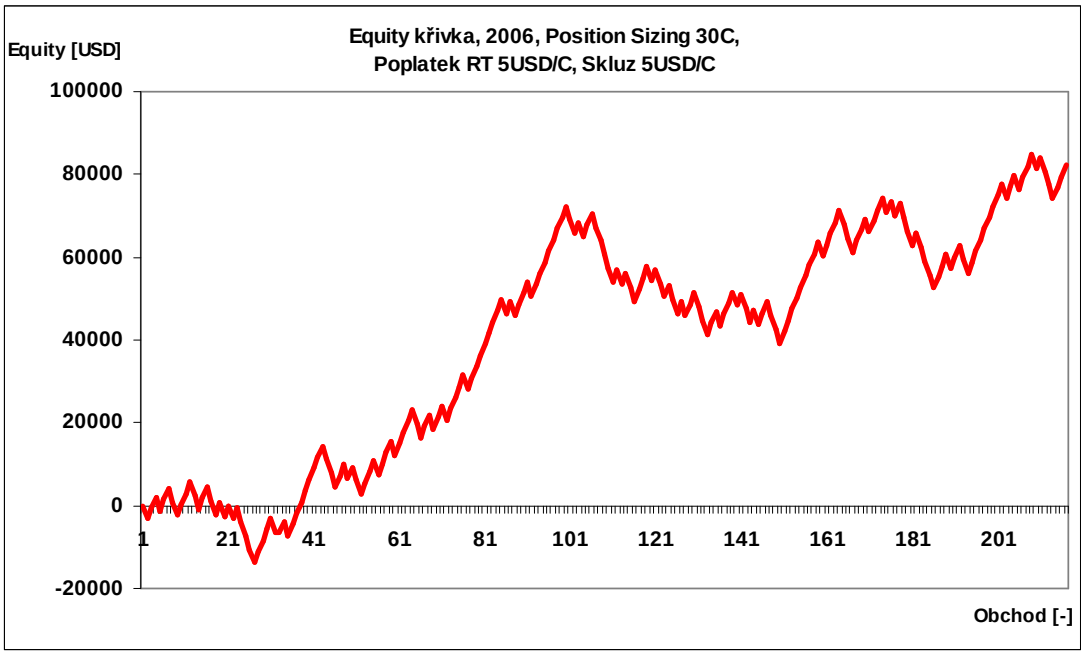
Obrázek 38: Četnosti nalezených ziskových a ztrátových vzorů *Korekce* - prodej

6.2.5 Equity křivka

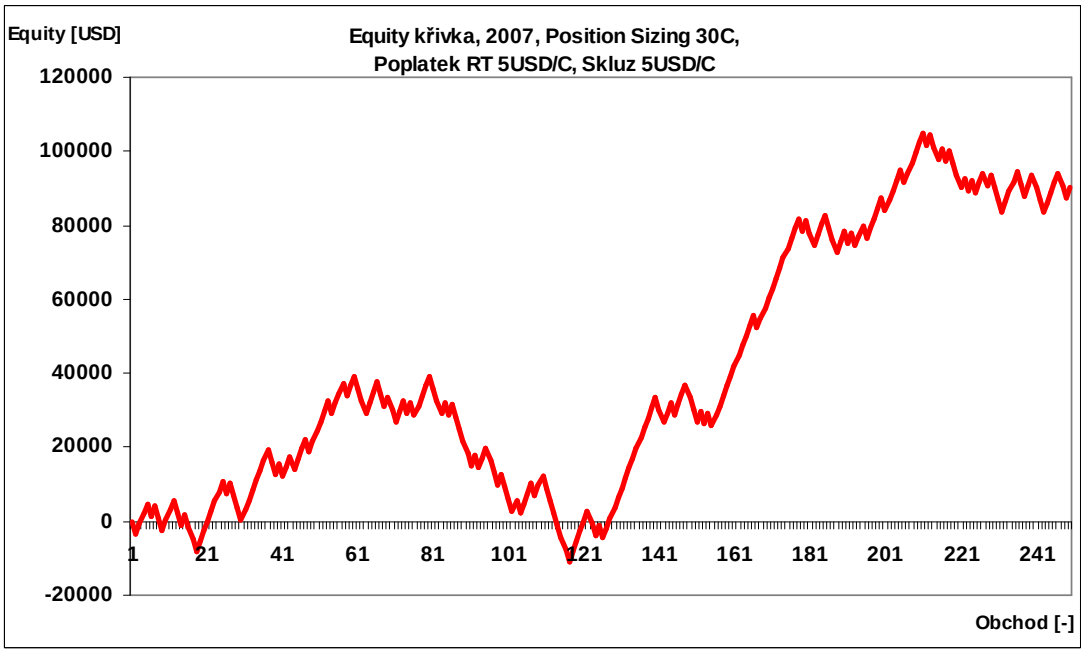
Podobně jako u vzoru pro nákup, budou platit podobná pravidla automatického obchodního systému:

