

Univerzita Pardubice
Fakulta ekonomicko-správní

Modelování predikce časové řady návštěvnosti web domény pomocí
dopředných neuronových sítí

Bc. Veronika Buriánková

Diplomová práce

2011

ZADÁNÍ DIPLOMOVÉ PRÁCE

(PROJEKTU, UMĚLECKÉHO DÍLA, UMĚLECKÉHO VÝKONU)

Jméno a příjmení: **Bc. Veronika BURIÁNKOVÁ**
Osobní číslo: **E090483**
Studijní program: **N6209 Systémové inženýrství a informatika**
Studijní obor: **Informatika ve veřejné správě**
Název tématu: **Modelování predikce časové řady návštěvnosti web domény pomocí dopředných neuronových sítí**
Zadávací katedra: **Ústav systémového inženýrství a informatiky**

Z á s a d y p r o v y p r a c o v á n í :

Analyzujte vstupní data (parametry) pro následující predikci.
Charakterizujte dopředné neuronové sítě z hlediska aproximace a predikce.
Navrhňte model na predikci návštěvnosti web domény.
Verifikujte navržený model.
Uskutečňte analýzu výsledků.

Rozsah grafických prací:

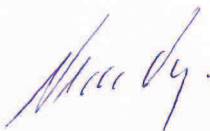
Rozsah pracovní zprávy:

Forma zpracování diplomové práce: **tištěná/elektronická**

Seznam odborné literatury:

KVASNIČKA, V. a kol.: Úvod do teorie neuronových sítí. Iris, Bratislava, 1997.

HAYKIN, S.: Neural Networks: A Comprehensive Foundation. 2nd edition, New Jersey, Prentice-Hall, Inc., 1999, 842s.



Vedoucí diplomové práce:

prof. Ing. Vladimír Olej, CSc.

Ústav systémového inženýrství a informatiky

Datum zadání diplomové práce: **5. října 2010**

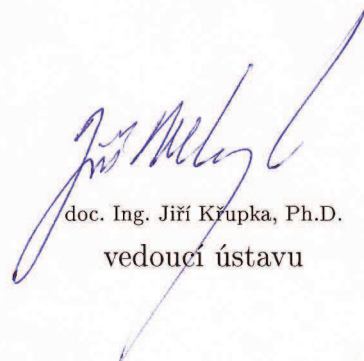
Termín odevzdání diplomové práce: **6. května 2011**



doc. Ing. Renáta Myšková, Ph.D.

děkanka

L.S.



doc. Ing. Jiří Křupka, Ph.D.

vedoucí ústavu

V Pardubicích dne 5. října 2010

Prohlášení autora

Prohlašuji:

Tuto práci jsem vypracovala samostatně. Veškeré literární prameny a informace, které jsem v práci využila, jsou uvedeny v seznamu použité literatury.

Byla jsem seznámena s tím, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorský zákon, zejména se skutečností, že Univerzita Pardubice má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle § 60 odst. 1 autorského zákona, a s tím, že pokud dojde k užití této práce mnou nebo bude poskytnuta licence o užití jinému subjektu, je Univerzita Pardubice oprávněna ode mne požadovat přiměřený příspěvek na úhradu nákladů, které na vytvoření díla vynaložila, a to podle okolností až do jejich skutečné výše.

Souhlasím s prezenčním zpřístupněním své práce v Univerzitní knihovně.

V Pardubicích dne 28.3.2011

Bc. Veronika Buriánková

Poděkování

Touto cestou bych ráda poděkovala vedoucímu mé práce prof. Ing. Vladimírovi Olejovi, CSc. za jeho cenné rady, připomínky a jeho podporu při vypracování této diplomové práce. Dále bych ráda poděkovala rodičům za jejich celoživotní podporu, trpělivost a také za to, že mi umožnili studovat.

ANOTACE

Diplomová práce se zabývá návrhem modelu pro predikci návštěvnosti web domény upce.cz pomocí dopředných neuronových sítí. V první části je popsána základní charakteristika web miningu a zásadní poznatky z oblasti dopředných neuronových sítí. Další část práce zahrnuje popis předzpracování dat. Pro učení neuronové sítě je použita časová řada návštěvnosti webu upce.cz, která je pro potřeby učení upravena pomocí indikátorů technické analýzy. Poslední částí je samotné učení neuronové sítě a následná analýza získaných výsledků.

KLÍČOVÁ SLOVA

web mining, dopředná neuronová síť, frontální neuronová síť, časová řada, predikce, JavaNNS

TITLE

Modelling of Prediction Time Series Visit of Web Domains by Feed-forward Neural Networks

ANNOTATION

This thesis describes the design model for predicting visit of web domains by feed-forward neural networks. The first section describes the basic characteristics of web mining and fundamental knowledge of the feed-forward neural networks. Another part includes a description of the preprocessing. For learning of the neural network is used time series visit of web domains upce.cz, which is adjusted using the learning needs of technical analysis indicators. The last part is the learning of neural network and the subsequent analysis of the results.

KEYWORDS

web mining, feed-forward neural network, frontal neural network, time series, prediction, JavaNNS

Obsah

| | |
|---|----|
| Obsah | 7 |
| Úvod | 9 |
| 1 Web mining | 10 |
| 1.1 Web Usage Mining | 11 |
| 1.2 Web Structure Mining | 15 |
| 1.3 Web Content Mining | 16 |
| 1.4 Metody a algoritmy modelování v oblasti web miningu | 17 |
| 1.5 Dílčí závěr | 19 |
| 2 Dopředné neuronové sítě | 21 |
| 2.1 Základní charakteristika | 21 |
| 2.2 Proces učení | 23 |
| 2.3 Neuronová síť jako univerzální aproximátor | 25 |
| 2.4 Frontální neuronová síť | 27 |
| 2.5 Dílčí závěr | 29 |
| 3 Návrh modelu pro predikci návštěvnosti webu upce.cz | 30 |
| 3.1 Získání vstupních dat | 32 |
| 3.2 Předzpracování dat | 34 |
| 3.3 Rozdělení dat na trénovací a testovací množinu | 35 |
| 3.4 Návrh struktury neuronové sítě | 37 |
| 3.5 Formát dat | 38 |
| 3.6 Proces učení neuronové sítě | 38 |
| 3.7 Dílčí závěr | 41 |
| 4 Analýza navrženého modelu | 42 |
| 4.1 Analýza výsledků krátkodobé časové řady | 42 |
| 4.2 Analýza výsledků střednědobé časové řady | 43 |
| 4.3 Analýza výsledků dlouhodobé časové řady | 44 |
| 4.4 Porovnání jednotlivých časových řad | 45 |
| 4.5 Dílčí závěr | 48 |
| 5 Návrh a analýza frontální neuronové sítě | 49 |
| 5.1 Analýza výsledků krátkodobé časové řady | 50 |
| 5.2 Analýza výsledků střednědobé časové řady | 51 |

| | | |
|-----|---|----|
| 5.3 | Analýza výsledků dlouhodobé časové řady..... | 51 |
| 5.4 | Porovnání jednotlivých časových řad..... | 52 |
| 5.5 | Porovnání dopředné a frontální neuronové sítě..... | 55 |
| 5.6 | Dílčí závěr..... | 56 |
| | Závěr..... | 58 |
| | Seznam použité literatury..... | 59 |
| | Seznam použitých zkratk..... | 61 |
| | Seznam obrázků..... | 62 |
| | Seznam tabulek..... | 63 |
| | Seznam příloh..... | 63 |
| | Příloha 1 – Indikátory technické analýzy..... | 64 |
| | Příloha 2 – Rozdělení dat na trénovací a testovací množinu..... | 69 |
| | Příloha 3 – Výsledky učení krátkodobé časové řady..... | 71 |
| | Příloha 4 – Výsledky učení střednědobé časové řady..... | 76 |
| | Příloha 5 – Výsledky učení dlouhodobé časové řady..... | 81 |
| | Příloha 6 – Výsledky učení frontální neuronové sítě..... | 86 |

Úvod

Dnešní moderní společnost a rychlý růst informačních technologií podporuje prudký nárůst internetových aplikací. Různé internetové aplikace, zejména elektronické podnikání a obchodování poskytuje řadu otevřených, či skrytých informací, které jsou následně využívány pro další analýzy. Obor, který se zabývá vyhledáváním a zpracováním informací v prostředí webu, se nazývá web mining.

Diplomová práce je rozdělena do 5 kapitol. V první kapitole je uvedena základní charakteristika a objasnění pojmu web mining. Podrobně jsou popsány jednotlivé části web miningu a metody a algoritmy, které jsou typické pro jeho využití.

Další kapitola charakterizuje dopředné neuronové sítě a objasňuje proces učení v těchto sítích. V této kapitole je také uvedena základní charakteristika frontální neuronové sítě.

Třetí kapitola popisuje zpracování a přípravu dat návštěvnosti webu upce.cz pro modelování. Definiuje algoritmus postupu. Zaměřuje se na použití indikátorů technické analýzy a rozdělení dat na trénovací a testovací množiny.

Čtvrtá kapitola obsahuje návrh modelu pro predikci návštěvnosti webu upce.cz pomocí klasické dopředné neuronové sítě. Snaží se nalézt vhodnou strukturu neuronové sítě s vhodnými parametry.

Poslední kapitola se zabývá návrhem modelu pro predikci pomocí frontální neuronové sítě. V závěru je popsána analýza získaných výsledků a porovnání mezi klasickou dopřednou a frontální neuronovou sítí.

Cílem diplomové práce je popis základní charakteristiky web miningu, návrh vhodné struktury dopředné neuronové sítě pro predikci návštěvnosti webu upce.cz a interpretace získaných výsledků.

1 Web mining

Web mining je poměrně nová, stále se vyvíjející disciplína, která využívá a zpracovává nejrůznější informace z prostředí webu. Je soustředěn na nejdynamičtěji se rozvíjející zdroj informací současnosti, na Internet a především na službu world wide web (www) [11]. Pomocí různých programových prostředků se snaží nalézt požadované informace, jejich vzájemné vztahy a závislosti, které mohou být dále využity pro modelování či analýzu definovaného problému.

Web mining v českém překladu jako dolování dat na webu, je proces výběru, prohledávání a modelování ve velkých objemech dat sloužících k odhalení dříve neznámých vztahů mezi daty za účelem získání obchodní výhody [1]. Web mining je tedy hojně využíván zejména v soukromém sektoru. Zdroje dat, ze kterých jsou čerpány dané informace, jsou zejména [1]:

- webové logovací soubory
- statistiky aplikací
- statistiky provozu sítí
- statistiky provozu serverů.

Webové logovací soubory, ze kterých se získává nejvíce informací o daných uživateli, mohou pocházet z několika zdrojů [3]:

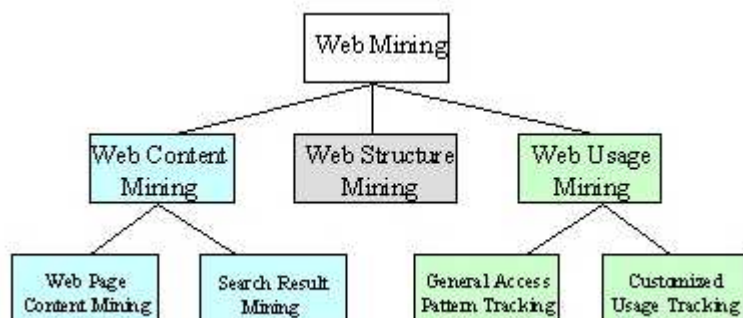
- Server www – soubory zahrnují údaje o přístupech různých uživatelů na stránky právě tohoto serveru.
- Klient www (prohlížeč) – pro tento účel je obvykle potřeba samotného klienta upravit. Metoda umožňuje shromáždit údaje o uživateli jediného počítače. Zahrnuti jsou také přístupy do celého prostoru www.
- Proxy server – ukládány jsou údaje o všech uživateli daného proxy serveru a všech jejich přístupech na internetové stránky.

První přístup je typický pro podnikové systémy a pro business intelligence systémy využívané v elektronickém obchodování. Analýza se soustředí pouze na jediný server, ve kterém zkoumá chování uživatelů. Výhodou tohoto přístupu bývá dobrá znalost struktury zkoumaného webového prostoru (jediný server). Další postupy, které jsou uvedeny pod třetí odrážkou, poskytují všechny potřebné údaje pro podporu konkrétních uživatelů a neomezuje se pouze na jeden webový server [3].

Web mining se snaží získávat informace především o návštěvnících webových stránek v soukromém sektoru, tedy zejména o zákaznících, kde zaznamenává jednotlivé kroky a vlastnosti každého uživatele. Web mining zde analyzuje především [2]:

- odkud návštěvníci přicházejí
- jak se návštěvníci chovají
- jaké jsou typické sekvence průchodu stránkami
- při jaké sekvenci průchodů došlo ke koupi či rezervaci
- jak dlouho se návštěvníci zdržují na stránkách
- jak a odkud dané stránky opouštějí.

Taxonomii web miningu zachycuje obr. 1-1.



Obr. 1-1: Taxonomie web miningu, zdroj [10]

Web mining je dělen na tři části [10]:

- Web Usage Mining (WUM) – analýza chování uživatele
 - General Access Pattern Tracking – obecně přístupná metoda sledování
 - Customized Usage Tracking – využití vlastního sledování
- Web Structure Mining (WSM) – získávání informací ze struktury www prostoru
- Web Content Mining (WCM) – zpracování obsahu www stránek
 - Web Page Content Mining – dolování dat z obsahu webové stránky
 - Search Result Mining – výsledky vyhledávání.

1.1 Web Usage Mining

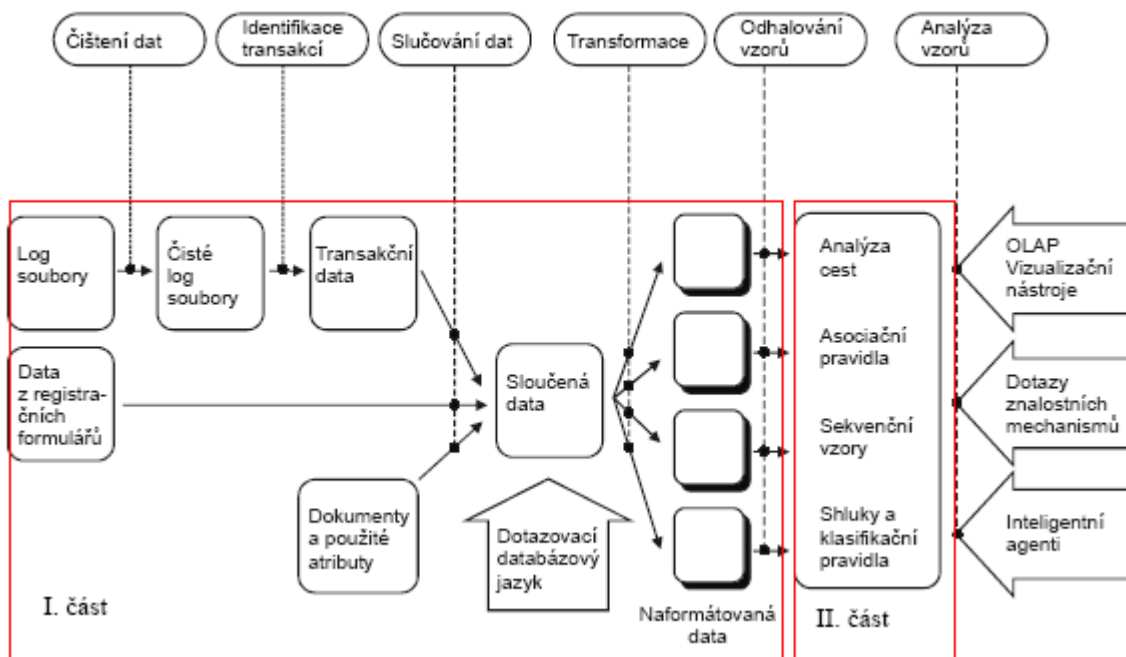
Web Usage Mining získává údaje o klientech na jednom nebo více webových místech. Typickými zdroji dat jsou [4]:

- automaticky generované údaje uložené na serveru v podobě záznamů o přístupech, záznamů o doporučení, klientských záznamů a cookies
- uživatelské profily
- meta data – atributy stránek, obsah atributů a využití dat.

Mezi nejčastěji využívané techniky pro WUM patří [5],[6] statistické metody, asociační pravidla a metody segmentace – ta se používá pro seskupování uživatelů s podobnými vzory chování nebo se seskupují stránky, které jsou navštívené stejnou skupinou uživatelů.

Další používanou technikou může být [3] např. metoda FGS (Frequent Generalized Sessions). Metoda umožňuje z údajů o připojení uživatele k síti extrahovat obvyklé vzory jeho chování. Vstupem do této metody je soubor vektorů, jejichž souřadnicemi jsou navštívené URL adresy seřazené podle pořadí, v němž byla navštívena. Výstupem budou n-prvkové posloupnosti, kde každá definuje jeden často se opakující vzor chování uživatele. Prvky uvedených posloupností jsou buď konkrétní URL nebo zástupný znak „*“ označující jeden nebo více navštívených URL. Zástupný znak ovšem nesmí být na první a poslední pozici v posloupnosti. Algoritmus metody FGS spočívá v iterativní tvorbě množiny posloupností od nejkratších ($n=1$) po nejdelší (maximální hodnotu n zadává uživatel). Potom každá posloupnost zařazená do výstupní množiny musí splnit podmínku překročení určitého prahu v počtu výskytů ve vstupních datech. Výsledné posloupnosti lze využívat při predikci chování uživatele.

Web Usage Mining získává data z Log souborů nebo pomocí kombinací cookies a JavaScriptu. Architektura WUM se zdrojem dat z Log souboru se dělí na dvě části. První část tvoří transformaci webových dat do podoby využitelné pro zpracování. Druhá část obsahuje konkrétní metody data miningu (DM) jako součást systémového DM. Architekturu zachycuje obr. 1-2 [5].



Obr. 1-2: Architektura WUM se zdrojem dat z Log souboru, zdroj [5]

Log soubor – obsahuje záznamy o všech požadavcích, které server zpracoval. Standardní formát Log souboru je CLF (Common Log Format). V tomto souboru je zaznamenána IP adresa, datum a čas návštěvy, frekvence návštěv a další informace [5].

Čištění dat – je prvním krokem prováděným ve WUM, který se snaží eliminovat vliv irelevantních položek na výsledek analýzy.

Identifikace transakcí – vytváří smysluplné uskupení odkazů pro každého uživatele.

Analýza cest – analyzuje nejkratší cesty, kterými uživatelé procházejí.

Odhalování vzorů – patří mezi hlavní části WUM, do které směřují algoritmy a techniky z DM, statistiky a strojového učení. Zjišťují jednotlivé vzory chování.

Analýzy vzorů – konečná fáze WUM. Hlavním cílem tohoto procesu je odstranit nadbytečná pravidla a vzory.

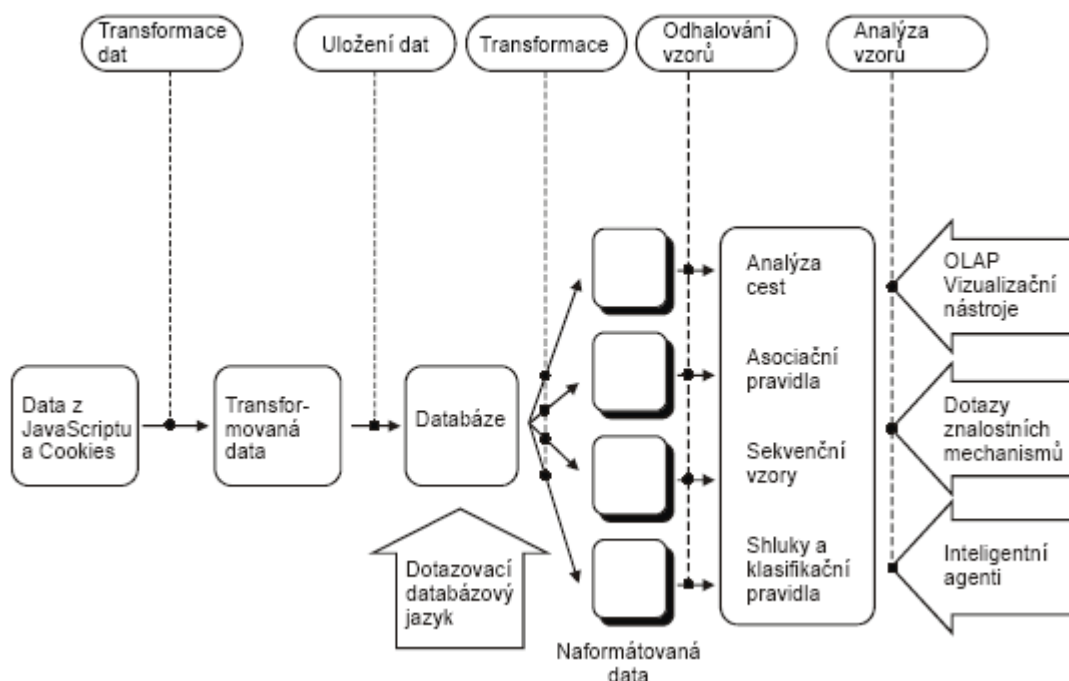
Asociační pravidla – technika používaná k objevování závislostí mezi nalezenými položkami v transakčních databázích.

Sekvenční vzory – pomáhají předpovídat budoucí vývoj.

Shluky a klasifikační pravidla – shlukování je technika používaná pro seskupování datových položek s podobnými charakteristikami a klasifikace rozděluje datové položky do jedné z několika předem definovaných tříd [6].

Získávání dat pomocí JavaScriptu a cookies umožňují získání podrobnějších informací o uživateli a umožňují provádět hlubší analýzu. Data z JavaScriptu jsou např. informace o předchozí navštívené stránce, prohlížeči či operačním systému uživatele.

Z cookies je možné sledovat unikátní kroky uživatele (např. při nákupu, uložení zboží do nákupního košíku a následné zaplacení u pokladny). Všechny údaje jsou agregovány do databáze na straně serveru. Nevýhodou bývá velký počet operací s databází. Tuto technologii sběru zachycuje obr. 1-3.



Obr. 1-3: Architektura WUM se zdrojem dat z JavaScriptu a cookies, zdroj [5]

Tento model je daleko jednodušší než předcházející, protože neobsahuje komplikovanou transformaci webových dat a následné formátování, pouze se napojuje do databáze např. pomocí SQL [5].

Hlavní aplikace WUM jsou [9]:

- oslovení potenciálních zákazníků pro elektronickou komerci
- zvýšení kvality a poskytování internetových služeb informování koncových uživatelů
- zlepšení výkonosti webového serveru
- identifikování potenciální reklamní lokality
- umožnění personalizace/ adaptivnosti webových stránek
- zlepšení designu stránek
- detekce průniku a podvodů
- předvídání akcí uživatelů (umožnění předběžného načítání).

1.2 Web Structure Mining

Web Structure Mining je proces objevování struktury z webových stránek. Může být prováděn buď na úrovni dokumentu, nebo na úrovni hypertextového odkazu. Ten slouží k dvěma hlavním účelům: buď jako čistá navigace nebo přejítí na stránky se stejným tématem, která obsahuje daný odkaz [4].

Web Structure Mining je zaměřen na množinu stránek, počínaje jedinou webovou stránkou až po web jako celek. Důležitým krokem je zde identifikovat relativní důležitost odlišných stránek, které se zobrazují stejně. Jako příklad lze uvést vyhledávače tematických odkazů, které analyzují topografii odkazu jako nalezení autoritativního informačního zdroje, pro obsáhlé vyhledávané téma. Informace je tudíž obsažena na autoritativních stránkách., které jsou definovány v relaci k Hubům. Huby nám představují stránky, které odkazují do mnoha souvisejících autorit.

Např. úspěšný vyhledávač Google využívá algoritmus PageRank, který funguje na principu: důležitost stránky se zvyšuje s tím, kolik stránek je odkazuje z jiných relevantních stránek. Tento algoritmus přiřazuje číselné váhy na webovou stránku. Numerická váha se přiřadí k příslušnému elementu E, který se označuje PR(E) – PageRank E. Hodnota PageRank stránky U závisí na hodnotách PageRank pro každou stránku ve stanoveném BU (tato sada obsahuje všechny stránky odkazující na stránku U), dělený počtem L(V) odkazů od stránky V [7],[8].

Matematický zápis je

$$PR(U) = \sum_{V \in BU} \frac{PR(V)}{L(V)}. \quad (1.1)$$

Používá se pro zvýšení účinnosti vyhledávačů na základě počtu zpětných odkazů [8].

Hojně využívaný nástroj WSM je konstrukce a analýza hierarchických stromových struktur reprezentujících www prostor. Vstupem pro tuto metodu je soubor navštívených URL stránek, výstupem je poté jejich uspořádání do stromové struktury pomocí rozkladu URL na jednotlivé logické části. Každá cesta stromem od kořene k listům vyjadřuje zápis jedné unikátní URL. Uzly mohou nést také statistické informace o počtu návštěv, případně údaje o celkové době strávené na stránkách obsahující danou část URL.

Další data pro následnou WSM analýzu jsou vazby dané stránky na stránky další. Www prostor je modelován jako orientovaný graf, jehož uzly zastupují konkrétní URL adresy

a hrany jejich vzájemná propojení pomocí hypertextových vazeb. Ze vzájemných hypertextových vazeb www stránek lze usuzovat na jejich tematickou podobnost. Základním předpokladem hypertextových odkazů bývá, že propojují stránky mající i obsahovou vazbu [3],[5].

1.3 Web Content Mining

Web Content Mining je proces extrakce užitečné informace z obsahu webových dokumentů (text, hypertext, obrázky, audio, video nebo strukturované záznamy, jako jsou seznamy a tabulky, také několik jednoduchých statistik slov a frází v dokumentu) [4],[6].

Web Content Mining umí pracovat s částečně strukturovanými daty. WCM se pokouší odvozovat závislosti ze struktury webové stránky se záměrem převést ji do databázové struktury, která umožňuje snadnější správu informací a dotazování se nad nimi. Dotazovací mechanismus umožňuje dotaz na místní úložiště na několika úrovních abstrakce. V důsledku tohoto dotazu si může systém vyžádat informace o stránkách z webu podrobněji podle potřeby [6],[7].

Využívají se tedy zejména metody určené pro zpracování dokumentů. Pracuje se zde s pojmem „term“ jako základní jednotkou pro popis dokumentu. Termem se označují jednotlivá slova či víceslovní spojení, která jsou v daném dokumentu významně zastoupena nebo jsou pro dokument charakteristická. Pro zpracování textového podkladu se využívají metody, které používají vektorovou reprezentaci dokumentů. Sada dokumentů a termů tvoří matici, ve které jsou následně uvedeny hodnoty významnosti termů v jednotlivých dokumentech. Významnou oblastí je výběr relevantních termů jednoslovných i víceslovných. Relevance bývá poměřována tzv. TFIDF vahou, která je založena na frekvenci termů v jednotlivých dokumentech a inverzní frekvenci termů v celém souboru dokumentů.

V prostředí webu lze pro extrakci termů využít nejen prostý text obsažený na webových stránkách, ale také další textové informace, např. z hlavičky stránky (titulek, klíčová slova, atd.) nebo jiných URL (texty odkazů ukazujících na danou stránku). Pro vyhledávání příbuzných termů se využívají pravděpodobnostní přístupy, např. bayesovské pravděpodobnostní sítě. Hlavním cílem uživatele na Internetu je nalezení relevantní informace. Cestou k objevení může být i zařazení dané stránky podle klíčových termů do již existující a dostupné ontologické struktury a následné zkoumání „nejbližších sousedů“ v této struktuře. Postup se snaží nahradit chybějící ontologický popis stránky. Výsledek ontologického zařazení se používá pro objevení podobně zaměřených stránek, pro shromáždění více termů, které lze využít ve vyhledávání pomocí některých z běžných

vyhledávacích služeb. Využívají se především nové nástroje, jako rozhraní Google s direktivou define, která rovněž zahrnuje WordNet, ale seskupuje výsledky z více zdrojů [3].

Web Content Mining tedy řeší zejména vyhledávání a meta-vyhledávání (hledání stránek relevantních k dotazu uživatele), kategorizaci dokumentů (shlukování stránek podle obsahu), filtrování (rozpoznání stránek relevantních k uživatelskému profilu) a dobývání znalostí „skrytých ve stránkách“ (extrakce informace nebo zodpovídání dotazu).

Nejpoužívanější světové vyhledávače jsou AltaVista (<http://www.altavista.com>), Yahoo! (<http://www.yahoo.com>). Mezi naše vyhledávače patří Seznam (<http://www.seznam.cz>), Kompas (<http://kompas.seznam.cz>), Atlas (<http://hledej.atlas.cz>), Alenka (<http://www.alenka.cz>), Centrum (<http://www.centrum.cz>). Mezi speciální vyhledávače lze zařadit Ahoy! (<http://ahoy.cs.washington.edu:6060/>), který je určen k vyhledávání domovských stránek osob. Meta-vyhledávací systémy umožňují přístup k mnoha vyhledávacím strojům z jednoho místa, např. All-in-One (<http://www.albany.net/allinone>) – uživatel ručně volí, který vyhledávací stroj použije a musí pro něj zformulovat dotaz. MetaCrawler (<http://www.metacrawler.com>), HuskySearch (<http://huskysearch.es.washington.edu>) – vytváří z nalezených odkazů shluky dokumentů na základě podobnosti témat, SavySearch (<http://guaraldi.cs.colostate.edu>) – rozhoduje, do kterého vyhledávacího stroje má uživatel poslat dotaz, AskJeeves (<http://www.askjeeves.com>) – kombinuje dotazování s vyhledáváním v databázi přímých odpovědí na běžné otázky. Pro extrakci informací se využívají systémy, které procházejí on-line katalogy firem s cílem nalézt cenově nejvýhodnější nabídku daného produktu. Systémy jsou označeny jako shopbots (jde o roboty pohybující se v obchodech). Příkladem lze uvést NetBot Jango (<http://www.jango.com>) – systém se naučí vyhledávat zboží v on-line katalogích, rozpozná název produktu, cenu a další informace a doporučí uživateli nejlevnějšího dodavatele [11].

