

**TECHNICKÁ UNIVERZITA V KOŠICIACH**  
**FAKULTA ELEKTROTECHNIKY A INFORMATIKY**

**Tvorba odporúčaní v procesoch hospodárskej praxe**  
**Diplomová práca**

**2016**

**Bc. Róbert Magyar**

**TECHNICKÁ UNIVERZITA V KOŠICIACH**  
**FAKULTA ELEKTROTECHNIKY A INFORMATIKY**

**Tvorba odporúčaní v procesoch hospodárskej praxe**  
**Diplomová práca**

Študijný program: Hospodárska informatika  
Študijný odbor: Hospodárska informatika(9.2.10)  
Školiace pracovisko: Katedra kybernetiky a umelej inteligencie (KKUI)  
Školiteľ: prof. Ing. Ján Paralič, PhD.  
Konzultant: Ing. Cecília Havrilová

**2016 Košice**

**Bc. Róbert Magyar**

## **Abstrakt v SJ**

Predmetom diplomovej práce je znalostná podpora procesu obchodovania na devízovom trhu. V prvej, teoretickej časti podávame prehľad problematiky obchodovania na devízovom trhu, navrhujeme systém pre podporu rozhodovania a určujeme, ktoré aspekty obchodovania sa budú podieľať v rozhodovacom procese. Praktická časť je venovaná dolovaniu znalostí, tvorbe rozhodovacieho procesu, implementácii získaných znalostí do platformy a testovaním systému ako celku. V závere porovnáваме jednotlivé nastavenia systému a dosiahnuté výsledky z hľadiska stanoveného cieľa.

## **Kľúčové slova v SJ**

Forex, devízový trh, dolovanie znalostí, systém na podporu rozhodovania, menové páry

## **Abstrakt v AJ**

This master thesis is focused on knowledge-based support for trading on foreign exchange market. We give in the first, theoretical part an overview of major issues with trading on foreign exchange market, we further design a system to support decision making and determine which aspects of trading will be involved in the decision making process. The second, practical part is devoted to data mining, decision making process, the implementation of the acquired knowledge to the platform and testing the system as a whole. In the final part we compare each system setup and results achieved in terms of our main goal.

## **Kľúčové slova v AJ**

Forex, Foreign Exchange, data mining, decision support system, currency pairs

**TECHNICKÁ UNIVERZITA V KOŠICIACH**  
**FAKULTA ELEKTROTECHNIKY A INFORMATIKY**  
Katedra kybernetiky a umelej inteligencie

**ZADANIE**  
**DIPLOMOVEJ PRÁCE**

Študijný odbor: **9.2.10 Hospodárska informatika**

Študijný program: **Hospodárska informatika**

Názov práce:

**Tvorba odporúčaní v procesoch hospodárskej praxe**  
Making recommendations in economic practice

Študent: **Bc. Róbert Magyar**  
Školiteľ: **prof. Ing. Ján Paralič, PhD.**  
Školiace pracovisko: **Katedra kybernetiky a umelej inteligencie**  
Konzultant práce: **Ing. Cecília Havrilová**  
Pracovisko konzultanta: **Katedra kybernetiky a umelej inteligencie**

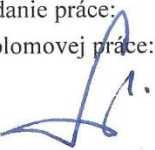
Pokyny na vypracovanie diplomovej práce:

1. Podat' stručný prehľad problematiky obchodovania na devízovom trhu, najmä z pohľadu využiteľnosti metód dolovania v dátach.
2. Navrhnuť systém pre podporu rozhodovania pri obchodovaní na devízových trhoch, založený na znalostiach získaných dolovaním v dátach.
3. Implementovať získané znalosti do vybranej platformy pre obchodovanie na devízovom trhu.
4. Realizovať a vyhodnotiť vlastné experimenty na zvolenej množine dát s cieľom verifikovať použiteľnosť vytvoreného systému na podporu obchodovania.
5. Vypracovať dokumentáciu podľa pokynov vedúceho práce (hlavná časť práce v rozsahu 50-70 strán, prílohy - používateľská a systémová príručka, DVD s textami, obrázkami a softvérovými výstupmi aplikácie, tlačaná forma v nerozoberateľnej väzbe).


Jazyk, v ktorom sa práca vypracuje: slovenský

Termín pre odovzdanie práce: 29.04.2016

Dátum zadania diplomovej práce: 31.10.2015

  
.....  
prof. Ing. Peter Sinčák, CSc.  
vedúci garantujúceho pracoviska



07   
.....  
prof. Ing. Liberios Vokorokos, PhD.  
dekan fakulty

## **Čestné vyhlásenie**

Vyhlasujem, že som celú diplomovú prácu vypracoval samostatne s použitím uvedenej odbornej literatúry.

Košice, 27. apríla 2016

.....

## **Podakovanie**

V prvom rade dakujem môjmu školiteľovi prof. Ing. Jánovi Paraličovi PhD a konzultantke Ing. Cecílii Havrilovej, za pomoc, rady a informácie pri písaní mojej práce. Ďalej chcem poďakovať všetkým prednášajúcim a cvičiacim, ktorí ma v inžinierskom štúdiu učili.

## Obsah

Zoznam obrázkov .....	9
Zoznam tabuliek .....	10
Zoznam symbolov a skratiek .....	11
Slovník termínov.....	12
Úvod .....	13
1. Teoretická časť .....	15
1.1. Postup tvorby odporúčacieho systému.....	15
1.2. CRISP-DM .....	16
1.3. Analýza súčasného stavu.....	18
1.4. Výber vhodného algoritmu .....	20
1.4.1. Rozhodovací strom tvorený C4.5 a C 5.0 algoritmom.....	20
1.4.2. CART .....	22
1.5. Fundamentálne správy.....	23
1.6. Nástroje technickej analýzy.....	24
1.7. Výber obchodnej platformy a brokera.....	25
2. Praktická časť .....	27
2.1. Dolovanie v dátach.....	27
2.1.1. Pochopenie problému .....	27
2.1.2. Pochopenie dát .....	28
2.1.3. Predspracovanie dát.....	35
Vyplniť chýbajúce hodnoty.....	35
Odstránenie výrazne odchýlených hodnôt .....	35
Vyhladzovanie šumu .....	35
Redukcia dimenzii.....	40
Redukcia početnosti dát.....	40
2.1.4. Návrh experimentov.....	42
Určenie modelovacích techník .....	42

Určenie vyhodnocovacej techniky .....	43
2.1.5. Výsledky experimentov .....	43
2.1.6. Vyhodnotenie experimentov .....	44
2.1.7. Analýza ekonomických správ .....	44
2.1.8. Analýza výsledku .....	46
2.2. Predstavenie konceptu odporúčacieho systému .....	46
2.2.1. Vstupné parametre .....	48
2.2.2. Rozhodovací proces.....	50
2.2.3. Váhovanie pravidiel.....	51
2.2.4. Spoľahlivosť vybraných pravidiel.....	51
2.2.5. Príklad najlepšieho pravidla .....	54
2.2.6. Modifikácia výberu pravidiel .....	55
2.2.7. Analýza výkonnosti systému týždenne.....	56
Záver.....	58
Zoznam použitej literatúry .....	59
Prílohy .....	61



## Zoznam obrázkov

Obrázok 1 Postup tvorby odporúčacieho systému. ....	15
Obrázok 2 Šesť fáz CRISP-DM (prevzaté z [1]).....	17
Obrázok 3 Tabuľka kvality dát pre EURUSD. ....	34
Obrázok 4 Tabuľka kvality dát pre databázu správ. ....	34
Obrázok 5 Výsledok predspracovania dát pre databázu menových párov.....	41
Obrázok 6 Výsledok predspracovania dát pre databázu správ.....	42
Obrázok 7 Početnosť záznamov výsledného atribútu result. ....	42
Obrázok 8 Počet správne odhadnutých teoretických následkov k reálnym.145 241 .....	45
Obrázok 9 Chybová matica.....	46
Obrázok 10 Vstupy odporúčacieho systému.....	49
Obrázok 11 Proces rozhodovania.....	50
Obrázok 12 Váhovanie pravidla. ....	51
Obrázok 13 Jednotlivé obchody pravidla 8 z obdobia 31.7.2015 do 28.8.2015 na menovom páre EURUSD. ....	54

## Zoznam tabuliek

Tabuľka 1 Jednotlivé fázy a ich úlohy (spracované podľa [[3]). .....	18
Tabuľka 2 Fundamentálne správy a ich zvyčajný efekt podľa Forex Factory [13].....	23
Tabuľka 3 Prehľad brokerov a platforiem, ktoré poskytujú. ....	25
Tabuľka 4 Podrobný popis atribútov v databázach pojednávajúcich o vývoji kurzov. ....	29
Tabuľka 5 Podrobný popis atribútov v databáze News. ....	30
Tabuľka 6 Základné charakteristiky atribútu close.....	30
Tabuľka 7 Kontingenčná tabuľka.....	31
Tabuľka 8 Hodnoty $\chi^2$ -test nezávislosti pre všetky vstupné atribúty a cieľový atribút Result. ...	34
Tabuľka 9 Vytvorené atribúty.....	40
Tabuľka 10 Návrh experimentov.....	43
Tabuľka 11 Výsledky modelovania. ....	44
Tabuľka 12 Ziskové pravidlá z kategórie popisujúcej posilňovanie amerického doláru. ....	52
Tabuľka 13 Ziskové pravidlá z kategórie popisujúcej oslabovanie amerického doláru. ....	53
Tabuľka 14 Podrobnejší pohľad na najlepšie pravidlo .....	55
Tabuľka 15 Dosiahnuté výsledky pred a po modifikácii pravidiel. ....	56
Tabuľka 16 Analýza výkonnosti systému týždenne.....	57

## Zoznam symbolov a skratiek

Fx	Forex
EUR	Euro
USD	Americký dolár
JPY	Japonský jen
NZD	Novozélandský dolár
AUD	Austrálsky dolár
CAD	Kanadský dolár
GBP	Britská libra
CHF	Švajčiarsky frank

## Slovník termínov

**Spred** (spread) je rozdiel medzi kurzom na trhu a kurzom, ktorý poskytuje broker obchodníkovi.

**Pip** je najmenšia zmena menového kurzu s ktorou sa obchoduje.

**Lot** je množstvo (objem) nakúpenej/predanej meny.

**Devízová arbitráž** je nákup (predaj) devízových mien na jednom trhu s cieľom profitovať na rozdiel cien pri predaji (nákupe) danej meny na inom trhu.

**Devízová špekulácia** je nákup (predaj) devízových mien s cieľom profitovať na rozdiel cien pri predaji (nákupe) niekedy v budúcnosti.

**RSI** je indikátor, ktorý porovnáva predošlé zisky ku predošlým stratám s cieľom určiť zónu predaja a nákupu devízy.

**Stop Loss** je obchodný príkaz, ktorý slúži na eliminovanie strát v obchodovaní.

**Take profit** je obchodný príkaz, ktorý slúži na realizáciu ziskov.

## Úvod

Devízový trh (Forex, t.j. Foreign Exchange) je miesto kde sa stretáva dopyt a ponuka po zahraničných menách a kde sa tvorí menový kurz, jednotná cena, ktorá umožňuje prevod peňažných prostriedkov z jednej meny na druhú. Je to voľne prístupný trh, ktorý je najmenej regulovaný a nemá vymedzené miesto obchodu ani hodiny obchodovania. Výhodou trhu je možnosť obchodovania dvadsaťštyri hodín denne, okrem soboty a nedele.

Základnou funkciou tohto trhu je výmena peňažných prostriedkov z jednej meny do druhej na investovanie, cestovanie, zahraničný obchod atď. Potreba tejto funkcie devízového trhu vychádza z množstva štátov, ktoré majú rôznu národnú menu a ktoré potrebujú medzi sebou obchodovať z rozličných dôvodov. Medzi sekundárne funkcie tohto trhu patrí zabezpečovanie sa proti kurzovému riziku, devízová arbitráž a devízové špekulácie.

Na Forexe sa stretávajú tri veľké skupiny obchodníkov. Sú to finančné inštitúcie, díleri a nefinanční zákazníci, hlavne firmy [2]. Medzi finančné inštitúcie patria banky, inštitucionálni investori, obchodné firmy a hedge fondy. Z aprílových výsledkov banky pre medzinárodné platby (BIS) je vidieť, že 53% účastníkov sú finančné inštitúcie, 39% sú díleri a iba 9% sú nefinanční zákazníci [2].

Táto práca je zameraná na vytvorenie odporúčacieho systému na devízovom trhu, ktorý využíva možnosť špekulácie na tomto trhu a podľa pravidiel, ekonomických správ a technickej analýzy odporúča smer obchodovania. Jadrom tohto systému sú pravidlá, ktoré sa získajú z forexových dát využitím procesu objavovania znalostí [1]. Tento proces je veľmi dôležitý, pretože na získaných pravidlách bude z veľkej časti postavený odporúčací systém. Okrem toho bude systém využívať vybrané nástroje technickej analýzy. Ak sa experimentmi potvrdí, že existujú ekonomické správy, ktorých výsledok vieme dobre predpovedať a vplyvom zaradenia týchto správ do rozhodovacieho systému získame lepšie odporúčacie výsledky, tak systém obohatíme aj o fundamentálny aspekt. Výsledok práce by mal byť systém, ktorý odporúča obchody na Forexe podľa pravidiel, ktoré sa získali procesom objavovania znalostí, ktorý berie do úvahy fundamentálne správy a indikátory technickej analýzy a ktorý sa okrem toho bude vedieť aj adaptovať na zmenu prehodnocovaním jednotlivých pravidiel.

Prácu teda môžeme rozdeliť na dve časti. Prvá časť sa zaoberá dolovaním znalostí z forexových dát, ktoré najskôr abstrahujeme na trend, stanovíme hypotézu a použijeme CRISP-DM metodiku

pri dolovaní v týchto dátach. Výsledkom tejto fázy je súbor pravidiel. Druhá fáza bude v znamení vytvárania automatického odporúčacieho systému (decision support system), ktorý bude odporúčať pomocou filtrovania znalostí na základe ohraničení. Dôležitou časťou bude tzv. back-testing a forward-testing, pomocou ktorých určíme do akej miery sa môžeme na daný systém spoľahnúť v rozhodovaní.