1. Výstup s maximální ztrátou na 1 kontrakt = 110 USD (100 USD + poplatky 5 USD + skluz 5 USD).
2. Výstup se ziskem na 1 kontrakt = 95 USD (100 USD – poplatky 5 USD).
3. Risk Management = risk maximálně 3% kapitálu na 1 obchod.
4. Money Management = 4 000 USD na 1 kontrakt.
5. Position Sizing = konst = 30 kontraktů (likvidita trhu YM toto množství kontraktů uspokojuje).
6. Počáteční obchodní kapitál = 150 000 USD.

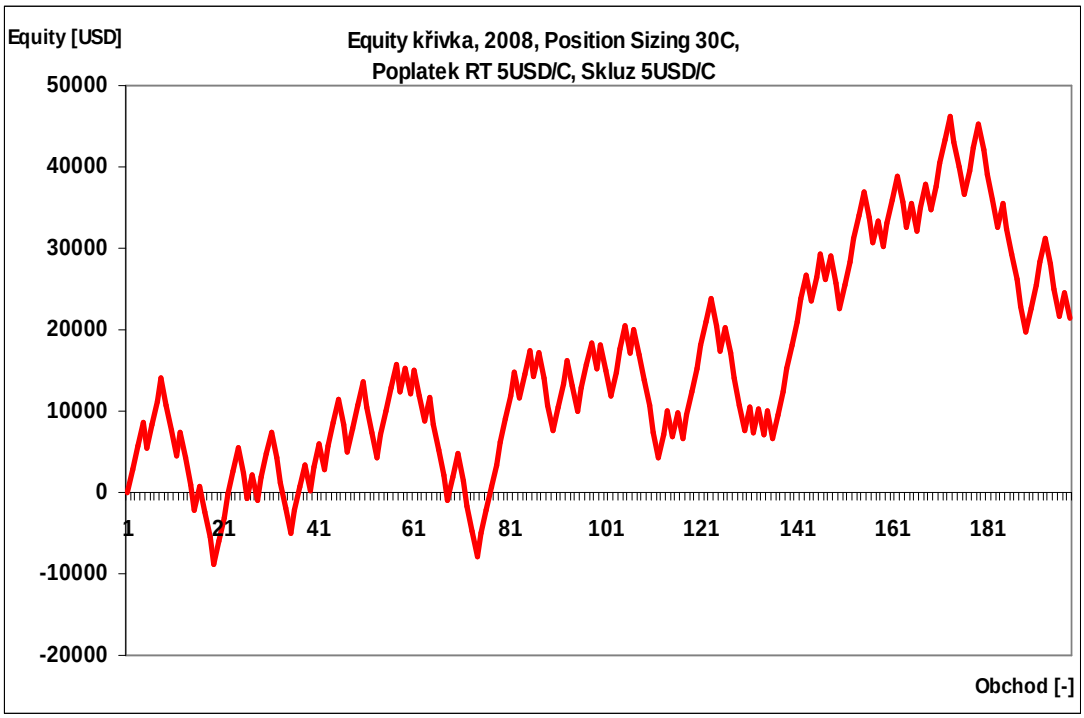
Pro tyto parametry bude výsledná equity křivka vypadat dle *Obrázek 39*, *Obrázek 40* a *Obrázek 41*.