1.4 Metody a algoritmy modelování v oblasti web miningu

Základní metody a algoritmy modelování v oblasti web miningu, stejně jako u klasického data miningu jsou [11]:

- regresní metody
- diskriminační analýza
- shluková analýza
- tvorba rozhodovacích stromů
- tvorba rozhodovacích pravidel
- tvorba asociačních pravidel

- neuronové sítě
- genetické algoritmy
- bayesovské sítě
- učení založené na analogii
- induktivní logické programování.

Regresní analýza zkoumá parametry závislosti mezi dvěma veličinami. Pro dvojice hodnot $[x_i, y_i]$ můžeme parametry rovnice odhadnout na základě metody nejmenších čtverců. Metoda se snaží minimalizovat rozdíly mezi požadovanou hodnotou y a očekávanou hodnotou $\hat{y} = f(x)$.

Diskriminační analýza hledá závislosti jedné nominální veličiny (určující příslušnost ke třídě) na dalších m numerických veličin. Jedná se o klasifikaci příkladů do předem zadaných tříd.

Shlukování se snaží rozdělit uživatele do jednotlivých shluků podle jejich blízkých vlastností či chování na webu. Nejčastěji je shlukování využíváno v elektronickém obchodování pro analýzu zákazníků. Vzdálenost mezi dvěma shluky je nejčastěji charakterizována euklidovskou vzdáleností

$$d_e(x_1, x_2) = \sqrt{\sum_{j=1}^m (x_{1j} - x_{2j})^2}. \quad (1.2)$$

Nejčastěji využívané metody shlukové analýzy jsou hierarchické shlukování a metoda k-means. Hierarchické shlukování postupuje metodou „zdola nahoru“. Každý uživatel tedy tvoří jeden samostatný shluk. Postupně se shluky spojují, končí se jedním shlukem, který obsahuje všechny uživatele. Spojování shluků se provádí na základě vzdáleností, kterou lze stanovit metodou nejbližšího souseda, metodou nejvzdálenějšího souseda, metodou průměrné vzdálenosti či centroidní metodou. Jednotlivé vztahy jsou uvedeny v [11]. Metoda k-means musí mít předem určený počet shluků. Obecný algoritmus této metody říká: náhodně zvol rozklad do k shluků, urči centroid pro všechny shluky v aktuálním rozkladu. Pro každého uživatele x urči vzdálenosti $d(x, c_k)$, $k = 1, 2, \dots, k$, kde c_k je centroid k -tého shluku. Nechť $d(x, c_l) = \min d(x, c_k)$. Není-li x součástí shluku l (k jehož centroidu c_l má nejbliže), přesuň x do tohoto shluku, došlo-li k nějaké změně, pokračuj druhým krokem, jinak ukonči algoritmus.

Rozhodovací stromy se používají hlavně pro klasifikaci získaných dat do dvou nebo více tříd. Postupuje se zde metodou rozděl a panuj. Trénovací data jsou postupně rozdělována do menších podmnožin. Využívá se algoritmus TDIDT (top down induction of decision trees): zvol jeden atribut jako kořen dílčího stromu, rozděl data v tomto uzlu na podmnožiny podle hodnot zvoleného atributu a přidej uzel pro každou podmnožinu. Existuje-li uzel, pro který nepatří všechna data do téže třídy, pro tento uzel opakuj postup od začátku, jinak skonči.

Asociační pravidla slouží pro hledání vzájemných vazeb mezi různými druhy dat. V závislosti na web miningu se hledají závislosti mezi různými přístupy na server od různých uživatelů. Nejznámějším algoritmem pro hledání asociačních pravidel je algoritmus apriori. Základem algoritmu je hledání často se opakujících množin přístupů na server. Jedná se o kombinace různých kategorií, které dosahují předem zadané četnosti v datech. Příkladem lze uvést např. 60 % studentů, kteří navštívili stránku <http://portal.upce.cz/jetspeed/portal/moje-studium/studijni-materialy.psml>, navštívili také stránku <http://portal.upce.cz/jetspeed/portal/moje-studium/zkousky.psml>.

Neuronové sítě jsou hojně využívanou metodou pro web mining. Používají se klasické dopředbě NS jak pro predikci, tak pro klasifikaci. Pro klasifikaci se využívá převážně metoda Support Vector Machine (SVM) a Kohonenovy samoorganizující se mapy (KSOFM).

Evoluční a genetické algoritmy jsou založeny na myšlence darwinovského principu evoluce. Hledá se optimální řešení na základě soutěže v rámci populace postupně se vyvíjejících řešení. Využívá se zde genetické programování, které využívá operátory selekce, křížení a mutace stejně jako genetické algoritmy.

Mezi metody založené na analogii patří případové usuzování (Case-Based Reasoning, CBR), pravidlo nejbližšího souseda (nearest neighbour rule), učení založené na instancích (Instance-Based Learning, IBL), líné učení (lazy learning), paměťové učení (Memory-Based Learning), shlukování (clustering). Všechny tyto metody lze použít pro deskriptivní úlohy (segmentace, shlukování) i pro klasifikační úlohy [11],[13],[14].

1.5 Dílčí závěr

Web mining je v současné době hojně se rozvíjející disciplína, využívající nespočetně mnoho informací z prostředí webu. Strukturu web miningu tvoří tři hlavní pilíře: Web Usage Mining, Web Structure Mining a Web Content Mining.

Za nejdůležitější část je považován WUM, jehož hlavním cílem je získání všech potřebných informací o uživateli, případně klientech webových stránek.

Web Structure Mining získává informace o jednotlivých internetových stránkách, zjišťuje jejich důležitost a vzájemné propojení stránek pomocí hypertextových odkazů.

Web Content Mining slouží pro analyzování obsahu webových stránek. Využívá nejrůznější metody, které jsou určeny pro zpracování dokumentů.

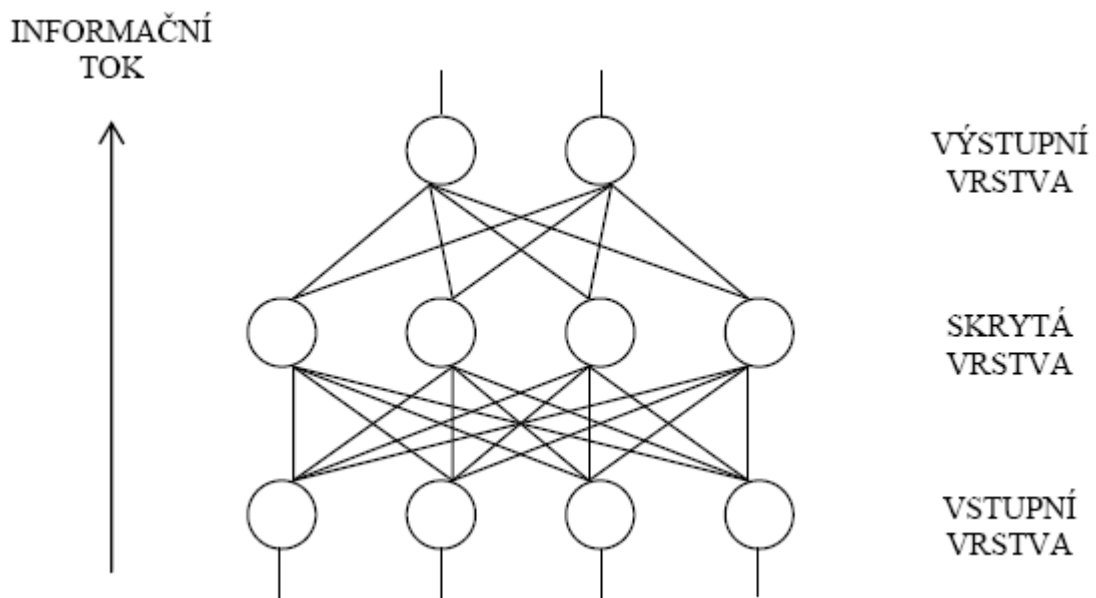
2 Dopředné neuronové sítě

Tato kapitola se zabývá charakteristikou dopředné neuronové sítě a jejích základních pojmů. Definuje strukturu dopředné neuronové sítě a postup jejího učení. Hlavní využití dopředných neuronových sítí je predikce a klasifikace dat, jejichž struktury jsou také zahrnuty v této kapitole. Je zde také uvedena základní charakteristika frontální neuronové sítě.

2.1 Základní charakteristika

Základními stavebními prvky neuronové sítě (NS) jsou neurony, které jsou v síti mezi sebou vzájemně propojeny ohodnocenými váhami. Tyto váhy se nazývají synapsie. NS je popsána jako dvojice množin (U, H) , kde U je množina uzlů (neuronů, vstupů a výstupů) a H je podmnožina množiny $N \times N$. Tato množina určuje synapsie mezi jednotlivými uzly. Synapsím jsou přiřazovány váhy $w \in R$. Každý neuron nám představuje nelineární typicky analogovou jednotku, které jsou málo porovnatelné s moderními číslicovými obvody. Hlavní úlohou neuronu je agregace signálů ze vstupních synapsí a realizace pomocí nelineární funkce [16].

Schéma dopředné NS je možno vidět na obr. 2-1. Jedná se o znázornění trojvrstvé NS s dopředným šířením.



Obr. 2-1: Schéma trojvrstvé NS s dopředným šířením určena pro klasifikaci, zdroj [17]

NS vždy obsahuje vstupní vrstvu, jednu nebo více skrytých vrstev a výstupní vrstvu. Neurony v této struktuře jsou mezi dvěma vrstvami vždy propojeny každý s každým.

Vstupy jsou číselné hodnoty, které mohou být [15]:

- kvalitativní – nabývají hodnot z diskrétních číselných množin, např. (-1,1)
- fuzzy kvalitativní vstupy jako fuzzy hodnoty jazykových proměnných
- kvantitativní – číselné reprezentace systémových proměnných.

Každý vstup je modifikován vahou w_{ij} , která představuje citlivost, s jakou působí na výstup neuronu. Výpočty těchto hodnot a jejich postupné ladění je hlavní částí učících algoritmů sítí. Na samém počátku jsou váhy nastaveny na defaultové hodnoty podle uživatele či automaticky. V celé síti se zpravidla používají stejné typy vstupů a výstupů.

Aktivační funkce agreguje hodnoty vstupů a a posouvá je vzhledem k určitému aktivačnímu prahu. Lineární aktivační funkce je ve tvaru

$$a_i(k) = \sum_{j=1}^n w_{ij}(k) * x_{ij}(k) + x_{io} = \sum_{g=1}^{ng} w_{ig}(k) * y_{ig}(k) + x_{io} = \sum_{g=1}^{ng} w_{ig}(k) * y_{ig}(k), \quad (2.1)$$

kde $a_i(k)$ je hodnota aktivační funkce i -tého neuronu pro k -tý trénovací obrazec, x_{ij} jsou hodnoty vstupů i -tého neuronu, y_{ig} jsou hodnoty výstupů neuronu, $w_{ij}(k)$ jsou váhy na vstupech neuronu, x_{io} je aktivační práh.

Přenosová funkce převádí hodnotu výstupu aktivační funkce do hodnoty výstupu neuronu. Používá se zejména jednoduchá lineární funkce, lineární prahová funkce, skoková funkce, sigmoidální funkce a gaussovská funkce [15].

Výstup dopředné NS lze vyjádřit pomocí vztahu [16]

$$Y = \sum_{k=1}^K \alpha_k \times d \left(\sum_{j=1}^J \beta_{jk} \times X_{jk} \right), \quad (2.2)$$

kde:

- Y je výstup NS,
- α je vektor vah synapsí mezi neurony ve skryté vrstvě a výstupním neuronem,
- β je vektor vah synapsí mezi vstupními neurony a neurony ve skryté vrstvě,

- k je index neuronu ve skryté vrstvě,
- K je počet neuronů ve skryté vrstvě,
- d je aktivační funkce,
- j je index vstupního neuronu,
- J je počet vstupních neuronů na jeden neuron ve skryté vrstvě,
- X je vstupní vektor NS.

2.2 Proces učení

V procesu učení NS dochází k úpravě hodnot vah synapsí mezi neurony. Cílem je vytvořit takový stav NS, v kterém bude schopna klasifikovat všechny vstupní vzorky, nebo je správným způsobem zpracovat. Vzhledem ke složité struktuře NS, v současné době neexistuje žádná známá metoda, která by byla schopná jednoznačně určit hodnoty vah synapsí definované úlohy [16].

Dopřednou neuronovou síť lze klasifikovat jako učení s učitelem. Postup učení lze definovat pomocí následujících procedur [17]:

- Vstupní vzor je představený síti. Vstup je potom šířen dopředně sítí až do okamžiku, než aktivace dosáhne výstupní vrstvy. Tato fáze se nazývá dopředné šíření signálu.
- Výstup z výstupní vrstvy je porovnáván s učicí hodnotou. Chyby δ_j mezi skutečným výstupem o_j a učicí hodnotou t_j neuronu j je užita spolu s výstupem o_i předchozího neuronu i k výpočtu změn vah synapsí w_{ij} . K výpočtu chyb výstupů vnitřních neuronů, pro které není učicí vstup dostupný (neuronů ve skrytých vrstvách) se využívají chyby následující vrstvy δ_k , které jsou již vypočteny. Tyto chyby jsou také zpětně šířeny.
- V průběhu učení jsou, po každém trénovacím vzoru, na síť aplikovány změny vah synapsí w_{ij} . Změny vah synapsí se pro všechny vzory v trénovacím souboru sčítají a součet všech změn je aplikovaný po jednom cyklu na trénovací soubor.

Nejčastěji používaným typem u dopředných NS jsou [16] gradientní metody. Z nichž nejpoužívanější je algoritmus zpětného šíření – Backpropagation algorithm. Je založen na minimalizaci součtu čtverců chyb s využitím poznatků o průběhu nelineární funkce v neuronech. Požaduje, aby nelineární funkce byla spojitá.

V algoritmu Back-propagation (BP) indexy i a j (váhy) představují indexy výstupu neuronu ve vyšší vrstvě a výstupu neuronu v nižší vrstvě. Součet kvadratických odchylek

mezi očekávanými výstupy m neuronů y_i^* ve výstupní vrstvě a skutečnými výstupy neuronů y_i v této vrstvě při přijetí k -tého obrazce se obvykle vyjadřuje ve tvaru

$$E(k) = \frac{1}{2} \sum_1^m (y_i^*(k) - y_i(k))^2 \quad (2.3)$$

a hodnota váhy upravená proti hodnotě $w_{ij}(k)$ o diferenci $\Delta w_{ij}(k)$ má tvar

$$w_{ij}(k+1) = w_{ij}(k) + \Delta w_{ij}(k) . \quad (2.4)$$

Pro výpočet korekce váhy $\Delta w_{ij}(k)$ se používá vztah z gradientových metod optimalizace

$$\Delta w_{ij}(k) = -\eta(\partial E(k) / \partial w_{ij}(k)) , \quad (2.5)$$

kde η je konstanta ($\eta > 0$), která ovlivňuje rychlost učení. Vztah závisí na tvarech aktivační a přenosové funkce i -tého neuronu výstupní vrstvy.

Postup se provádí tak, že se vypočítají odezvy sítě na vstupní obrazec a nastaví se všechny váhy v síti. Přijme se další vstupní obrazec a postup se opakuje. BP umožňuje upravovat váhy po každém vstupu obrazce z trénovací posloupnosti. Tento postup se uplatňuje tak dlouho, dokud suma odchylek $E(k)$ uvažovaná pro všechny obrazce trénovací posloupnosti ($k = 1, 2, \dots, m$) nebude menší než požadovaná přesnost [15].

Existuje celá řada modifikací tohoto algoritmu. Neznámější učící funkce jsou [17] Vanilla Back-propagation, Back-propagation Momentum, Quiet-propagation, atd. Kvalita učení se hodnotí např. pomocí střední čtvercové odchylky MSE. Tuto odchylku lze vypočítat podle vztahu

$$MSE = \frac{\sum_{p \in \text{patterns}} \sum_{j \in \text{output}} (t_{pj} - o_{pj})^2}{n - a} , \quad (2.6)$$

kde t_{pj} je výstup procesu učení dopředné neuronové sítě pro neuron j a vzor p , o_{pj} je skutečný výstup, n je počet vzorů a a je počet volných parametrů dopředné neuronové sítě (tj. počet synapsí mezi neurony) [17].

2.3 Neuronová síť jako univerzální aproximátor

Hecht-Nielsen v roce 1987 jako první ukázal, že trojvrstvé NS s dopředným šířením a s dostatečným počtem skrytých neuronů jsou schopné aproximovat s požadovanou přesností každé spojitě zobrazení.

Pokud máme spojitou funkci F , která nám zobrazuje n -rozměrný prostor R^n na otevřený interval $(0,1)$

$$F : R^n \rightarrow (0,1), \quad (2.7)$$

kde $y = F(x) = f(x_1, x_2, \dots, x_n)$. Trénovací množina A_{train} obsahuje r bodů z n -rozměrného prostoru R^n , $A_{train} = \{x_1, x_2, \dots, x_r\}$. Potom funkce

$$t : R \rightarrow (0,1) \quad (2.8)$$

je přechodová funkce spojitá a monotónně rostoucí, která vyhovuje podmínkám $t(-\infty) = 0, t(\infty) = 1$. Zároveň k této funkci musí existovat funkce inverzní $t^{-1} : (0,1) \rightarrow R$. Platí zde následující věta:

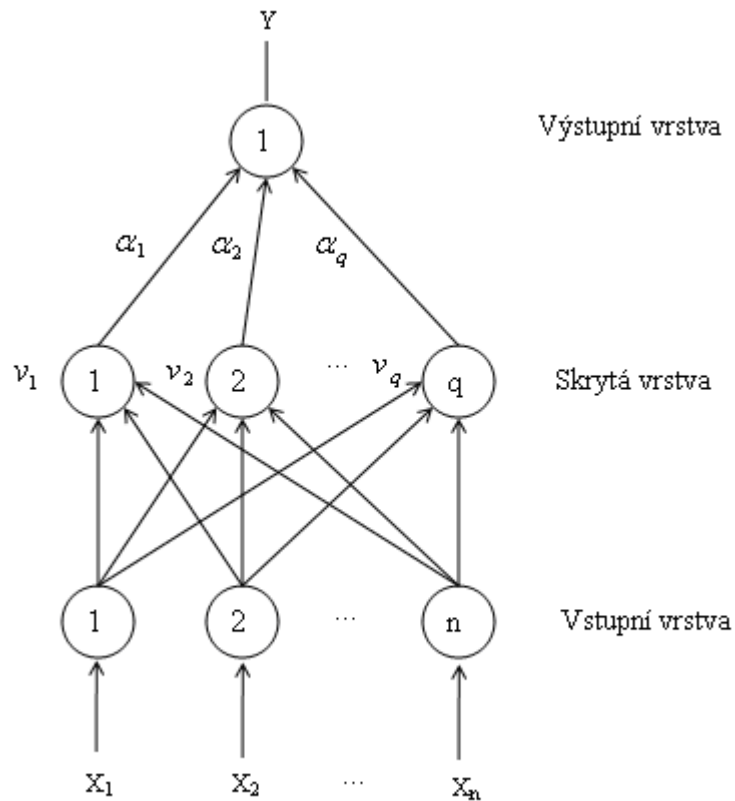
Pro každé $\varepsilon > 0$ existuje funkce

$$G(x) = \sum_{i=1}^q \alpha_i t(v_i + w_i \cdot x), \quad (2.9)$$

kde α_i a v_i jsou reálné koeficienty, $w_i = (w_1^{(i)}, w_2^{(i)}, \dots, w_n^{(i)})$ jsou vektory obsahující n reálných komponent a $x \cdot w_i = x_1 w_1^{(i)} + x_2 w_2^{(i)} + \dots + x_n w_n^{(i)}$ je skalární součin vektorů x a w_i , že

$$\sum_{k=1}^r |F(x_k) - G(x_k)| < \varepsilon. \quad (2.10)$$

Podle výše uvedené věty můžeme tvrdit, že funkce $F(x)$ je aproximovaná s přesností ε nad trénovací množinou A_{train} pomocí funkce $G(x)$ a všeobecné přechodové funkce $t(x)$. Funkce $G(x)$ je jednoduše interpretovatelná NS s dopředným šířením, která bude obsahovat jednu vrstvu q skrytých neuronů. Struktura této sítě je uvedena na obr. 2-2.



Obr. 2-2: NS s dopředným šířením určená pro predikci, zdroj [upraveno podle 12]

Koeficienty α_i jsou váhy spojů mezi skrytými neurony a výstupním neuronem. v_i jsou prahové koeficienty skrytých neuronů a w_i tvoří váhové koeficienty hran mezi i -tým skrytým neuronem a vstupními neurony. Musí zde dojít k modifikaci výše uvedené věty, kde podmínka bude mít tvar

$$\sum_{k=1}^r |F(x_k) - \tilde{G}(x_k)| < \varepsilon' \quad . \quad (2.11)$$

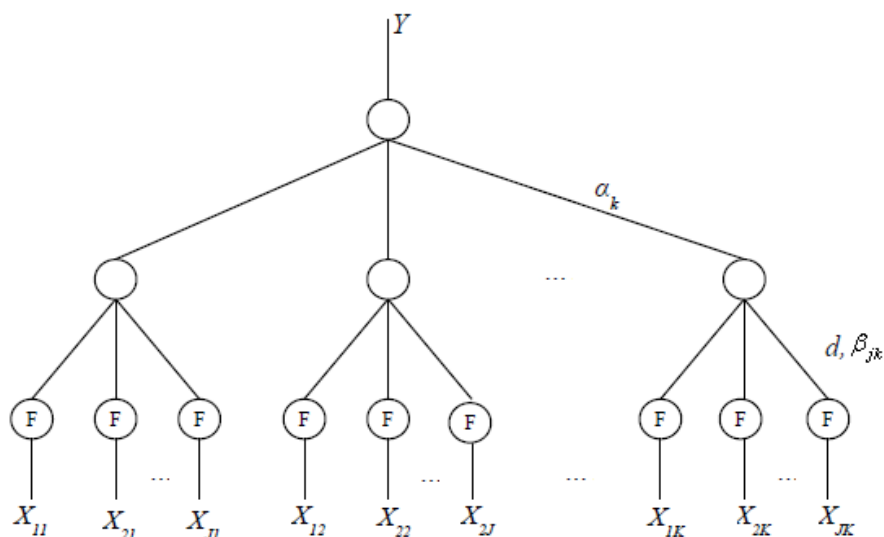
Nová funkce $\tilde{G}(x)$ bude ve tvaru

$$\tilde{G}(x) = t\left(\sum_{i=1}^q \alpha_i t(v_i + w_i \cdot x)\right). \quad (2.12)$$

Neuronová síť nám tedy představuje všeobecný prostředek pro regresní analýzu funkcí, kde počet skrytých neuronů, prahové a váhové koeficienty jsou regresní parametry. Hlavním cílem NS s dopředným šířením je však extrapolace funkčních hodnot mimo trénovací množiny, zejména problém predikce a klasifikace [12].

2.4 Frontální neuronová síť

Další typ neuronové sítě, která se využívá pro predikci, je frontální neuronová síť, nazývaná také jako Time Delay Neural Network (TDNN). Tato síť patří mezi speciální vícevrstvé dopředné NS, která dokáže využívat časové zpoždění při zpracování. Využívá tzv. filtry, které používají krátkodobou paměť. Čím hlubší je daný filtr, tím více informací o minulých stavech okolí uchovává. Celá paměť je situována na začátku NS jako určitá představa, která zpracovává časový kontext. Takto utvořenou NS je možno učit klasickým algoritmem zpětného šíření chyby (Back-propagation) nebo jiným algoritmem, který je využitelný v dopředných NS [16],[21]. Struktura frontální NS je uvedena na obr. 2-3.



Obr. 2-3: Příklad frontální neuronové sítě, zdroj [16]

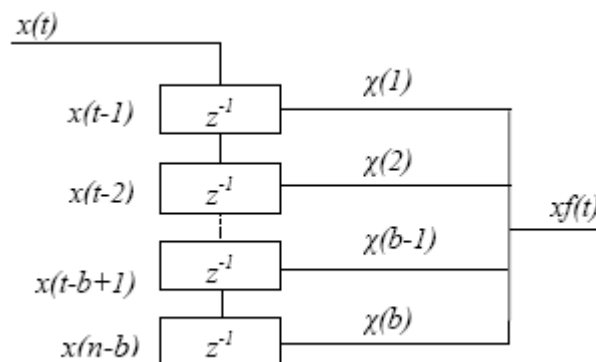
Výstup [16] frontální NS se stejnou hloubkou paměti všech filtrů, lineárním filtrem a lineárním výstupním neuronem je možné vyjádřit následujícím způsobem

$$Y = \sum_{k=1}^K \alpha_k \times d \left(\sum_{j=1}^J \beta_{jk} \times \sum_{i=1}^b \chi_{ijk} \times X_{ijk} \right), \quad (2.13)$$

kde:

- Y je výstup NS,
- α je vektor vah synapsí mezi neurony ve skryté vrstvě a výstupním neuronem,
- β je vektor vah synapsí mezi filtry a neurony ve skryté vrstvě,
- χ je vektor vah synapsí uvnitř filtru,
- k je index neuronu ve skryté vrstvě,
- K je počet neuronů ve skryté vrstvě,
- d je aktivační funkce,
- j je index filtru,
- J je počet filtrů na jeden neuron ve skryté vrstvě,
- i je index vstupu filtru,
- I je počet vstupů filtru,
- b je hloubka krátkodobé paměti filtru,
- X je vstupní vektor NS.

Filtr frontální NS je zobrazen na obr. 2-4.



Obr. 2-4: Filtr frontální NS, zdroj [16]

Kde, $x(t)$ představuje hodnotu vstupu filtru v čase t , b udává hloubku paměti filtru, χ jsou váhy synapsí v rámci filtru, z^{-1} je operátor časového zpoždění a $xf(t)$ představuje výstup filtru pro hodnotu vstupu v čase t .

2.5 Dílčí závěr

Dopředné neuronové sítě patří mezi základní a jednoduché NS, které se využívají zejména pro predikci dat. Při učení sítě se využívá řada algoritmů, z nichž nejrozšířenější je algoritmus Back-propagation. Speciálním případem dopředné NS je frontální neuronová síť, která dokáže používat krátkodobou paměť za pomoci využití filtrů. Hlavním cílem v procesu učení NS je dosažení nejnižší chyby na testovacích datech.

3 Návrh modelu pro predikci návštěvnosti webu upce.cz

V úvodu této kapitoly je na obr. 3-1 zobrazen algoritmus řešení, který ukazuje jednotlivé kroky návrhu modelu pro predikci návštěvnosti web domény upce.cz.

Algoritmus začíná získáním požadovaných vstupních dat pro predikci návštěvnosti webu upce.cz pomocí Google Analytics (GA).

Následuje předzpracování dat, což je považováno za jednu z nejdůležitějších částí algoritmu. Data se upravují pomocí standardizace a normalizace. Nedílnou součástí tohoto kroku je použití indikátorů technické analýzy.

Na základě výsledků z technické analýzy jsou vybrány vstupní parametry, které budou použity pro modelování. Dochází zde k rozdělení vybraných dat na trénovací a testovací množinu.

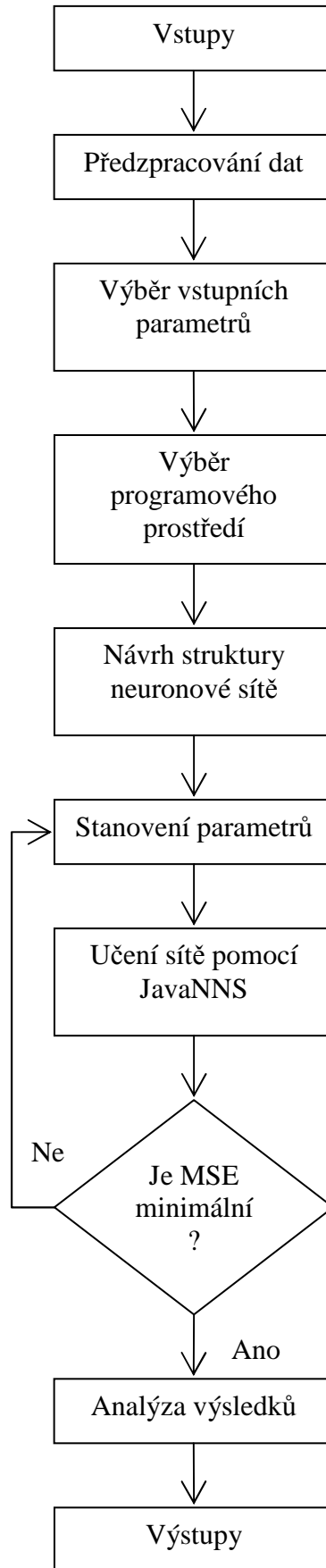
Dále se musí rozhodnout o použitém programovém prostředí pro vytvoření modelu. Pro modelování predikce návštěvnosti web domény upce.cz byl použit programový prostředek JavaNNS. Data musí být následně uložena v takovém formátu a s použitým kódem, s kterým tento software dokáže pracovat.

Již v JavaNNS se vytváří návrh struktury dopředné neuronové sítě pro predikci. Stanovuje se zde návrh vstupní, skryté a výstupní vrstvy a jejich počet.

V dalším kroku se nastavují jednotlivé hodnoty parametrů zvoleného učícího algoritmu.

Dále dochází k samotnému učení dopředné neuronové sítě podle zvolených hodnot parametrů a zjišťuje se MSE chyba. Zde dochází v algoritmu k opakovaným krokům, kde se upravují hodnoty jednotlivých parametrů tak, aby chyba MSE byla minimální.