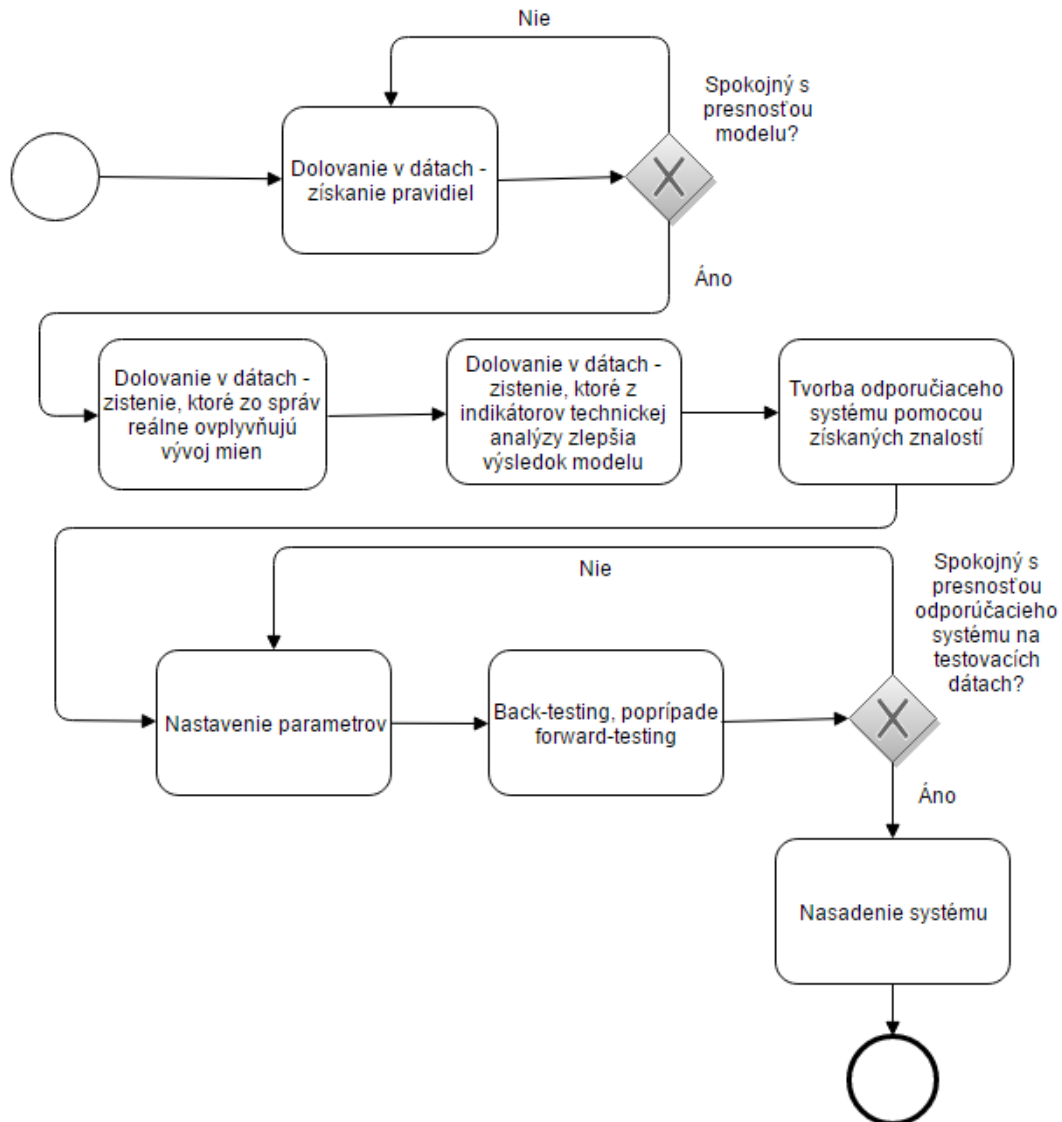
Vytváranie takéhoto systému má zmysel jednak z hľadiska edukatívneho, teda možnosti lepšieho pochopenia správania sa devízového trhu, ale aj z hľadiska biznisu, teda dosahovania zisku pri obchodovaní na forexovom trhu. Dosiahnutie zisku nie je zámer tejto práce, ale je nevyhnutnou podmienkou toho, aby bol navrhnutý systém úspešný.

Musíme ešte dodať, že všetky experimenty sa budú vykonávať len na majoritných menách [2], ktorými sú euro, americký dolár, libra, austrálsky dolár, novozélandský dolár, japonský jen, kanadský dolár a švajčiarsky frank. Dôvodom je ich vysoká likvidita oproti ostatným menám a relatívne nízke poplatky [2].

## 1. Teoretická časť

### 1.1. Postup tvorby odporúčacieho systému

Na nasledujúcom obrázku je zobrazený postup tvorby navrhovaného odporúčacieho systému.



Obrázok 1 Postup tvorby odporúčacieho systému.

Jadro systému sa skladá zo znalostí získaných pomocou dolovania v dátach. Získame tri typy znalostí:

- Rozhodovacie pravidlá, ktoré sú v podstate opísané vzory (javy), ktoré sa frekventovane dejú na trhu.
- Množinu fundamentálnych správ, na ktoré sa dá reálne spoľahnúť, teda vieme, že s veľkou pravdepodobnosťou po ich vyhlásení nastane určitý jav (cena bude rásť alebo klesať určitú dobu).
- Množinu indikátorov, ktoré zlepšili prediktívnu schopnosť modelu.

Znalosti sa potom naprogramujú do platformy pre obchodovanie, kde sa bude vykonávať back-testing a forward-testing. Následne sa zanalyzuje výsledok, či je dostatočne správny a či prináša výhodu na trhu.

Celkový výsledok by mal byť systém, ktorý beží v pozadí na platforme a pri určitých splnených podmienkach odporučí obchod. Výsledok z odporúčaného obchodu sa zaznamená a zhodnotí sa, či pravidlo, ktoré ho odporučilo je ešte stále validné. Jednotlivé pravidlá sa teda hodnotia váhou, pričom váha sa zvyšuje ak bolo odporúčanie úspešné. Ak odporúčanie nebolo úspešné, tak sa váha pre dané pravidlo znižuje o stanovené konštantné číslo.

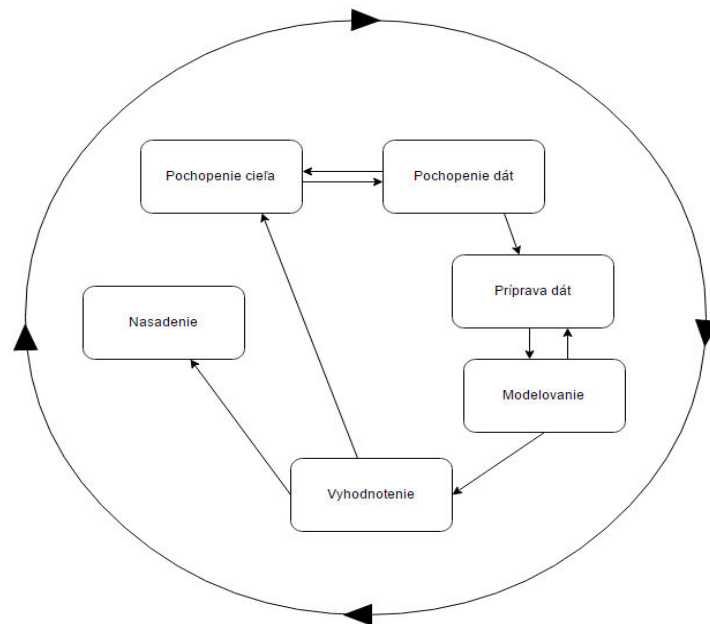
Systém musí pracovať s dvoma externými informáciami. Prvou je trhovú cenu menového páru, ktorá sa neustále mení a druhou fundamentálne správy, ktoré prichádzajú v určitý čas v určitom dni v mesiaci.

## 1.2. CRISP-DM

Pri narastajúcom množstve projektov zaoberajúcich sa dolovaním v dátach a hľadaním znalostí v dátach vzniklo niekoľko problémov medzi ktoré patrí zvyšujúca sa komplexnosť analyzovaných systémov, problémy s plánovaním, nesúlady finálneho cieľa projektu a cieľov organizácie, pre ktorú sa dolovanie vykonáva a mnohé iné. Neúspešnosť projektov dosahovala hodnotu až 50% [3]. Potreba vyvinutia štandardu pre proces dolovania v dátach viedla k vytvoreniu metodiky CRISP-DM (Cross industry standard process for Data mining) v roku 1996 [4]. CRISP-DM bol vytvorený na základe praktických skúseností expertov a vytvára akýsi hrubý procesný model, ktorý vedie k riešeniu identifikovaného biznis problému pomocou dolovania v dátach. Pri tvorení tejto metodiky sa požičiavali koncepty zo starších metodík ako 5 A's a Two Crows [3]. Táto metodika pozostáva zo šiestich fáz, ktoré vytvárajú postup v procese dolovania. Dôležité je ale povedať, že postup



v tomto procese nie je striktný a často je potrebné sa vracat' do predošlých fáz, čo naznačujú orientované hrany (pozri Obrázok 2) .



Obrázok 2 Šesť fáz CRISP-DM (prevzaté z [1]).

- Pochopenie cieľa – V tejto fáze sa majú pochopiť ciele organizácie, ktoré sa majú pretaviť do cieľov projektu dolovania znalostí a určiť kritériá úspechu. Kritériá úspechu sú dôležité z hľadiska vyhodnotenia úspešnosti projektu.
- Pochopenie dát – Fáza sa začína zberom dát a má za úlohu oboznámiť sa s dátami pomocou popisu, prieskumu kvality dát a pod.
- Príprava dát – Príprava dát slúži na formovanie surových dát do tvaru takého, aby sme s nimi mohli pracovať vo fáze modelovania, zahŕňa rôzne techniky predspracovania [1].
- Modelovanie – V tejto fáze sa určia modelovacie techniky, ich parametre a pokračuje sa tvorbou jednotlivých modelov, pričom je snaha dosiahnuť čo najlepšiu kvalitu výsledných modelov z pohľadu dolovania v dátach.
- Vyhodnotenie – Spraví sa analýza výsledkov a určí sa či sme dosiahli pomocou dolovania v dátach stanovený podnikový cieľ (dosiahli na začiatku dané kritériá úspechu).

- Nasadenie – Znalosti sa prezentujú zákazníkovi a určí sa postup nasadenia do praxe, teda jednotlivé znalosti získané z procesu dolovania sa aplikujú do praxe a monitoruje sa ich výkon.

Každá z jednotlivých fáz sa skladá z niekoľkých úloh, ktoré musia byť vykonané aby sa fáza mohla ukončiť a aby projekt mohol pokračovať ďalšou fázou [1][3][4].

Pochopenie cieľa	Pochopenie dát	Príprava dát	Modelovanie	Vyhodnotenie	Nasadenie
Pochopenie cieľov projektu	Zber dát	Výber dát	Výber techniky modelovania	Vyhodnotenie výsledkov	Plán nasadenia
Pochopenie cieľov dolovania znalostí	Popis dát	Čistenie dát	Návrh testovania	Kontrola procesu	Plán monitorovania a údržby
Navrhnutie plánu projektu	Prieskum dát	Konštrukcia dát	Tvorba modelu	Definovanie ďalších krokov	Záverečná správa
Stanovenie kritérií úspechu	Verifikácia kvality dát	Integrácia a formátovanie dát	Posúdenie správnosti modelu	--	Posúdenie

Tabuľka 1 Jednotlivé fázy a ich úlohy (spracované podľa [[3]).

Túto metodiku plne využijeme vo fáze dolovania znalostí pre odporúčací systém. Výsledkom vykonania týchto 6 fáz budú pravidlá, teda jadro pre náš systém.

### 1.3. Analýza súčasného stavu

Túto časť venujeme analyzovaniu obchodných systémov, ktoré nám poskytnú ucelený pohľad na štýl tvorby takéhoto typu systému.

V práci [10] Automatic stock market trading based on Technical Analysis od Fredrika Larsena z roku 2007 sa stretávame s obchodným systémom, ktorý sa zameriava na sledovanie vývoja 24 akcií. Pracuje s dátami z rokov 1970 až 1998. Obchodný systém sa skladá z 26 agentov. Každý

z agentov popisuje jeden z indikátorov poprípade cenovú formáciu z technickej analýzy. Agenty popisujú známe technické formácie ako Double Top/Bottom, Island Reversal, Price Envelope Band, Breakaway a Runaway Gap, Head and Shoulders, Hanging Man, Reversal Day a indikátory ako kĺzavý priemer (14, 40 a 200 dňové) a RSI (9, 14 a 21 dňové)[11]. Používa aj neurónové siete na určenie nákupu alebo predaja jednej z formácií a rozhodovacie stromy ID3 a JK8. Každý agent pracuje samostatne a vytvára signál, ktorý sa hodnotí váhou. Tieto váhy sú optimalizované genetickým algoritmom. Podľa vzorca sa potom váhy jednotlivých agentov spočítajú a určí sa smer predikcie. Výsledok je potom systém, ktorého každý agent má svoju presnosť a práve podľa tohto faktora sa jeho váha a teda sila hlasovacieho práva mení. Systém ale nie je reálne obchodovateľným, pretože jeho využitie sa nepotvrdilo na reálnych trhoch [11]. Taktiež neberú do úvahy obchodné parametre ako Stop Loss a Take profit [11] a fundamentálne správy.

Práca autorov K.K. Lai, Lean Yu, a Shouyang Wang [11] s názvom A Neural Network and Web-Based Decision Support System for Forex Forecasting and Trading z roku 2004 popisuje vývoj systému na podporu rozhodovania na Forexe. Tento systém sa skladá z dvoch modulov, konkrétnejšie z neurónovej siete (BPNN) a samotného systému na podporu rozhodovania situovanom na webe (DSS). Pomocou neurónových sietí sa natrénoval model, ktorý by mal predpovedať budúci vývoj na majoritných menových pároch. Následne sa pomocou integrácie modulov poskytne predikcia systému na podporu rozhodovania, ktorý na základe štyroch typov pravidiel odporúča smer pre investora. DSS pracuje s týmito typmi pravidiel:

- Prvý typ určuje stratégiu obchodovania veľmi jednoduchými pravidlami napr. Ak predikcia > aktuálna cena potom Nákup.
- Druhý typ selektuje vytvorené prípady z prvých pravidiel, aby sa neurčoval nákup a predaj pri veľmi nízkych rozdieloch v aktuálnej cene a predpovedanej cene.
- Tretí typ určuje pravdepodobnosť zisku pri nákupe a predaji.
- Štvrtý typ berie do úvahy riziko obchodu.

