**Univerzita Pardubice**
**Fakulta ekonomicko–správní**

# Modelování predikce časové řady návštěvnosti web domény pomocí RBF neuronových sítí

**Bc. Kateřina Štěpánková**

**Diplomová práce**
**2011**

**Univerzita Pardubice**
**Fakulta ekonomicko-správní**
Akademický rok: **2010/2011**

# ZADÁNÍ DIPLOMOVÉ PRÁCE

(PROJEKTU, UMĚLECKÉHO DÍLA, UMĚLECKÉHO VÝKONU)

| | |
|---|---|
| Jméno a příjmení: | **Bc. Kateřina ŠTĚPÁNKOVÁ** |
| Osobní číslo: | **E090493** |
| Studijní program: | **N6209 Systémové inženýrství a informatika** |
| Studijní obor: | **Informatika ve veřejné správě** |
| Název tématu: | **Modelování predikce časové řady návštěvnosti web domény pomocí RBF neuronových sítí** |
| Zadávající katedra: | **Ústav systémového inženýrství a informatiky** |

Zásady pro vypracování:

Analýza vstupních dat (parametrů) pro predikci.
Charakteristika RBF neuronové sítě z hlediska aproximace a predikce.
Návrh modelu na predikci návštěvnosti web domény.
Verifikace navrženého modelu.
Analýza výsledků.

Rozsah grafických prací:

Rozsah pracovní zprávy:

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

Seznam odborné literatury:

**HAYKIN, S. Neural Networks: A Comprehensive Foundation. 2nd edition, New Jersey : Prentice Hall, 1999.**

**KVASNIČKA, V. a kol. Úvod do teórie neurónových sietí. Bratislava : Iris, 2007.**

**OLEJ, V. Modelovanie ekonomických procesov na báze výpočtovej inteligencie. Hradec Králové : Miloš Vognar - M&V, 2003.**

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

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. dubna 2011

.....

Kateřina Štěpánková

# Poděkování

Na tomto místě bych ráda poděkovala především panu prof. Ing. Vladimíru Olejovi, CSc. za jeho pomoc, cenné rady, připomínky a podporu při zpracování této diplomové práce.

# Anotace

Diplomová práce se zabývá návrhem modelu pro predikci časové řady návštěvnosti domény upce.cz. Pro návrh modelu byly zvoleny neuronové sítě radiálně bazického typu. K predikci byly využity tři časové řady návštěvnosti různých délek. Cílem práce je navrhnout model pro predikci časové řady návštěvnosti www.upce.cz.

# Klíčová slova

Web mining, radiálně bazické funkce, RBF, neuronová síť, predikce, časová řada.

# Title

Modelling the Prediction of Time Series of the Website Traffic by RBF Neural Network

# Abstract

This thesis deals with modelling the prediction of time series of the upce.cz website traffic. Radial Basis Function were chosen for design models. Three time series of various length of the upce.cz website traffic were used for prediction. The aim is design models for prediction of time series of the upce.cz website traffic.

# Keywords

Web mining, radial basis function, RBF, neural network, prediction, time series.

# Obsah

# Úvod

Diplomová práce se zabývá návrhem neuronové sítě radiálně bazického typu pro predikci návštěvnosti domény upce.cz. Zahrnuje základní poznatky z oblasti dolování znalostí z webových stránek, tzv. Web miningu. Web mining je poměrně mladá disciplína, která může poskytnout cenné informace nejen vývojářům internetových stránek, ale také marketingovým odborníkům, nebo může manažerům pomoci v jejich rozhodování. Těmito cennými informacemi, získanými Web miningem, jsou informace týkající se ať už struktury nebo obsahu internetových stránek, tak převážně informace o chování jednotlivých návštěvníků, kteří tyto stránky navštěvují. Díky takovým informacím lze internetové stránky přizpůsobit tak, aby na nich návštěvník setrval a rychle našel to, kvůli čemu na tyto stránky přichází. Tím je možno dosáhnout v konečném důsledku většího zisku (například v případě internetového obchodu). Úvodem do Web miningu se bude zabývat první kapitola této práce.

Druhá kapitola bude věnována neuronovým sítím. Neuronové sítě se snaží matematicky napodobit lidský mozek, jehož největší předností je schopnost učit se. Tento fakt inspiroval odborníky natolik, že se snažili navrhnout takové počítačové aplikace, které by uměly pracovat podobně. Neuronové sítě mají v dnešní době velké uplatnění v mnoha odvětvích. Ať už se jedná o predikci, tedy předvídání budoucí hodnoty jakékoliv veličiny, nebo klasifikaci objektů do homogenních skupin. Významnou roli v oblasti neuronových sítí mají právě sítě typu RBF, které byly použity pro predikci návštěvnosti webové domény upce.cz.