Konečnou součástí algoritmu je analýza získaných výsledků v průběhu učení, podle kterých se rozhoduje o návrhu struktury dopředné NS s danými parametry.



Obr. 3-1: Algoritmus pro návrh modelu, zdroj [vlastní]

3.1 Získání vstupních dat

Konkrétní data pro predikci časové řady návštěvnosti web domény upce.cz byla získána pomocí Google Analytics (GA). GA patří mezi statistické programy, běžící na serverech Google. Uživatelům, kteří jsou registrovaní, umožňují vložit do jejich vlastních stránek měřicí kódy, pomocí kterých mohou zjišťovat různé statistické údaje [18]. Pomocí implementace JavaScriptového kódu, který je vložen do webové prezentace, nabízí široké spektrum provozních charakteristik, nazývaných také jako webové metriky. V GA lze webové metriky rozdělit do čtyř základních skupin [2],[19]:

- Návštěvy – zde se sleduje zejména počet návštěv, počet zhlédnutých stránek, poměr nových návštěvníků vůči těm, kteří se na stránky stále vrací. Sledovat se může také čas, který návštěvníci stráví na určité stránce. Ukazatel geolokace zjišťuje, z jaké země zákazníci nejčastěji přicházejí, což je jedna z důležitých informací pro jazykové mutace. Získávají se zde také informace o chování návštěvníků na stránkách, zda se v nich ztrácí, kde bloudí, či kudy procházejí.
- Zdroje přístupů.
- Obsah – sleduje se, o jaký obsah mají návštěvníci největší zájem, včetně poklesů a vzestupů.
- Konverze – konverze je chápána jako stav, kdy se návštěvník webové stránky stane zákazníkem, např. při prodeji, objednávce zboží, vyplnění formuláře, odeslání poptávky, stažení souboru, aj. Zde se sleduje, jak jsou efektivní odkazující média či zdroje kampaní vzhledem ke konverzím, kolik času a návštěv zabere návštěvníkovi, než se z něj stane zákazník, která klíčová slova vedou ke konverzi, které kampaně nebo sestavy zastavit a naopak, do kterých investovat.

Pro predikci návštěvnosti byl sledován ukazatel počet návštěv. Jedná se o základní ukazatel, který zobrazuje počet návštěv vždy za určité období. V GA je návštěva brána jako neopakovatelná kombinace IP adresy, cookies. Submetrikou lze zjistit také absolutně unikátní návštěva, která je definována jako neopakovatelná IP adresa a cookies v zadaném časovém období.

Příkladem lze uvést výčet základních informací, které je možno o webu upce.cz získat v GA například za květen roku 2009 [2]:

- Celková návštěvnost za daný měsíc klesá. Je zde vidět trend, kdy pondělí má nejvyšší návštěvnost, která klesá až po zbytek týdne, nejnižší je potom sobota.

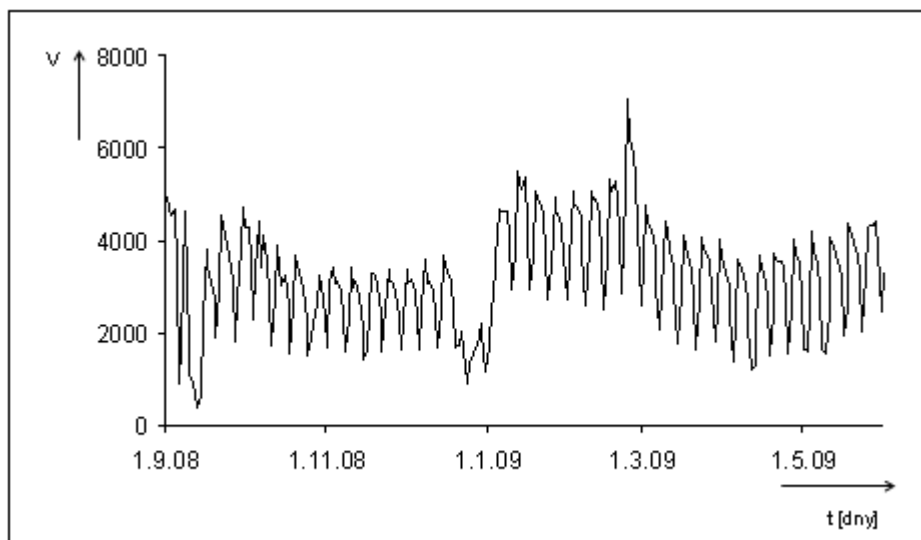
- Průměrný počet zhlédnutých stránek za návštěvu je více než tři.
- Na stránce se návštěvník zdrží v průměru pět a půl minuty.
- Míra nezaujaté návštěvy je přibližně šedesát procent.
- Návštěvníci přicházejí většinou přímo, což je velice dobré.
- Mezi nejnavštěvovanější stránky můžeme zařadit úvodní stranu, potom jsou to stránky fakulty ekonomicko-správní a fakulty filozofické.

Návštěvnost se měří v pravidelných ekvidistantních rovnoměrných časových jednotkách. Tedy např. po sekundách, dnech, měsících či letech. Naměřená data zobrazují časovou řadu, kterou lze chápat jako posloupnost věcně a prostorově srovnatelných pozorování, která jsou jednoznačně uspořádaná z hlediska času [2].

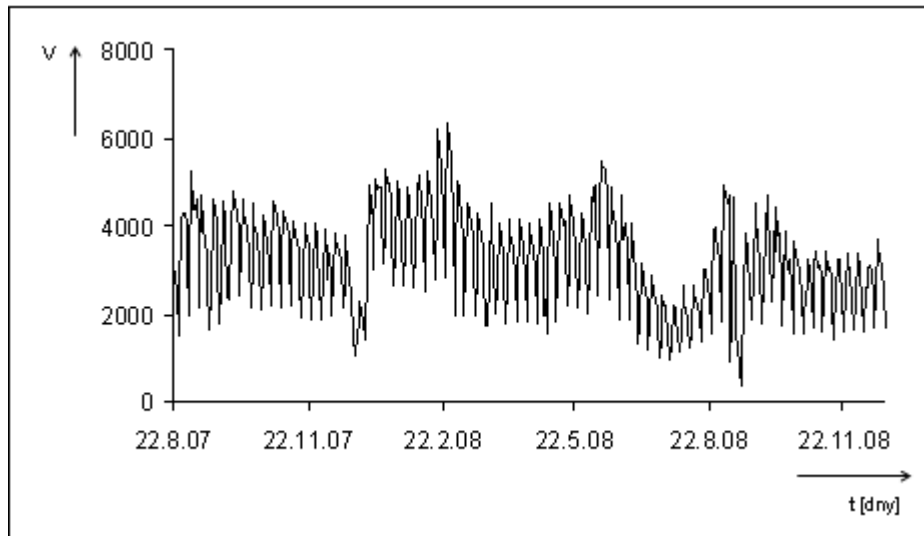
Měřená data návštěvnosti webu upce.cz jsou zaznamenána po jednotlivých dnech a sledována za tři časová období:

- Krátkodobá časová řada (soubor S) od 1.9.2008 do 31.5.2009 (obr. 3-2)
- Střednědobá časová řada (soubor I) od 22.8.2007 do 21.12.2008 (obr. 3-3)
- Dlouhodobá časová řada (soubor L) od 21.8.2007 do 20.9.2009 (obr. 3-4).

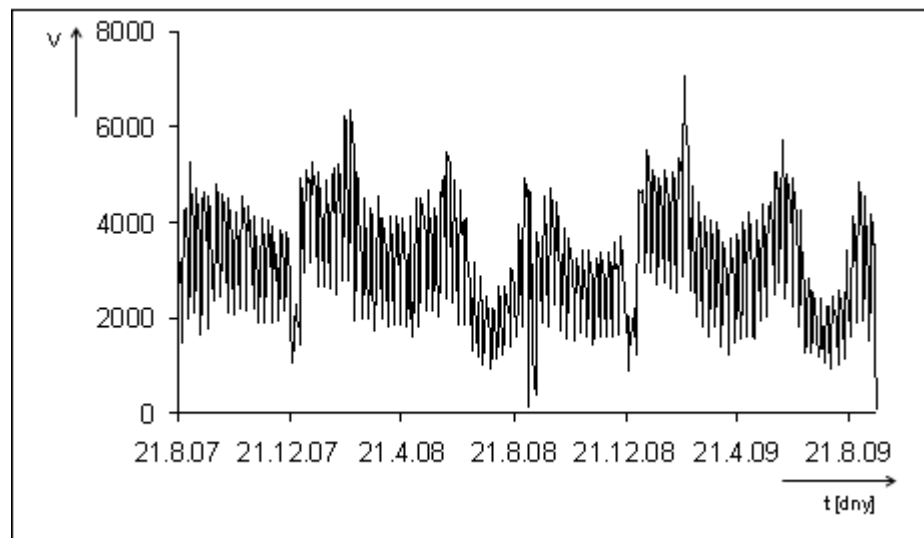
U níže uvedených obrázků osa x představuje jednotlivé dny a osa y počet návštěvníků.



Obr. 3-2: Návštěvnost web domény upce.cz souboru S, zdroj [vlastní]



Obr. 3-3: Návštěvnost web domény upce.cz souboru I, zdroj [vlastní]



Obr. 3-4: Návštěvnost web domény upce.cz souboru L, zdroj [vlastní]

3.2 Předzpracování dat

Před modelováním musí být data vhodně ošetřena. První provedenou úpravou je standardizace dat. Používá se pro úpravu dat, aby všechny znaky byly souměřitelné.

Nechť je dána matice dat $Z = (z_{ij})$ typu $n \times p$, jejíž řádky jsou p -rozměrné vektory čísel, které charakterizují n objektů. Standardizace se provede ve dvou krocích [20]:

- výpočet střední hodnoty \bar{z}_j j -tého znaku z_j a směrodatné odchylky s_j pro $j = 1, 2, \dots, p$ podle vztahů

$$\bar{z}_j = \frac{1}{n} \sum_{i=1}^n z_{ij} , \quad (3.1)$$

$$s_j = \left[\frac{1}{n} \sum_{i=1}^n (z_{ij} - \bar{z}_j)^2 \right]^{1/2}. \quad (3.2)$$

- poté jsou původní hodnoty z_{ij} j-tého znaku i-tého objektu přepočteny na tzv. standardizované hodnoty podle vztahu

$$x_{ij} = \frac{z_{ij} - \bar{z}_j}{s_j}. \quad (3.3)$$

Všechny standardizované hodnoty mají střední hodnotu rovnu 0 a rozptyl 1.

Dalším krokem je úprava dat pomocí indikátorů technické analýzy. Přehled použitých indikátorů zobrazuje tab. 3-1. Průběhy jednotlivých indikátorů jsou uvedeny v příloze 1.

Tab. 3-1: Indikátory technické analýzy, zdroj [upraveno podle 2]

| Druhy indikátorů technické analýzy | Charakteristika |
|---|-------------------------------------|
| Jednoduché klouzavé průměry (JKP) | 5, 7, 9 denní |
| Centrované klouzavé průměry (CKP) | 4, 6, 8 denní |
| Klouzavé mediány (KM) | 5, 7, 9 denní |
| Jednoduché exponenciální vyrovnání (JEV) | Pro $\alpha = 0.1$ a $\alpha = 0.2$ |
| Dvojitě exponenciální vyrovnání (DEV) | Pro $\alpha = 0.7$ a $\alpha = 0.9$ |

3.3 Rozdělení dat na trénovací a testovací množinu

Upravená data musí být pro potřeby modelování rozdělena na trénovací a testovací množinu. Podle [12] lze popsat jednoduchý způsob rozkladu množiny objektů A , $A = A_{train} \cup A_{test}$. Pomocí některé z metod shlukové analýzy dojde k rozložení množiny A na disjunktí podmnožiny, tzv. shluky, kde každý ze shluků obsahuje podobné objekty

$$A = C_1 \cup C_2 \cup \dots \cup C_p, \quad (3.4)$$

kde i -tý shluk C_i obsahuje n_i objektů z A ,

$$C_i = \{o_1^{(i)}, o_2^{(i)}, \dots, o_{n_i}^{(i)}\} \subset A, \quad (3.5)$$

kde objekt $o_1^{(i)} \in C_i$ je z i -tého shluku C_i , který leží nejbližší k jeho centru. O tomto objektu můžeme následně prohlásit, že je reprezentant objektů ze shluku C_i . Trénovací a testovací množina je tedy určena objekty

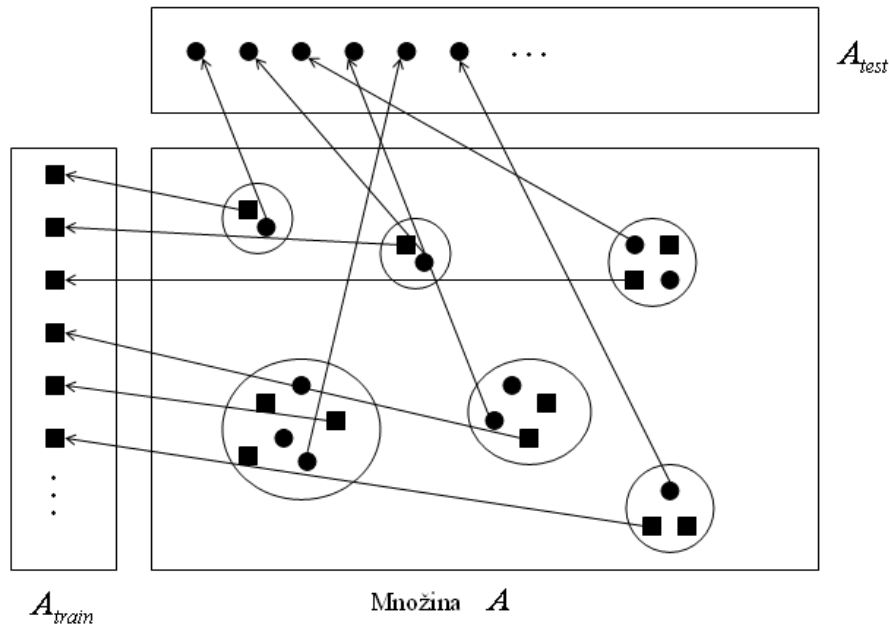
$$A_{train} = \{o_1^{(1)}, o_1^{(2)}, \dots, o_1^{(p)}\}, \quad (3.6)$$

$$A_{test} = (C_1 - \{o_1^{(1)}\}) \cup (C_2 - \{o_1^{(2)}\}) \cup \dots \cup (C_p - \{o_1^{(p)}\}). \quad (3.7)$$

Podle výše uvedených vzorců vyplývá, že trénovací množina je složená ze všech reprezentantů daných shluků a testovací množina obsahuje všechny ostatní objekty. Počet objektů v trénovací množině je stejný s počtem shluků, $|A_{train}| = p$ a $|A_{test}| = |A| - p$.

Schematické znázornění je zachyceno na obr. 3-5. Jedná se o rozklad množiny objektů A na trénovací a testovací množinu pomocí rozkladu A na shluky C_1, C_2, \dots, C_p .

Objekty, které jsou v obrázku znázorněny čtverci, leží nejbližší ke středům příslušných shluků. Ostatní objekty jsou představovány jednotlivými kruhy.



Obr. 3-5: Rozklad množiny objektů na trénovací a testovací, zdroj [upraveno podle 12]

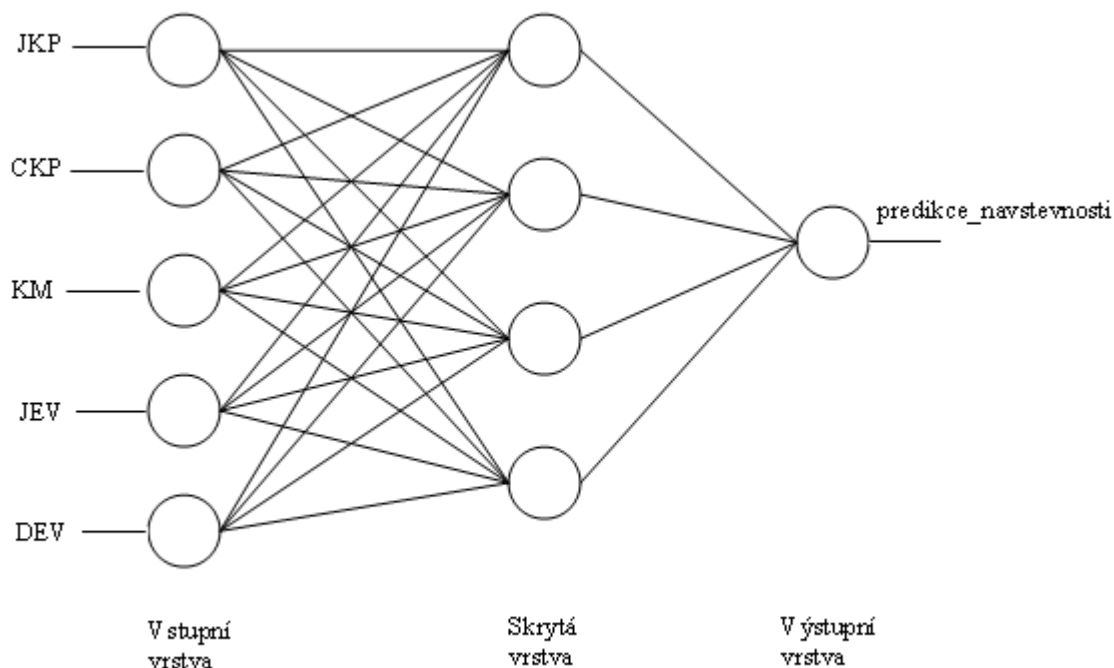
Data pro predikci návštěvnosti webu upce.cz byla rozdělena pesimistickým způsobem, tzv. zcela náhodně. Rozdělení množiny dat na trénovací a testovací je realizováno v programovém prostředí SPSS Clementine. Všechny tři soubory S, I a L jsou rozdělena v poměrech 50:50; 60:40; 70:30; 80:20 a 90:10 (trénovací:testovací). Ukázka toho rozdělení je uvedena v příloze 2.

3.4 Návrh struktury neuronové sítě

Návrh struktury neuronové sítě pro predikci návštěvnosti webu upce.cz byl realizován v programovém prostředí JavaNNS (Java Neural Network Simulator). Vstupem do neuronové sítě je časová řada návštěvnosti web domény upce.cz zpracována na základě indikátorů technické analýzy. Použit je:

- jednoduchý klouzavý průměr (JKP)
- Centrovaný klouzavý průměr (CKP)
- Klouzavý medián (KM)
- Jednoduché exponenciální vyrovnání (JEV)
- Dvojitě exponenciální vyrovnání (DEV).

Struktura NS (obr. 3-6) má tedy 5 vstupních neuronů, 4 skryté neurony a 1 výstupní neuron.



Obr. 3-6: Struktura navržené neuronové sítě, zdroj [vlastní]

3.5 Formát dat

Pro návrh modelu v programu JavaNNS je zapotřebí získaná data převést do vhodného formátu. JavaNNS dokáže pracovat pouze se soubory ve formátu pat. Data musí být tedy převedena do textového editoru, kde jsou dále doplněna příslušným kódem. Návrh kódu je zobrazen na obr. 3-7. Každý soubor musí mít v záhlaví zapsán počet záznamů, počet vstupních a výstupních hodnot. Poté je soubor uložen s koncovkou pat.

```
SNNS pattern definition file V3.2
generated at Mon Sep 4 15:43:27 1995

No. of patterns : 42
No. of input units : 5
No. of output units : 1
```

Obr. 3-7: Popis záhlaví pro formát pat., zdroj [vlastní]

3.6 Proces učení neuronové sítě

Před učením neuronové sítě musí být nastaveny všechny funkce potřebné pro učení. Inicializační funkcí byla zvolena funkce Random Weights, která nastavuje váhy synapsí zcela náhodně na hodnoty z intervalu od -1.0 do 1.0.

Pro zpracování byla použita přenosová funkce Binary. Tato funkce nemůže být zvolena, je automaticky nastavena na hodnotu 0.5. Funkce rozděluje hodnoty vstupních vzorů na dvě skupiny, ty které jsou vyšší než 0.5 jsou nastaveny na hodnotu 1, ostatní nižší hodnoty na 0.

Pro samotné učení byl zvolen algoritmus Back-propagation momentum, kde mohou být voleny následující parametry:

η – rychlost učení,

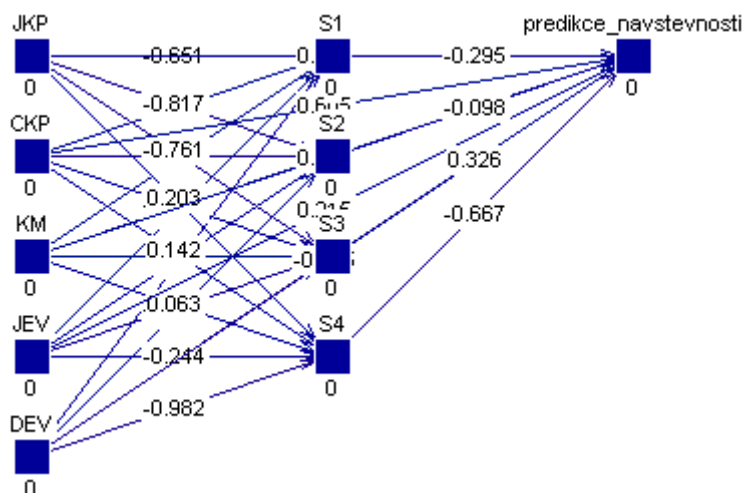
μ - parametr momentu,

c – eliminace tzv. plochých bodů,

d_{\max} – maximální chyba, tzn. maximální rozdíl mezi učící hodnotou a výsledkem výstupního neuronu.

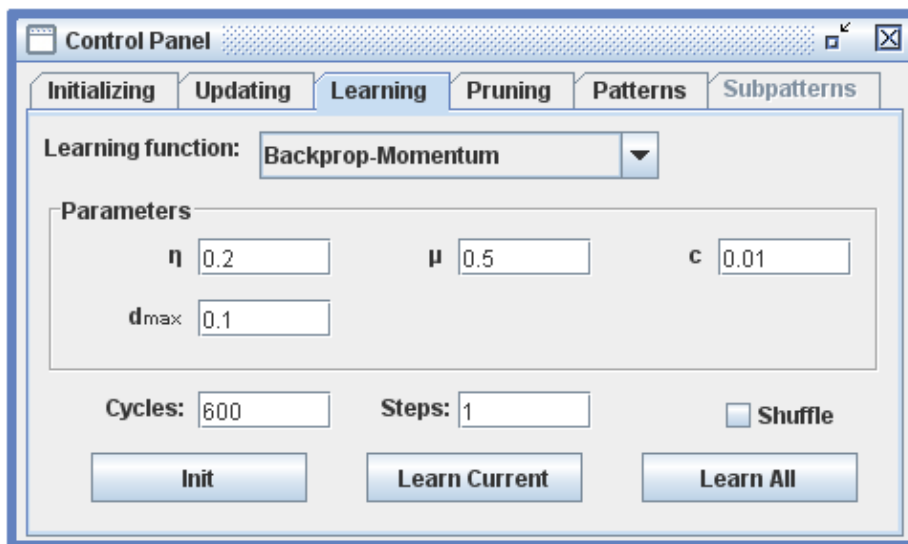
Dále je zvolen cyklus učení, který byl pro všechny soubory jednoznačně stanoven na 600 cyklů. Volba hodnot jednotlivých parametřů je provedena experimentálně. Nastavení všech funkcí a hodnot jednotlivých parametřů je pro soubory S, I a L stejné.

Následující popis uvádí ukázkou proces učení a výstupní MSE chyby pro krátkodobou časovou řadu při rozdělení dat v poměru 50:50. Celý proces učení začíná nastavením vah synapsí, které je uvedeno na obr. 3-8.



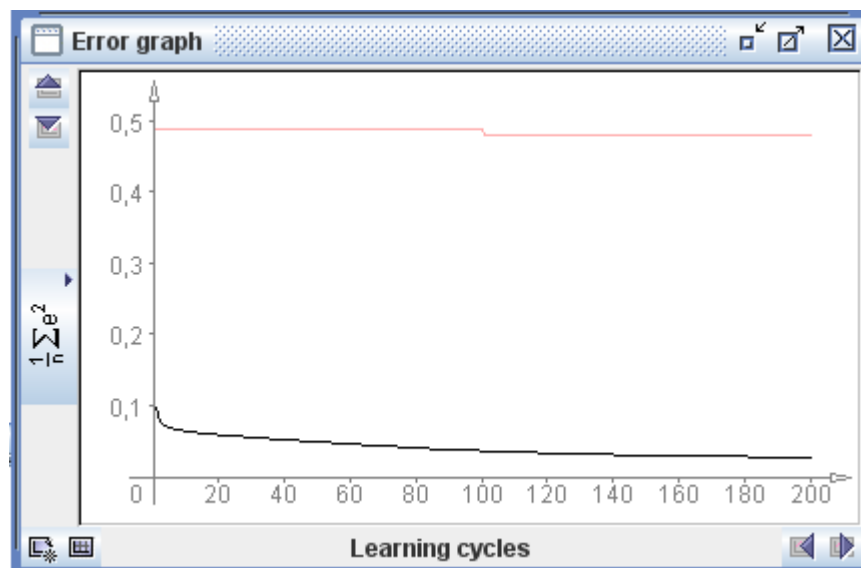
Obr. 3-8: Nastavení vah, zdroj [vlastní]

Dále jsou definovány všechny potřebné funkce a následně jednotlivé parametry učícího algoritmu. Panel pro zvolení hodnot parametřů je uveden na obr. 3-9.



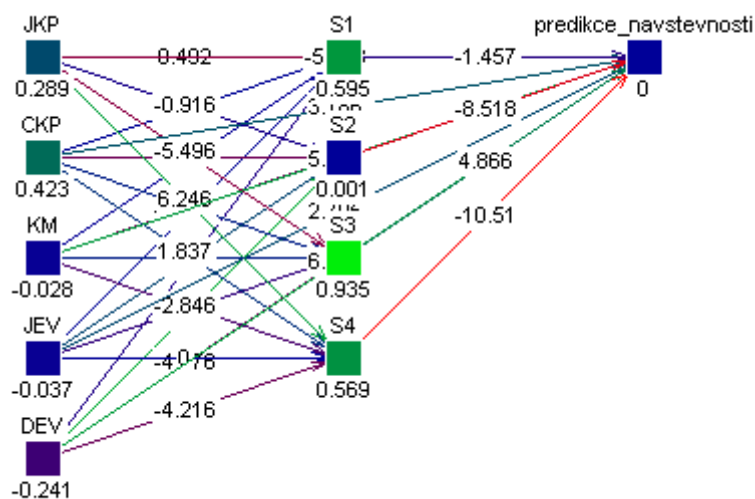
Obr. 3-9: Nastavení hodnot parametřů algoritmu Backprop-Momentum, zdroj [vlastní]

Po procesu učení je zobrazena MSE chyba testovacích (červená křivka) a trénovacích (černá křivka) dat (obr. 3-10).



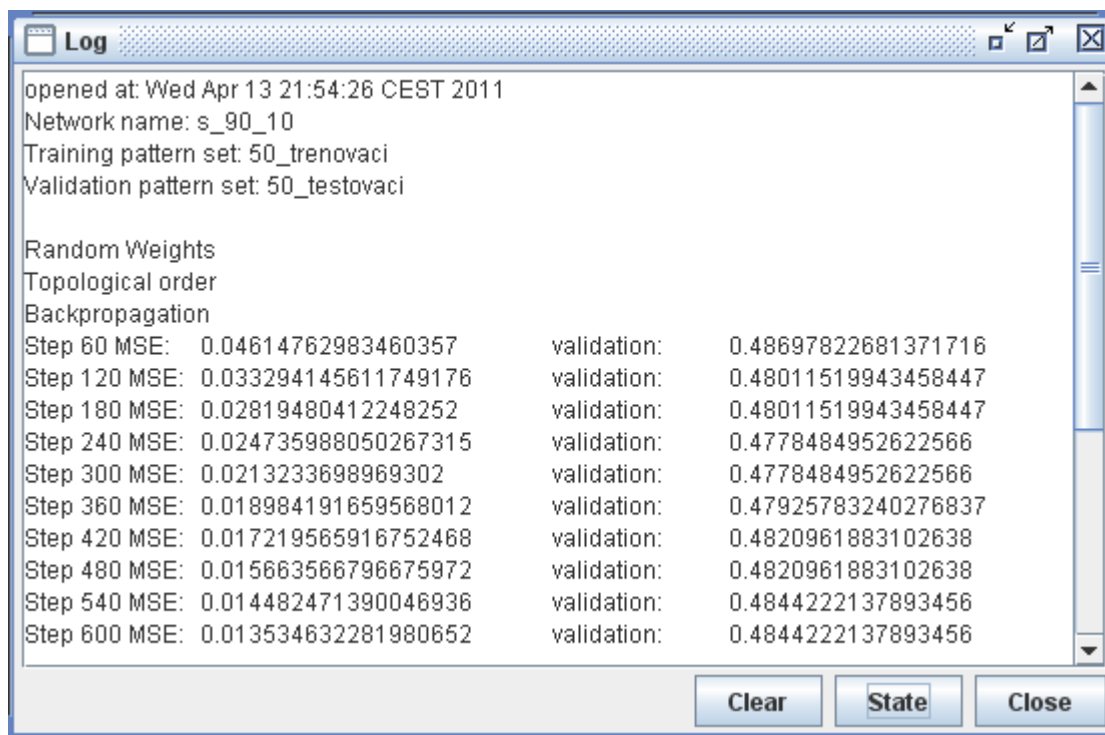
Obr. 3-10: Průběh MSE chyby, zdroj [vlastní]

Váhy po procesu učení jsou zobrazeny na obr. 3-11.