Cieľom autorov bolo vytvoriť odporúčací systém pre investorov, ktorý si pomocou neho môžu uľahčiť rozhodovanie [11]. Presnosť predpovede systému nebola uvedená.

V práci [12] Automated Foreign Exchange Trading System od Jaymin R. Mehta, Marcus D. Menghini a Daniel A. Sarafconn z roku 2011 je popísaný obchodný systém, ktorý bol založený na

testovaní jednoduchých dobre známých stratégií technickej analýzy, ktoré sú postavené na indikátoroch ako kĺzavý priemer, Bollinger Bands, Keltner Channel, Commodity Channel Index, Volume Oscillator a Volume Ratio [13]. Celý systém sa vytvoril na platforme TradeStation, ktorej programovanie je založené na jazyku EasyLanguage, čo je v podstate zjednodušený programovací jazyk s vopred naprogramovanými indikátormi. Pomocou indikátorov sa vytvorili štyri stratégie, ktoré sa testovali a sledoval sa ich výkon. Následne sa vybrali dve najlepšie a z nich sa vytvoril obchodný systém. Systém bol optimalizovaný na vzorke dát z dvoch mesiacov a testovaný na nasledujúcom mesiaci. Najvýkonnejšia stratégia založená na Commodity Channel Index a Volume Oscillator dosiahla úspešnosť 72,73% pri ôsmich profitabilných a troch stratových obchodoch.

Po preštudovaní rôznych obchodných systémov môžeme tvrdiť, že mnoho z nich využíva indikátory a poznatky technickej analýzy na tvorbu pravidiel. Väčšine z nich ale chýba fundamentálna analýza. Tieto systémy sú založené na integrácii modulov, v ktorých sa systémom hlasovania určuje finálny predikovaný smer vývoja kurzu daného menového páru. Každý z nich najskôr testuje správnosť predikcie a neskôr na tejto predikcii stavia odporúčanie, resp. rozhodnutie o kúpe/predaji. Všimol som si aj prílišnú jednoduchosť v systémoch, ktoré sa reálne obchodujú a tie komplexnejšie sú využívané len na experimenty. Mnoho systémov ale neposkytuje informácie o ich štruktúre z dôvodov obchodného tajomstva.

#### **1.4. Výber vhodného algoritmu**

Umelá inteligencia nám poskytuje množstvo kvalitných algoritmov, ktoré by sme mohli použiť k riešeniu nášho problému [6]. Naším výberovým kritériom bude algoritmus, ktorý nám poskytne ako výstup pravidlá. Práve preto sme sa rozhodli pre využitie rozhodovacích stromov pre dosiahnutie nášho cieľu.

##### **1.4.1. Rozhodovací strom tvorený C4.5 a C 5.0 algoritmom**

Algoritmus C 4.5 dokáže generovať klasifikátor vo forme rozhodovacieho stromu. Algoritmus , ktorý vytvoril Ross Quinlan [5], je vylepšenou verziou ID3 a je založený na metóde “rozdeľuj a panuj” [7].

V rozhodovacom strome list znamená priradenie do jednej z cieľových tried, medziľahlý uzol je vybraný atribút a hrana je test atribútu. Algoritmus využíva informačnú teóriu na výber najlepšieho testovacieho atribútu v danom nelistovom uzle, konkrétnejšie pomerový informačný

zisk. Funkciu pre výber najlepšieho testovacieho atribútu  $A$  v nelistovom uzle  $S$  možno v prípade algoritmu C4.5 vyjadriť matematicky takto:

$$I_p(S, A) = \frac{I(S, A)}{H_p(S, A)}$$

$$H_p(S, A) = - \sum_{j=1}^m p(a_j) * \log_2(p(a_j))$$

$$I(S, A) = H(S) - H(S, A)$$

$$H(S) = - \sum_{j=1}^n p(T_j) * \log_2(p(T_j))$$

$$H(S, A) = \sum_{j=1}^m p(a_j) * H(S_j)$$

$I_p(S, A)$  = pomerový informačný zisk pri výbere atribútu  $A$  na vetvenie v danom nelistovom uzle  $S$ ,

$H_p(S, A)$  = pomerová entropia v uzle  $S$  pri výbere atribútu  $A$ ,

$H(S, A)$  = entropia v uzle  $S$  použitím atribútu  $A$  na vetvenie v danom uzle,

$H(S)$  = entropia v uzle  $S$ ,

$I(S, A)$  = informačný zisk pri výbere atribútu  $A$  na vetvenie v danom uzle,

$T_j$  =  $j$ -tá trieda ,

$a_j$  =  $j$ -ta hodnota atribútu  $A$ .

Algoritmus C5.0 je vylepšená verzia algoritmu C4.5, ktorá bola vytvorená pre komerčné účely. Najväčšími výhodami C 5.0 oproti C4.5 podľa tvorca tohto algoritmu Rulequest [7] sú:

- Rýchlosť pri spracovaní veľkého množstva dát čo potvrdzuje aj tento výskum [8].
- Väčšia presnosť [7].

- Efektívnejšie využitie pamäte.
- Rozmerovo menšie rozhodovacie stromy, čo znamená menší počet pravidiel.
- Vstavaný boosting, váhovanie a orezávanie.

Keďže algoritmus C5.0 poskytuje väčšiu presnosť [7] a tvorí rozmerovo menší strom [7] je vhodnejší pre riešenie nášho problému ako C 4.5.

#### 1.4.2. CART

CART (Classification and Regression Trees) je algoritmus, ktorý generuje binárny rozhodovací strom. Tento algoritmus vytvorili L. Breiman, J.H. Friedman, R.A. Olshen a C.J. Stone [9]. Pracuje s reálnymi aj symbolickými atribútmi.

Úlohou algoritmu je nájsť optimálne delenie, čo sa deje pomocou dosiahnutia najväčšieho poklesu nečistoty rozdelenie príkladov. Algoritmus využíva na rozdelenie buď Gini index alebo Twoing [9].

Pokles nečistoty  $\Delta i(S)$  po rozdelení príkladov podľa hodnôt nejakého testovacieho atribútu v aktuálnom uzle  $S$  na dva podstromy  $S_L$  a  $S_P$  sa vypočíta nasledovne:

$$\Delta i(S) = i(S) - p \cdot i(S_L) - p \cdot i(S_R)$$

Pričom miera nečistoty sa môže počítať buď podľa deliaceho pravidla Gini:

$$i(S) = \sum_{i \neq j} p(i|S) \cdot p(j|S)$$

$\Delta i(S) = - \sum_{i=1}^K p^2(i|S) + p_L \sum_{i=1}^K p^2(i|S_L) + p_P \sum_{i=1}^K p^2(i|S_P)$  alebo podľa pravidla Twoing:

$$\Delta i(S) = p_L p_R \left[ \sum_{i=1}^K |p(i|S_L) - p(i|S_R)| \right]^2$$

, kde:  $i(S)$  = výpočet funkcie nečistoty (impurity) v uzle  $S$  pre Gini index,

$p(i|S)$  = pravdepodobnosť výskytu triedy  $i$  v uzle  $S$

$K$  je počet tried

$p_L = p(S_L)$  = pravdepodobnosť, že prípad pôjde do ľavého potomka uzla  $S$ , t.j.  $S_L$  (pravdepodobnosť ľavého podstromu  $S$ )

$p_R = p(S_R)$  = pravdepodobnosť, že prípad pôjde do pravého potomka uzla  $S$ , t.j.  $S_R$  (pravdepodobnosť pravého podstromu  $S$ )

### 1.5. Fundamentálne správy

Ak chceme aby systém pracoval s fundamentálnymi správami musíme experimentálne určiť, na ktoré z množiny správ sa vieme spoľahnúť z hľadiska výsledku správy (akcia) a následnej zmeny devízového kurzu (reakcia). Potrebujeme teda vedieť v koľkých prípadoch sa po správe, ktorá má vplývať na vývoj ceny negatívne (pozitívne), sa tento negatívny (pozitívny) vývoj aj naozaj prejaví. Podľa nasledujúcej tabuľky budeme hodnotiť správy a následný vývoj ceny, aby sme zistili, ktoré správy sa oplatí sledovať a ktoré nie, pričom sledujeme práve spoľahlivosť príčiny (príchod správy s určitým obsahom) a následku (vývoj menového kurzu). Nasledujúca tabuľka obsahuje zvyčajný efekt správ ovplyvňujúci americký dolár podľa interného výskumu Forex Factory [13]. Pričom R značí aktuálnu hodnotu a K konsenzus.

Tabuľka 2 Fundamentálne správy a ich zvyčajný efekt podľa Forex Factory [13].

Názov fundamentálnej správy	Zvyčajný efekt		
	Pozitívny vývoj	Záporný vývoj	Neutrálny vývoj
Unemployment Rate	R<K	R>K	R=K
Consumer Price Index Ex Food & Energy (YoY)	R>K	R<K	R=K
Fed Interest Rate De	R>K	R<K	R=K
ISM Non-Manufacturing PMI	R>K	R<K	R=K
Durable Goods Orders	R>K	R<K	R=K
Consumer Confidence	R>K	R<K	R=K
Gross Domestic Product Annualized	R>K	R<K	R=K
Reuters/Michigan Consumer Sentiment Index	R>K	R<K	R=K
ISM Manufacturing PMI	R>K	R<K	R=K
Consumer Price Index (YoY)	R>K	R<K	R=K
Retail Sales (MoM)	R>K	R<K	R=K

Nonfarm Payrolls	R>K	R<K	R=K
------------------	-----	-----	-----

Výsledkom takéhoto experimentu bude množina správ, ktoré s veľkou pravdepodobnosťou ovplyvňujú vývoj menového kurzu zakaždým rovnako.

### 1.6. Nástroje technickej analýzy

Z množstva rôznych indikátorov, ktoré ponúka technická analýza sme vybrali Bollinger Bands a kľzavý priemer, ktorých implementáciou by sa mal zabezpečiť lepší výsledok.

Jednoduchý kľzavý priemer je aritmetický priemer cien, ktorý sa počíta ako súčet zatvárajúcich cien menového páru za určité obdobie podelený veľkosťou tohto obdobia. Pomocou tohto nástroja zachytíme krátkodobý trend. SMA je jednoduchý kľzavý priemer (Simple Moving Average) a je možné ho vypočítať nasledovne:

$$SMA = \frac{\sum_i^t c_i}{t}$$

$c_i$  = zatváracia cena v čase  $i$

$t$  = dĺžka obdobia

Bollinger bands (Bollingerové pásma) je indikátor, ktorý pomocou troch kriviek dynamicky určuje pásmo, v ktorom cena osciluje [15]. Stredná krivka je jednoduchý kľzavý priemer, ktorý je základom pre hornú a dolnú krivku. Tieto sa zužujú a rozťahujú vzhľadom na jednoduchý kľzavý priemer podľa volatility daného menového páru [15]. V princípe tento indikátor určuje spodnú a hornú hranicu ceny. Štandardnú odchýlku  $\sigma$  možno vypočítať nasledovne:

$$\sigma = \sqrt{\frac{\sum_i^t (c_i - SMA)^2}{t}}$$

Ďalšie indikátory sú založené na tzv. Bollingerovom pásme (BB z anglického Bollinger Band). Toto pásmo je definované hornou ( $BB_U$ ) a dolnou ( $BB_L$ ) hranicou nasledovne:

$$BB_U = SMA + 2 * \sigma$$

$$BB_L = SMA - 2 * \sigma$$



Tento indikátor pomôže určiť situácie kedy je výhodnejšie nakupovať, a to ak cena klesne pod dolné Bollingerovho pásmo a predávať ak cena zasa porastie nad horné pásmo. Je to určitá výpomoc pri zisťovaní obratu trendu ceny. V praktickej časti zistíme, či tieto indikátory napomohli pri presnosti predpovede.

### 1.7. Výber obchodnej platformy a brokera

Každý broker, ktorý poskytuje služby obchodovania na finančných trhoch ponúka obchodnú platformu, ktorá vykonáva obchodné príkazy zadávané užívateľom. Z hľadiska Slovenska a Česka máme niekoľko brokerov, ktorí poskytujú rôzne obchodné platformy a našim cieľom je si vybrať najlepšiu pre vlastné účely.

Pri výbere takejto platformy je dôležité, aby nielen poskytovala služby zadarmo, historické dáta, back-testing a grafické rozhranie ale aj možnosť programovania vlastných obchodných systémov. V nasledujúcej tabuľke je prehľad slovenských a českých brokerov a poskytovaných platforiem.

Názov brokera	Poskytuje obchodovanie na devízovom trhu?	Minimálny vklad v eurách	Minimálne poplatky v eurách	Platforma	Možnosť programovanie obchodných systémov
X-Trade Brokers	Áno	500	8	Metatrader 4	Áno
Trim broker	Áno	1000	15	Saxotrader a Metatrader 4	Áno ale iba Metatrader 4
Patria direct	Nie	neuvedené	14.90	Vlastná aplikácia	neuvedené
Capital markets	Nie	neuvedené	25	Vlastná aplikácia	neuvedené
Brokerjet	Áno	1000	11.95	Vlastná aplikácia	neuvedené

Tabuľka 3 Prehľad brokerov a platforiem, ktoré poskytujú.

Z hľadiska poplatkov, minimálneho vkladu a poskytovania služieb na devízovom trhu sa nám zúži výber jednotlivých brokerov. Keďže dôležitý aspekt výberu je práve možnosť programovania obchodných systémov, tak sa nám výber zúži ešte viac a to na platformu Metatrader 4. Táto platforma má všetky hore uvedené vlastnosti, ktoré nám pomôžu dosiahnuť náš cieľ.