Třetí kapitola je věnována návrhu modelů pro predikci časové řady. Pro predikci je využito třech časových řad různé délky. V kapitole jsou popsány jednotlivé modely pro každou časovou řadu zvlášť. Poté následuje zhodnocení a porovnání všech těchto časových řad.

Cílem této diplomové práce je navrhnout model pro predikci časové řady, kterou tvoří návštěvy domény upce.cz. Návrhy modelu budou realizovány v programovém prostředí SPSS Clementine 10.1, což je mocný data miningový nástroj. Předmětem návrhu zde bude hledání takových parametrů RBF neuronové sítě, při kterých bude síť nejlépe naučena, tj. bude dosaženo nejmenší chyby na testovacích datech. Výstupem práce je vybraná časová řada a návrh modelu neuronové sítě typu RBF, který vykazoval nejlepší výsledky.

# 1 Web mining

V dnešní době si již život bez vyhledávání informací přes internetové stránky lze jen těžko představit. Každý návštěvník webové stránky se snaží co nejrychleji nalézt informaci, kterou hledá. S rostoucím počtem uživatelů internetových stránek a informací, které jsou k získání z těchto stránek, bylo zapotřebí navrhnout užitečné nástroje k nalezení požadovaných informací a také ke sledování a analýze chování uživatelů. Velmi podobnou úlohu v hledání znalostí v datech hraje data mining. Data mining hledá souvislosti mezi daty a používá širokou škálu technik, kterými tato data zpracovává za účelem získání hlubších znalostí nebo souvislostí. Tyto znalosti dále slouží k podpoře rozhodování, stanovení marketingových plánů atd. Pojem Web mining byl podle [1] poprvé použit v roce 1996, avšak v nynější době se dostává čím dál více do popředí z důvodu rozsáhlého využívání internetových stránek nejen k podpoře podnikání. Podle [2] je možné si pod pojmem Web mining představit techniky, které se používají k procházení webových stránek za účelem shromažďování užitečných informací, které používají ať už jednotlivci či společnosti k podpoře podnikání, přizpůsobení marketingových nebo výrobních plánů atd.

Dle [3] Web mining neshromažďuje pouze běžná statistická data jako je například počet návštěvníků webové stránky za den. Web mining může pomoci zodpovědět otázky typu:

- Odkud přicházejí návštěvníci.
- Jak se návštěvníci na webové stránce chovají.
- Jaké jsou typické sekvence průchodu stránkami.
- Při jaké sekvenci průchodu došlo k nákupu zboží nebo rezervaci.
- Jak dlouho se na stránkách návštěvníci zdrží.
- Jak a odkud návštěvníci opouští stránky.
- Jaký internetový prohlížeč návštěvníci používají.
- Jaký operační systém návštěvníci používají.

Na těchto datech může být aplikováno rozsáhlé množství data miningových technik, které vedou k úpravě nebo vytvoření webových stránek tak, aby jejich návštěvníci rychle dostali požadované informace, a pokud se jedná např. o internetový obchod, tak aby si návštěvník objednal zboží.