Obr. 3-11: Váhy po procesu učení, zdroj [vlastní]

Log panel, který je uveden na obr. 3-12, zobrazuje protokol o učení. Jsou zde uvedeny soubory testovacích a trénovacích dat, zvolené funkce pro učení a dosažené MSE chyby jednotlivých cyklů pro trénovací i testovací množinu.



Obr. 3-12: Protokol o učení, zdroj [vlastní]

3.7 Dílčí závěr

Před samotným návrhem modelu NS je potřeba získaná data vhodným způsobem předzpracovat. Využita byla standardizace dat a indikátory technické analýzy. Zpracovaná data byla rozdělena pesimistickým způsobem na trénovací a testovací množinu. Pro tvorbu modelu byl zvolen programový prostředek JavaNNS a byla vytvořena NS s 5 vstupními, 4 skrytými a 1 výstupním neuronem.

4 Analýza navrženého modelu

Analýza navrženého modelu byla realizována v programovém prostředí JavaNNS. Tento softwarový prostředek je volně dostupný. Cílem této kapitoly je pomocí experimentů a volených hodnot parametrů navrhnout odpovídající model pro predikci návštěvnosti upce.cz.

4.1 Analýza výsledků krátkodobé časové řady

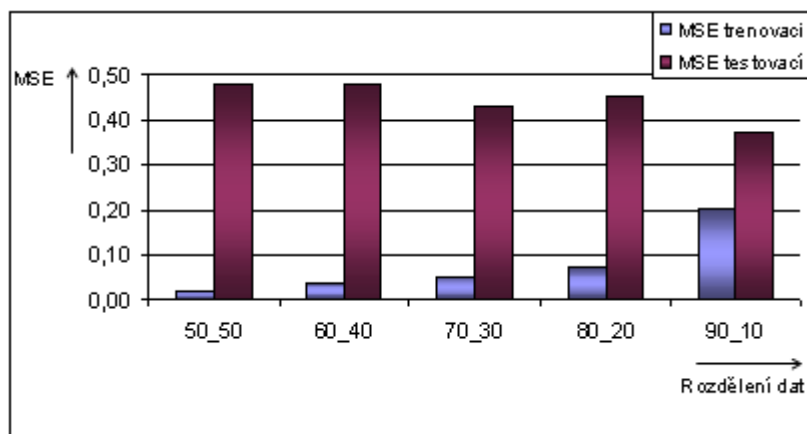
Jak již bylo uvedeno v předešlé kapitole, krátkodobá časová řada je rozdělena v pěti různých poměrech dat na trénovací a testovací množinu. Pro učení NS byla data převedena do vhodného formátu. Pro samotné učení byly nastaveny všechny potřebné funkce. Jako učící algoritmus byl zvolen Back-propagation momentum.

Prvním krokem byla provedena simulace na změnu parametru c při zachování hodnot všech ostatních parametrů (ostatní parametry zůstaly ve výchozím nastavení). Podle výsledků experimentů je pro další zpracování zvolena hodnota parametru c taková, při které je dosaženo nejnižší chyby na testovacích datech. Druhým krokem byla zjišťována změna parametru μ , u kterého se opět sledují nejnižší chyby na testovacích datech. Třetí parametr algoritmu Back-propagation momentum je parametr d_{\max} . Podle nejnižší chyby zachycené na testovacích datech je volena hodnota tohoto parametru. Posledním parametrem zvoleného algoritmu je rychlost učení η , u kterého se také sleduje nejnižší dosažená chyba testovacích dat.

Jednotlivé výsledky pro všechny poměry rozdělení dat jsou uvedeny v příloze 3. Naměřené střední čtvercové chyby (MSE) jsou z důvodu přehlednosti zaokrouhleny na 5 desetinných míst.

Celkové výsledky, při kterých je dosaženo nejnižší MSE chyby na testovacích datech při všech pěti rozdělení dat, jsou zobrazeny na obr. 4-1 a jejich přesné hodnoty jsou uvedeny v tab. 4-1.

Krátkodobá časová řada jasně vykazuje nejnižší MSE chybu na testovacích datech při rozdělení dat v poměru 90:10. Dosáhla tedy nejlepších výsledků při nejvyšším počtu trénovacích dat.



Obr. 4-1: Porovnání výsledků pro jednotlivá rozdělení dat souboru S , zdroj [vlastní]

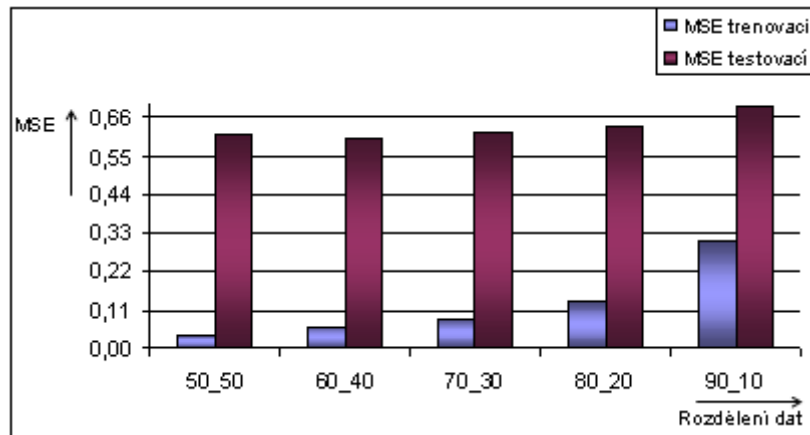
Tab. 4-1: Celkové výsledky pro soubor S , zdroj [vlastní]

| Rozdělení | μ | c | d_{\max} | η | MSE trénovací | MSE testovací |
|-----------|-------|------|------------|--------|---------------|---------------|
| 50:50 | 0.9 | 0.01 | 0.1 | 0.2 | 0.01353 | 0.48442 |
| 60:40 | 0.7 | 0.1 | 0.1 | 0.6 | 0.03837 | 0.47806 |
| 70:30 | 0.5 | 0.01 | 0.1 | 0.2 | 0.04981 | 0.43003 |
| 80:20 | 0.5 | 0.1 | 0.001 | 0.2 | 0.06832 | 0.45103 |
| 90:10 | 0.3 | 0.1 | 0.1 | 0.2 | 0.20266 | 0.36969 |

4.2 Analýza výsledků střednědobé časové řady

Střednědobá časová řada je rozdělena do různých poměrů rozdělení dat na trénovací a testovací množinu. I zde, stejně jako u krátkodobé časové řady, je zvolen algoritmus Back-propagation momentum, který obsahuje čtyři volitelné parametry. Učení začíná volbou parametru c , při jehož volbě je snahou dosáhnout nejnižší MSE chyby na testovacích datech. Dále následuje změna parametru μ , kde se zjišťuje nejnižší MSE chyba na testovacích datech. Poté je zjišťována změna parametru d_{\max} a nakonec parametru η , kde je opět sledována chyba na testovacích datech.

Dosažené výsledky pro všechny typy rozdělení dat jsou uvedeny v příloze 4. Nejlepší výsledky všech poměrů rozdělení dat zobrazuje obr. 4-2 a přesné hodnoty těchto výsledků jsou uvedeny v tab. 4-3.



Obr. 4-2: Porovnání výsledků pro jednotlivá rozdělení dat souboru I, zdroj [vlastní]

Podle tab. 4-3 je minimální MSE chyby na testovacích datech dosaženo při rozdělení dat v poměru 60:40.

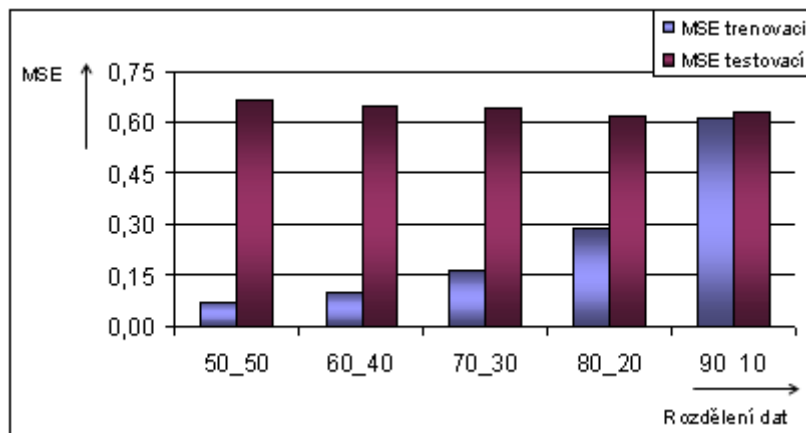
Tab. 4-3: Celkové výsledky pro soubor I, zdroj [vlastní]

| Rozdělení | μ | c | d_{\max} | η | MSE trénovací | MSE testovací |
|-----------|-------|--------|------------|--------|---------------|---------------|
| 50:50 | 0.3 | 0.001 | 0.1 | 0.2 | 0.03952 | 0.61064 |
| 60:40 | 0.3 | 0.0001 | 0.1 | 0.2 | 0.06119 | 0.59912 |
| 70:30 | 0.3 | 0.01 | 0.1 | 0.2 | 0.08636 | 0.62066 |
| 80:20 | 0.3 | 0.01 | 0.1 | 0.2 | 0.13196 | 0.63456 |
| 90:10 | 0.7 | 0.1 | 0.1 | 0.6 | 0.30841 | 0.69737 |

4.3 Analýza výsledků dlouhodobé časové řady

Data dlouhodobé časové řady byla také rozdělena v pěti poměrech dat na trénovací a testovací množiny. Pro učení sítě jsou zvoleny stejné funkce a algoritmus Back-propagation momentum jako v předešlých dvou řadách. První experimenty jsou provedeny na změnu parametru c . Podle nejnižší dosažené MSE chyby na testovacích datech je hodnota parametru použita pro další zpracování. Následuje volba parametru μ , s jejíž nastavenou hodnotou je dále pokračováno podle dosažené MSE chyby. U třetího parametru d_{\max} je opět snaha o dosažení nejnižší MSE chyby na testovacích datech. Poslední provedenou změnou je volba parametru η , která je nastavena podle nejnižší dosažené MSE chyby testovacích dat.

Dílejší výsledky pro jednotlivá rozdělení dat dlouhodobé časové řady jsou uvedeny v příloze 5. Nejlepší dosažené MSE chyby na testovacích datech pro jednotlivá rozdělení dat jsou zobrazeny na obr. 4-3. Přesné hodnoty minimálních dosažených MSE chyb jsou uvedeny v tab. 4-5.



Obr. 4-3: Porovnání výsledků pro jednotlivá rozdělení dat souboru L , zdroj [vlastní]

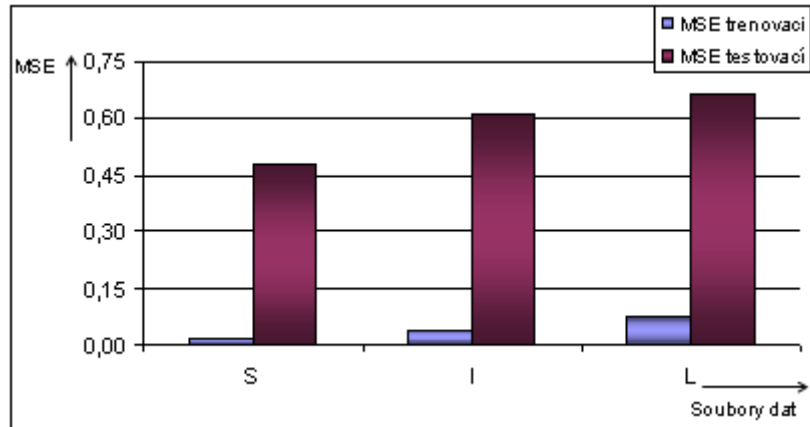
Podle tab. 4-5 je nejlepšími výsledky dosaženo při rozdělení dat v poměru 80:20.

Tab. 4-5: Celkové výsledky pro soubor L , zdroj [vlastní]

| Rozdělení | μ | c | d_{\max} | η | MSE trénovací | MSE testovací |
|-----------|-------|--------|------------|--------|---------------|---------------|
| 50:50 | 0.7 | 0.01 | 0.0001 | 0.2 | 0.07383 | 0.66630 |
| 60:40 | 0.5 | 0.1 | 0.001 | 0.2 | 0.10022 | 0.64777 |
| 70:30 | 0.7 | 0.001 | 0.1 | 0.2 | 0.16643 | 0.63762 |
| 80:20 | 0.3 | 0.001 | 0.0001 | 0.6 | 0.28697 | 0.62016 |
| 90:10 | 0.7 | 0.0001 | 0.0001 | 0.2 | 0.61376 | 0.63133 |

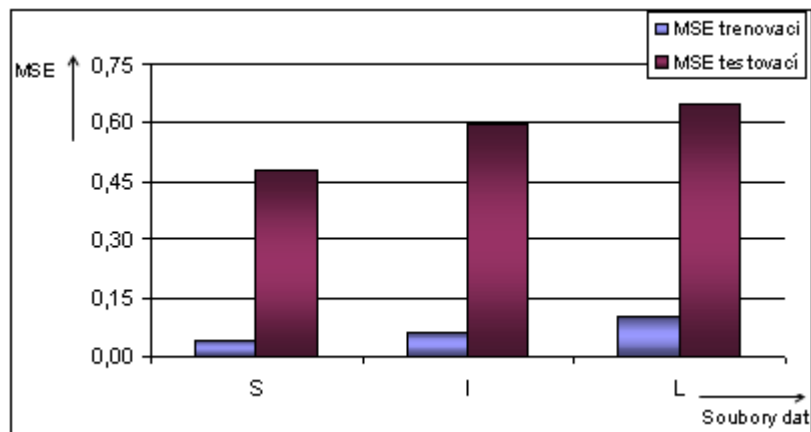
4.4 Porovnání jednotlivých časových řad

Jednotlivé časové řady byly rozděleny ve stejných poměrech dat na trénovací a testovací množiny. Porovnání krátkodobé, střednědobé a dlouhodobé časové řady v závislosti na rozdělení dat v poměru 50:50 zobrazuje obr. 4-4, z kterého jasně vyplývá, že nejnižší MSE chyby na testovacích datech při tomto rozdělení dat dosáhla krátkodobá časová řada, vyšší chyby byly zaznamenány na střednědobé časové řadě a nejhorších výsledků dosáhla dlouhodobá časová řada.



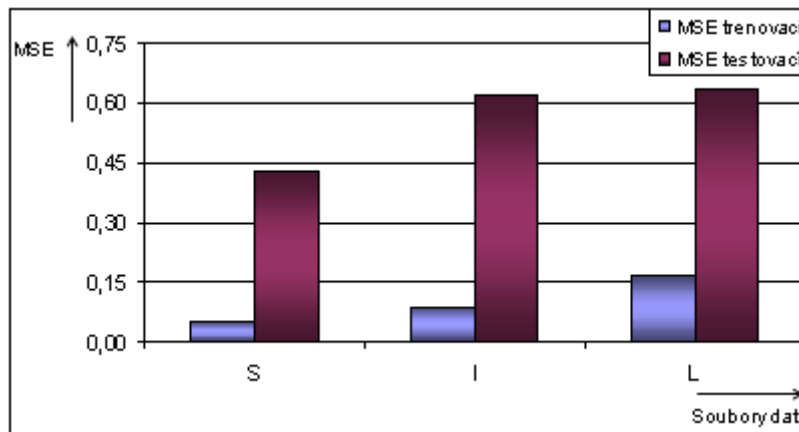
Obr. 4-4: Porovnání jednotlivých souborů dat pro rozdělení v poměru 50:50, zdroj [vlastní]

Porovnání jednotlivých časových řad pro rozložení dat v poměru 60:40 zobrazuje obr. 4-5. Nejnižší MSE chyby na testovacích datech je dosaženo u krátkodobé časové řady a nejvyšší MSE chyby na testovacích datech je dosaženo u dlouhodobé časové řady.



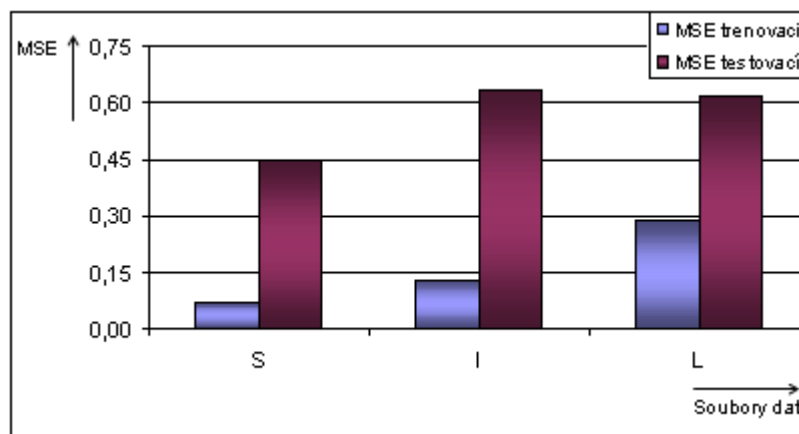
Obr. 4-5: Porovnání jednotlivých souborů dat pro rozdělení v poměru 60:40, zdroj [vlastní]

Pro rozdělení dat v poměru 70:30 je porovnání jednotlivých souborů zobrazeno na obr. 4-6. Nejnižší chyby je opět dosaženo v souboru S a nejhorší chyby poskytuje soubor L.



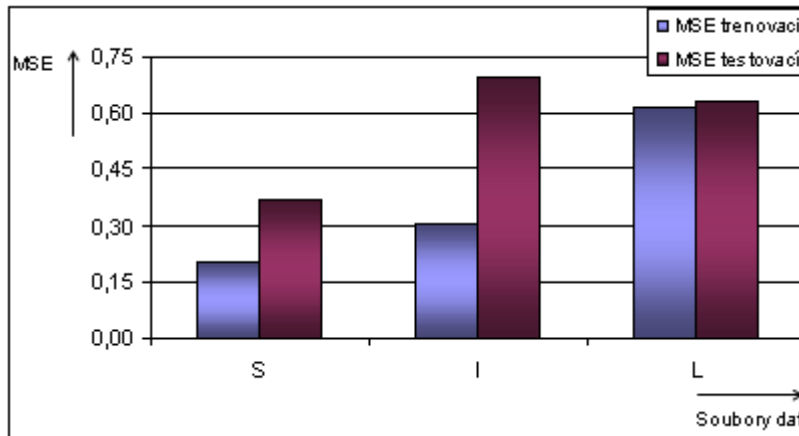
Obr. 4-6: Porovnání jednotlivých souborů dat pro rozdělení v poměru 70:30, zdroj [vlastní]

Porovnání nejlepších výsledků pro jednotlivé soubory dat pro rozdělení v poměru 80:20 je zobrazeno na obr. 4-7. Nejnižší MSE chyby na testovacích datech je dosaženo u krátkodobé časové řady. Nejhorších výsledků je pak zachyceno u střednědobé časové řady.



Obr. 4-7: Porovnání jednotlivých souborů dat pro rozdělení v poměru 80:20, zdroj [vlastní]

Pro rozdělení dat v poměru 90:10 jsou výsledky jednotlivých souborů dat zobrazeny na obr. 4-8. Nejnižší chyby dosahuje krátkodobá časová řada, nejhorších výsledků dosahuje střednědobá časová řada.



Obr. 4-8: Porovnání jednotlivých souborů dat pro rozdělení v poměru 90:10, zdroj [vlastní]

4.5 Dílčí závěr

Pro učení neuronové sítě u krátkodobé, střednědobé i dlouhodobé časové řady a různých rozdělení poměrů dat byl zvolen učící algoritmus Back-propagation momentum. Tento algoritmus obsahuje čtyři parametry, které byly voleny experimentálně.

Při porovnání jednotlivých časových řad jednoznačně vyplývá, že nejlepších výsledků dosahuje krátkodobá časová řada při všech poměrech rozdělení dat na trénovací a testovací množinu. Nejvyšší MSE chyby jsou dosaženy při poměrech dat 50:50, 60:40 a 70:30 u dlouhodobé časové řady, u poměrů dat 80:20 a 90:10 u střednědobé časové řady.

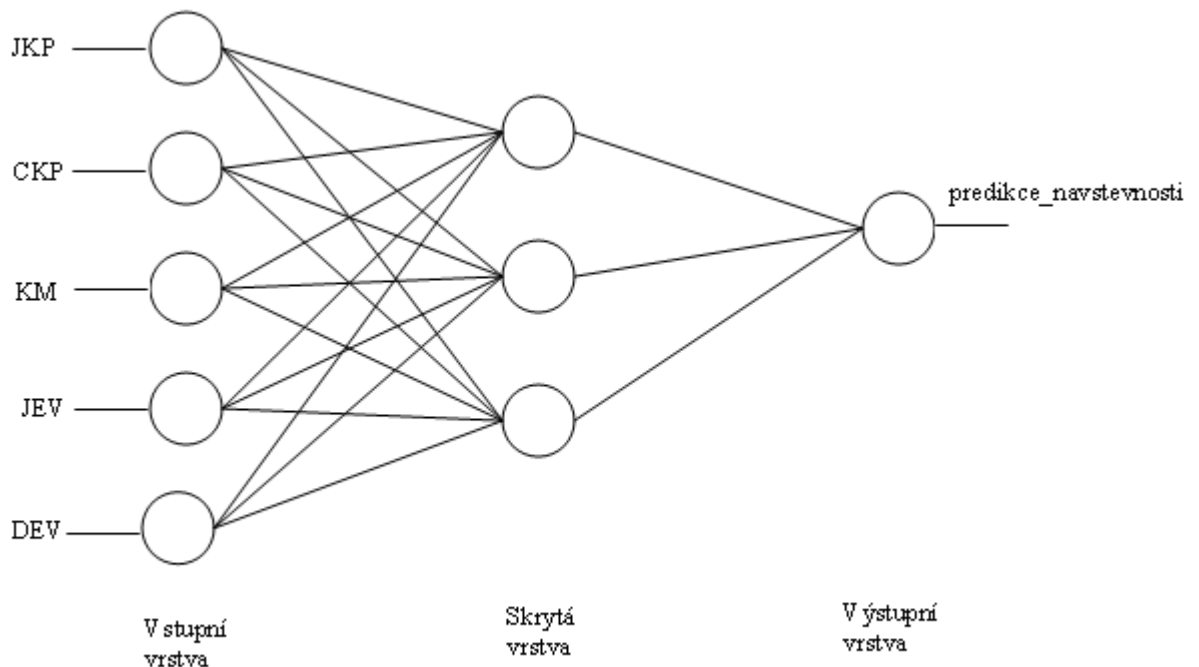
U krátkodobé časové řady je nejnižší MSE chyby dosaženo při rozdělení dat v poměru 90:10 se zvolenými parametry algoritmu Back-propagation momentum v hodnotách $c = 0.1$; $\mu = 0.3$; $d_{\max} = 0.1$ a $\eta = 0.2$. Takto navržená NS dosahuje MSE chyby na trénovacích datech 0.20266 a MSE chyby na testovacích datech 0.36969.

U střednědobé časové řady je nejnižší MSE chyby dosaženo při rozdělení dat v poměru 60:40 a zvolenými parametry v hodnotách $c = 0.0001$; $\mu = 0.3$; $d_{\max} = 0.1$ a $\eta = 0.2$. Takto navržená NS dosahuje MSE chyby na trénovacích datech 0.06119 a MSE chyby na testovacích datech 0.59912.

Dlouhodobá časová řada dosahuje nejnižší MSE chyby při rozdělení dat v poměru 80:20 a zvolenými parametry v hodnotách $c = 0.001$; $\mu = 0.3$; $d_{\max} = 0.0001$ a $\eta = 0.6$. Takto navržená NS dosahuje MSE chyby na trénovacích datech 0.28697 a MSE chyby na testovacích datech 0.62016.

5 Návrh a analýza frontální neuronové sítě

Frontální neuronová síť v tomto případě stejně jako klasická dopředná NS má definován pevný počet neuronů ve vstupní vrstvě. Návrh struktury NS spočívá v definování délky filtru, který udává počet minulých vstupů. Pro tento experiment byla zvolena hloubka filtru $b = 3$, počet neuronů ve vstupní vrstvě 5, počet neuronů ve skryté vrstvě 3 a počet neuronů na výstupu 1. Experimenty jsou opět provedeny na všech souborech S, I a L v poměru rozdělení dat 50:50, 60:40, 70:30, 80:20, 90:10. vstupem do NS jsou opět indikátory technické analýzy uvedeny v kapitole 3.4. Struktura frontální neuronové sítě je zobrazena na obr. 5-1.



Obr. 5-1: Struktura frontální NS, zdroj [vlastní]

Inicializační funkcí byla zvolena funkce Random Weights a jako přenosová funkce byla zvolena funkce Binary. Pro tento typ sítě se dále využívá funkce TimeDelay Order, která se využívá během časového zpoždění sítě, je tedy závislá na struktuře NS. Pro učení byl zvolen algoritmus Time-Delay Backprop, u kterého volíme dva parametry:

- η – rychlost učení, byly voleny hodnoty 0.1; 0.2; 0.3.
- d_{\max} – maximální chyba, byly voleny hodnoty 0; 0.1; 0.01.

Pro učení neuronové sítě musí být stanoven počet cyklů. Pro tento typ dat bylo zvoleno 600 cyklů. Volba jednotlivých parametrů je provedena experimentálně.

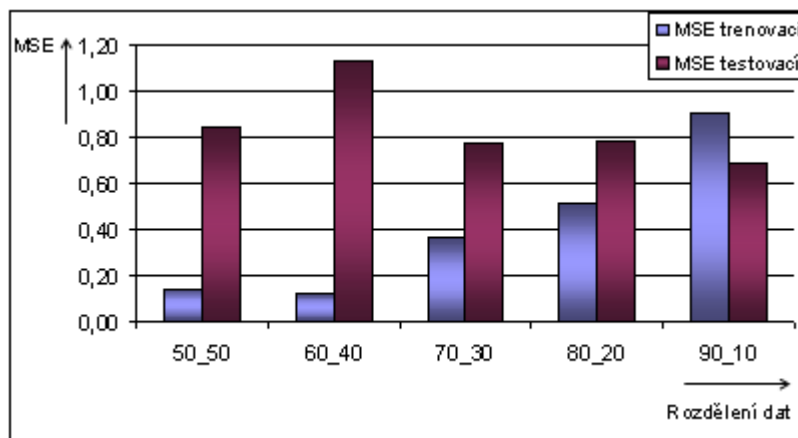
5.1 Analýza výsledků krátkodobé časové řady

Krátkodobá časová řada u frontální NS je rozdělena do pěti poměrů dat na trénovací a testovací množiny stejně jako u klasické dopředné NS. Pro učení byl zvolen algoritmus Time-Delay Backprop.

V procesu učení byl nejprve sledován vliv změny parametru d_{\max} na MSE chybu testovacích dat. Podle nejnižší dosažené chyby na testovacích datech je hodnota parametru zvolena pro další zpracování. V druhém kroku je sledována změna MSE chyby testovacích dat podle hodnot parametru η . Snahou je také dosáhnout nejnižší chyby na testovacích datech.

Jednotlivé výsledky všech typů rozdělení dat jsou uvedeny v příloze 6. Byla zde sledována střední čtvercová chyba (MSE) a výsledky jsou z důvodu lepší přehlednosti zaokrouhleny na 5 desetinných míst.

Celkové nejlepší výsledky všech typů rozdělení dat jsou zachycena na obr. 5-2 a jejich přesné hodnoty jsou uvedeny v tab. 5-1. Krátkodobá časová řada dosáhla nejnižší MSE chyby na testovacích datech při rozdělení dat v poměru 90:10.



Obr. 5-2: Porovnání výsledků pro jednotlivá rozdělení dat souboru S , zdroj [vlastní]

Tab. 5-1: Celkové výsledky pro soubor S , zdroj [vlastní]

| Rozdělení | d_{\max} | η | MSE trénovací | MSE testovací |
|-----------|------------|--------|---------------|---------------|
| 50:50 | 0.01 | 0.3 | 0.14422 | 0.83974 |
| 60:40 | 0.1 | 0.1 | 0.11831 | 1.13322 |
| 70:30 | 0.1 | 0.1 | 0.36276 | 0.77513 |
| 80:20 | 0.1 | 0.1 | 0.51687 | 0.78215 |
| 90:10 | 0.1 | 0.1 | 0.90683 | 0.68334 |

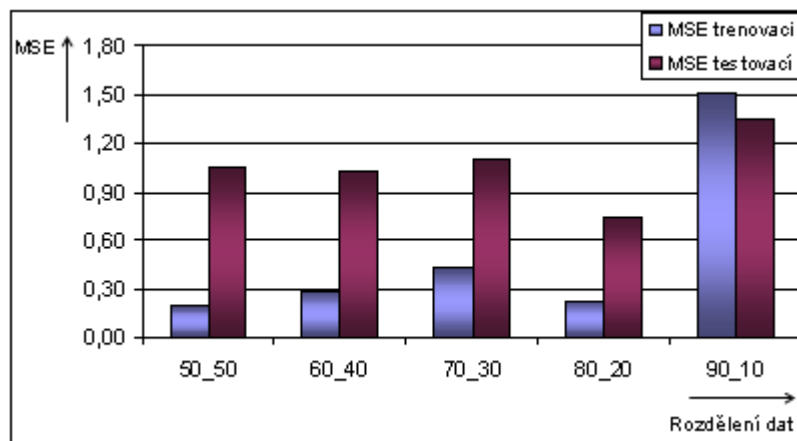
5.2 Analýza výsledků střednědobé časové řady

Střednědobá časová řada je opět rozdělena v pěti poměrech dat na trénovací a testovací množiny. Prvním krokem je sledován vliv změny parametru d_{\max} na výslednou MSE chybu u testovacích dat. S hodnotou parametru při nejnižší dosažené chybě se pokračuje v dalších experimentech. Druhým krokem je sledován vliv změny parametru η na MSE chybu u testovacích dat. Opět je sledována nejnižší dosažená MSE chyba na testovacích datech.