## 2. Praktická časť

### 2.1. Dolovanie v dátach

#### 2.1.1. Pochopenie problému

Zmyslom práce s forexovými dátami je predikcia budúceho vývoja. Dostatočne správna predikcia znamená veľkú výhodu pre participantov na trhu. Problémom stratégií, ktoré sú založené na technickej analýze je, že nekonkretizujú za akých podmienok sa má daná stratégia použiť. Výsledkom je vysoká fluktuácia presnosti predikcie budúceho vývoja, ktorá je zapríčinená zmenou podmienok, teda dopytu a ponuky po danej mene. Hlavnou požiadavkou je určiť, ktorá mena je globálne silná (trenduje), ktorá je globálne slabá a používať zvolené stratégie na menový pár, ktorý sa vytvorí zo silnej a slabej meny. Finálnym výsledkom by malo byť obchodovanie pri najlepších podmienkach a teda pri najvyššej šanci na úspech.

Dôležité je zistenie, či určovanie globálnej sily meny má zmysel z hľadiska budúceho vývoja tým, že smer vývoja meny bude naďalej rovnaký. Cieľom práce je potvrdiť hypotézu, že na tomto trhu sa dá určiť smer meny podľa pohybu menových párov v ktorých daná mena vystupuje. Zo všetkých menových párov danej meny sa určí, či je mena silná, slabá alebo má neutrálny výhľad, podľa čoho sa bude odvíjať ďalší vývoj. Hlavnou myšlienkou je abstrahovať cenu na trend, určiť silu meny a z tohto bodu odvíjať predpoveď.

Kritériom úspechu bude, ak výsledný model klasifikuje s aspoň 55% presnosťou budúci vývoj na základe určenia sily meny. Okrem toho sa musí preukázať, že sila meny má zotrvačnosť a určovanie tohto globálneho faktora v danej hodine má zmysel z hľadiska výsledkov.

Zameraním práce nie je nájsť model, ktorý bude perfektne klasifikovať smer, ale skôr nájsť pravidlá (javy), ktoré sa s veľkou pravdepodobnosťou dejú.

Okrem hlavnej úlohy popísanej vyššie, chcem aby dolovanie znalostí zodpovedalo nasledujúce otázky.

1. Ak je v danú hodinu stanovený trend na menovom páre, s akou pravdepodobnosťou bude trend pokračovať v rovnakom smere? Cieľom je zistiť, či sa tento jav vyskytuje dostatočne často, aby sme ho mohli aplikovať v hlavnej úlohe.

2. Aká je pravdepodobnosť, že negatívna ekonomická správa ovplyvní trh negatívne a pozitívna pozitívne? Cieľom je zistiť, či sa oplatí sledovať ekonomické správy.
3. Ako veľmi ovplyvní výsledok aplikovanie nástrojov technickej analýzy, konkrétne Bollinger Bands a troch kĺzavých priemerov (3,5 a 10 hodinové)? Cieľom je zistiť, či nástroje technickej analýzy prinesú väčšiu presnosť modelu, teda výhodu.
4. Ktoré ekonomické správy je dôležité sledovať a ktoré nie?

Využitím metodiky CRISP-DM rozdelíme prácu do 6 fáz. Vyhodnocovanie modelu bude podľa jeho presnosti.

## **2.1.2. Pochopenie dát**

### **2.1.2.1. Popis dát**

Pracujeme s ôsmimi dátovými zdrojmi. Sedem dátových zdrojov obsahuje dáta o jednej zo 7 mien (euro, libra, austrálsky dolár, novozélandský dolár, kanadský dolár, švajčiarsky frank a japonský jen) a americkom dolári. Posledný zdroj obsahuje ekonomické správy, ktoré vychádzali počas stanoveného obdobia. Počet záznamov je rôzny a preto bolo potrebné zlúčiť údaje a určiť jednotný časový interval pre všetky menové páry. Výsledkom je databáza, ktorá popisuje cenový vývoj menových párov vzhľadom na americký dolár od 1.10.2014 po 17.4.2015 a databáza ekonomických správ.

Všetky databázy pojednávajúce o vývoji kurzov majú rovnakú štruktúru dát. Riadok v databáze stanovuje dátum, čas, objem, otváraciu, zatváraciu, minimálnu a maximálnu cenu v danej hodine. Dokopy 7 numerických atribútov, ktoré popisujú vývoj menového páru za hodinu.

Názov atribútu	Krátky popis	Možné hodnoty
<b>date</b>	dátum	2014.10.01-2015.4.17
<b>time</b>	čas v hodinách	00:00:00-23:00:00
<b>open</b>	otváracia cena	<i>Numerické(real), rozpätie záleží od menového páru</i>
<b>maximum</b>	maximálna cena	<i>Numerické(real), rozpätie záleží od menového páru</i>
<b>minimum</b>	minimálna cena	<i>Numerické(real), rozpätie záleží od menového páru</i>
<b>close</b>	zatváracia cena	<i>Numerické(real), rozpätie záleží od menového páru</i>
<b>volume</b>	objem za hodinu	<i>Numerické (integer), rozpätie záleží od menového páru</i>

Tabuľka 4 Podrobný popis atribútov v databázach pojednávajúcich o vývoji kurzov.

Databáza ekonomických správ obsahuje dátum a čas správy, názov, štát, ktorý vydáva správu, stupeň volatility, aktuálnu, predošlú hodnotu a konsenzus experta.

Názov atribútu	Krátky popis	Možné hodnoty
<b>DateTime</b>	dátum a čas udalosti	2014.10.01 03:00:00-2015.3.11 00:50:00
<b>Name</b>	názov správy	<i>Nominálne, názvy jednotlivých udalostí.</i>
<b>Country</b>	krajina, ktorú ovplyvňuje správa	<i>Nominálne: Austrália, Canada, China, European Monetary Union, France, Germany, Greece, Italy, Japan, New Zeland,Switzerland, United Kingdom, United States</i>
<b>Volatility</b>	vplyv na pohyb men. páru	<i>Numerické(integer): 2 alebo 3.</i>
<b>Actual_news</b>	aktuálna hodnota	<i>Nominálne</i>
<b>Previous_news</b>	predošlá hodnota	<i>Nominálne</i>

<b>Consensus_news</b>	predpovedaná hodnota expertom	<i>Nominálne</i>
-----------------------	-------------------------------	------------------

Tabuľka 5 Podrobný popis atribútov v databáze News.

### 2.1.2.2. Základné charakteristiky atribútov

Najskôr preskúmame atribúty pre každú databázu samostatne. Pre ukážku som zvolil atribút close. (pozri Tabuľku č. 3).

<b>Mena/USD</b>	<b>Počet</b>	<b>Priemer</b>	<b>Min</b>	<b>Max</b>	<b>Rozpätie</b>	<b>Variancia</b>	<b>št. odch.</b>	<b>Medián</b>	<b>Modus</b>
<b>EUR</b>	2669	1.205	1.075	1.284	0.209	0.003	0.057	1.229	1.131
<b>GBP</b>	2670	1.558	1.497	1.623	0.127	0.001	0.035	1.562	1.517
<b>AUD</b>	3254	0.818	0.755	0.890	0.135	0.002	0.042	0.814	0.781
<b>NZD</b>	3237	0.766	0.719	0.801	0.081	0.000	0.019	0.772	0.754
<b>CAD</b>	3255	1.192	1.108	1.282	0.173	0.003	0.058	1.183	1.248
<b>CHF</b>	3245	0.959	0.812	1.022	0.210	0.001	0.031	0.961	0.954
<b>JPY</b>	3254	116.903	105.67	121.91	16.235	18.748	4.330	118.572	118.958

Tabuľka 6 Základné charakteristiky atribútu close.

Z hľadiska tejto analýzy je vidieť, že každý menový pár má konkrétne cenové hranice, ktoré ohraničujú cenové pásmo obchodovania. Táto informácia sa môže použiť ako vodítko pri obchodovaní, keďže blízko pri minime je oveľa rizikovejšie predávať menu ako pri maxime. To isté platí pri nakupovaní, keď nakupovanie blízko pri maxime je oveľa rizikovejšie ako nakupovanie pri minime.

### 2.1.2.3. Závislosť medzi atribútmi

Z logického hľadiska nemá význam určovať závislosť medzi atribútmi close, open, min a max, keďže musia byť korelované. Tieto atribúty nám len pomôžu určiť krátkodobý trend, s ktorým sa bude ďalej pracovať.

Závislosť medzi vstupnými atribútmi a cieľovým atribútom Result vypočítame pomocou  $\chi^2$ -testu nezávislosti. Vytvoríme si kontingenčnú tabuľku pre vybraný atribút trend\_EURUSD a cieľový atribút result.

Result/trend_EURUSD	-1	0	1	$\Sigma$
Loss(obrat trendu)	432	175	443	1050
Strength(S)	77	65	481	623
Weakness(W)	418	73	66	557
Neutral(null)	133	78	142	353
$\Sigma$	1060	391	1132	2583

Tabuľka 7 Kontingenčná tabuľka.

*Hypotézy:*

$H_0$ : Ak platí  $h_{ij} = \frac{h_i \cdot h_j}{n}$ , potom platí nezávislosť atribútov Result a trend\_EURUSD.

$H_1$ : Ak platí  $h_{ij} \neq \frac{h_i \cdot h_j}{n}$ , potom platí závislosť atribútov Result a trend\_EURUSD na hladine významnosti  $\alpha=0.05$ .

Pričom  $h_i$  a  $h_j$  sú okrajové početnosti.

Testovacie kritérium:

$$G = \sum_i^k \sum_j^m \frac{(h_{ij} - \frac{h_{i.} * h_{.j}}{n})^2}{\frac{h_{i.} * h_{.j}}{n}} \quad (1)$$

Stupeň voľnosti je  $(k - 1) * (m - 1)$ .

Kritická hodnota je  $\chi^2_{1-\alpha, (k-1)*(m-1)}$ .

Po dosadení hodnôt z tabuľky č.7 dostaneme  $G=596.387$  so stupňom voľnosti 6. Kritickú hodnotu vyčítame z tabuľky pre daný test a dostaneme číslo 18.548.

Keďže kritická hodnota je menšia ako hodnota  $G$  tak hypotézu  $H_0$  zamietame a teda test nezávislosti ukázal, že s 99.5% pravdepodobnosťou sú atribúty Result a trend\_EURUSD závislé.

Tento test sa vykonal pre všetky vstupné symbolické atribúty .

Nasledujúca tabuľka obsahuje hodnoty  $\chi^2$ -test nezávislosti pre všetky symbolické vstupné atribúty a cieľový atribút Result.

Názov symbolického atribútu	G štatistika	Pravdepodobnosť v %	Stupeň voľnosti	Kritická hodnota	Závislosť
trend_EURUSD	596.387	99.5	6	18.548	Áno
trend_GBPUSD	475.313	99.5	6	18.548	Áno
trend_AUDUSD	494.680	99.5	6	18.548	Áno
trend_NZDUSD	451.080	99.5	6	18.548	Áno
trend_USDCHF	335.842	99.5	6	18.548	Áno
trend_USDCAD	355.045	99.5	6	18.548	Áno
trend_USDJPY	273.826	99.5	6	18.548	Áno



Count CC	2579.217	99.5	42	69.336	Áno
Time of the day	16.761	92.5	9	15.630	Áno
strength/weakness	654.671	99.5	6	18.548	Áno
Range_mean_BIN	10.063	90	9	14.684	Nie
MA10_EURUSD	5.505	90	3	6.251	Nie
MA5_EURUSD	18.926	90	3	6.251	Áno
MA3_EURUSD	52.128	90	3	6.251	Áno
UBB_EURUSD	5.570	90	3	6.251	Nie
LBB_EURUSD	21.408	90	3	6.251	Áno
MA10_GBPUSD	4.081	90	3	6.251	Nie
MA5_GBPUSD	12.175	90	3	6.251	Áno
MA3_GBPUSD	33.221	90	3	6.251	Áno
UBB_GBPUSD	4.531	90	3	6.251	Nie
LBB_GBPUSD	3.924	90	3	6.251	Nie
MA10_AUDUSD	5.352	90	3	6.251	Nie
MA5_AUDUSD	6.691	90	3	6.251	Áno
MA3_AUDUSD	5.756	90	3	6.251	Nie
UBB_AUDUSD	6,841	90	3	6.251	Áno
LBB_AUDUSD	2.283	90	3	6.251	Nie
MA10_NZDUSD	7.251	90	3	6.251	Áno
MA5_NZDUSD	1.205	90	3	6.251	Nie
MA3_NZDUSD	0.407	90	3	6.251	Nie
UBB_NZDUSD	9.175	90	3	6.251	Áno
LBB_NZDUSD	1.959	90	3	6.251	Nie
MA10_USDCHF	12.667	90	3	6.251	Áno
MA5_USDCHF	10.283	90	3	6.251	Áno
MA3_USDCHF	13.948	90	3	6.251	Áno
UBB_USDCHF	16.32	90	3	6.251	Áno
LBB_USDCHF	6.484	90	3	6.251	Áno
MA10_USDCAD	2.813	90	3	6.251	Nie
MA5_USDCAD	3.511	90	3	6.251	Nie
MA3_USDCAD	1.485	90	3	6.251	Nie
UBB_USDCAD	5.955	90	3	6.251	Nie

LBB_USDCAD	0.684	90	3	6.251	Nie
MA10_USDJPY	0.749	90	3	6.251	Nie
MA5_USDJPY	0.291	90	3	6.251	Nie
MA3_USDJPY	1.139	90	3	6.251	Nie
UBB_USDJPY	0.551	90	3	6.251	Nie
LBB_USDJPY	4.072	90	3	6.251	Nie

Tabuľka 8 Hodnoty  $\chi^2$ -test nezávislosti pre všetky vstupné atribúty a cieľový atribút Result.

Z tabuľky č.8 je vidieť, že nie všetky symbolické atribúty sú závislé s cieľovým atribútom Result. S atribútmi, ktoré sú s cieľovým atribútom závislé sa bude ďalej pracovať a ostatné nebudeme v ďalšej analýze brať do úvahy.