Obrázek 39: Equity křivka obchodů, pozice short, rok 2006



Obrázek 40: Equity křivka obchodů, pozice short, rok 2007



Obrázek 41: Equity křivka obchodů, pozice short, rok 2008

Tabulka 6 udává ziskovost nalezených vzorů během jednotlivých roků.

Rok	Zisk	
	[USD]	[%]
2006	82200	54,8
2007	90300	60,2
2008	21450	14,3

Tabulka 6: Ziskovost nalezených vzorů – pozice short

7. ZÁVĚR

V této diplomové práci bylo hlavním cílem provést rešerši v oblasti popisu a analýzy časově proměnných dat a vyhledávání předem vybraných vzorů v časových řadách. Časové řady byly probrány ve druhé kapitole. Základním úkolem analýzy časových řad je snaha porozumět principům, na základě kterých se generují hodnoty časové řady. Tedy pokud se podaří tento mechanismus odhalit, je možné předpovídat budoucí vývoj analyzované časové řady a v některých případech je dokonce možné řídit a optimalizovat její vývoj.

Následující kapitola *Technická analýza* se specializuje na analýzu specifického druhu časových řad, a to na burzovní grafy. Technická analýza je aplikována na všechny finanční produkty za účelem predikce budoucích cenových pohybů. Predikce probíhá na základě analyzování a vyhodnocování minulých a aktuálních burzovních dat. V této kapitole byla podrobněji uvedena velmi účinná metoda *Pure Price Action*, která se používá pro vyhledávání cenových vzorů a k efektivní predikci cenového vývoje, jak ukazuje praktický příklad z *Kapitoly 3.5*.

Problematika rozeznávání vzorů je uvedena v samostatné kapitole. *Pattern Recognition* tedy usiluje o klasifikaci dat na základě statistických informací získaných ze vzorů nebo také na základě apriorní znalosti. Cílem vytvořeného algoritmu pro vyhledávání vzorů je tedy pomocí trénování přesně klasifikovat vzor a dále nalézt a určit správné rozhodovací meze. Dalším důležitým požadavkem tohoto algoritmu je minimalizovat nákladovou funkci pomocí upravování parametrů systému.

Kapitola *Učení založené na instancích* představila velmi efektivní metodu, která se používá pro vyhledávání vzorů. Tato metoda byla použita v praktické části *Programová implementace* při vyhledávání cenových vzorů v burzovních datech.

Úspěšně vytvořený program v prostředí MATLAB s vysokou přesností vyhledává burzovní vzory *Korekce* pro nákup i prodej. Na testovacích datech USA indexu *Dow Jones* algoritmus dokázal sám vyhledal 298 výskytů tohoto vzoru pro nákup a 662 vzorů pro prodej daného titulu během roků 2006, 2007 a 2008.

Nalezené cenové vzory se podobaly definovanému vzoru, jak ukazují *Obrázek 15* až *Obrázek 18* pro nákup a *Obrázek 30* až *Obrázek 33* pro prodej.

Dosažená přesnost predikce u nalezených cenových vzorů je graficky vyjádřena pomocí *Equity křivek* na *Obrázek 26* až *Obrázek 28* pro nákup a *Obrázek 39* až *Obrázek 41* pro prodej. *Equity křivka* graficky vyjadřuje úspěšnost obchodního systému, který byl navrhnut na základě vlastností cenových průběhů u nalezených vzorů. Byl vytvořen zjednodušený automatický obchodní systém, který definuje vstupy a výstupy u hypotetických obchodů. Jednotlivé obchody generují zisky a ztráty a dle jejich distribuce se vyvíjí equity křivka. Na těchto křivkách je vidět, že navržený obchodní systém je ziskový během testovacího období (rok 2006, 2007, 2008). Dle *Tabulky 3* je procentuální ziskovost obchodního systému pro nákup 88,8% (2006), 50,2% (2007) a 35% (2008) původního kapitálu. *Tabulka 6* udává procentuální ziskovost nalezených vzorů pro prodej během jednotlivých roků: 54,8% (2006), 60,2% (2007) a 14,3% (2008) původního kapitálu. Významně lepších výsledků predikce bude dosaženo při dohlížení experta na obchodní rozhodnutí automatického obchodního systému při samotném procesu obchodování. Expert s použitím citu a intuice dokáže filtrovat mnoho ztrátových obchodů, umí uzavírat ztrátové obchody s menší ztrátou než samotný automatický systém, ale zejména dokáže sofistikovaněji vystupovat ze ziskových pozic.