Jednotlivé výsledky při změnách parametrů všech typů rozdělení dat jsou uvedeny v příloze 6. Výsledky s nejnižší chybou na testovacích datech u všech rozdělení dat jsou zobrazeny na obr. 5-3. Konkrétně naměřené hodnoty těchto chyb jsou uvedeny v tab. 5-2.

Tab. 5-2: Celkové výsledky pro soubor I, zdroj [vlastní]

| Rozdělení | d_{\max} | η | MSE trénovací | MSE testovací |
|-----------|------------|--------|---------------|---------------|
| 50:50 | 0.1 | 0.1 | 0.19651 | 1.04864 |
| 60:40 | 0.1 | 0.1 | 0.28397 | 1.03130 |
| 70:30 | 0.1 | 0.1 | 0.43403 | 1.09977 |
| 80:20 | 0 | 0.1 | 0.22091 | 0.74048 |
| 90:10 | 0.1 | 0.1 | 1.50946 | 1.35651 |



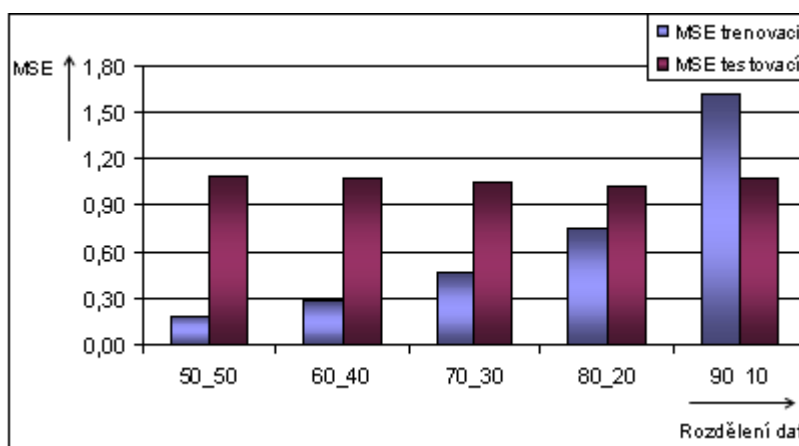
Obr. 5-3: Porovnání výsledků pro jednotlivá rozdělení dat souboru I, zdroj [vlastní]

5.3 Analýza výsledků dlouhodobé časové řady

Dlouhodobá časová řada byla rozdělena ve stejných poměrech dat jako předešlé časové řady. Pro učení této frontální NS byl také zvolen algoritmus Time-Delay Backprop. U tohoto algoritmu byl nejdříve sledován parametr d_{\max} , který byl volen zcela experimentálně, a jeho hodnoty měly různý vliv na výsledné MSE chyby u testovacích dat. Snahou bylo

dosáhnout do nejnižší chyby u testovacích dat. Druhým sledovaným parametrem je η , u kterého je sledována nejnižší chyba testovacích dat.

Provedené experimenty a jejich dílčí výsledky u všech typů rozdělení dat jsou uvedeny v příloze 6. Celkové výsledky jednotlivých poměrů dat s nejnižšími chybami na testovacích datech jsou zobrazeny na obr. 5-4, jejich přesné hodnoty jsou uvedeny v tab. 5-3.



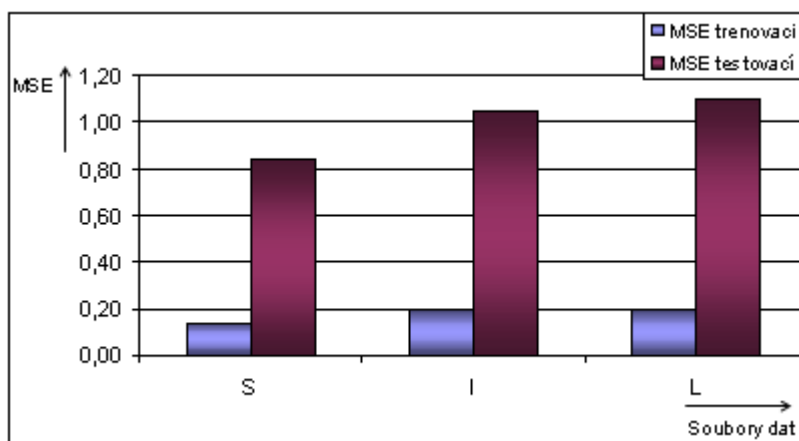
Obr. 5-4: Porovnání výsledků pro jednotlivá rozdělení dat souboru L , zdroj [vlastní]

Tab. 5-3: Celkové výsledky pro soubor L , zdroj [vlastní]

| Rozdělení | d_{\max} | η | MSE trénovací | MSE testovací |
|-----------|------------|--------|---------------|---------------|
| 50:50 | 0.1 | 0.1 | 0.19642 | 1.09790 |
| 60:40 | 0.1 | 0.1 | 0.28395 | 1.08210 |
| 70:30 | 0.1 | 0.1 | 0.47291 | 1.05870 |
| 80:20 | 0.1 | 0.1 | 0.75957 | 1.03264 |
| 90:10 | 0.1 | 0.1 | 1.62025 | 1.07689 |

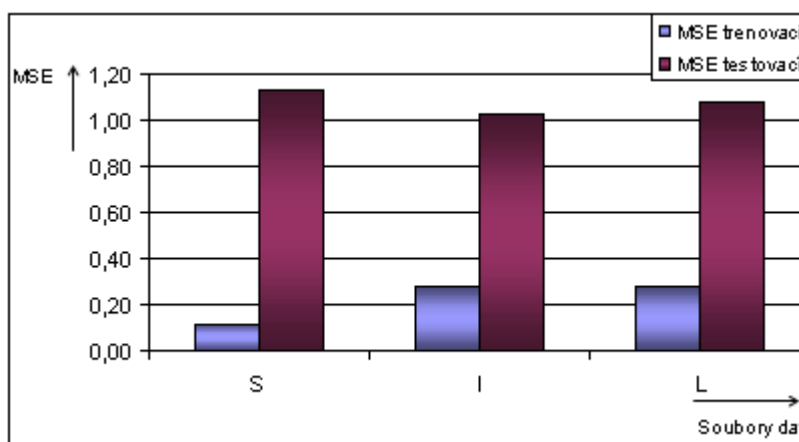
5.4 Porovnání jednotlivých časových řad

Krátkodobá, střednědobá i dlouhodobá časová řada je rozdělena v pěti poměrech dat na trénovací a testovací množiny. Porovnání nejlepších výsledků v těchto řadách v závislosti na rozdělení dat v poměru 50:50 je zobrazeno na obr. 5-5. Nejnižší MSE chyby na testovacích datech je dosaženo u krátkodobé časové řady, vyšší chyby jsou zachycena u střednědobé časové řady a nejhorších výsledků dosahuje dlouhodobá časová řada.



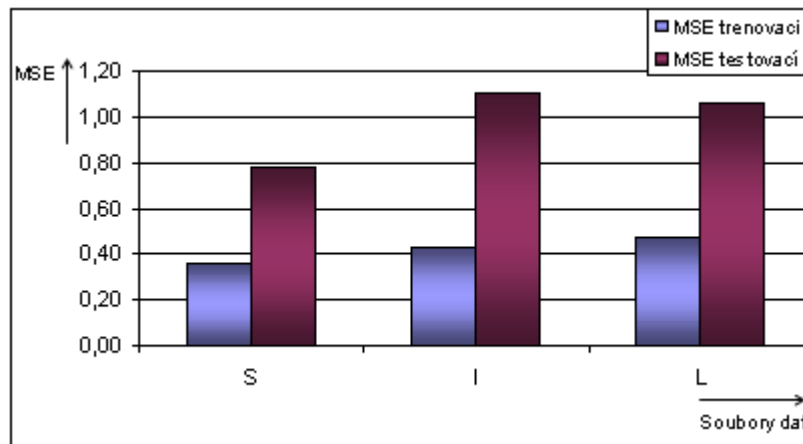
Obr. 5-5: Porovnání jednotlivých souborů dat pro rozdělení v poměru 50:50, zdroj [vlastní]

Porovnání jednotlivých souborů časových řad pro rozložení dat v poměru 60:40 zobrazuje obr. 5-6. Nejnížší MSE chyby na testovacích datech je dosaženo u střednědobé časové řady a nejvyšší MSE chyby na testovacích datech je dosaženo u krátkodobé časové řady.



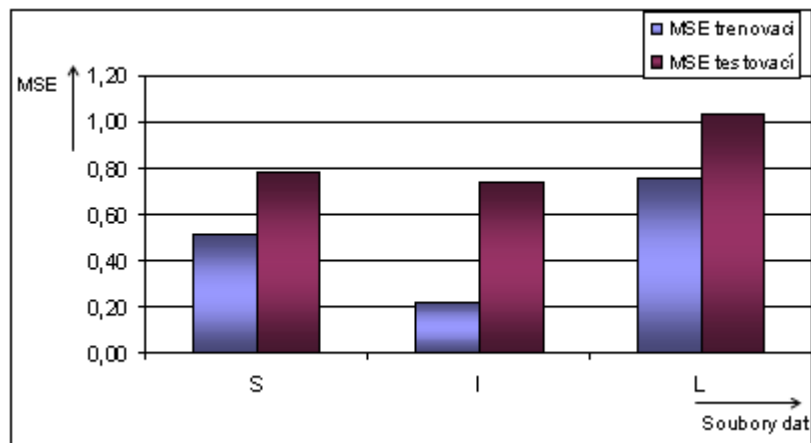
Obr. 5-6: Porovnání jednotlivých souborů dat pro rozdělení v poměru 60:40, zdroj [vlastní]

Na obr. 5-7 je zobrazeno porovnání jednotlivých časových řad pro rozdělení dat v poměru 70:30. Nejlepších výsledků na testovacích datech dosáhla krátkodobá časová řada. Nejhorších výsledků je dosaženo u střednědobé časové řady.



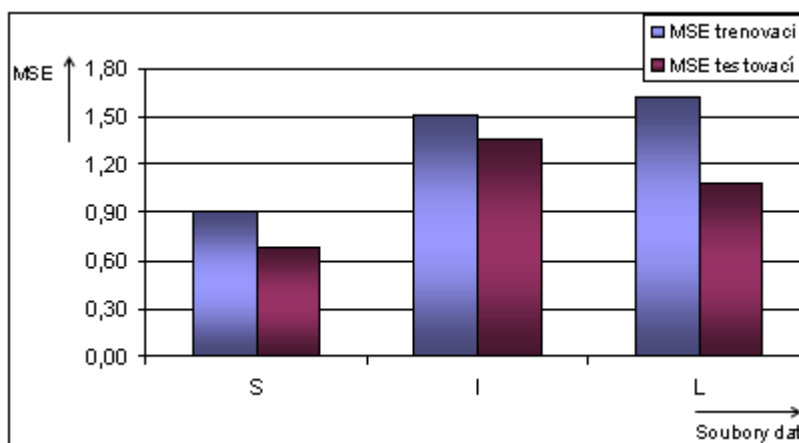
Obr. 5-7: Porovnání jednotlivých souborů dat pro rozdělení v poměru 70:30, zdroj [vlastní]

Porovnání nejlepších výsledků pro jednotlivé časové řady pro rozdělení dat v poměru 80:20 je zobrazeno na obr. 5-8. Nejlepších výsledků pro MSE chybu na testovacích datech je dosaženo u krátkodobé časové řady. Nejhorších výsledků dosahuje dlouhodobá časová řada.



Obr. 5-8: Porovnání jednotlivých souborů dat pro rozdělení v poměru 80:20, zdroj [vlastní]

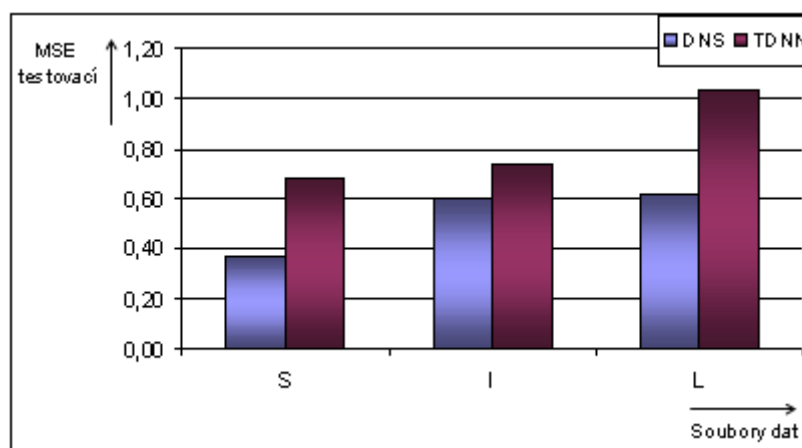
Pro rozdělení dat v poměru 90:10 jsou výsledky jednotlivých souborů dat zobrazeny na obr. 5-9. MSE chyba u testovacích dat pro toto rozdělení dosahuje nejlepších výsledků u krátkodobé časové řady, poté následuje dlouhodobá časová řada a nejvyšší MSE chyba je zachycena u střednědobé časové řady.



Obr. 5-9: Porovnání jednotlivých souborů dat pro rozdělení v poměru 90:10, zdroj [vlastní]

5.5 Porovnání dopředné a frontální neuronové sítě

Porovnání nejlepších výsledků u jednotlivých souborů dat pro dopřednou neuronovou síť (DNS) a frontální neuronovou síť (TDNN) zachycuje obr. 5-10. Podle tohoto obrázku je jednoznačně vidět, že všechny soubory dat dosahují nižších MSE chyb u testovacích dat v klasické dopředné NS. TDNN dosahuje u všech souborů S, I a L chyb vyšších. Konkrétní hodnoty dosažených výsledků jsou uvedeny v tab. 5-4.



Obr. 5-10: Porovnání DNS a TDNN podle jednotlivých souborů dat, zdroj [vlastní]

Tab. 5-4: Porovnání nejlepších výsledků DNS a TDNN, zdroj [vlastní]

| Soubor dat | Rozdělení | μ | c | d_{\max} | η | MSE trénovací | MSE testovací |
|------------|-----------|-------|--------|------------|--------|---------------|---------------|
| DNS – S | 90:10 | 0.3 | 0.1 | 0.1 | 0.2 | 0.20266 | 0.36969 |
| DNS – I | 60:40 | 0.3 | 0.0001 | 0.1 | 0.2 | 0.06119 | 0.59912 |
| DNS – L | 80:20 | 0.3 | 0.001 | 0.0001 | 0.6 | 0.28697 | 0.62016 |
| TDNN – S | 90:10 | - | - | 0.1 | 0.1 | 0.90683 | 0.68334 |
| TDNN – I | 80:20 | - | - | 0 | 0.1 | 0.22091 | 0.74048 |
| TDNN – L | 80:20 | - | - | 0.1 | 0.1 | 0.75957 | 1.03264 |

Na základě výše uvedených experimentů byly vytvořeny klasické dopředné neuronové sítě. Poměr trénovacích a testovacích dat a nastavení jednotlivých parametrů pro soubor S je uveden v tab. 5-5. Při tomto nastavení dosáhla MSE chyba trénovacích dat 0.20266 a testovacích dat 0.36969.

Tab. 5-5: Konečné nastavení parametrů pro soubor S, zdroj [vlastní]

| Rozdělení | μ | c | d_{\max} | η | MSE trénovací | MSE testovací |
|-----------|-------|-----|------------|--------|---------------|---------------|
| 90:10 | 0.3 | 0.1 | 0.1 | 0.2 | 0.20266 | 0.36969 |

Nastavení parametrů a rozdělení dat pro soubor I je uveden v tab. 5-6. V tomto nastavení dosáhla MSE chyba trénovacích dat 0.06119 a testovacích dat 0.59912.

Tab. 5-6: Konečné nastavení parametrů pro soubor I, zdroj [vlastní]

| Rozdělení | μ | c | d_{\max} | η | MSE trénovací | MSE testovací |
|-----------|-------|--------|------------|--------|---------------|---------------|
| 60:40 | 0.3 | 0.0001 | 0.1 | 0.2 | 0.06119 | 0.59912 |

Konečné nastavení parametrů pro soubor L a poměr rozdělení dat na trénovací a testovací množinu zobrazuje tab. 5-7. Při tomto nastavení dosáhla MSE chyba trénovacích dat 0.28697 a testovacích dat 0.62016.

Tab. 5-7: Konečné nastavení parametrů pro soubor L, zdroj [vlastní]

| Rozdělení | μ | c | d_{\max} | η | MSE trénovací | MSE testovací |
|-----------|-------|-------|------------|--------|---------------|---------------|
| 80:20 | 0.3 | 0.001 | 0.0001 | 0.6 | 0.28697 | 0.62016 |

5.6 Dílčí závěr

V této kapitole byly provedeny experimenty na model frontální neuronové sítě. Síť, která byla vytvořena, obsahuje 5 vstupních neuronů, 3 skryté neurony a 1 výstupní neuron,

hloubka filtru takto vytvořené sítě je 3. Pro učení NS byl u všech typů rozdělení dat a všech souborů zvolen algoritmus Time-Delay Backprop. Cílem provedených experimentů je dosažení nejnižší MSE chyby u testovacích dat.

V krátkodobé časové řadě je nejnižší chyby na testovacích datech dosaženo při rozdělení dat v poměru 90:10 pro nastavení parametrů na hodnoty $d_{\max} = 0.1$ a $\eta = 0.1$. Na trénovacích datech bylo dosaženo MSE chyby 0.90683 a testovacích dat 0.68334.

Pro střednědobou časovou řadu je nejnižší chyby dosaženo při rozdělení dat v poměru 80:20 pro nastavení parametrů na hodnoty $d_{\max} = 0$ a $\eta = 0.1$. Při tomto nastavení je dosaženo MSE chyby na trénovacích datech 0.22091 a testovacích datech 0.74048.

Dlouhodobá časová řada dosahuje nejnižší chyby na testovacích datech při poměru dat 80:20 a nastavení parametru na hodnoty $d_{\max} = 0.1$ a $\eta = 0.1$. Velikost MSE chyby u trénovacích dat je 0.75957 a testovacích dat je 1.03264.

Při porovnání klasické dopředné NS a TDNN jasně vyplývá, že nejlepších výsledků ve všech souborech dosahuje klasická dopředná NS.

Závěr

Cílem diplomové práce bylo navrhnout vhodný model pro predikci návštěvnosti webu upce.cz pomocí dopředných neuronových sítí.

Vstupem do NS byla předzpracovaná časová řada návštěvnosti webu upce.cz pomocí indikátorů technické analýzy. Takto upravené časové řady byly rozděleny v pěti různých poměrech dat na trénovací a testovací množiny.

Procesy učení byly realizovány pomocí klasické dopředné neuronové sítě a frontální neuronové sítě v programovém prostředí JavaNNS. Jednotlivé modely a jejich nastavené parametry byly hodnoceny pomocí střední kvadratické odchylky MSE. Důraz při učení NS byl kladen na to, aby chyba MSE na testovacích datech byla minimální.

Podle provedených experimentů a pevného počtu vstupů (5) není v tomto případě vhodné použít frontální NS, protože poskytuje vyšší MSE chyby na testovacích datech. Pro soubor S, I a L byla vytvořena klasická dopředná NS.

Soubor S byl rozdělen v poměru dat 90:10 s hodnotami parametrů $\mu = 0,3$; $c = 0,1$; $d_{\max} = 0,1$; $\eta = 0,2$. MSE trénovacích dat je 0,20266 a testovacích dat je 0,36969.

Soubor I byl rozdělen v poměru dat 60:40 s hodnotami parametrů $\mu = 0,3$; $c = 0,0001$; $d_{\max} = 0,1$; $\eta = 0,2$. MSE trénovacích dat je 0,06119 a testovacích dat je 0,59912.

Soubor L byl rozdělen v poměru dat 80:20 s hodnotami parametrů $\mu = 0,3$; $c = 0,001$; $d_{\max} = 0,0001$; $\eta = 0,6$. MSE trénovacích dat je 0,28697 a testovacích dat je 0,62016.

Závěrem lze říci, že cíle práce byly splněny.

Seznam použité literatury

- [1] HAVLÍČEK, J. Elektronické inteligence [online]. © 2003 [cit. 2011-03-19]. Elektronické inteligence: tvorba znalostí na základě využití dat z elektronického podnikání a obchodování. Dostupné z WWW: <<http://si.vse.cz/archive/proceedings/2000/elektronicke-inteligence-tvorba-znalosti-na-zaklade-vyuziti-dat-z-elektronickeho-podnikani-a-obchodovani.pdf>>.
- [2] OLEJ, V., Hájek, P., Filipová, J. Modelling of Web Domain Visits by IF-Inference Systém. WSEAS Transactions on Computers, WSEAS Press, Isme 10, Vol.9, October 2010, pp. 1170-1180, ISSN 1790-5079
- [3] Jelínek, J. Uživatelská podpora v prostředí WWW. INFORUM 2004, 10. konference o profesionálních informačních zdrojích. [online]. 2004 [cit. 2011-03-19].
- [4] SRIVASTAVA, J., DESIKAN, P. Web Mining - Accomplishments & Future Directions [online]. 2004 [cit. 2011-02-17]. Web mining. Dostupné z WWW: <<http://www.ieee.org.ar/downloads/Srivastava-tut-pres.pdf>>.
- [5] FILIPOVÁ, J., MICHÁLEK, K., PETR, P. Identifikace automatických přístupů internetových obchodů s využitím metod web usage miningu [online]. 2006 [cit. 2011-03-19]. Dostupné z WWW:<http://dspace.upce.cz/bitstream/10195/32234/1/CL639.pdf>.
- [6] WAAS, M.D.N. Knowledge Sharing [online]. 2006 [cit. 2011-03-19]. Web Mining. Dostupné z WWW: <<http://www.christiealwis.com/KnowledgeSharing.aspx>>.
- [7] MALIK, F. Extrakce informací z hypertextu. Brno, 2007. 54 s. Masarykova univerzita, Fakulta informatiky. Diplomová práce.
- [8] AKSHAT, Saxena. Web mining [online]. © 2011 [cit. 2011-02-20]. Web mining. Dostupné z WWW: <<http://www.slideshare.net/guestb73ec6/web-mining>>.
- [9] VELÁSQUEZ, Juan D. Mining web data: Techniques for understanding the user [online]. © 2006 [cit. 2011-02-20]. Mining web data. Dostupné z WWW: <<http://www.slideshare.net/Tommy96/mining-web-data-techniques-for-understanding-the-user>>.
- [10] DR. ZAIANE, Osmar R. Principles of Knowledge Discovery in Databases [online]. 1999 [cit. 2011-02-20]. Web mining. Dostupné z WWW: <<http://webdocs.cs.ualberta.ca/~zaiane/courses/cmput690/slides/Chapter9/sld001.htm>>.
- [11] BERKA, Pavel. Dobývání znalostí z databází [online]. 2006 [cit. 2011-02-20]. Dobývání znalostí z databází - mnohostranná interpretace dat. Dostupné z WWW: <<http://sorry.vse.cz/~berka/docs/4iz450/KDDprehled.pdf>>.

- [12] KVASNIČKA, V. a kol. Úvod do teórie neurónových sietí. Bratislava : Iris, 1997. 262 s.
- [13] KUBÍK, A. Inteligentní agenty : Tvorba aplikačního software na bázi multiagentových systémů. Brno : Computer Press, 2004. 280 s. ISBN 80-251-0323-4.
- [14] HYNEK, J. Genetické algoritmy a genetické programování. Praha : Grada, 2008. 182 s. ISBN 978-80-247-2695-3.
- [15] BÍLA, J. Umělá inteligence a neuronové sítě v aplikacích. Praha : ČVUT, 1998. 135 s. ISBN 80-01-01769-9.
- [16] OLEJ, V. Modelovanie ekonomických procesov na báze výpočtovej inteligencie. Hradec Králové : Česká republika, 2003. 160 s. ISBN 80-90324-9-1.
- [17] OLEJ, V., Hájek, P., VOMOČIL, M. Modelování bonity obcí pomocí dopředných neuronových sítí. Univerzita Pardubice, Fakulta ekonomicko-správní, Ústav systémového inženýrství a informatiky. 2007. 10 s.
- [18] JANOVSÝ, Dušan. Jak používat Google Analytics [online]. ©2011, 22. února 2011. [cit. 2011-03-09]. Google Analytics. Dostupné z WWW: <<http://www.jakpsatweb.cz/seo/google-analytics.html>>.
- [19] Google Analytics: Návod pro webmastery, marketéry i podnikatele - [online]. (c) Copyright 2001 - 2011 [cit. 2011-03-09]. Google Analytics: Návod pro webmastery, marketéry i podnikatele. Dostupné z WWW: <<http://webovaanalytika.robertainemec.com/google-analytics-navod/>>.
- [20] KELBEL, J., ŠILHÁN, D. Shluková analýza [online]. 2008 [cit. 2011-03-25]. Dostupné z WWW: <<http://staff.utia.cas.cz/nagy/skola/Projekty/Classification/ShlukovaAnalyza.pdf>>.
- [21] HAYKIN, S. Neural networks : a comprehensive foundation. 2nd edition. Upper Saddle River, New Jersey : Prentice Hall, 1999. 842 s. ISBN 0-13-273350-1.

Seznam použitých zkratek

| Zkratka | Anglický význam | Český význam |
|---------|---|--|
| WWW | World Wide Web | Internetové stránky |
| WUM | Web Usage Mining | Analýza chování uživatele |
| WSM | Web Structure Mining | Získávání informací ze struktury www |
| WCM | Web Content Mining | Zpracování obsahu www |
| FGS | Frequent Generalized Sessions | Metoda extrakce obvyklých vzorů chování |
| URL | Uniform Resource Locator | Jednotný lokátor zdrojů |
| DM | Data Mining | Dolování dat |
| CLF | Common Log Format | Standardní textový formát |
| IP | Internet Protokol | Internetový protokol |
| SQL | Structured Query Language | Strukturovaný dotazovací jazyk |
| TFIDF | Term Frequency Inverse Document Frequency | Váha, založená na frekvenci termů v dokumentu |
| TDIDT | Top Down Induction of Decision Trees | Algoritmus rozhodovacích stromů metodou shora dolů |
| SVM | Support Vector Machine | Metoda strojového učení |
| KSOFM | kohonen self-organizing maps | Kohonenovy samoorganizující se mapy |
| CBR | Case-Based Reasoning | Případové usuzování |
| IBL | Instance-Based Learning | Učení založené na instancích |
| NS | Neural Network | Neuronová síť |
| BP | Back-propagation | Algoritmus zpětného šíření chyby |
| TDNN | Time Delay Neural Network | Frontální neuronová síť |
| GA | Google Analytics | Program pro zachycení informací na internetu |
| JavaNNS | Java Neural Network Simulator | Simulátor pro neuronové sítě |
| JKP | Simple Moving Average | Jednoduchý klouzavý průměr |
| CKP | Central Moving Average | Centrovaný klouzavý průměr |
| KM | Moving Median | Klouzavý medián |
| JEV | Simple Exponential Smoothing | Jednoduché exponenciální vyrovnání |
| DEV | Double Exponential Smoothing | Dvojitě exponenciální vyrovnání |
| MSE | Root Mean Square Deviation | Průměrná kvadratická odchylka |

Seznam obrázků

| | |
|--|----|
| Obr. 1-1: Taxonomie web miningu, zdroj [10] | 11 |
| Obr. 1-2: Architektura WUM se zdrojem dat z Log souboru, zdroj [5]..... | 13 |
| Obr. 1-3: Architektura WUM se zdrojem dat z JavaScriptu a cookies, zdroj [5] | 14 |
| Obr. 2-1: Schéma trojvrstvé NS s dopředným šířením určena pro klasifikaci, zdroj [17] | 21 |
| Obr. 2-2: NS s dopředným šířením určená pro predikci, zdroj [upraveno podle 12]..... | 26 |
| Obr. 2-3: Příklad frontální neuronové sítě, zdroj [16]..... | 27 |
| Obr. 2-4: Filtr frontální NS, zdroj [16]..... | 28 |
| Obr. 3-1: Algoritmus pro návrh modelu, zdroj [vlastní] | 31 |
| Obr. 3-2: Návštěvnost web domény upce.cz souboru S, zdroj [vlastní] | 33 |
| Obr. 3-3: Návštěvnost web domény upce.cz souboru I, zdroj [vlastní] | 34 |
| Obr. 3-4: Návštěvnost web domény upce.cz souboru L, zdroj [vlastní] | 34 |
| Obr. 3-5: Rozklad množiny objektů na trénovací a testovací, zdroj [upraveno podle 12]..... | 36 |
| Obr. 3-6: Struktura navržené neuronové sítě, zdroj [vlastní] | 37 |
| Obr. 3-7: Popis záhlaví pro formát pat., zdroj [vlastní]..... | 38 |
| Obr. 3-8: Nastavení vah, zdroj [vlastní] | 39 |
| Obr. 3-9: Nastavení hodnot parametrů algoritmu Backprop-Momentum, zdroj [vlastní] | 39 |
| Obr. 3-10: Průběh MSE chyby, zdroj [vlastní]..... | 40 |
| Obr. 3-11: Váhy po procesu učení, zdroj [vlastní] | 40 |
| Obr. 3-12: Protokol o učení, zdroj [vlastní] | 41 |
| Obr. 4-1: Porovnání výsledků pro jednotlivá rozdělení dat souboru S, zdroj [vlastní]..... | 43 |
| Obr. 4-2: Porovnání výsledků pro jednotlivá rozdělení dat souboru I, zdroj [vlastní]..... | 44 |
| Obr. 4-3: Porovnání výsledků pro jednotlivá rozdělení dat souboru L, zdroj [vlastní]..... | 45 |
| Obr. 4-4: Porovnání jednotlivých souborů dat pro rozdělení v poměru 50:50, zdroj [vlastní] | 46 |
| Obr. 4-5: Porovnání jednotlivých souborů dat pro rozdělení v poměru 60:40, zdroj [vlastní] | 46 |
| Obr. 4-6: Porovnání jednotlivých souborů dat pro rozdělení v poměru 70:30, zdroj [vlastní] | 47 |
| Obr. 4-7: Porovnání jednotlivých souborů dat pro rozdělení v poměru 80:20, zdroj [vlastní] | 47 |
| Obr. 4-8: Porovnání jednotlivých souborů dat pro rozdělení v poměru 90:10, zdroj [vlastní] | 48 |
| Obr. 5-1: Struktura frontální NS, zdroj [vlastní] | 49 |
| Obr. 5-2: Porovnání výsledků pro jednotlivá rozdělení dat souboru S, zdroj [vlastní]..... | 50 |
| Obr. 5-3: Porovnání výsledků pro jednotlivá rozdělení dat souboru I, zdroj [vlastní]..... | 51 |
| Obr. 5-4: Porovnání výsledků pro jednotlivá rozdělení dat souboru L, zdroj [vlastní]..... | 52 |
| Obr. 5-5: Porovnání jednotlivých souborů dat pro rozdělení v poměru 50:50, zdroj [vlastní] | 53 |

| | |
|--|----|
| Obr. 5-6: Porovnání jednotlivých souborů dat pro rozdělení v poměru 60:40, zdroj [vlastní] | 53 |
| Obr. 5-7: Porovnání jednotlivých souborů dat pro rozdělení v poměru 70:30, zdroj [vlastní] | 54 |
| Obr. 5-8: Porovnání jednotlivých souborů dat pro rozdělení v poměru 80:20, zdroj [vlastní] | 54 |
| Obr. 5-9: Porovnání jednotlivých souborů dat pro rozdělení v poměru 90:10, zdroj [vlastní] | 55 |
| Obr. 5-10: Porovnání DNS a TDNN podle jednotlivých souborů dat, zdroj [vlastní] | 55 |