#### 2.1.2.4. Overenie kvality dát

Field	Type	% Complete	Valid Records	Null Value
date	Range	100	2669	0
time	Range	100	2669	0
open	Range	100	2669	0
maximum	Range	100	2669	0
minimum	Range	100	2669	0
close	Range	100	2669	0
volume	Range	100	2669	0

Obrázok 3 Tabuľka kvality dát pre EURUSD.

Po prezretí dát o vývoji kurzov sa nenašli žiadne redundantné dáta, všetky atribúty sú úplné a neobsahujú žiadnu chýbajúcu hodnotu. Spolu validných záznamov je 2669 pre každú databázu pojednávajúcu o vývoji kurzu.

Field	Type	% Complete	Valid Records	Null Value
Name	Set	100	72	0
Country	Flag	100	72	0
Volatility	Range	100	72	0
date	Range	100	72	0
time	Range	100	72	0
Actual_news	Range	81,944	59	13
Previous_ne...	Range	83,333	60	12
Consensus_...	Range	80,556	58	14

Obrázok 4 Tabuľka kvality dát pre databázu správ.

Rozdielny príbeh vidíme v databáze správ, kde atribúty obsahujú prázdne hodnoty. Pri atribútoch Actual\_news a Previous\_news to je skoro 20 %. Po analýze sa zistilo, že ekonomické správy s null hodnotou sú rôzne porady a vyhlásenia centrálnych bánk. Vplyv týchto správ nevieme podchytiť, keďže vyvolávajú nepravidelnú volatilitu na trhu a trvajú spravidla len niekoľko hodín.

### 2.1.3. Predspracovanie dát

#### 2.1.3.1. Čistenie dát

##### Vyplniť chýbajúce hodnoty

Množstvo chýbajúcich hodnôt sa vyriešilo pomocou nahradenia hodnoty null za neutral z databázy správ. Výsledkom bolo iba 72 validných záznamov, čo je dôsledkom selektovania správ z hľadiska volatility, ktorú prinášajú (pracujeme so správami, ktoré vždy ovplyvňujú trh). Výsledný počet záznamov je nízky, pretože väčšina takto dôležitých správ sa zverejňuje raz za mesiac a v našom prípade sledujeme správy v 5-mesačnom intervale.

##### Odstránenie výrazne odchýlených hodnôt

Atribút Volume, ktorý obsahoval najväčší počet odchýlených hodnôt sa odstránil. Tento atribút nemá výpovednú hodnotu, čo vyplýva z podstaty fungovania forexového trhu, ktorý nemá burzu, ktorá by objem obchodov počítala. Jednotlivé hodnoty v databázach sú len objemy brokera (XTB) od ktorého sú dáta, čo je z reálneho objemu len čiastková hodnota.

##### Vyhľadzovanie šumu

Aby sme sa vyhli šumu, ktorý vychádza z nesprávneho určenia hodinového trendu a to vtedy, keď otváracia a zatváracia cena sú veľmi blízko seba, musíme určiť pre každý menový pár minimálne cenové rozpätie. Minimálne cenové rozpätie je teda číslo, ktoré je naviazané na menový pár a udáva o minimálne koľko sa musí líšiť otváracia cena od zatváracej v absolútnej hodnote, aby sa mohol určiť rastúci alebo klesajúci trend. Ak je teda minimálne cenové rozpätie menšie, tak sa určí neutrálny trend.

$$MCR = \frac{\sum_{i=1}^n \frac{|o_i - c_i|}{n}}{5}$$

MCR= minimálne cenové rozpätie

n= počet príkladov

$o_i$ = otváracia cena menového páru v danej hodine  $i$

$c_i$  zatváracia cena menového páru v danej hodine  $i$

Po vyhladení tohto šumu sa zlepšil výsledok celého algoritmu o 7,02% z 89,16% na 96,18% na trénovacej množine.

Ďalej sa použil binning na vytvorený atribút `range_mean`, kde sa vytvorili 4 koše.

### 2.1.3.2. Transformácia

Pre ďalšiu prácu bolo potrebné vytvoriť niekoľko nových atribútov. Pre uľahčenie sa ďalej používa názov mena všade tam, kde popis platí pre všetky menové páry (EURUSD, GBPUSD, AUDUSD, NZDUSD, USDCHF, USDCAD, USDJPY).

Názov nového atribútu	Odvođený od	Typ nového atribútu	Popis nového atribútu	Vzorec
<b>trend_mena</b>	open,close	Flag	Určuje trend v danej hodine.	ak <code>open &lt; close</code> potom <code>up</code> inak <code>down</code>
<b>mena_range</b>	open,close	Numeric	Rozdiel medzi otváracou a zatváracou cenou.	<code>abs(open-close)</code>

<b>strength/weakness</b>	trend_mena	Set	Popisuje silu meny v danej hodine.	ak (trendGBPUSD = "up" and trendEURUSD = "up" and trendAUDUSD = "up" and trendNZDUSD = "up" and trendUSDCAD = "down" and trendUSDCHF = "down" and trendUSDJPY = "down") potom weak inak ak (trendGBPUSD = "down" and trendEURUSD = "down" and trendAUDUSD = "down" and trendNZDUSD = "down" and trendUSDCAD = "up" and trendUSDCHF = "up" and trendUSDJPY = "up") potom strong inak no
<b>Count CC</b>	trend_mena	Set	Count CC je bodové ohodnotenie sily trendu. Môže nadobúdať hodnoty od -7 do 7. Každá trend_mena obsahuje hodnotu -1 alebo 1 ,podľa toho kde smeruje. Súčtom dostaneme silu meny.	ak trendGBPUSD +trendEURUSD+trendAUDUSD+trendNZDUSD+trendUSDCAD+trendUSDCHF+trendUSDJPY=7 tak 7 atď.
<b>result</b>	strength/weakness	Set	Určuje výsledok predpovede (cieľový atribút). Ak sme určili že USD je silný a následnú hodinu sa to potvrdilo, tak sme boli úspešný, inak neúspešný.	ak strength/weakness=strength and strength/weaknessNextHour=strength potom úspech inak strata.
<b>range_mean</b>	mena_range	Numeric	Aritmetický priemer rozpätia všetkých menových párov v danej hodine.	(P_GBPUSDrange+P_EURUSDrange+P_AUDUSDrange+P_NZDUSDrange+P_USDC

				ADrange+P_USDCHFrangle +P_USDJPYrange)/7
<b>time of the day</b>	time	Set	Rozdelenie dňa na 4 časti(ráno, poobede, večer, noc)	ak time>10 and time<=16 potom poobedie...atď.
<b>Result_news</b>	Actual_news ,Consensus_ news	Set	Určuje, či daná správa má z teoretického hľadiska oslabovať alebo posilňovať menu.	Záleží od správy: Napríklad. Ak správa= Unemployment rate a Actual_news>Consensus_news potom bad for currency.  Inak Ak správa= Unemployment rate a Actual_news <Consensus_news potom good for currency.  Inak neutral. Pozri 1.5 a 2.1.7. z kapitoly 2.1.6.
<b>range_mean_BIN</b>	range_mean	Set	Priemerné rozpätie vyhladené do 4 košov.	No. of bins 4.
<b>date(news databáza)</b>	datetime	Set	Vytiahnutý dátum z poľa, ktoré obsahovalo aj dátum a čas pohromade.	--
<b>time(news databáza)</b>	datetime	Set	Vytiahnutý čas z poľa, ktoré obsahovalo aj dátum a čas pohromade.	--
<b>MA3_mena(3-hodinový kľzavý priemer)</b>	close	Flag	Určuje, či 3-hodinový kľzavý priemer je pod alebo nad cenou.	Ak MA3 >= open potom above inak under.  $MA3=(close\_1+close\_2+close\_3)/3$

<b>MA5_mena(5-hodinový kľzavý priemer)</b>	close	Flag	Určuje, či 5-hodinový kľzavý priemer je pod alebo nad cenou.	Ak $MA5 \geq open$ potom above inak under.  $MA5 = (close\_1 + close\_2 + close\_3 + close\_4 + close\_5) / 10$
<b>MA10_mena(10-hodinový kľzavý priemer)</b>	close	Flag	Určuje, či 10-hodinový kľzavý priemer je pod alebo nad cenou.	Ak $MA10 \geq open$ potom above inak under.  $MA10 = (close\_1 + close\_2 + close\_3 + close\_4 + close\_5 + close\_6 + close\_7 + close\_8 + close\_9 + close\_10) / 10$
<b>UBB_mena(horná hranica Bollingerovho pásma)</b>	close	Flag	Určuje, či horná hranica Bollingerovho pásma je pod alebo nad cenou.	Ak $open \leq (MA10 + (sdev\_n(@FIELDS\_BETWEEN(close\_1, close\_10)) * 2))$ potom above inak under.
<b>LBB_mena(dolná hranica Bollingerovho pásma)</b>	close	Flag	Určuje, či dolná hranica Bollingerovho pásma je pod alebo nad cenou.	Ak $open > (MA10 - (sdev\_n(@FIELDS\_BETWEEN(close\_1, close\_10)) * 2))$ potom above inak under.
<b>strength/weakness_1</b>	strength/weakness	Set	Budúca hodnota.	-
<b>MCR(minimálne cenové rozpätie)</b>	close, open	Numeric	Určuje minimálne cenové rozpätie pre menový pár. Menšie hodnoty ako táto značia šum.	Pozri kapitolu 2.1.3.1 (Vyhladzovanie šumu)
<b>Count CC_1</b>	trend_mena	Set	Predošlá hodnota Count CC.	ak trendGBPUSD+trendEURUSD+trendAUDUSD+trendNZDUSD+trend

				USDCAD+trendUSDCHF+tr endUSDJPY=7 tak 7 atď.
<b>trend_mena_1</b>	open,close	Flag	Predošlá hodnota atribútu trend_mena.	ak open<close potom up inak down

Tabuľka 9 Vytvorené atribúty.

### 2.1.3.3. Redukcia dát

#### Redukcia dimenzií

Odstránil sa celý atribút Volume, ktorého dáta boli nepoužiteľné. Po vytvorení atribútu trend\_mena sa odstránili atribúty close, open, minimum a maximum. Po vytvorení atribútu date a time sa z databázy správ odstránil atribút DateTime. Po vytvorení atribútu range\_mean sa odstránil atribút mena\_range. Atribút Country obsahuje iba hodnotu United States, pretože sledujeme iba správy, ktoré globálne ovplyvňujú americký dolár. Tieto správy sú z kategórie, ktoré najviac ovplyvňujú vývoj meny (Volatility=3) a preto aj atribút Volatility má iba jednu hodnotu. Najskôr sme uvažovali zaradenie aj správ s nižšou volatilitou, ale ukázalo sa, že vplyvom veľkého šumu (veľa prázdnych hodnôt) veľmi skresľovali výsledky a vyšlo, že správy neovplyvňujú vývoj meny, čo sa po odstránení takýchto zašumených dát ukázalo ako nezmysel. Teda atribúty Country a Volatility sa odstránili.

Odstránili sa symbolické atribúty, ktoré neboli závislé na cieľovom atribúte Result. Pozri kapitolu 2.3

#### Redukcia početnosti dát

Do ďalšej fázy sa nedostali tie záznamy z databázy ekonomických správ, ktoré obsahovali neutral hodnotu v atribúte Result\_news (porady a vyhlásenia centrálnych bánk).



### 2.1.3.4. Výsledok predspracovania dát

Výsledok predspracovania pre databázu menových párov je na obrázku č. 3 a pre databázu správ na obrázku č.4.




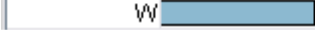
Field	Type	Values	Missing	Check	Direction
Count CC	Set	-7,-6,-5,-4,-3,-2,-1,0...		None	In
Count CC__1	Set	-7,-6,-5,-4,-3,-2,-1,0...		None	In
date	Range	[2014-10-01,2015-...		None	None
LBB_AUDUSD	Flag	under/above		None	None
LBB_EURUSD	Flag	under/above		None	In
LBB_GBPUSD	Flag	under/above		None	None
LBB_NZDUSD	Flag	under/above		None	None
LBB_USDCAD	Flag	under/above		None	None
LBB_USDCHF	Flag	under/above		None	In
LBB_USDJPY	Flag	under/above		None	None
MA10_AUDUSD	Flag	under/above		None	None
MA10_EURUSD	Flag	under/above		None	None
MA10_GBPUSD	Flag	under/above		None	None
MA10_NZDUSD	Flag	under/above		None	In
MA10_USDCAD	Flag	under/above		None	None
MA10_USDCHF	Flag	under/above		None	In
MA10_USDJPY	Flag	under/above		None	None
MA3_AUDUSD	Flag	under/above		None	None
MA3_EURUSD	Flag	under/above		None	In
MA3_GBPUSD	Flag	under/above		None	In
MA3_NZDUSD	Flag	under/above		None	None
MA3_USDCAD	Flag	under/above		None	None
MA3_USDCHF	Flag	under/above		None	In
MA3_USDJPY	Flag	under/above		None	None
MA5_AUDUSD	Flag	under/above		None	In
MA5_EURUSD	Flag	under/above		None	In
MA5_GBPUSD	Flag	under/above		None	In
MA5_NZDUSD	Flag	under/above		None	None
MA5_USDCAD	Flag	under/above		None	None
MA5_USDCHF	Flag	under/above		None	In
MA5_USDJPY	Flag	under/above		None	None
range_mean	Range	[1 2857142857163...		None	In
range_mean_BIN	Set	1,2,3,4		None	None
result	Set	Loss,S,W,null		None	Out
sila/oslabovanie	Set	no,strength,weakn...		None	None
sila/oslabovanie__1	Set	no,strength,weakn...		None	None
time	Range	[00:00:00,23:00:00]		None	None
time of the day	Set	noc,poobedie,rano...		None	In
trendAUDUSD	Set	"-1","0","1"		None	In
trendAUDUSD__1	Set	"-1","0","1"		None	In
trendEURUSD	Set	"-1","0","1"		None	In
trendEURUSD__1	Set	"-1","0","1"		None	In
trendGBPUSD	Set	"-1","0","1"		None	In
trendGBPUSD__1	Set	"-1","0","1"		None	In
trendNZDUSD	Set	"-1","0","1"		None	In
trendNZDUSD__1	Set	"-1","0","1"		None	In
trendUSDCAD	Set	"-1","0","1"		None	In
trendUSDCAD__1	Set	"-1","0","1"		None	In
trendUSDCHF	Set	"-1","0","1"		None	In
trendUSDCHF__1	Set	"-1","0","1"		None	In
trendUSDJPY	Set	"-1","0","1"		None	In
trendUSDJPY__1	Set	"-1","0","1"		None	In
UBB_AUDUSD	Flag	under/above		None	In
UBB_EURUSD	Flag	under/above		None	None
UBB_GBPUSD	Flag	under/above		None	None
UBB_NZDUSD	Flag	under/above		None	In
UBB_USDCAD	Flag	under/above		None	None
UBB_USDCHF	Flag	under/above		None	In
UBB_USDJPY	Flag	under/above		None	None

Obrázok 5 Výsledok predspracovania dát pre databázu menových párov.