Druhý hlavní úkol této diplomové práce, tedy ověření a vyhodnocení vybrané techniky, byl úspěšně splněn. Metoda *Učení založené na instancích* se osvědčila při rozeznávání vzorů. Také metoda *Pure Price Action*, určená pro analýzu a predikci burzovních grafů, je velmi účinná. Vytvořený algoritmus je naučen vyhledávat definovaný cenový vzor s velmi vysokou přesností a vyhodnotit přesnost predikce.

8. POUŽÍTÁ LITERATURA

[1] Sang C. Suh, Dan Li, Jingmiao Gao: A novel chart pattern recognition approach: A case study on Cup with Handle.
http://istg.tamu-commerce.edu/Suh-Papers/ANNIE2004.pdf
http://istg.tamu-commerce.edu/Suh-Papers/IDPT2005.pdf

[2] Silas N. Onyango: On the pattern recognition of Verhulst-logistic Ito Processes in Market Price Data.
http://www.strathmore.edu/research/nairobi_stock_exchange_study.pdf

[3] Robi Polikar: Pattern Recognition.
http://users.rowan.edu/~polikar/RESEARCH/PUBLICATIONS/wiley06.pdf

[4] Invariant PR using Bayesian Inference on Hierarchical Sequences.
http://www.stanford.edu/~dil/RNI/DilJeffTechReport.pdf

[5] Encyklopedie Wikipedia – Pattern Recognition.
http://en.wikipedia.org/wiki/Pattern_recognition

[6] Odkazy na www stránky na téma rozpoznávání vzorů:
http://cgm.cs.mcgill.ca/~godfried/teaching/pr-web.html

[7] Statistical Pattern Recognition.
http://media.wiley.com/product_data/excerpt/39/04708451/0470845139.pdf

[8] A probabilistic Hough Transform.
http://www.cs.technion.ac.il/~freddy/papers/Papers39.PDF

[9] Introduction to Time Series Analysis.
http://www.itl.nist.gov/div898/handbook/pmc/section4/pmc4.htm

[10] Encyklopedie Wikipedia – Time Series.
http://en.wikipedia.org/wiki/Time_series

[11] David MacKay: Information Tudory, Inference, and Learning Algorithms
http://www.inference.phy.cam.ac.uk/itprnn/book.pdf

[12] Pattern Recognition from One Example by Chopping.
http://ida.first.fraunhofer.de/~blanchard/publi/FleBla05.pdf

[13] Honzík Petr: Strojové učení. FEKT VUT, 2006.

[14] D. Dutta Majumder: Fuzzy sets in pattern recognition
http://dml.cz/bitstream/handle/10338.dmlcz/104148/AplMat_30-1985-4_2.pdf

[15] Švarc Ivan: Automatizace. FSI VUT, 2005.

[16] Kvasnička Michal, Vašíček Osvald: Úvod do analýzy časových řad. ESF MU Brno, 2001.
http://merkur.econ.muni.cz/~qasar/vyuka/emm2/skriptaemmii.pdf

[17] Hančlová Jana, Tvrdý Lubor: Úvod do analýzy časových řad. EF VŠB-TU Ostrava, 2003.
http://gis.vsb.cz/pan/Skoleni_Texty/TextySkoleni/AnalyzaCasRad.pdf

[18] Encyklopedie Wikipedia – Časová řada.
http://cs.wikipedia.org/wiki/%C4%8Casov%C3%A1_%C5%99ada#Modely_n.C3.A1hodn.C3.BDch_proch.C3.A1zek