Seznam tabulek

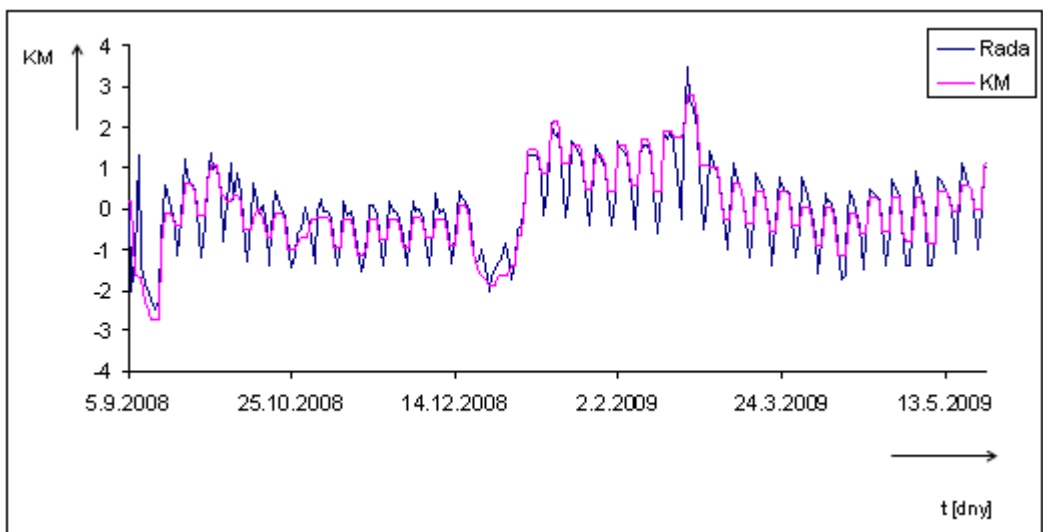
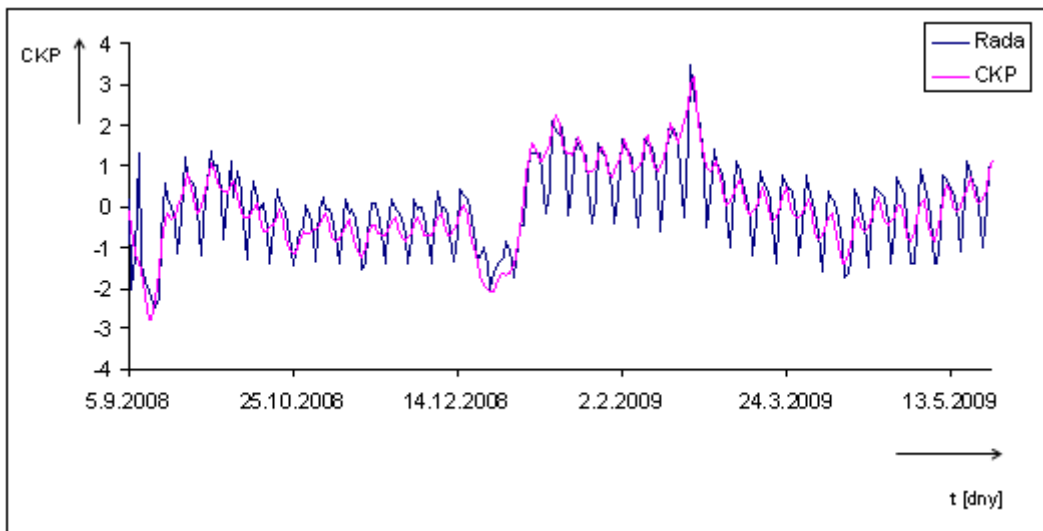
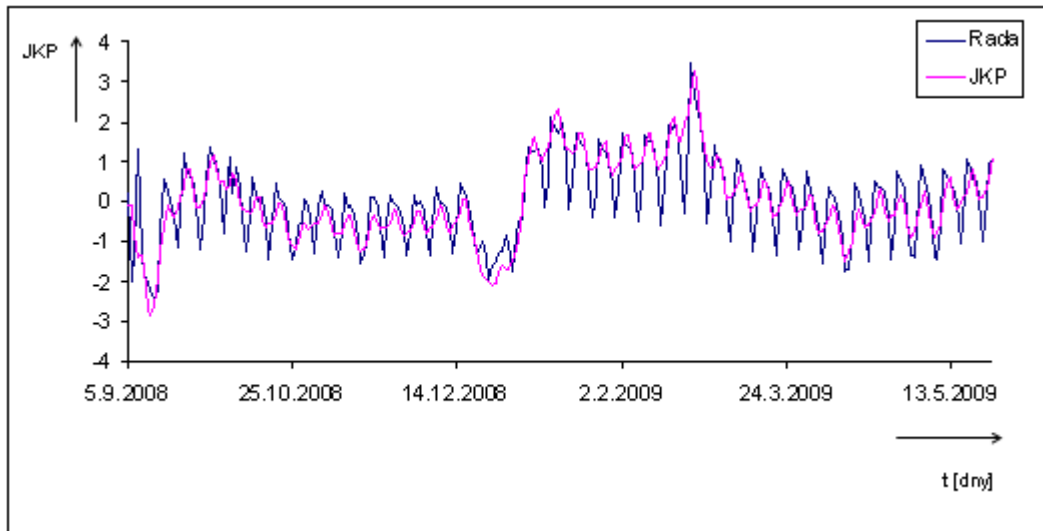
| | |
|---|----|
| Tab. 3-1: Indikátory technické analýzy, zdroj [upraveno podle 2] | 35 |
| Tab. 4-1: Celkové výsledky pro soubor S, zdroj [vlastní] | 43 |
| Tab. 4-3: Celkové výsledky pro soubor I, zdroj [vlastní] | 44 |
| Tab. 4-5: Celkové výsledky pro soubor L, zdroj [vlastní] | 45 |
| Tab. 5-1: Celkové výsledky pro soubor S, zdroj [vlastní] | 50 |
| Tab. 5-2: Celkové výsledky pro soubor I, zdroj [vlastní] | 51 |
| Tab. 5-3: Celkové výsledky pro soubor L, zdroj [vlastní] | 52 |
| Tab. 5-4: Porovnání nejlepších výsledků DNS a TDNN, zdroj [vlastní] | 56 |
| Tab. 5-5: Konečné nastavení parametrů pro soubor S, zdroj [vlastní] | 56 |
| Tab. 5-6: Konečné nastavení parametrů pro soubor I, zdroj [vlastní] | 56 |
| Tab. 5-7: Konečné nastavení parametrů pro soubor L, zdroj [vlastní] | 56 |

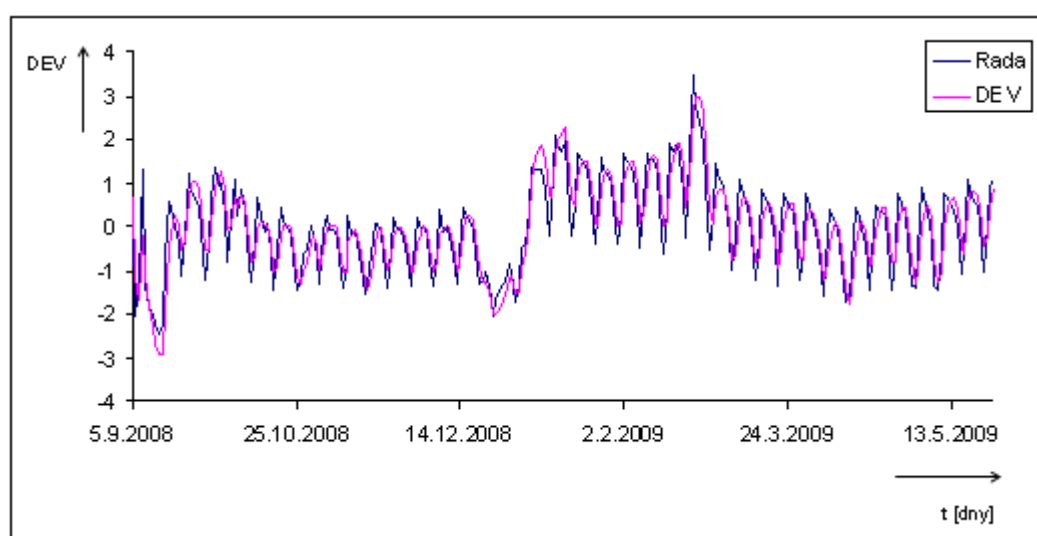
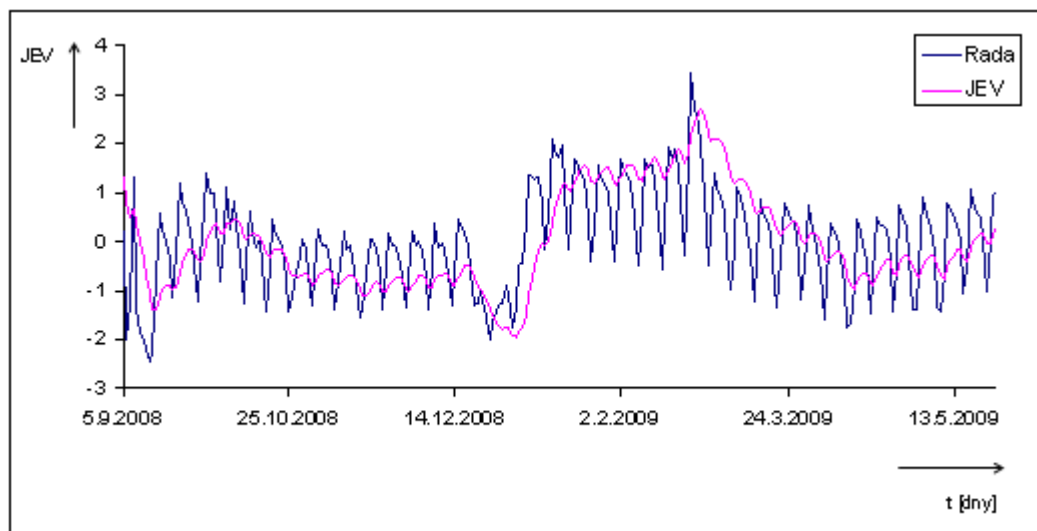
Seznam příloh

| | |
|--|----|
| Příloha 1 – Indikátory technické analýzy | 64 |
| Příloha 2 – Rozdělení dat na trénovací a testovací množinu | 69 |
| Příloha 3 – Výsledky učení krátkodobé časové řady | 71 |
| Příloha 4 – Výsledky učení střednědobé časové řady | 76 |
| Příloha 5 – Výsledky učení dlouhodobé časové řady | 81 |
| Příloha 6 – Výsledky učení frontální neuronové sítě | 86 |

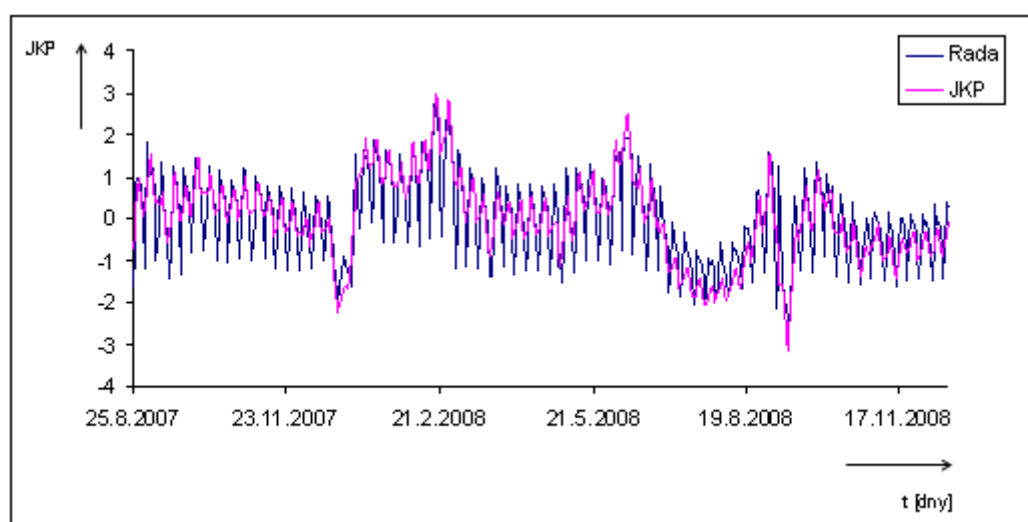
Příloha 1 – Indikátory technické analýzy

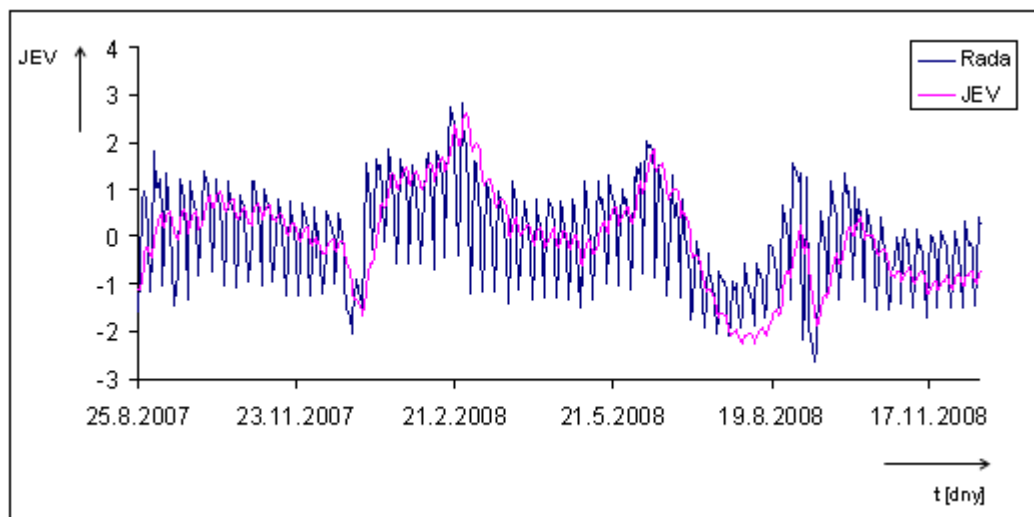
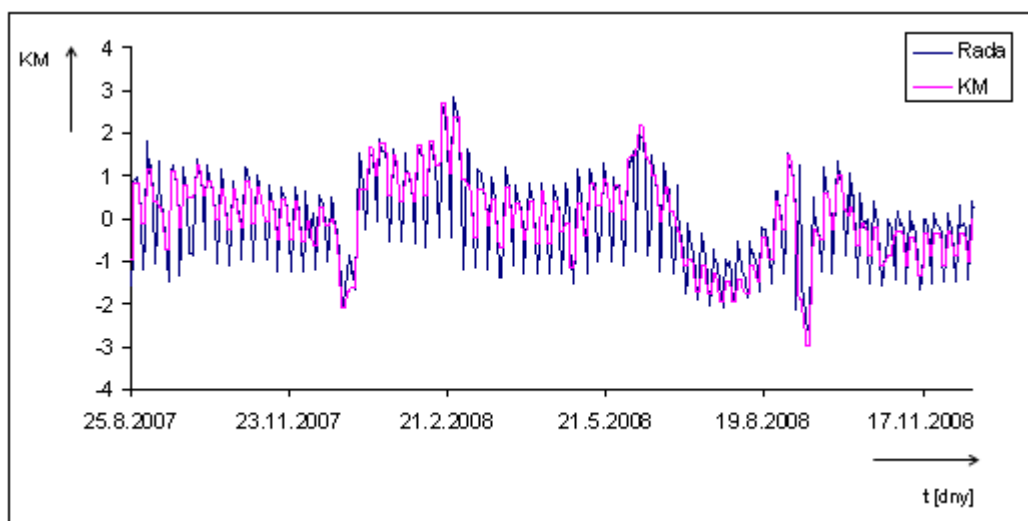
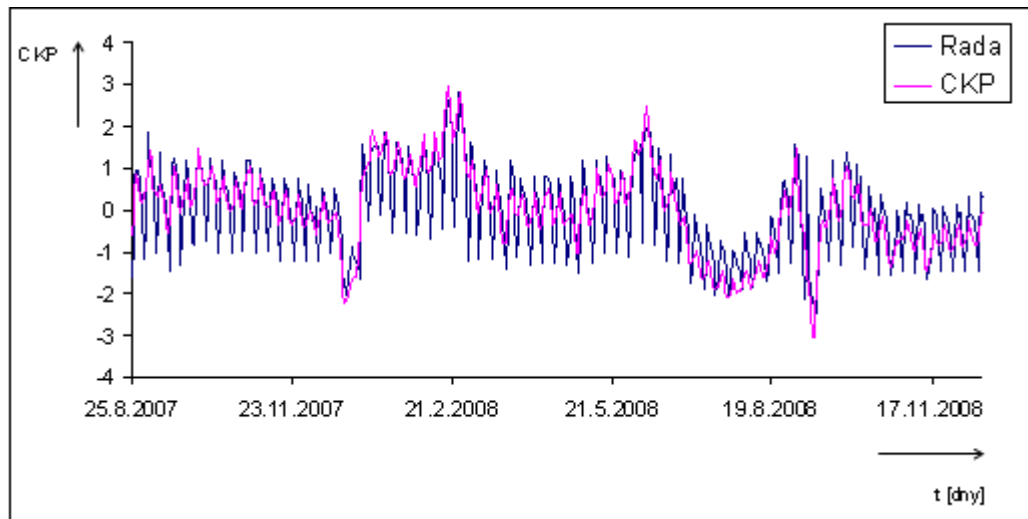
Indikátory technické analýzy souboru S

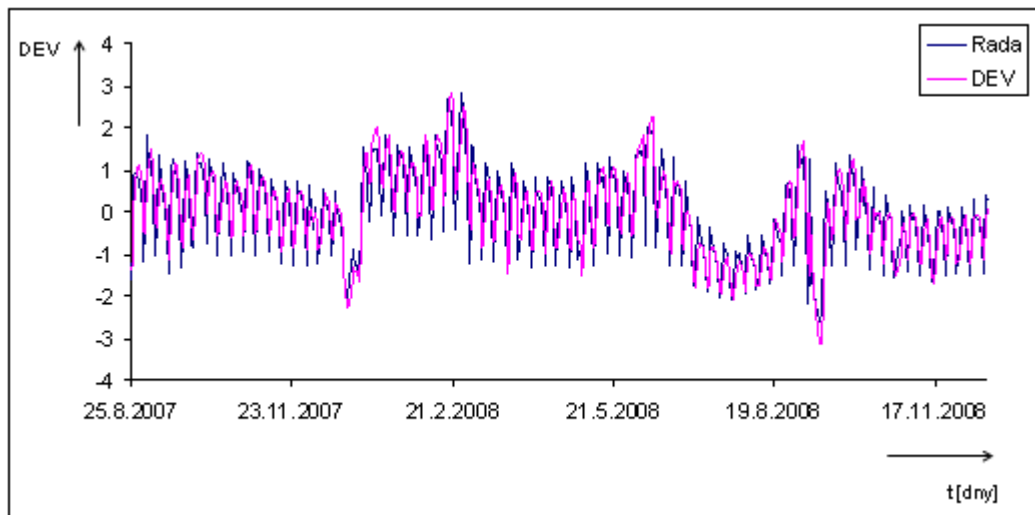




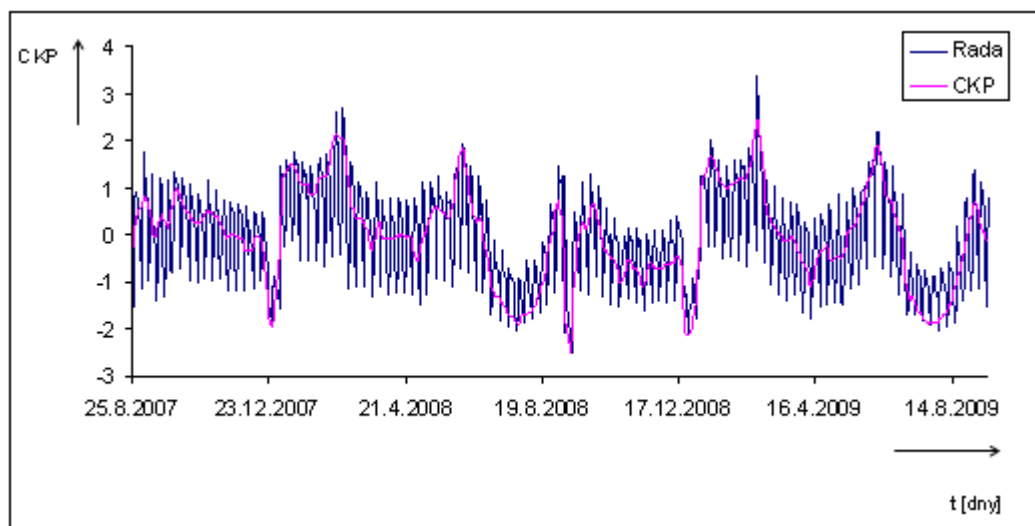
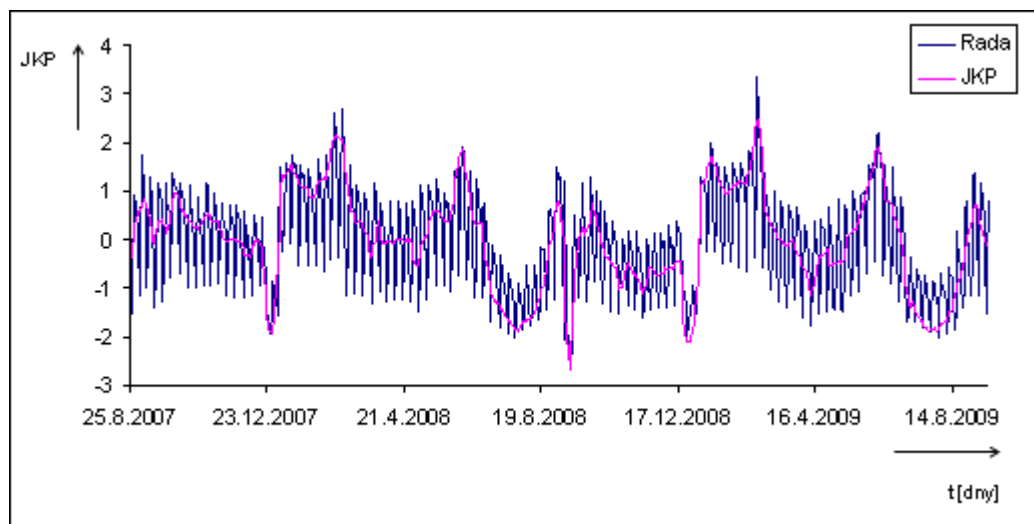
Indikátory technické analýzy souboru I

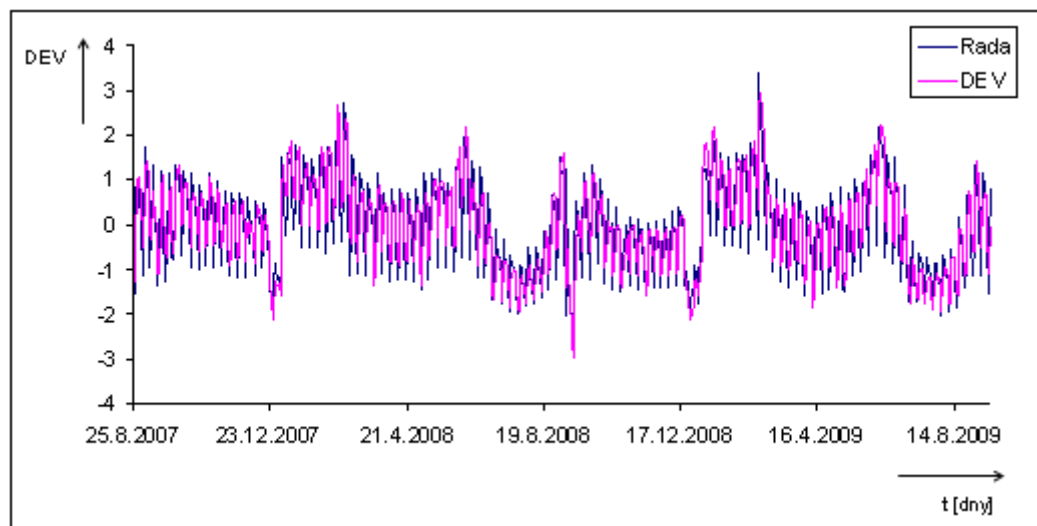
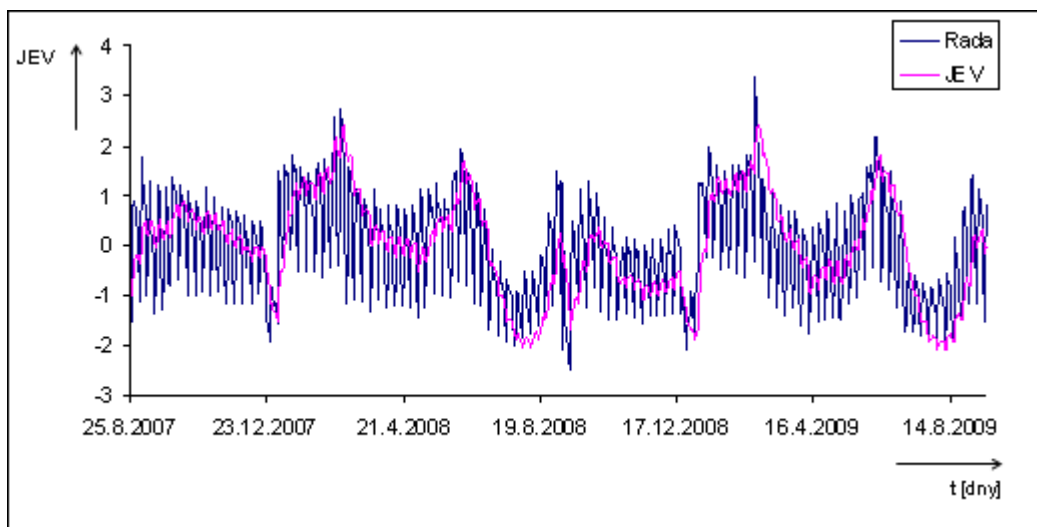
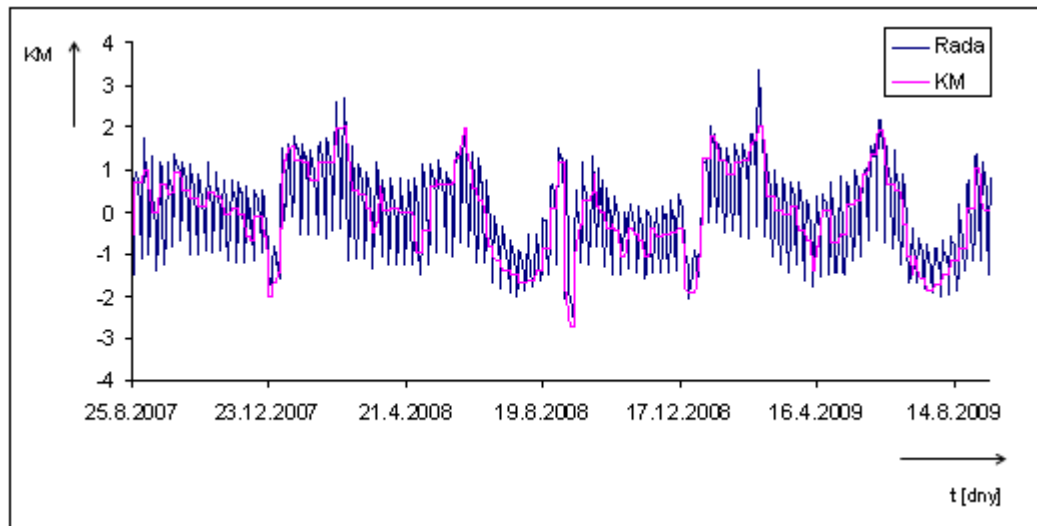