Field	Type	Values	Missing	Check	Direction
Name	Set	"Consumer Conf..."		None	In
Country	Flag	"United States"/..."		None	In
Volatility	Range	[2,3]		None	In
date	Set	"2014-01-02","20..."		None	In
time	Set	"13:00:00","14:00..."		None	In
Actual_news	Range	[-18.3,353.0]		None	In
Previous_ne...	Range	[-18.3,351.0]		None	In
Consensus_...	Range	[-18.0,338.0]		None	In
wysledok	Set	bad,default,good...		None	In

Obrázok 6 Výsledok predspracovania dát pre databázu správ.

Výsledný atribút result je nevyvážený (Obr.č.5). Preto sa pred modelovaním spraví over-sampling v tréningovej množine.

Value	Proportion	%	Count
Loss		40,65	1050
null		13,67	353
S		24,12	623
W		21,56	557

Obrázok 7 Početnosť záznamov výsledného atribútu result.

Aby bola množina dát vyvážená, bolo potrebné vynásobiť null 3-krát, W (weakness) 1.8-krát a S (strenght) 1.8-krát.

#### 2.1.4. Návrh experimentov

##### Určenie modelovacích techník

Pre modelovanie sme vybrali rozhodovacie stromy CART a C5.0. Nasledujúca tabuľka obsahuje návrh experimentov s danými modelmi. Jednotlivé modely sa trénujú na tréningových dátach a následne sa otestuje ich presnosť na testovacích dátach. Percentuálne rozdelenie množín dát som stanovil na 80 % z celkových dát na tréningovanie modelov a 20 % na testovanie.

Experiment	Model	Predspracovanie dát	Použité dáta	Zvyšovanie presnosti
1	C5.0	-	Všetky	-
2	C5.0	Redukcia atribútov	Odstránené všetky okrem trend_mena	-
3	C5.0	-	Všetky	Boosting-10
4	CART	-	Všetky	-
5	CART	Redukcia atribútov	Odstránené všetky okrem trend_mena	-

Tabuľka 10 Návrh experimentov.

### Určenie vyhodnocovacej techniky

Pre určenie, ktorý z klasifikátorov je lepší sme porovnali ich presnosť zatriedenia prípadov na testovacej množine. Pričom platí že ten klasifikátor, ktorý má väčšiu presnosť je lepší. Výsledky najlepšieho klasifikátora sa potom porovnávajú s cieľmi a zhodnotí sa, či sme splnili úlohu a celkové dolovanie v dátach malo význam.

#### 2.1.5. Výsledky experimentov

Experiment	Model	Predspracovanie dát	Použité dáta	Zvyšovanie presnosti	Výsledky na testovacích dátach
1	C5.0	-	Všetky	-	54.55%
2	C5.0	Redukcia	Odstránené	-	53.25%

		atribútov	všetky okrem trend_mena		
3	C5.0	-	Všetky	Boosting-10	56.03%
4	CART	-	Všetky	-	52,08%
5	CART	Redukcia atribútov	Odstránené všetky okrem trend_mena	-	44,16%

Tabuľka 11 Výsledky modelovania.

### 2.1.6. Vyhodnotenie experimentov

Našiel sa model, ktorý kategorizuje testovacie príklady s pravdepodobnosťou 55,66%. Modely vytvorené rozhodovacím stromom C5.0 boli výkonovo lepšie ako modely vytvorené pomocou CART.

Boosting, ktorý bol nastavený na hodnotu 10 marginálne zlepšil predpoveď. Ukázalo sa, že prísnu redukciou atribútov sa zlepšila predpoveď a že vytvorené atribúty celkovo zväčšili presnosť.

### 2.1.7. Analýza ekonomických správ

Na otázku, na ktoré ekonomické správy sa vieme spoľahnúť, teda vieme dostatočne predpovedať ich vplyv na trh musíme vykonať analýzu a zhodnotiť navýšenie rizika, ktoré nastane pri zaradení správ do rozhodovacieho procesu.

Sledovali sme americké ekonomické správy, konkrétne ISM Manufacturing PMI, Nonfarm Payrolls, Unemployment Rate, ISM Non-Manufacturing PMI, Retail Sales, Reuters/Michigan Consumer Sentiment Index, Consumer Price Index, Consumer Price Index Ex Food & Energy, Durable Goods Orders, Consumer Confidence, Fed Interest Rate De, Gross Domestic Product Annualized v období od 2.1.2014 do 31.12.2015. Pomocou experimentov sme zistili, že približne 60% času sa reálny vývoj meny rovnal teoretickému následku ekonomickej správy.

	Nazov spravy	pocet spravne predpovedanych	pocet celkovych sprav	percentulne vyjadrenie
1	Consumer Confidence	14	24	58.333
2	Consumer Price Index (YoY)	5	15	33.333
3	Consumer Price Index Ex Food & Energy (YoY)	5	14	35.714
4	Durable Goods Orders	15	22	68.182
5	Fed Interest Rate De	3	3	100.000
6	Gross Domestic Product Annualized	13	21	61.905
7	ISM Manufacturing PMI	13	22	59.091
8	ISM Non-Manufacturing PMI	17	23	73.913
9	Nonfarm Payrolls	13	24	54.167
10	Retail Sales (MoM)	14	22	63.636
11	Reuters/Michigan Consumer Sentiment Index	21	34	61.765
12	Unemployment Rate	12	17	70.588

Obrázok 8 Počet správne odhadnutých teoretických následkov k reálnym.145 241

Z predchádzajúceho obrázku je vidieť, že so sto percentnou pravdepodobnosťou sa vieme spoľahnúť iba na vyhlásenia o úrokovej miere od amerického Fedu. To znamená, že ak by malo vyhlásenie o úrokovej miere oslabiť americký dolár, tak trh zareaguje a americký dolár bude so sto percentnou pravdepodobnosťou oslabovať. So skoro sedemdesiat percentnou pravdepodobnosťou vieme predpovedať mieru nezamestnanosti, objednávky tovarov dlhodobej spotreby a ISM sektor služieb.

Táto analýza pomohla v tom, že selektuje správy s volatilitou 3 (najdôležitejšie správy) na tie, ktoré sa oplatí sledovať a tie, ktoré sa neoplatí. Z hľadiska praktického využitia je jasné, že ak sa odporúča obchody pomocou ekonomických správ, tak sa zvýši riziko, pretože pri nákupe alebo predaji pol hodinu pred a pol hodinu po vyhlásení sa zvyšuje spread niekoľko násobne a to veľká volatilita spôsobí, že sa obchody nedokážu otvárať a ani uzatvárať na stanovených cenách (vzniká tzv. slippage).

### 2.1.8. Analýza výsledku

Nasledujúci obrázok ukazuje chybovú maticu (confusion matrix), ktorá nám pomôže lepšie analyzovať výsledok.

'Partition'	1_Training		2_Testing	
Correct	1 966	96,18%	302	56,03%
Wrong	78	3,82%	237	43,97%
Total	2 044		539	

Coincidence Matrix for \$C-result (rows show actuals)

'Partition'= 1_Training	Loss	S	W	null
Loss	767	47	25	0
S	4	477	0	0
W	2	0	434	0
null	0	0	0	288
'Partition'= 2_Testing	Loss	S	W	null
Loss	102	48	46	15
S	41	88	0	13
W	39	0	77	5
null	15	9	6	35

Obrázok 9 Chybová matica.

Chybová matica ukazuje v koľkých prípadoch systém predpovedal korektné vzhľadom na správnu odpoveď. Stĺpce chybovej matice ukazujú dosiahnuté výsledky a riadky predpovedané výsledky. Napríklad v testovacej množine systém predpovedal 88-krát správne, že americký dolár bude posilňovať a 57-krát nesprávne.

Pomocou dolovania znalostí sme dosiahli presnosť 56.03% na testovacej množine. Takáto presnosť je typická pre daný typ dát. Schopnosť systému správne odporúčať závisí z veľkej časti na pomere medzi rizikom a profitom a nie až tak na konkrétnej presnosti. V ďalšej časti sa budeme zaoberať práve týmto aspektom obchodovania.

## 2.2. Predstavenie konceptu odporúčacieho systému

Odporúčací systém alebo expertný systém (EA - Expert advisor) pre obchodovanie na Forexe je systém, ktorý vplyvom zmeny cien na trhu odporúča, poprípade vykonáva obchody na základe určitého rozhodovacieho procesu.

Systém môže pracovať semi-automaticky alebo plne automaticky. V prvom prípade sa užívateľ zapája do procesu rozhodovania a robí konečné rozhodnutia, pričom v druhom prípade systém pracuje nezávisle od rozhodovania užívateľa [14].

Odporúčací systém sme vytvorili pomocou programovacieho jazyka Mql4 v MetaEditore. Výhodou jazyka Mql4 oproti ostatným jazykom je skutočnosť, že je zostavený tak, aby priamo podporoval tvorbu obchodných systémov [14]. Vytvorený EA sa spúšťa v obchodnej platforme MetaTrader, pomocou ktorej sa posielajú žiadosti o otvorenie a zatvorenie obchodov priamo brokerovi.

EA sa skladá z troch hlavných funkcií, inicializačnej, deinicializačnej a tikovej (tick - zmena ceny). Pomocou inicializačnej a deinicializačnej funkcie sa nastaví vstupné a výstupné parametre poprípade sa spustia alebo uzavrujú podporné programy. V tikovej funkcii sa vykonáva jadro programu, rozhodovanie, váhovanie pravidiel, porovnávanie jednotlivých faktorov, nákup, predaj a iné. Hlavná funkcia sa vykonáva len pri zmene ceny, čo znamená, že program pracuje síce iteratívne ale nespúšťa sa v rovnakých časových okamihoch.

Pomocou dolovania znalostí z kapitoly 2.1.6 na strane 44 sme vytvorili model, ktorého výsledkom sú pravidlá pre štyri typy obchodov, a síce stratové, neziskové, ziskové, pomocou ktorých sa predáva mena (americký dolár oslabuje) a ziskové pomocou ktorých sa nakupuje mena (americký dolár posilňuje). Z hľadiska obchodovania nás zaujímajú len posledné dva typy obchodov, preto pravidlá pre prvé dva typy neuvažujeme. Po usporiadaní pravidiel z hľadiska spoľahlivosti máme teda jadro rozhodovacieho procesu v expertnom systéme.

Takýto typ systémov neprežije ak sa nebude prispôsobovať meniacemu sa trhu. Aby sa nemusel zakaždým vytvárať nový klasifikačný model a z neho extrahovať pravidlá, rozhodli sme sa využiť mechanizmus váhovania pravidiel podľa ich úspešnosti. Váha odzrkadľuje spoľahlivosť pravidla v čase a jej zmena nastáva v závislosti od toho, či pravidlo prinieslo zisk alebo stratu. Neúspešné pravidlá zaniknú ak ich váha klesne pod určitú hranicu. Tieto pravidlá sa už nebudú spúšťať a tým sa eliminuje strata a podporí zisk úspešných pravidiel, pretože sa naďalej budú vyhodnocovať iba úspešné pravidlá. Z dlhodobého hľadiska môžu nastať tieto prípady. V prvom prípade sa vytvorí skupina dlhodobo profitabilných pravidiel. V druhom prípade žiadne z pravidiel nebude dlhodobo profitabilné a vplyvom času už nebudú relevantné, teda nastane situácia, kedy je potrebné model znovu natrénovať na dátach, na ktorých bol neúspešný.

V praxi by mal systém vyzerať nasledovne. Obchodník spustí platformu a v nej si nastaví vstupné parametre pre odporúčací systém. Program pri zmene ceny prejde všetkými pravidlami a podľa ich váhy a toho či dané pravidlo platí vykoná nákup v prípade ak sa jedná o pravidlo, ktoré hovorí

o posilňovaní amerického dolára alebo vykoná predaj v prípade ak sa jedná o pravidlo, ktoré hovorí o oslabovaní amerického dolára. Podľa vopred vykonanej optimalizácie rizika a cieľného zisku sa nastaví parametre nákupu a predaja. Po uzavretí obchodu vplyvom dosiahnutia cieľného zisku alebo straty sa zhodnotí pravidlo a aktualizuje sa mu váha. Pri ďalšej zmene ceny sa rozhodovací proces opakuje so zmenenými váhami.

### **2.2.1. Vstupné parametre**

Dôležitou úlohou užívateľa odporúčacieho systému je nastaviť vstupné parametre systému. Pomocou týchto parametrov sa vykonávajú obchody a priamo ovplyvňuje výsledok. Väčšina parametrov sa optimalizuje, no používateľ má možnosť zvoliť niektoré parametre pomocou vlastnej preferencie (napr. nižšie riziko). V ďalšej časti popíšem jednotlivé parametre.

Objem jednotlivých obchodov je množstvo nakupovanej alebo predávanej meny. Určuje sa v tzv. lotoch. Nákup alebo predaj väčšieho objemu meny priamo úmerne zvyšuje riziko, poprípade profit. Z tohto dôvodu je objem obchodu popri cieľnom profite a riziku jeden z najdôležitejších faktorov pri obchodovaní.

Cieľný zisk je množstvo kapitálu, ktoré by používateľ chcel dosiahnuť pri kontrolovanom riziku. Stanovuje sa cena menového páru, ktorá je systémom predpovedaná ako dosiahnuteľná. Určuje množstvo pipov (najmenšia zmena ceny) od otváracej ceny obchodu a uzatváracej ceny obchodu. Rozumný cieľný zisk môže zmeniť neziskovú stratégiu na ziskovú. Hodnota cieľného zisku sa získava pomocou optimalizácie.

Riziko je množstvo kapitálu, ktoré je ochotný používateľ riskovať aby dosiahol zisk. Stanovuje sa trhovú cenu, pri ktorej je jasné, že predpoveď systémom bola nesprávna. Správne nastavenie rizika je jeden z najdôležitejších faktorov obchodovania a väčšinou sa nastavuje optimalizáciou.

Počet paralelných obchodov stanovuje množstvo obchodov, ktoré je možné otvoriť súčasne. V prípade, že v rozhodovacom procese bude viacero pravidiel platných, musíme stanoviť limit, koľko obchodov sa otvorí, poprípade koľko obchodov sa otvorí kvôli platnosti určitého pravidla.

Začiatková váha určuje štartovaciu pozíciu každého pravidla pred váhovaním. Ak sa inicializuje počiatková váha veľmi blízko hranice, kedy sa už neberie pravidlo do rozhodovacieho procesu, môže sa stať, že sa veľké množstvo pravidiel odpíše po jednom negatívnom obchode. Ak sa



stanoví počiatočná váha veľmi ďaleko od tejto hranice, môže nastať prípad, že stratové pravidlá sa ponechajú príliš dlho v rozhodovacom procese a význam váhovania sa stratí.

Hranica je taká deliaca hodnota, ktorá určuje, či sa s pravidlom bude alebo nebude počítat v rozhodovacom procese.

Inkrement váhy určuje prírastok o ktorý sa navýši váha pravidla po profitabilnom obchode vplyvom daného pravidla.

Dekrement váhy určuje úbytok o ktorý sa zníži váha pravidla po stratovom obchode vplyvom daného pravidla.

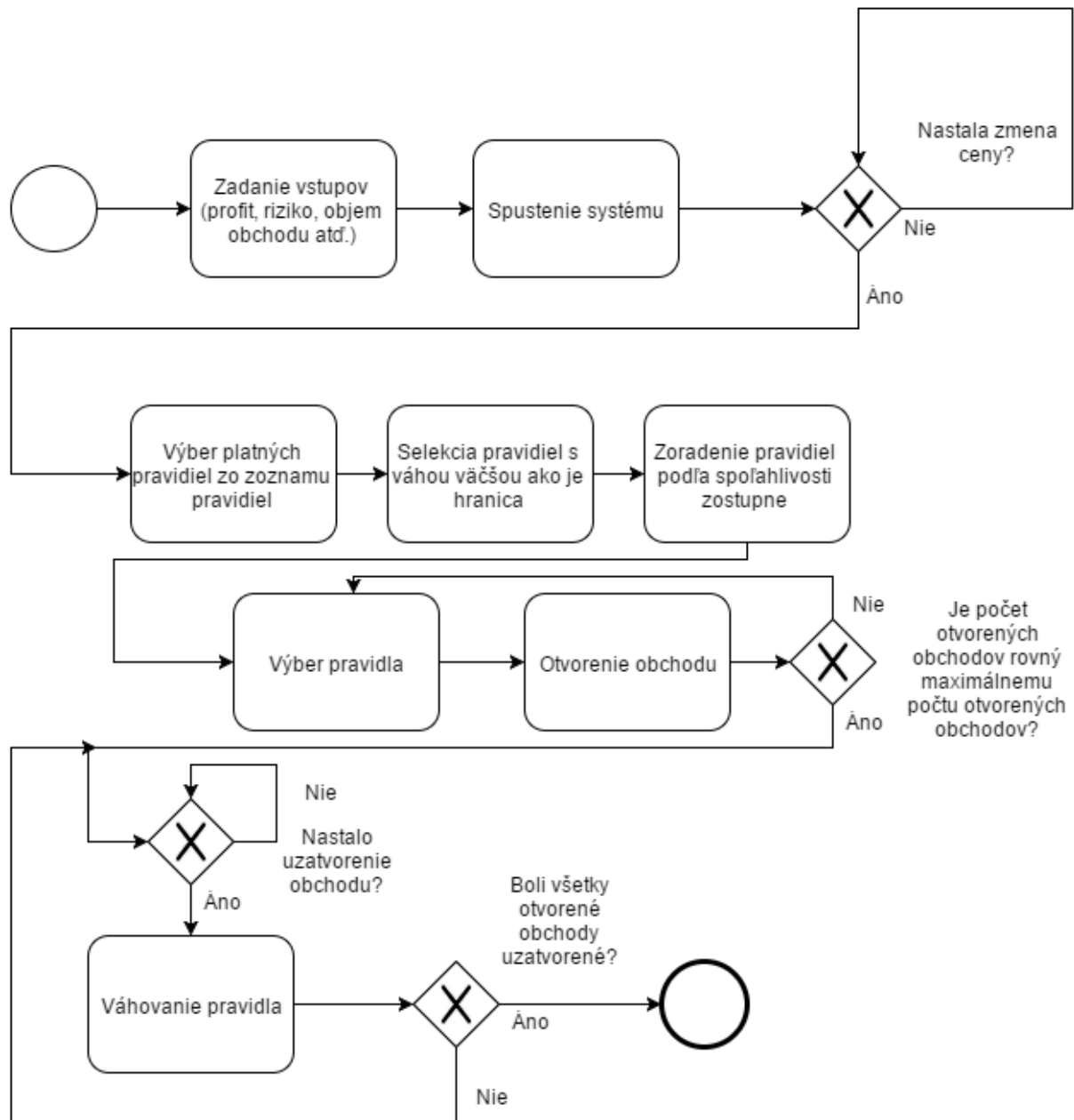
Optimalizácia týchto faktorov je kľúčový proces, ktorý zásadne ovplyvňuje spoľahlivosť pravidiel. Určité pravidlo môže byť profitabilné pri určitých hodnotách týchto faktorov a stratové pri iných.

Na obrázku 11. je zobrazené možné nastavenie odporúčacieho systému.

Premenná	Hodnota	Štart	Krok	Stop
<input checked="" type="checkbox"/> lots	0.1	0.01	0.01	1.0
<input checked="" type="checkbox"/> stopLoss	20	10	10	350
<input checked="" type="checkbox"/> takeProfit	330	10	10	350
<input checked="" type="checkbox"/> maxTrades	1	1	1	3
<input checked="" type="checkbox"/> zaciatočna_vaha	0.5	0.5	0.1	1.0
<input checked="" type="checkbox"/> hranica	0.1	0.1	0.1	1.0
<input checked="" type="checkbox"/> inkrement	0.1	0.1	0.1	1.0
<input checked="" type="checkbox"/> dekrement	0.1	0.1	0.1	1.0

Obrázok 10 Vstupy odporúčacieho systému.

### 2.2.2. Rozhodovací proces

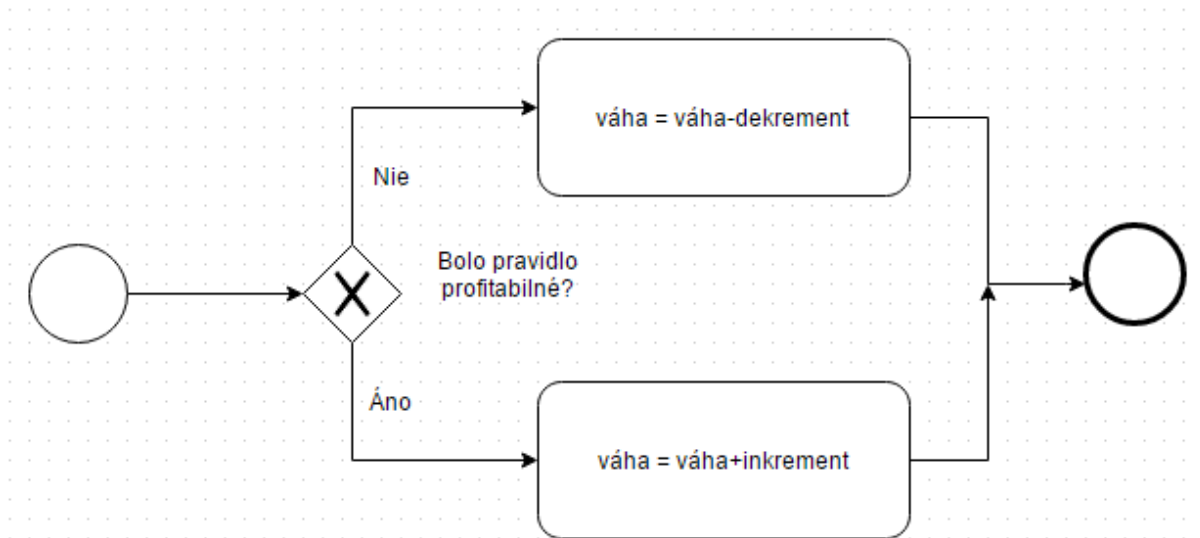


Obrázok 11 Proces rozhodovania.

Odporúčací systém pracuje iteratívne na pozadí platformy. Pri každej zmene ceny sa vykoná celý proces rozhodovania (pozri Obrázok 12). Tento proces sa začína stanovením vstupných parametrov. Po nastavení parametrov sa program dá do spiacieho stavu, kedy čaká na externý faktor. Ak nejaký nastane, tak systém určí zoznam platných pravidiel, vyselektuje tie, ktorých váha

je väčšia ako stanovená hranica a zoradí pravidlá podľa spoľahlivosti zostupne. Pre každé pravidlo potom iteratívne otvorí obchod, pričom počet otvorených obchodov je limitovaný parametrom maximálny počet otvorených obchodov. Po uzavretí obchodu sa vykoná analýza výsledku a určí sa nová váha pravidla. Uzavretie obchodu je jav, ktorý nemusí nastať pri každej iterácii.

### 2.2.3. Váhovanie pravidiel



Obrázok 12 Váhovanie pravidla.

Vo fáze váhovania sa vplyvom výsledku pravidla pripočítava k váhe pravidla inkrement alebo odpočítava dekrement.

Vo fáze testovania sme zvolili nastavenie začiatkovej váhy pravidla na 0.5, hranicu na hodnote 0.1, hodnoty inkrementu a dekrementu zhodne na 0.1.

### 2.2.4. Spoľahlivosť vybraných pravidiel

Pomocou dolovania znalostí sme dostali celkovo 263 pravidiel (pozri stranu 44), ktoré vieme rozdeliť do 4 skupín:

1. Pravidlá, ktoré opisujú stratové obchody (80 pravidiel).
2. Pravidlá, ktoré popisujú nerozhodnosť amerického dolára (55 pravidiel).
3. Pravidlá, ktoré popisujú oslabovanie amerického dolára (59 pravidiel).
4. Pravidlá, ktoré popisujú posilňovanie amerického dolára (69 pravidiel).

Pre našu prácu, pravidiel prvého a druhého typu nie sú žiadnym prínosom, pretože my hľadáme pravidlá, ktoré hovoria o tendencii amerického dolára pokračovať v trende.

Je dôležité si uvedomiť, že množstvo otvorených obchodov priamo ovplyvňuje množstvo pravidiel, pri ktorých bude systém sledovať ich platnosť. Keďže jednotlivé obchody nie sú limitované časom ale cieľným ziskom a rizikom, môže nastať prípad, že sa dosiahne maximálny počet otvorených obchodov a teda ďalší obchod sa neotvorí aj keď bude jedno z pravidiel platné. Z tohto dôvodu vyberieme 10 najlepších pravidiel zo skupín 3 a 4.

V tabuľkách 13 a 14 je zoznam ziskových pravidiel z oboch dôležitých skupín 3 a 4, pričom váhovanie je nastavené pre všetky pravidlá rovnako (hranica = 0.1, začiatková váha = 0.5, inkrement = 0.1, dekrement = 0.1, maximálny počet otvorených obchodov = 1 a objem obchodu = 0.1 lotu). Optimalizáciu sme robili na dátach EURUSD od 1.7.2015 do 31.10.2015.

Pravidlo	Riziko(pip)	Cieľný zisk(pip)	Najväčší pokles kapitálu (EUR)	Počet otvorených obchodov	Zisk (EUR)
1	350	90	251.44	6	478.21
7	10	350	127.18	7	256.60
9	240	240	168.17	1	219.39
3	260	330	624.22	5	124.92
4	340	10	208.12	14	120.53
5	10	50	0	1	45.28
8	320	10	12.49	1	9.02

Tabuľka 12 Ziskové pravidlá z kategórie popisujúcej posilňovanie amerického doláru.

Po optimalizácii vstupných parametrov sa do tabuľky 13 nedostali pravidlá 2, 6 a 10, pretože mali iba stratové výsledky. Do rozhodovacieho procesu sa teda nedostanú stratové pravidlá a pravidlá,

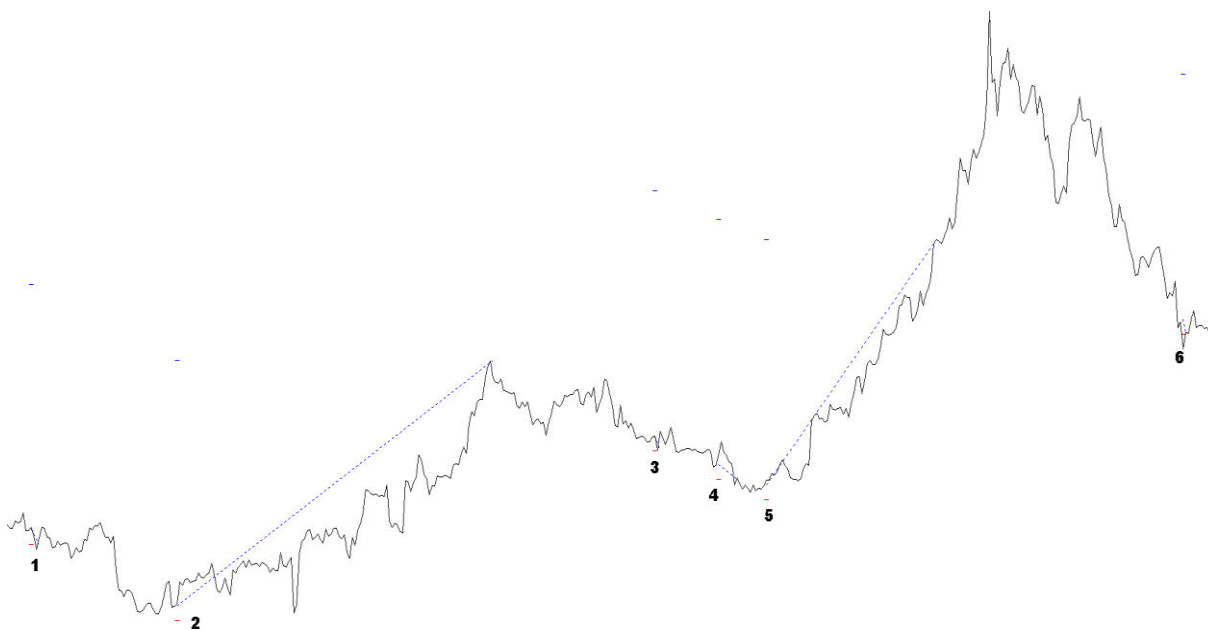
ktorých frekvencia platnosti je nízka (pravidlo 9, 5 a 8). Pravidlo 3 má vysoký pokles kapitálu takže sa s ním tiež nebude počítať.