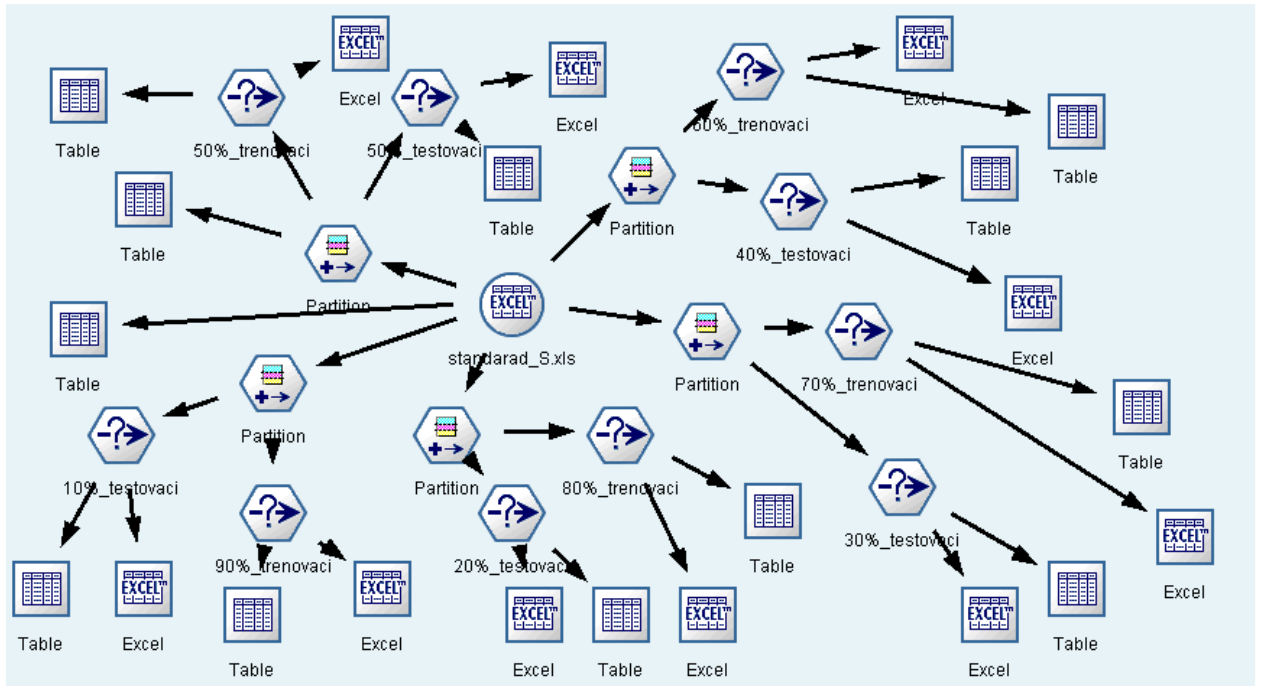
Indikátory technické analýzy souboru L



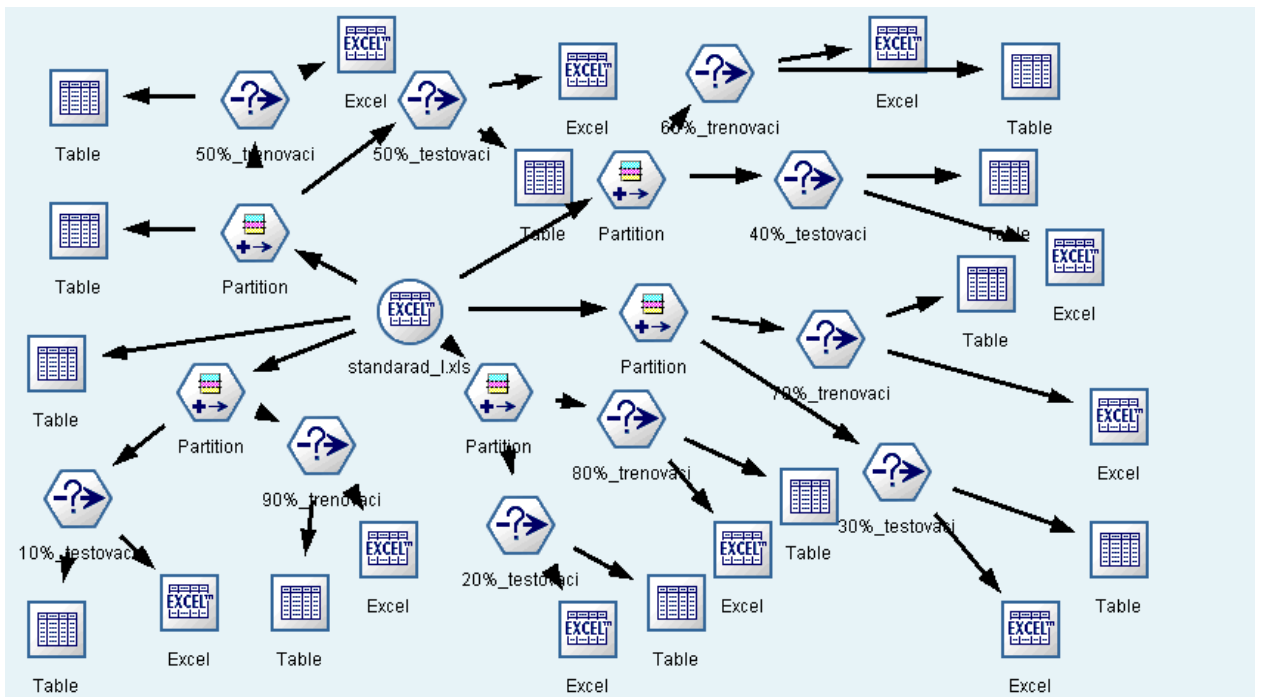


Příloha 2 – Rozdělení dat na trénovací a testovací množinu

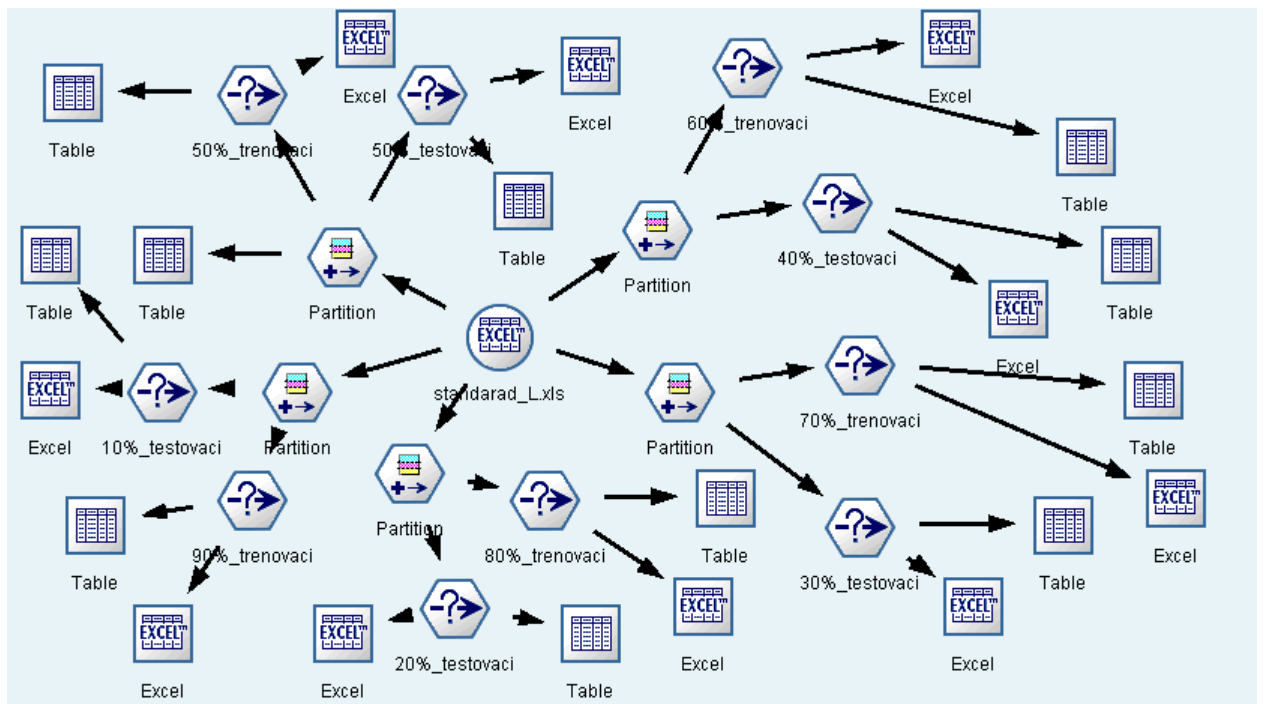
Rozdělení dat na trénovací a testovací množinu je realizováno v programovém prostředí SPSS Clementine 11.1.



Obr. 1: Rozdělení dat na trénovací a testovací množinu souboru S, zdroj [vlastní]



Obr. 2: Rozdělení dat na trénovací a testovací množinu souboru I, zdroj [vlastní]



Obr. 3: Rozdělení dat na trénovací a testovací množinu souboru L, zdroj [vlastní]

Příloha 3 – Výsledky učení krátkodobé časové řady

Poměr dat 50:50

Tab. 1: Změna parametru c pro soubor S (50:50), zdroj [vlastní]

| μ | c | d_{\max} | η | MSE trénovací | MSE testovací |
|------------|-------------|------------|------------|----------------|----------------|
| 0.5 | 0.1 | 0.1 | 0.2 | 0.01012 | 0.50247 |
| 0.5 | 0.01 | 0.1 | 0.2 | 0.01254 | 0.49513 |
| 0.5 | 0.001 | 0.1 | 0.2 | 0.01376 | 0.50421 |
| 0.5 | 0.0001 | 0.1 | 0.2 | 0.00973 | 0.49742 |

Tab. 2: Změna parametru μ pro soubor S (50:50), zdroj [vlastní]

| μ | c | d_{\max} | η | MSE trénovací | MSE testovací |
|------------|-------------|------------|------------|----------------|----------------|
| 0.1 | 0.01 | 0.1 | 0.2 | 0.01771 | 0.48595 |
| 0.3 | 0.01 | 0.1 | 0.2 | 0.03383 | 0.52121 |
| 0.5 | 0.01 | 0.1 | 0.2 | 0.01353 | 0.48442 |
| 0.7 | 0.01 | 0.1 | 0.2 | 0.00910 | 0.49500 |
| 0.9 | 0.01 | 0.1 | 0.2 | 0.00774 | 0.48778 |

Tab. 3: Změna parametru d_{\max} pro soubor S (50:50), zdroj [vlastní]

| μ | c | d_{\max} | η | MSE trénovací | MSE testovací |
|------------|-------------|------------|------------|----------------|----------------|
| 0.5 | 0.01 | 0.1 | 0.2 | 0.01353 | 0.48442 |
| 0.5 | 0.01 | 0.01 | 0.2 | 0.01386 | 0.49088 |
| 0.5 | 0.01 | 0.001 | 0.2 | 0.01078 | 0.49863 |
| 0.5 | 0.01 | 0.0001 | 0.2 | 0.01223 | 0.50900 |

Tab. 4: Změna parametru η pro soubor S (50:50), zdroj [vlastní]

| μ | c | d_{\max} | η | MSE trénovací | MSE testovací |
|------------|-------------|------------|------------|----------------|----------------|
| 0.5 | 0.01 | 0.1 | 0.2 | 0.01353 | 0.48442 |
| 0.5 | 0.01 | 0.1 | 0.4 | 0.00990 | 0.49129 |
| 0.5 | 0.01 | 0.1 | 0.6 | 0.00786 | 0.50391 |
| 0.5 | 0.01 | 0.1 | 0.8 | 0.00819 | 0.51333 |

Poměr dat 60:40

Tab. 5: Změna parametru c pro soubor S (60:40), zdroj [vlastní]

| μ | c | d_{\max} | η | MSE trénovací | MSE testovací |
|------------|------------|------------|------------|----------------|----------------|
| 0.5 | 0.1 | 0.1 | 0.2 | 0.02558 | 0.48239 |
| 0.5 | 0.01 | 0.1 | 0.2 | 0.02550 | 0.50245 |
| 0.5 | 0.001 | 0.1 | 0.2 | 0.02764 | 0.50300 |
| 0.5 | 0.0001 | 0.1 | 0.2 | 0.02394 | 0.50850 |

Tab. 6: Změna parametru μ pro soubor S (60:40), zdroj [vlastní]

| μ | c | d_{\max} | η | MSE trénovací | MSE testovací |
|------------|------------|------------|------------|----------------|----------------|
| 0.1 | 0.1 | 0.1 | 0.2 | 0.02849 | 0.48604 |
| 0.3 | 0.1 | 0.1 | 0.2 | 0.06151 | 0.51118 |
| 0.5 | 0.1 | 0.1 | 0.2 | 0.02558 | 0.48239 |
| 0.7 | 0.1 | 0.1 | 0.2 | 0.02633 | 0.47992 |
| 0.9 | 0.1 | 0.1 | 0.2 | 0.00592 | 0.49485 |

Tab. 7: Změna parametru d_{\max} pro soubor S (60:40), zdroj [vlastní]

| μ | c | d_{\max} | η | MSE trénovací | MSE testovací |
|------------|------------|------------|------------|----------------|----------------|
| 0.7 | 0.1 | 0.1 | 0.2 | 0.02633 | 0.47992 |
| 0.7 | 0.1 | 0.01 | 0.2 | 0.02579 | 0.48014 |
| 0.7 | 0.1 | 0.001 | 0.2 | 0.01646 | 0.50327 |
| 0.7 | 0.1 | 0.0001 | 0.2 | 0.01401 | 0.48643 |

Tab. 8: Změna parametru η pro soubor S (60:40), zdroj [vlastní]

| μ | c | d_{\max} | η | MSE trénovací | MSE testovací |
|------------|------------|------------|------------|----------------|----------------|
| 0.7 | 0.1 | 0.1 | 0.2 | 0.02633 | 0.47992 |
| 0.7 | 0.1 | 0.1 | 0.4 | 0.01295 | 0.48555 |
| 0.7 | 0.1 | 0.1 | 0.6 | 0.03837 | 0.47806 |
| 0.7 | 0.1 | 0.1 | 0.8 | 0.04013 | 0.50534 |

Poměr dat 70:30

Tab. 9: Změna parametru c pro soubor S (70:30), zdroj [vlastní]

| μ | c | d_{\max} | η | MSE trénovací | MSE testovací |
|------------|-------------|------------|------------|----------------|----------------|
| 0.5 | 0.1 | 0.1 | 0.2 | 0.03930 | 0.44601 |
| 0.5 | 0.01 | 0.1 | 0.2 | 0.04159 | 0.44486 |
| 0.5 | 0.001 | 0.1 | 0.2 | 0.04077 | 0.45545 |
| 0.5 | 0.0001 | 0.1 | 0.2 | 0.04863 | 0.44800 |

Tab. 10: Změna parametru μ pro soubor S (70:30), zdroj [vlastní]

| μ | c | d_{\max} | η | MSE trénovací | MSE testovací |
|------------|-------------|------------|------------|----------------|----------------|
| 0.1 | 0.01 | 0.1 | 0.2 | 0.04791 | 0.44646 |
| 0.3 | 0.01 | 0.1 | 0.2 | 0.08754 | 0.46461 |
| 0.5 | 0.01 | 0.1 | 0.2 | 0.04981 | 0.43003 |
| 0.7 | 0.01 | 0.1 | 0.2 | 0.03174 | 0.46278 |
| 0.9 | 0.01 | 0.1 | 0.2 | 0.02104 | 0.43056 |

Tab. 11: Změna parametru d_{\max} pro soubor S (70:30), zdroj [vlastní]

| μ | c | d_{\max} | η | MSE trénovací | MSE testovací |
|------------|-------------|------------|------------|----------------|----------------|
| 0.5 | 0.01 | 0.1 | 0.2 | 0.04981 | 0.43003 |
| 0.5 | 0.01 | 0.01 | 0.2 | 0.03437 | 0.45463 |
| 0.5 | 0.01 | 0.001 | 0.2 | 0.03455 | 0.45931 |
| 0.5 | 0.01 | 0.0001 | 0.2 | 0.04149 | 0.46253 |

Tab. 12: Změna parametru η pro soubor S (70:30), zdroj [vlastní]

| μ | c | d_{\max} | η | MSE trénovací | MSE testovací |
|------------|-------------|------------|------------|----------------|----------------|
| 0.5 | 0.01 | 0.1 | 0.2 | 0.04981 | 0.43003 |
| 0.5 | 0.01 | 0.1 | 0.4 | 0.05357 | 0.44688 |
| 0.5 | 0.01 | 0.1 | 0.6 | 0.04462 | 0.45547 |
| 0.5 | 0.01 | 0.1 | 0.8 | 0.01567 | 0.45371 |

Poměr dat 80:20

Tab. 13: Změna parametru c pro soubor S (80:20), zdroj [vlastní]

| μ | c | d_{\max} | η | MSE trénovací | MSE testovací |
|------------|------------|------------|------------|----------------|----------------|
| 0.5 | 0.1 | 0.1 | 0.2 | 0.07181 | 0.45249 |
| 0.5 | 0.01 | 0.1 | 0.2 | 0.06500 | 0.47066 |
| 0.5 | 0.001 | 0.1 | 0.2 | 0.06865 | 0.47390 |
| 0.5 | 0.0001 | 0.1 | 0.2 | 0.05874 | 0.47568 |

Tab. 14: Změna parametru μ pro soubor S (80:20), zdroj [vlastní]

| μ | c | d_{\max} | η | MSE trénovací | MSE testovací |
|------------|------------|------------|------------|----------------|----------------|
| 0.1 | 0.1 | 0.1 | 0.2 | 0.07326 | 0.46658 |
| 0.3 | 0.1 | 0.1 | 0.2 | 0.10863 | 0.46917 |
| 0.5 | 0.1 | 0.1 | 0.2 | 0.10008 | 0.46291 |
| 0.7 | 0.1 | 0.1 | 0.2 | 0.10357 | 0.46915 |
| 0.9 | 0.1 | 0.1 | 0.2 | 0.10283 | 0.48151 |

Tab. 15: Změna parametru d_{\max} pro soubor S (80:20), zdroj [vlastní]

| μ | c | d_{\max} | η | MSE trénovací | MSE testovací |
|------------|------------|--------------|------------|----------------|----------------|
| 0.5 | 0.1 | 0.1 | 0.2 | 0.10008 | 0.46291 |
| 0.5 | 0.1 | 0.01 | 0.2 | 0.05614 | 0.48222 |
| 0.5 | 0.1 | 0.001 | 0.2 | 0.06832 | 0.45103 |
| 0.5 | 0.1 | 0.0001 | 0.2 | 0.07847 | 0.46843 |

Tab. 16: Změna parametru η pro soubor S (80:20), zdroj [vlastní]

| μ | c | d_{\max} | η | MSE trénovací | MSE testovací |
|------------|------------|--------------|------------|----------------|----------------|
| 0.5 | 0.1 | 0.001 | 0.2 | 0.06832 | 0.45103 |
| 0.5 | 0.1 | 0.001 | 0.4 | 0.06988 | 0.48213 |
| 0.5 | 0.1 | 0.001 | 0.6 | 0.04397 | 0.48522 |
| 0.5 | 0.1 | 0.001 | 0.8 | 0.08101 | 0.47839 |

Poměr dat 90:10

Tab. 17: Změna parametru c pro soubor $S(90:10)$, zdroj [vlastní]

| μ | c | d_{\max} | η | MSE trénovací | MSE testovací |
|------------|------------|------------|------------|----------------|----------------|
| 0.5 | 0.1 | 0.1 | 0.2 | 0.12114 | 0.37991 |
| 0.5 | 0.01 | 0.1 | 0.2 | 0.13950 | 0.38169 |
| 0.5 | 0.001 | 0.1 | 0.2 | 0.14644 | 0.38833 |
| 0.5 | 0.0001 | 0.1 | 0.2 | 0.14156 | 0.39067 |

Tab. 18: Změna parametru μ pro soubor $S(90:10)$, zdroj [vlastní]

| μ | c | d_{\max} | η | MSE trénovací | MSE testovací |
|------------|------------|------------|------------|----------------|----------------|
| 0.1 | 0.1 | 0.1 | 0.2 | 0.16092 | 0.37417 |
| 0.3 | 0.1 | 0.1 | 0.2 | 0.20266 | 0.36969 |
| 0.5 | 0.1 | 0.1 | 0.2 | 0.19205 | 0.38359 |
| 0.7 | 0.1 | 0.1 | 0.2 | 0.14832 | 0.37711 |
| 0.9 | 0.1 | 0.1 | 0.2 | 0.11093 | 0.39544 |

Tab. 19: Změna parametru d_{\max} pro soubor $S(90:10)$, zdroj [vlastní]

| μ | c | d_{\max} | η | MSE trénovací | MSE testovací |
|------------|------------|------------|------------|----------------|----------------|
| 0.3 | 0.1 | 0.1 | 0.2 | 0.20266 | 0.36969 |
| 0.3 | 0.1 | 0.01 | 0.2 | 0.10397 | 0.39197 |
| 0.3 | 0.1 | 0.001 | 0.2 | 0.12862 | 0.37324 |
| 0.3 | 0.1 | 0.0001 | 0.2 | 0.16969 | 0.38013 |

Tab. 20: Změna parametru η pro soubor $S(90:10)$, zdroj [vlastní]

| μ | c | d_{\max} | η | MSE trénovací | MSE testovací |
|------------|------------|------------|------------|----------------|----------------|
| 0.3 | 0.1 | 0.1 | 0.2 | 0.20266 | 0.36969 |
| 0.3 | 0.1 | 0.1 | 0.4 | 0.16918 | 0.37010 |
| 0.3 | 0.1 | 0.1 | 0.6 | 0.15334 | 0.40064 |
| 0.3 | 0.1 | 0.1 | 0.8 | 0.14477 | 0.38916 |

Příloha 4 – Výsledky učení střednědobé časové řady

Poměr dat 50:50

Tab. 21: Změna parametru c pro soubor I (50:50), zdroj [vlastní]

| μ | c | d_{\max} | η | MSE trénovací | MSE testovací |
|------------|--------------|------------|------------|----------------|----------------|
| 0.5 | 0.1 | 0.1 | 0.2 | 0.02617 | 0.61976 |
| 0.5 | 0.01 | 0.1 | 0.2 | 0.02455 | 0.62105 |
| 0.5 | 0.001 | 0.1 | 0.2 | 0.01949 | 0.61136 |
| 0.5 | 0.0001 | 0.1 | 0.2 | 0.02973 | 0.61364 |

Tab. 22: Změna parametru μ pro soubor I (50:50), zdroj [vlastní]

| μ | c | d_{\max} | η | MSE trénovací | MSE testovací |
|------------|--------------|------------|------------|----------------|----------------|
| 0.1 | 0.001 | 0.1 | 0.2 | 0.03618 | 0.62432 |
| 0.3 | 0.001 | 0.1 | 0.2 | 0.03952 | 0.61064 |
| 0.5 | 0.001 | 0.1 | 0.2 | 0.02462 | 0.65976 |
| 0.7 | 0.001 | 0.1 | 0.2 | 0.01862 | 0.63337 |
| 0.9 | 0.001 | 0.1 | 0.2 | 0.01405 | 0.65683 |

Tab. 23: Změna parametru d_{\max} pro soubor I (50:50), zdroj [vlastní]

| μ | c | d_{\max} | η | MSE trénovací | MSE testovací |
|------------|--------------|---------------|------------|----------------|----------------|
| 0.3 | 0.001 | 0.1 | 0.2 | 0.03952 | 0.61064 |
| 0.3 | 0.001 | 0.01 | 0.2 | 0.02822 | 0.62263 |
| 0.3 | 0.001 | 0.001 | 0.2 | 0.02732 | 0.61271 |
| 0.3 | 0.001 | 0.0001 | 0.2 | 0.03337 | 0.60962 |

Tab. 24: Změna parametru η pro soubor I (50:50), zdroj [vlastní]

| μ | c | d_{\max} | η | MSE trénovací | MSE testovací |
|------------|--------------|------------|------------|----------------|----------------|
| 0.3 | 0.001 | 0.1 | 0.2 | 0.03952 | 0.61064 |
| 0.3 | 0.001 | 0.1 | 0.4 | 0.02262 | 0.62105 |
| 0.3 | 0.001 | 0.1 | 0.6 | 0.01965 | 0.61943 |
| 0.3 | 0.001 | 0.1 | 0.8 | 0.01954 | 0.62595 |

Poměr dat 60:40

Tab. 25: Změna parametru c pro soubor I (60:40), zdroj [vlastní]

| μ | c | d_{\max} | η | MSE trénovací | MSE testovací |
|------------|--------------|------------|------------|----------------|----------------|
| 0.5 | 0.1 | 0.1 | 0.2 | 0.04487 | 0.61227 |
| 0.5 | 0.01 | 0.1 | 0.2 | 0.04043 | 0.61296 |
| 0.5 | 0.001 | 0.1 | 0.2 | 0.03407 | 0.59308 |
| 0.5 | 0.0001 | 0.1 | 0.2 | 0.04094 | 0.60719 |

Tab. 26: Změna parametru μ pro soubor I (60:40), zdroj [vlastní]

| μ | c | d_{\max} | η | MSE trénovací | MSE testovací |
|------------|--------------|------------|------------|----------------|----------------|
| 0.1 | 0.001 | 0.1 | 0.2 | 0.06215 | 0.59972 |
| 0.3 | 0.001 | 0.1 | 0.2 | 0.06119 | 0.59912 |
| 0.5 | 0.001 | 0.1 | 0.2 | 0.04311 | 0.63805 |
| 0.7 | 0.001 | 0.1 | 0.2 | 0.03667 | 0.62243 |
| 0.9 | 0.001 | 0.1 | 0.2 | 0.02866 | 0.62176 |

Tab. 27: Změna parametru d_{\max} pro soubor I (60:40), zdroj [vlastní]

| μ | c | d_{\max} | η | MSE trénovací | MSE testovací |
|------------|--------------|------------|------------|----------------|----------------|
| 0.3 | 0.001 | 0.1 | 0.2 | 0.06119 | 0.59912 |
| 0.3 | 0.001 | 0.01 | 0.2 | 0.04732 | 0.60519 |
| 0.3 | 0.001 | 0.001 | 0.2 | 0.04889 | 0.60622 |
| 0.3 | 0.001 | 0.0001 | 0.2 | 0.05947 | 0.59996 |

Tab. 28: Změna parametru η pro soubor I (60:40), zdroj [vlastní]

| μ | c | d_{\max} | η | MSE trénovací | MSE testovací |
|------------|---------------|------------|------------|----------------|----------------|
| 0.3 | 0.0001 | 0.1 | 0.2 | 0.06119 | 0.59912 |
| 0.3 | 0.0001 | 0.1 | 0.4 | 0.03694 | 0.61730 |
| 0.3 | 0.0001 | 0.1 | 0.6 | 0.03394 | 0.61550 |
| 0.3 | 0.0001 | 0.1 | 0.8 | 0.03103 | 0.62173 |

Poměr dat 70:30

Tab. 29: Změna parametru c pro soubor I (70:30), zdroj [vlastní]

| μ | c | d_{\max} | η | MSE trénovací | MSE testovací |
|------------|-------------|------------|------------|----------------|----------------|
| 0.5 | 0.1 | 0.1 | 0.2 | 0.05122 | 0.66021 |
| 0.5 | 0.01 | 0.1 | 0.2 | 0.06699 | 0.65122 |
| 0.5 | 0.001 | 0.1 | 0.2 | 0.05319 | 0.66458 |
| 0.5 | 0.0001 | 0.1 | 0.2 | 0.06370 | 0.66104 |

Tab. 30: Změna parametru μ pro soubor I (70:30), zdroj [vlastní]

| μ | c | d_{\max} | η | MSE trénovací | MSE testovací |
|------------|-------------|------------|------------|----------------|----------------|
| 0.1 | 0.01 | 0.1 | 0.2 | 0.06997 | 0.64985 |
| 0.3 | 0.01 | 0.1 | 0.2 | 0.08636 | 0.62066 |
| 0.5 | 0.01 | 0.1 | 0.2 | 0.06607 | 0.68792 |
| 0.7 | 0.01 | 0.1 | 0.2 | 0.05037 | 0.65746 |
| 0.9 | 0.01 | 0.1 | 0.2 | 0.04564 | 0.66238 |

Tab. 31: Změna parametru d_{\max} pro soubor I (70:30), zdroj [vlastní]

| μ | c | d_{\max} | η | MSE trénovací | MSE testovací |
|------------|-------------|------------|------------|----------------|----------------|
| 0.3 | 0.01 | 0.1 | 0.2 | 0.08636 | 0.62066 |
| 0.3 | 0.01 | 0.01 | 0.2 | 0.06673 | 0.65725 |
| 0.3 | 0.01 | 0.001 | 0.2 | 0.06854 | 0.65234 |
| 0.3 | 0.01 | 0.0001 | 0.2 | 0.06854 | 0.64457 |

Tab. 32: Změna parametru η pro soubor I (70:30), zdroj [vlastní]

| μ | c | d_{\max} | η | MSE trénovací | MSE testovací |
|------------|-------------|------------|------------|----------------|----------------|
| 0.3 | 0.01 | 0.1 | 0.2 | 0.08636 | 0.62066 |
| 0.3 | 0.01 | 0.1 | 0.4 | 0.05736 | 0.65771 |
| 0.3 | 0.01 | 0.1 | 0.6 | 0.06369 | 0.65372 |
| 0.3 | 0.01 | 0.1 | 0.8 | 0.04924 | 0.67994 |

Poměr dat 80:20

Tab. 33: Změna parametru c pro soubor I (80:20), zdroj [vlastní]

| μ | c | d_{\max} | η | MSE trénovací | MSE testovací |
|------------|-------------|------------|------------|----------------|----------------|
| 0.5 | 0.1 | 0.1 | 0.2 | 0.07148 | 0.69075 |
| 0.5 | 0.01 | 0.1 | 0.2 | 0.10258 | 0.66543 |
| 0.5 | 0.001 | 0.1 | 0.2 | 0.09100 | 0.68349 |
| 0.5 | 0.0001 | 0.1 | 0.2 | 0.10062 | 0.68018 |

Tab. 34: Změna parametru μ pro soubor I (80:20), zdroj [vlastní]

| μ | c | d_{\max} | η | MSE trénovací | MSE testovací |
|------------|-------------|------------|------------|----------------|----------------|
| 0.1 | 0.01 | 0.1 | 0.2 | 0.11365 | 0.66243 |
| 0.3 | 0.01 | 0.1 | 0.2 | 0.13196 | 0.63456 |
| 0.5 | 0.01 | 0.1 | 0.2 | 0.09941 | 0.65815 |
| 0.7 | 0.01 | 0.1 | 0.2 | 0.07300 | 0.67167 |
| 0.9 | 0.01 | 0.1 | 0.2 | 0.07454 | 0.68726 |

Tab. 35: Změna parametru d_{\max} pro soubor I (80:20), zdroj [vlastní]

| μ | c | d_{\max} | η | MSE trénovací | MSE testovací |
|------------|-------------|------------|------------|----------------|----------------|
| 0.3 | 0.01 | 0.1 | 0.2 | 0.13196 | 0.63456 |
| 0.3 | 0.01 | 0.01 | 0.2 | 0.10059 | 0.67875 |
| 0.3 | 0.01 | 0.001 | 0.2 | 0.11107 | 0.66650 |
| 0.3 | 0.01 | 0.0001 | 0.2 | 0.08712 | 0.66431 |

Tab. 36: Změna parametru η pro soubor I (80:20), zdroj [vlastní]

| μ | c | d_{\max} | η | MSE trénovací | MSE testovací |
|------------|-------------|------------|------------|----------------|----------------|
| 0.3 | 0.01 | 0.1 | 0.2 | 0.13196 | 0.63456 |
| 0.3 | 0.01 | 0.1 | 0.4 | 0.09845 | 0.67926 |
| 0.3 | 0.01 | 0.1 | 0.6 | 0.09736 | 0.67512 |
| 0.3 | 0.01 | 0.1 | 0.8 | 0.08820 | 0.68312 |

Poměr dat 90:10

Tab. 37: Změna parametru c pro soubor I (90:10), zdroj [vlastní]

| μ | c | d_{\max} | η | MSE trénovací | MSE testovací |
|------------|------------|------------|------------|----------------|----------------|
| 0.5 | 0.1 | 0.1 | 0.2 | 0.18308 | 0.69029 |
| 0.5 | 0.01 | 0.1 | 0.2 | 0.20693 | 0.70358 |
| 0.5 | 0.001 | 0.1 | 0.2 | 0.19162 | 0.69736 |
| 0.5 | 0.0001 | 0.1 | 0.2 | 0.21354 | 0.70209 |

Tab. 38: Změna parametru μ pro soubor I (90:10), zdroj [vlastní]

| μ | c | d_{\max} | η | MSE trénovací | MSE testovací |
|------------|------------|------------|------------|----------------|----------------|
| 0.1 | 0.1 | 0.1 | 0.2 | 0.24058 | 0.71067 |
| 0.3 | 0.1 | 0.1 | 0.2 | 0.20812 | 0.73949 |
| 0.5 | 0.1 | 0.1 | 0.2 | 0.18351 | 0.70342 |
| 0.7 | 0.1 | 0.1 | 0.2 | 0.17302 | 0.70009 |
| 0.9 | 0.1 | 0.1 | 0.2 | 0.39024 | 0.78626 |

Tab. 39: Změna parametru d_{\max} pro soubor I (90:10), zdroj [vlastní]

| μ | c | d_{\max} | η | MSE trénovací | MSE testovací |
|------------|------------|------------|------------|----------------|----------------|
| 0.7 | 0.1 | 0.1 | 0.2 | 0.17302 | 0.70009 |
| 0.7 | 0.1 | 0.01 | 0.2 | 0.16468 | 0.71122 |
| 0.7 | 0.1 | 0.001 | 0.2 | 0.14405 | 0.74481 |
| 0.7 | 0.1 | 0.0001 | 0.2 | 0.16542 | 0.75566 |

Tab. 40: Změna parametru η pro soubor I (90:10), zdroj [vlastní]

| μ | c | d_{\max} | η | MSE trénovací | MSE testovací |
|------------|------------|------------|------------|----------------|----------------|
| 0.7 | 0.1 | 0.1 | 0.2 | 0.17302 | 0.70009 |
| 0.7 | 0.1 | 0.1 | 0.4 | 0.31964 | 0.74291 |
| 0.7 | 0.1 | 0.1 | 0.6 | 0.30841 | 0.69737 |
| 0.7 | 0.1 | 0.1 | 0.8 | 0.45377 | 0.72660 |

Příloha 5 – Výsledky učení dlouhodobé časové řady

Poměr dat 50:50

Tab. 41: Změna parametru c pro soubor L (50:50), zdroj [vlastní]

| μ | c | d_{\max} | η | MSE trénovací | MSE testovací |
|------------|-------------|------------|------------|----------------|----------------|
| 0.5 | 0.1 | 0.1 | 0.2 | 0.07276 | 0.67487 |
| 0.5 | 0.01 | 0.1 | 0.2 | 0.06960 | 0.67183 |
| 0.5 | 0.001 | 0.1 | 0.2 | 0.06989 | 0.69702 |
| 0.5 | 0.0001 | 0.1 | 0.2 | 0.06623 | 0.69209 |

Tab. 42: Změna parametru μ pro soubor L (50:50), zdroj [vlastní]

| μ | c | d_{\max} | η | MSE trénovací | MSE testovací |
|------------|-------------|------------|------------|----------------|----------------|
| 0.1 | 0.01 | 0.1 | 0.2 | 0.06956 | 0.68006 |
| 0.3 | 0.01 | 0.1 | 0.2 | 0.07137 | 0.67192 |
| 0.5 | 0.01 | 0.1 | 0.2 | 0.07278 | 0.66842 |
| 0.7 | 0.01 | 0.1 | 0.2 | 0.07299 | 0.66642 |
| 0.9 | 0.01 | 0.1 | 0.2 | 0.07077 | 0.68345 |

Tab. 43: Změna parametru d_{\max} pro soubor L (50:50), zdroj [vlastní]

| μ | c | d_{\max} | η | MSE trénovací | MSE testovací |
|------------|-------------|---------------|------------|----------------|----------------|
| 0.7 | 0.01 | 0.1 | 0.2 | 0.07299 | 0.66642 |
| 0.7 | 0.01 | 0.01 | 0.2 | 0.06793 | 0.66964 |
| 0.7 | 0.01 | 0.001 | 0.2 | 0.06903 | 0.69737 |
| 0.7 | 0.01 | 0.0001 | 0.2 | 0.07383 | 0.66630 |

Tab. 44: Změna parametru η pro soubor L (50:50), zdroj [vlastní]

| μ | c | d_{\max} | η | MSE trénovací | MSE testovací |
|------------|-------------|---------------|------------|----------------|----------------|
| 0.7 | 0.01 | 0.0001 | 0.2 | 0.07383 | 0.66630 |
| 0.7 | 0.01 | 0.0001 | 0.4 | 0.05890 | 0.69342 |
| 0.7 | 0.01 | 0.0001 | 0.6 | 0.06623 | 0.69971 |
| 0.7 | 0.01 | 0.0001 | 0.8 | 0.06774 | 0.69289 |

Poměr dat 60:40

Tab. 45: Změna parametru c pro soubor L (60:40), zdroj [vlastní]

| μ | c | d_{\max} | η | MSE trénovací | MSE testovací |
|------------|------------|------------|------------|----------------|----------------|
| 0.5 | 0.1 | 0.1 | 0.2 | 0.10795 | 0.66509 |
| 0.5 | 0.01 | 0.1 | 0.2 | 0.10673 | 0.66540 |
| 0.5 | 0.001 | 0.1 | 0.2 | 0.11054 | 0.66616 |
| 0.5 | 0.0001 | 0.1 | 0.2 | 0.09946 | 0.67989 |

Tab. 46: Změna parametru μ pro soubor L (60:40), zdroj [vlastní]

| μ | c | d_{\max} | η | MSE trénovací | MSE testovací |
|------------|------------|------------|------------|----------------|----------------|
| 0.1 | 0.1 | 0.1 | 0.2 | 0.10861 | 0.67984 |
| 0.3 | 0.1 | 0.1 | 0.2 | 0.10614 | 0.67292 |
| 0.5 | 0.1 | 0.1 | 0.2 | 0.10795 | 0.66509 |
| 0.7 | 0.1 | 0.1 | 0.2 | 0.10448 | 0.67928 |
| 0.9 | 0.1 | 0.1 | 0.2 | 0.13826 | 0.71969 |

Tab. 47: Změna parametru d_{\max} pro soubor L (60:40), zdroj [vlastní]

| μ | c | d_{\max} | η | MSE trénovací | MSE testovací |
|------------|------------|--------------|------------|----------------|----------------|
| 0.5 | 0.1 | 0.1 | 0.2 | 0.10795 | 0.66509 |
| 0.5 | 0.1 | 0.01 | 0.2 | 0.11016 | 0.66903 |
| 0.5 | 0.1 | 0.001 | 0.2 | 0.10022 | 0.64777 |
| 0.5 | 0.1 | 0.0001 | 0.2 | 0.09553 | 0.68143 |

Tab. 48: Změna parametru η pro soubor L (60:40), zdroj [vlastní]

| μ | c | d_{\max} | η | MSE trénovací | MSE testovací |
|------------|------------|--------------|------------|----------------|----------------|
| 0.5 | 0.1 | 0.001 | 0.2 | 0.10022 | 0.64777 |
| 0.5 | 0.1 | 0.001 | 0.4 | 0.10961 | 0.66229 |
| 0.5 | 0.1 | 0.001 | 0.6 | 0.11348 | 0.67187 |
| 0.5 | 0.1 | 0.001 | 0.8 | 0.12165 | 0.70690 |

Poměr dat 70:30

Tab. 49: Změna parametru c pro soubor L (70:30), zdroj [vlastní]

| μ | c | d_{\max} | η | MSE trénovací | MSE testovací |
|------------|--------------|------------|------------|----------------|----------------|
| 0.5 | 0.1 | 0.1 | 0.2 | 0.18191 | 0.64444 |
| 0.5 | 0.01 | 0.1 | 0.2 | 0.16600 | 0.66226 |
| 0.5 | 0.001 | 0.1 | 0.2 | 0.17020 | 0.63871 |
| 0.5 | 0.0001 | 0.1 | 0.2 | 0.17114 | 0.66381 |

Tab. 50: Změna parametru μ pro soubor L (70:30), zdroj [vlastní]

| μ | c | d_{\max} | η | MSE trénovací | MSE testovací |
|------------|--------------|------------|------------|----------------|----------------|
| 0.1 | 0.001 | 0.1 | 0.2 | 0.19159 | 0.65034 |
| 0.3 | 0.001 | 0.1 | 0.2 | 0.17967 | 0.64671 |
| 0.5 | 0.001 | 0.1 | 0.2 | 0.18710 | 0.66638 |
| 0.7 | 0.001 | 0.1 | 0.2 | 0.16643 | 0.63762 |
| 0.9 | 0.001 | 0.1 | 0.2 | 0.18631 | 0.64128 |

Tab. 51: Změna parametru d_{\max} pro soubor L (70:30), zdroj [vlastní]

| μ | c | d_{\max} | η | MSE trénovací | MSE testovací |
|------------|--------------|------------|------------|----------------|----------------|
| 0.7 | 0.001 | 0.1 | 0.2 | 0.16643 | 0.63762 |
| 0.7 | 0.001 | 0.01 | 0.2 | 0.19668 | 0.65688 |
| 0.7 | 0.001 | 0.001 | 0.2 | 0.17710 | 0.64558 |
| 0.7 | 0.001 | 0.0001 | 0.2 | 0.17734 | 0.64913 |

Tab. 52: Změna parametru η pro soubor L (70:30), zdroj [vlastní]

| μ | c | d_{\max} | η | MSE trénovací | MSE testovací |
|------------|--------------|------------|------------|----------------|----------------|
| 0.7 | 0.001 | 0.1 | 0.2 | 0.16643 | 0.63762 |
| 0.7 | 0.001 | 0.1 | 0.4 | 0.17823 | 0.65912 |
| 0.7 | 0.001 | 0.1 | 0.6 | 0.20333 | 0.64959 |
| 0.7 | 0.001 | 0.1 | 0.8 | 0.18141 | 0.65851 |

Poměr dat 80:20

Tab. 53: Změna parametru c pro soubor L (80:20), zdroj [vlastní]

| μ | c | d_{\max} | η | MSE trénovací | MSE testovací |
|------------|--------------|------------|------------|----------------|----------------|
| 0.5 | 0.1 | 0.1 | 0.2 | 0.27895 | 0.63290 |
| 0.5 | 0.01 | 0.1 | 0.2 | 0.27555 | 0.65500 |
| 0.5 | 0.001 | 0.1 | 0.2 | 0.28187 | 0.62775 |
| 0.5 | 0.0001 | 0.1 | 0.2 | 0.26486 | 0.63035 |

Tab. 54: Změna parametru μ pro soubor L (80:20), zdroj [vlastní]

| μ | c | d_{\max} | η | MSE trénovací | MSE testovací |
|------------|--------------|------------|------------|----------------|----------------|
| 0.1 | 0.001 | 0.1 | 0.2 | 0.30971 | 0.65039 |
| 0.3 | 0.001 | 0.1 | 0.2 | 0.29850 | 0.63588 |
| 0.5 | 0.001 | 0.1 | 0.2 | 0.29903 | 0.65273 |
| 0.7 | 0.001 | 0.1 | 0.2 | 0.30898 | 0.64117 |
| 0.9 | 0.001 | 0.1 | 0.2 | 0.27685 | 0.66275 |

Tab. 55: Změna parametru d_{\max} pro soubor L (80:20), zdroj [vlastní]

| μ | c | d_{\max} | η | MSE trénovací | MSE testovací |
|------------|--------------|---------------|------------|----------------|----------------|
| 0.3 | 0.001 | 0.1 | 0.2 | 0.29850 | 0.63588 |
| 0.3 | 0.001 | 0.01 | 0.2 | 0.26622 | 0.64574 |
| 0.3 | 0.001 | 0.001 | 0.2 | 0.28785 | 0.64276 |
| 0.3 | 0.001 | 0.0001 | 0.2 | 0.28378 | 0.62807 |

Tab. 56: Změna parametru η pro soubor L (80:20), zdroj [vlastní]

| μ | c | d_{\max} | η | MSE trénovací | MSE testovací |
|------------|--------------|---------------|------------|----------------|----------------|
| 0.3 | 0.001 | 0.0001 | 0.2 | 0.28378 | 0.62807 |
| 0.3 | 0.001 | 0.0001 | 0.4 | 0.28887 | 0.64073 |
| 0.3 | 0.001 | 0.0001 | 0.6 | 0.28697 | 0.62016 |
| 0.3 | 0.001 | 0.0001 | 0.8 | 0.28215 | 0.65729 |

Poměr dat 90:10

Tab. 57: Změna parametru c pro soubor $L(90:10)$, zdroj [vlastní]

| μ | c | d_{\max} | η | MSE trénovací | MSE testovací |
|------------|---------------|------------|------------|----------------|----------------|
| 0.5 | 0.1 | 0.1 | 0.2 | 0.60283 | 0.64334 |
| 0.5 | 0.01 | 0.1 | 0.2 | 0.65650 | 0.63391 |
| 0.5 | 0.001 | 0.1 | 0.2 | 0.58894 | 0.63798 |
| 0.5 | 0.0001 | 0.1 | 0.2 | 0.58907 | 0.63189 |

Tab. 58: Změna parametru μ pro soubor $L(90:10)$, zdroj [vlastní]

| μ | c | d_{\max} | η | MSE trénovací | MSE testovací |
|------------|---------------|------------|------------|----------------|----------------|
| 0.1 | 0.0001 | 0.1 | 0.2 | 0.62787 | 0.63785 |
| 0.3 | 0.0001 | 0.1 | 0.2 | 0.61580 | 0.63813 |
| 0.5 | 0.0001 | 0.1 | 0.2 | 0.59113 | 0.63557 |
| 0.7 | 0.0001 | 0.1 | 0.2 | 0.62471 | 0.63480 |
| 0.9 | 0.0001 | 0.1 | 0.2 | 0.58427 | 0.69049 |

Tab. 59: Změna parametru d_{\max} pro soubor $L(90:10)$, zdroj [vlastní]

| μ | c | d_{\max} | η | MSE trénovací | MSE testovací |
|------------|---------------|---------------|------------|----------------|----------------|
| 0.7 | 0.0001 | 0.1 | 0.2 | 0.62471 | 0.63480 |
| 0.7 | 0.0001 | 0.01 | 0.2 | 0.59673 | 0.63893 |
| 0.7 | 0.0001 | 0.001 | 0.2 | 0.58082 | 0.63963 |
| 0.7 | 0.0001 | 0.0001 | 0.2 | 0.61376 | 0.63133 |

Tab. 60: Změna parametru η pro soubor $L(90:10)$, zdroj [vlastní]

| μ | c | d_{\max} | η | MSE trénovací | MSE testovací |
|------------|---------------|---------------|------------|----------------|----------------|
| 0.7 | 0.0001 | 0.0001 | 0.2 | 0.61376 | 0.63133 |
| 0.7 | 0.0001 | 0.0001 | 0.4 | 0.62785 | 0.65326 |
| 0.7 | 0.0001 | 0.0001 | 0.6 | 0.66319 | 0.69701 |
| 0.7 | 0.0001 | 0.0001 | 0.8 | 0.65605 | 0.64462 |

Příloha 6 – Výsledky učení frontální neuronové sítě

Soubor S - Rozdělení dat 50:50

Tab. 61: Změna parametru d_{\max} pro soubor S (50:50), zdroj [vlastní]

| η | d_{\max} | MSE trénovací | MSE testovací |
|------------|-------------|----------------|----------------|
| 0.2 | 0 | 0.23249 | 1.02333 |
| 0.2 | 0.1 | 0.27951 | 1.07060 |
| 0.2 | 0.01 | 0.15124 | 0.85433 |

Tab. 62: Změna parametru η pro soubor S (50:50), zdroj [vlastní]

| η | d_{\max} | MSE trénovací | MSE testovací |
|------------|-------------|----------------|----------------|
| 0.1 | 0.01 | 0.21037 | 0.99099 |
| 0.2 | 0.01 | 0.27951 | 1.07060 |
| 0.3 | 0.01 | 0.14422 | 0.83974 |

Rozdělení dat 60:40

Tab. 63: Změna parametru d_{\max} pro soubor S (60:40), zdroj [vlastní]

| η | d_{\max} | MSE trénovací | MSE testovací |
|------------|------------|----------------|----------------|
| 0.2 | 0 | 0.20988 | 1.53999 |
| 0.2 | 0.1 | 0.18871 | 1.45999 |
| 0.2 | 0.01 | 0.22953 | 1.60279 |

Tab. 64: Změna parametru η pro soubor S (60:40), zdroj [vlastní]

| η | d_{\max} | MSE trénovací | MSE testovací |
|------------|------------|----------------|----------------|
| 0.1 | 0.1 | 0.11831 | 1.13322 |
| 0.2 | 0.1 | 0.18871 | 1.45999 |
| 0.3 | 0.1 | 0.16293 | 1.37120 |

Rozdělení dat 70:30

Tab. 65: Změna parametru d_{\max} pro soubor S (70:30), zdroj [vlastní]

| η | d_{\max} | MSE trénovací | MSE testovací |
|------------|------------|----------------|----------------|
| 0.2 | 0 | 0.61490 | 0.91463 |
| 0.2 | 0.1 | 0.54378 | 0.89053 |
| 0.2 | 0.01 | 0.64853 | 0.91467 |

Tab. 66: Změna parametru η pro soubor S (70:30), zdroj [vlastní]

| η | d_{\max} | MSE trénovací | MSE testovací |
|------------|------------|----------------|----------------|
| 0.1 | 0.1 | 0.36276 | 0.77513 |
| 0.2 | 0.1 | 0.54378 | 0.89053 |
| 0.3 | 0.1 | 0.49389 | 0.86985 |

Rozdělení dat 80:20

Tab. 67: Změna parametru d_{\max} pro soubor S (80:20), zdroj [vlastní]

| η | d_{\max} | MSE trénovací | MSE testovací |
|------------|------------|----------------|----------------|
| 0.2 | 0 | 0.86172 | 0.92016 |
| 0.2 | 0.1 | 0.76423 | 0.89652 |
| 0.2 | 0.01 | 0.90801 | 0.91883 |

Tab. 68: Změna parametru η pro soubor S (80:20), zdroj [vlastní]

| η | d_{\max} | MSE trénovací | MSE testovací |
|------------|------------|----------------|----------------|
| 0.1 | 0.1 | 0.51687 | 0.78215 |
| 0.2 | 0.1 | 0.76423 | 0.89652 |
| 0.3 | 0.1 | 0.69625 | 0.87608 |

Rozdělení dat 90:10

Tab. 69: Změna parametru d_{\max} pro soubor S (90:10), zdroj [vlastní]

| η | d_{\max} | MSE trénovací | MSE testovací |
|------------|------------|----------------|----------------|
| 0.2 | 0 | 1.51067 | 0.79194 |
| 0.2 | 0.1 | 1.34019 | 0.77451 |
| 0.2 | 0.01 | 1.59152 | 0.78983 |

Tab. 70: Změna parametru η pro soubor S (90:10), zdroj [vlastní]

| η | d_{\max} | MSE trénovací | MSE testovací |
|------------|------------|----------------|----------------|
| 0.1 | 0.1 | 0.90683 | 0.68334 |
| 0.2 | 0.1 | 1.34019 | 0.77451 |
| 0.3 | 0.1 | 1.22056 | 0.76036 |

Soubor I - Rozdělení dat 50:50

Tab. 71: Změna parametru d_{\max} pro soubor I (50:50), zdroj [vlastní]

| η | d_{\max} | MSE trénovací | MSE testovací |
|------------|------------|----------------|----------------|
| 0.2 | 0 | 0.27315 | 1.30546 |
| 0.2 | 0.1 | 0.25273 | 1.24654 |
| 0.2 | 0.01 | 0.28276 | 1.32798 |

Tab. 72: Změna parametru η pro soubor I (50:50), zdroj [vlastní]

| η | d_{\max} | MSE trénovací | MSE testovací |
|------------|------------|----------------|----------------|
| 0.1 | 0.1 | 0.19651 | 1.04864 |
| 0.2 | 0.1 | 0.25273 | 1.24654 |
| 0.3 | 0.1 | 0.24033 | 1.19781 |

Rozdělení dat 60:40

Tab. 73: Změna parametru d_{\max} pro soubor I (60:40), zdroj [vlastní]

| η | d_{\max} | MSE trénovací | MSE testovací |
|------------|------------|----------------|----------------|
| 0.2 | 0 | 0.39741 | 1.27069 |
| 0.2 | 0.1 | 0.36684 | 1.21751 |
| 0.2 | 0.01 | 0.41168 | 1.28961 |

Tab. 74: Změna parametru η pro soubor I (60:40), zdroj [vlastní]

| η | d_{\max} | MSE trénovací | MSE testovací |
|------------|------------|----------------|----------------|
| 0.1 | 0.1 | 0.28397 | 1.03130 |
| 0.2 | 0.1 | 0.36684 | 1.21751 |
| 0.3 | 0.1 | 0.34815 | 1.17342 |

Rozdělení dat 70:30

Tab. 75: Změna parametru d_{\max} pro soubor I (70:30), zdroj [vlastní]

| η | d_{\max} | MSE trénovací | MSE testovací |
|------------|------------|----------------|----------------|
| 0.2 | 0 | 0.59563 | 1.35501 |
| 0.2 | 0.1 | 0.55213 | 1.29758 |
| 0.2 | 0.01 | 0.61627 | 1.37573 |

Tab. 76: Změna parametru η pro soubor I (70:30), zdroj [vlastní]

| η | d_{\max} | MSE trénovací | MSE testovací |
|------------|------------|----------------|----------------|
| 0.1 | 0.1 | 0.43403 | 1.09977 |
| 0.2 | 0.1 | 0.55213 | 1.29758 |
| 0.3 | 0.1 | 0.52627 | 1.24932 |

Rozdělení dat 80:20

Tab. 77: Změna parametru d_{\max} pro soubor I (80:20), zdroj [vlastní]

| η | d_{\max} | MSE trénovací | MSE testovací |
|------------|------------|----------------|----------------|
| 0.2 | 0 | 0.22150 | 0.75213 |
| 0.2 | 0.1 | 0.21265 | 0.80316 |
| 0.2 | 0.01 | 0.22179 | 0.75289 |

Tab. 78: Změna parametru η pro soubor I (80:20), zdroj [vlastní]

| η | d_{\max} | MSE trénovací | MSE testovací |
|------------|------------|----------------|----------------|
| 0.1 | 0 | 0.22091 | 0.74048 |
| 0.2 | 0 | 0.22150 | 0.75213 |
| 0.3 | 0 | 0.22277 | 0.75946 |

Rozdělení dat 90:10

Tab. 79: Změna parametru d_{\max} pro soubor I (90:10), zdroj [vlastní]

| η | d_{\max} | MSE trénovací | MSE testovací |
|------------|------------|----------------|----------------|
| 0.2 | 0 | 1.72402 | 1.43215 |
| 0.2 | 0.1 | 1.58984 | 1.39169 |
| 0.2 | 0.01 | 1.78905 | 1.44114 |

Tab. 80: Změna parametru η pro soubor I (90:10), zdroj [vlastní]

| η | d_{\max} | MSE trénovací | MSE testovací |
|------------|------------|----------------|----------------|
| 0.1 | 0.1 | 1.50946 | 1.35651 |
| 0.2 | 0.1 | 1.58984 | 1.39169 |
| 0.3 | 0.1 | 1.16640 | 1.19844 |

Soubor L - Rozdělení dat 50:50

Tab. 81: Změna parametru d_{\max} pro soubor L (50:50), zdroj [vlastní]

| η | d_{\max} | MSE trénovací | MSE testovací |
|------------|------------|----------------|----------------|
| 0.2 | 0 | 0.25685 | 1.36442 |
| 0.2 | 0.1 | 0.24624 | 1.30216 |
| 0.2 | 0.01 | 0.27554 | 1.38762 |

Tab. 82: Změna parametru η pro soubor L (50:50), zdroj [vlastní]

| η | d_{\max} | MSE trénovací | MSE testovací |
|------------|------------|----------------|----------------|
| 0.1 | 0.1 | 0.19642 | 1.09790 |
| 0.2 | 0.1 | 0.24624 | 1.30216 |
| 0.3 | 0.1 | 0.23481 | 1.24960 |

Rozdělení dat 60:40

Tab. 83: Změna parametru d_{\max} pro soubor L (60:40), zdroj [vlastní]

| η | d_{\max} | MSE trénovací | MSE testovací |
|------------|------------|----------------|----------------|
| 0.2 | 0 | 0.39078 | 1.33735 |
| 0.2 | 0.1 | 0.35980 | 1.28088 |
| 0.2 | 0.01 | 0.40633 | 1.35633 |

Tab. 84: Změna parametru η pro soubor L (60:40), zdroj [vlastní]

| η | d_{\max} | MSE trénovací | MSE testovací |
|------------|------------|----------------|----------------|
| 0.1 | 0.1 | 0.28395 | 1.08210 |
| 0.2 | 0.1 | 0.35980 | 1.28088 |
| 0.3 | 0.1 | 0.34113 | 1.23263 |

Rozdělení dat 70:30

Tab. 85: Změna parametru d_{\max} pro soubor L (70:30), zdroj [vlastní]

| η | d_{\max} | MSE trénovací | MSE testovací |
|------------|------------|----------------|----------------|
| 0.2 | 0 | 0.65219 | 1.29935 |
| 0.2 | 0.1 | 0.60003 | 1.24775 |
| 0.2 | 0.01 | 0.67800 | 1.31529 |

Tab. 86: Změna parametru η pro soubor L (70:30), zdroj [vlastní]

| η | d_{\max} | MSE trénovací | MSE testovací |
|------------|------------|----------------|----------------|
| 0.1 | 0.1 | 0.47291 | 1.05870 |
| 0.2 | 0.1 | 0.60003 | 1.24775 |
| 0.3 | 0.1 | 0.56818 | 1.20368 |

Rozdělení dat 80:20

Tab. 87: Změna parametru d_{\max} pro soubor L (80:20), zdroj [vlastní]

| η | d_{\max} | MSE trénovací | MSE testovací |
|------------|------------|----------------|----------------|
| 0.2 | 0 | 1.04445 | 1.28340 |
| 0.2 | 0.1 | 0.96162 | 1.22782 |
| 0.2 | 0.01 | 1.08491 | 1.30301 |

Tab. 88: Změna parametru η pro soubor L (80:20), zdroj [vlastní]

| η | d_{\max} | MSE trénovací | MSE testovací |
|------------|------------|-----------------|----------------|
| 0.1 | 0.1 | 0.758957 | 1.03264 |
| 0.2 | 0.1 | 0.96162 | 1.22782 |
| 0.3 | 0.1 | 0.91126 | 1.18075 |

Rozdělení dat 90:10

Tab. 89: Změna parametru d_{\max} pro soubor L (90:10), zdroj [vlastní]

| η | d_{\max} | MSE trénovací | MSE testovací |
|------------|------------|----------------|----------------|
| 0.2 | 0 | 2.25510 | 1.28746 |
| 0.2 | 0.1 | 2.07027 | 1.24782 |
| 0.2 | 0.01 | 2.34371 | 1.29525 |

Tab. 90: Změna parametru η pro soubor L (90:10), zdroj [vlastní]

| η | d_{\max} | MSE trénovací | MSE testovací |
|------------|------------|----------------|----------------|
| 0.1 | 0.1 | 1.62025 | 1.07689 |
| 0.2 | 0.1 | 2.07027 | 1.24782 |
| 0.3 | 0.1 | 1.95620 | 1.21289 |