Pravidlo	Riziko(pip)	Cielený profit(pip)	Najväčší pokles kapitálu(EUR)	Počet otvorených obchodov	Zisk (EUR)
8	20	330	181.93	19	927.47
5	230	280	352.37	5	474.93
3	230	160	346.46	7	458.02
10	280	160	139.14	3	417.43
1	310	160	367.16	6	394.38
6	170	160	349.73	13	351.63
2	120	280	347.69	10	311
4	330	10	236.59	20	165.66
9	130	230	139.83	2	73.91

Tabuľka 13 Ziskové pravidlá z kategórie popisujúcej oslabovanie amerického doláru.

V tabuľke 14 chýba pravidlo 7, ktoré bolo celkovo stratové aj po optimalizácii parametrov. Pravidlo 8 je jednoznačne najlepším pravidlom v tejto kategórii, pretože má vysokú frekvenciu platnosti, nadpriemerný zisk, rozumný pomer medzi cieľným profitom a rizikom a relatívne nízky pokles kapitálu. Pravidlo 9 sa odstráni z rozhodovacieho procesu kvôli nízkej frekvencii a nízkemu profitu oproti ostatným pravidlám. Výsledkom je zoznam pravidiel, ktoré prinášajú zisk, pri ktorých je pokles kapitálu v norme a frekvencia platnosti postačujúca.

### 2.2.5. Príklad najlepšieho pravidla



Obrázok 13 Jednotlivé obchody pravidla 8 z obdobia 31.7.2015 do 28.8.2015 na menovom páre EURUSD.

Na obrázku č.14 je zobrazených šesť obchodov, ktoré odporučilo pravidlo 8 (pozri stranu 44) z kategórie, ktorá hovorí o oslabovaní amerického dolára. Na obrázku vidíme dva ziskové obchody (2 a 5) a štyri neziskové obchody (1, 3, 4, 6). Pravidlo 8 je významné, pretože v každom obchode sa trend trhu zmenil aspoň krátkodobou. V dvoch ziskových prípadoch sa vplyvom tohto pravidla otvoril obchod v najlepšiu možnú chvíľu. Výhodou tohto pravidla je postačujúca frekvencia platnosti a veľmi nízke riziko na jeden obchod.

Pravidlo 8 je v zásade veľmi jednoduché. Sleduje globálny trend amerického dolára, trend menového páru GBPUSD a kĺzavý priemer menového páru NZDUSD vypočítaný z desiatich zatváracích cien. V procese dolovania znalostí sa tomuto pravidlu udelila spoľahlivosť 0.926.

V nasledujúcej tabuľke sa pozrieme na výkonnosť tohto pravidla podrobnejšie. Sumy sú udávané v eurách.

Počiatočný vklad	10 000
Hrubý zisk	1185.84
Čistý zisk	927.47
Počet obchodov	19
Najväčší pokles kapitálu	181.93
Najväčší pokles kapitálu v %	1.81
Ziskové pozície	5
Najväčší ziskový obchod	292.13
Priemerný ziskový obchod	237.17
Maximálny počet ziskových obchodov nasledujúcich po sebe	1
Hrubá strata	-258.36
Stratové pozície	14
Najväčší stratový obchod	-20.48
Priemerný stratový obchod	-18.45
Maximálny počet stratových obchodov nasledujúcich po sebe	7
Priemerný počet stratových obchodov nasledujúcich po sebe	3

Tabuľka 14 Podrobnejší pohľad na najlepšie pravidlo

Pravidlo 8 (oslabovanie amerického dolára) má približne 26 % šancu, že prinesie zisk. Po ziskovom obchode by priemerne mali nasledovať 3 stratové. Dôsledkom vysokého pomeru rizika a cieľného profitu 1:16.5 sa zabezpečí ziskovosť pravidla.

### 2.2.6. Modifikácia výberu pravidiel

Pomocou zmeny výberu pravidiel a odstránenie tých, ktoré nie sú profitabilné podľa kapitoly 2.2.4 sme získali upravenú množinu pravidiel. V nasledujúcej tabuľke je výkon odporúčacieho systému po optimalizácii pred modifikáciou výberu pravidiel a po nej z obdobia 1.11.2015 po 22.2.2016 na menovom páre EURUSD.

	Prvotný výber pravidiel	Modifikovaný výber pravidiel
Hrubý zisk	896.61	572.95
Čistý celkový zisk	361.17	276.66
Všetky pozície	62	71
Maximálny drawdown v %	1.47	0.92
Ziskové pozície v % z celku	14.52	88.73
Výhry nasledujúce po sebe	1	7
Prehry nasledujúce po sebe	6	1
Riziko v pipoch	10	40
Očakávaný zisk v pipoch	100	10

Tabuľka 15 Dosiahnuté výsledky pred a po modifikácii pravidiel.

Pri optimalizácii sa určila maximálna hodnota rizika na 50 pipov a zisk na 150, pričom váhy sa zafixovali na hodnotách uvedených v kapitole 2.2.3.

Pri pôvodných pravidlách sa zabezpečí väčší čistý celkový zisk. Nevýhodou využívania týchto pravidiel je ich percentuálna úspešnosť 14.52%. Takýto systém je závislý na zopár vysoko profitabilných obchodoch. Keďže cieľom je nájsť spoľahlivý systém, tak priaznivejší systém je systém s modifikovanými pravidlami, pomocou ktorého dosiahneme úspešnosť 88.73%.

### 2.2.7. Analýza výkonnosti systému týždenne

Na to aby sme mohli dôkladnejšie zmerať výkonnosť systému je potrebné rozdeliť sledovaný časový interval na menšie časti. Obdobie od 1.11.2015 do 22.2.2016 rozdelíme na 16 týždňov. Táto analýza nám poskytne údaje o konzistencii výkonnosti systému a distribúciu frekvencií obchodov.

Týždeň	Čistý celkový zisk	Všetky pozície	Ziskové pozície v %
1.11-8.11	30.09	5	80
9.11-15.11	-9.47	4	75
16.11-22.11	-18.77	8	75
23.11-29.11	27.68	3	100
30.11-6.12	17.54	7	85.71
7.12-13.12	-56.12	5	40



14.12-20.12	63.31	7	100
21.12-27.12	0	0	-
28.12-3.1	8.90	6	83.33
4.1-10.1	27.53	3	100
11.1-17.1	36.73	4	100
18.1-24.1	-18.18	3	66.67
25.1-31.1	48.15	6	100
1.2-7.2	54.37	6	100
8.2-14.2	17.76	2	100
15.2-22.2	26.93	3	100

Tabuľka 16 Analýza výkonnosti systému týždenne.

Z predchádzajúcej tabuľky môžeme vidieť, že frekvencia obchodov sa v jednotlivých týždňoch pohybuje od dvoch obchodov po osem, ak nevezmeme do úvahy Vianoce. Jedenásť týždňov zo šestnástich bolo ziskových. Dalo by sa povedať, že okrem týždňa v dátume od 7.12 do 12.12 kde systém odporučil len 40% obchodov správne, bol systém úspešný. Rozdiel medzi čistým celkovým ziskom v tabuľke Tabuľka 15 a sumou čistých celkových ziskov v jednotlivých týždňoch v tabuľke Tabuľka 16 sú obchody, ktoré sa otvorili na konci týždňa a ukončili na začiatku nového týždňa.

## Záver

Predstavu systému, ktorú sme navrhli v kapitole 2.2 sme realizovali a otestovali a ciele, ktoré sme si stanovili v kapitole 2.1.1 sme úspešne splnili.

Vlastným prínosom v tejto práci bolo navrhnutie pohľadu na devízové dáta iným spôsobom a tento pohľad pretaviť do rozhodovacích pravidiel, ktoré brali do úvahy rôzne aspekty makro a mikro analýzy danej meny. Z rozhodovacích pravidiel sme následne navrhli odporúčací systém a optimalizovali jeho výsledky, pričom dlhšiu životnosť systému sme zabezpečili adaptáciou systému vplyvom zmeny váh pravidiel podľa ich výkonu.

V kapitole 2.1.1 sme stanovili hypotézu, že na devízovom trhu sa dá určiť smer meny podľa pohybu menových párov, v ktorých daná mena vystupuje. Výsledky z predchádzajúcej kapitoly ukazujú, že má význam určovať globálny smer meny, teda určovať, či mena je globálne silná alebo slabá voči ostatným majoritným menám, a že znalosť tejto informácie vplyva kladne na predpoveď vývoja meny.

Keďže výsledok tejto práce je tak praktický ako sa len dá, nie je úplne jasný ďalší krok vo vývoji tohto systému. Systém dokáže nielen odporúčať možné príležitosti užívateľovi ale aj automaticky obchodovať bez intervencie človeka. Prínosom pre systém by mohlo byť pridanie rôznych iných indikátorov, ale to by aj znamenalo vytvorenie komplexnejších pravidiel. Samozrejme veľká zmena by nastala pri zmene rozhodovacieho stromu na iný klasifikačný nástroj. Dôvodom použitia rozhodovacieho stromu bola ľahká čitateľnosť a rýchle pochopenie znalosti, ktoré pravidlá nesú.

## Zoznam použitej literatúry

- [1]. Paralič, Ján: Objavovanie znalostí v databázach[online]. Košice, 2003, Dostupné na internete: < <http://people.tuke.sk/jan.paralic/knihy/ObjavovanieZnalostivDB.pdf> >.
- [2]. Bank of international settlements: Trennial Central Bank survey [online]. Dostupné na internete:<<http://www.bis.org/publ/rpfx13fx.pdf>>.
- [3]. Oscar Marbán, Gonzolo Mariscal, Javier Segovia: A Data Mining & Knowledge Discovery Process Model [online]. Madrid, 2009, Dostupné na internete: < [http://cdn.intechopen.com/pdfs/5937/InTech-A\\_data\\_mining\\_amp\\_knowledge\\_discovery\\_process\\_model.pdf](http://cdn.intechopen.com/pdfs/5937/InTech-A_data_mining_amp_knowledge_discovery_process_model.pdf)>.
- [4]. Rüdiger Wirth, Jochen Hipp: CRISP-DM: Towards a Standard Process Model for Data Mining [online]. Nemecko, Dostupné na internete: < <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.198.5133&rep=rep1&type=pdf>>.
- [5]. J. Ross Quinlan: C 4.5: Programs for Machine Learning. California, 1993, ISBN 1-55860-238-0
- [6]. Machová, Kristína: Strojové učenie: Princípy a algoritmy[online]. Košice, 2002, Dostupné na internete: < <http://people.tuke.sk/kristina.machova/pdf/SU4.pdf> >.
- [7]. Rulequest Research: C5.0 and C4.5 comparison[online]. Február 2012, Dostupné na internete: < <http://www.rulequest.com/see5-comparison.html> >.
- [8]. Samuel A. Moore, Daniel M. D'Addario, James Kurinkas, Gary M. Weiss: Are Decision Trees Always Greener on the Open (Source) Side of the Fence [online]. New York, Dostupné na internete: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.158.4350&rep=rep1&type=pdf>
- [9]. Breiman, L., Friedman, J.H., Olshen, R., and Stone, C.J.: Classification and Regression Tree, California, 1984
- [10]. Fredrik Larsen: Automatic stock market trading based on Technical Analysis [online]. Trondheim, 2007, Dostupné na internete: < <http://www.diva-portal.org/smash/get/diva2:347439/FULLTEXT01.pdf> >.

- 
- [11]. K.K. Lai, Lean Yu, Shouyang Wang: A Neural Network and Web-Based Decision Support System for Forex Forecasting and Trading [online]. Hong Kong, 2004, Dostupné na internete: < <http://web.cenet.org.cn/upfile/50892.pdf> >.
- [12]. Jaymin R. Mehta, Marcus D. Menghini , Daniel A. Sarafconn: Automated Foreign Exchange Trading System [online]. Worcester, 2011, Dostupné na internete: <[https://www.wpi.edu/Pubs/E-project/Available/E-project-052311-230105/unrestricted/Automated\\_Foreign\\_Exchange\\_Trading\\_System.pdf](https://www.wpi.edu/Pubs/E-project/Available/E-project-052311-230105/unrestricted/Automated_Foreign_Exchange_Trading_System.pdf)>.
- [13]. Forex Factory Inc.: Usual Effect [online]. Dostupné na internete: <<http://www.forexfactory.com/calendar.php> >.
- [14]. Mql5 Ltd: Expert Advisor [online] Dostupné na internete: < <http://docs.mql4.com/>>.
- [15]. John A. Bollinger: Bollinger On Bollinger Bands, New York, 2001, ISBN 0-07-137368-3

## Prílohy

Príloha A: CD médium – diplomová práca v elektronickej podobe, prílohy v elektronickej podobe, aplikácia, súbory obsahujúce algoritmy v Clementine, dáta použité v objavovaní znalostí

Príloha B: Systémová príručka

Príloha C: Používateľská príručka