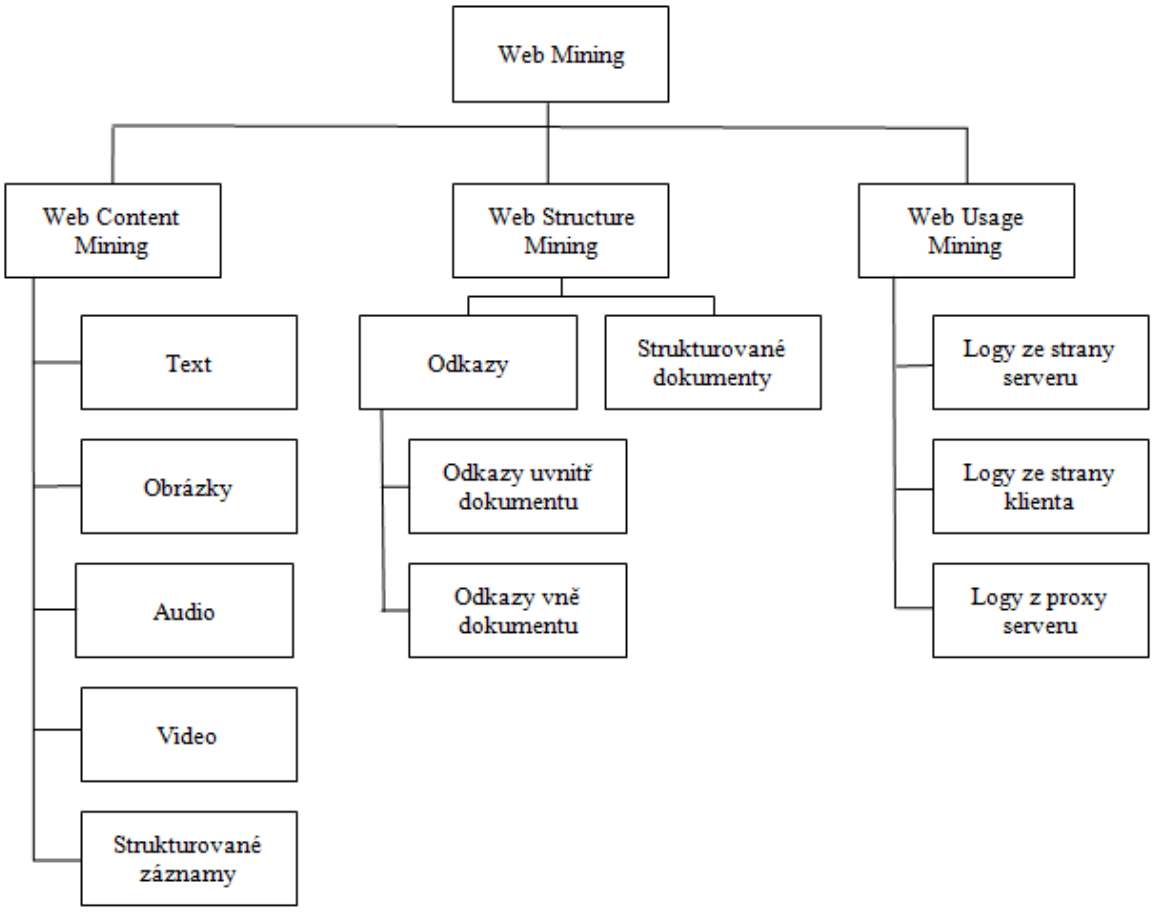
Technologie Web miningu tedy nejen, že shromažďují informace z webových stránek, které jsou následně vstupy např. do predikčních modelů, ale zároveň umožňují sledovat jednotlivé kroky návštěvníků.

## 1.1 Taxonomie Web miningu

Jak již bylo zmíněno v úvodu této kapitoly, Web mining je dolováním znalostí z internetových stránek. Je rozdělen do třech částí[1] [3], [4]:

- Web Content Mining – dobývání znalostí na základě obsahu webu.
- Web Structure Mining – dobývání znalostí na základě struktury webu.
- Web Usage Mining – dobývání znalostí na základě používání webu.

Tato taxonomie je zobrazena na obr. 1.1, kde je i přehledně zobrazeno, kterými částmi internetových stránek se jednotlivá část zaobírá:




**Obr. 1.1 - Taxonomie Web miningu [5]**

[1] Například v práci [6] je uvedená taxonomie Web miningu pouze jako Web Content Mining a Web Usage Mining.

## 1.2 Web Content Mining

Web Content Mining, neboli dobývání znalostí na základě obsahu webu (dále jen WCM), je extrahování užitečných informací z obsahu webových stránek. Předmětem zájmu jsou zde prvky webové stránky jako je text, obrázky, audio prvky, video prvky nebo strukturované záznamy. WCM je podle [4] velmi blízký text miningu. Text mining je speciálním typem dobývání znalostí z databází. Rozdíl je zde v tom, že v databázích se pracuje s daty, které jsou uloženy v pevné struktuře, kdežto v případě text miningu se pracuje s nestrukturovaným textem. V oblasti WCM data mohou být nestrukturovaná nebo polostrukturovaná. Cílem WCM je získat znalosti z webových stránek, které jsou v tomto smyslu chápány jako dokumenty. Předmětem zájmu WCM jsou následující úlohy [4]:

- Vyhledávání a metavyhledávání stránek internetovými vyhledávači.
- Shlukování stránek podle jejich obsahu.
- Filtrování stránek (v tomto kontextu je filtrování chápáno jako rozpoznání stránek relevantních k profilu uživatele).
- Dobývání skrytých znalostí ze stránek.

Cílem každého majitele webových stránek je přilákat co nejvíce návštěvníků a nabídnout jim informace, které hledají. Zároveň by měla být zajištěna co nejjednodušší orientace návštěvníka na stránkách. K tomu, aby návštěvník nalezl takovéto stránky, je využíváno vyhledávačů. V dnešní době je celosvětově velmi populární vyhledávač Google, který se již stal nejpoužívanějším i v České republice, čímž předstihl český vyhledávač Seznam.cz[2]. Všechny vyhledávače fungují tak, že na základě zadaného dotazu naleznou odkazy na stránky, které tomuto dotazu nejlépe odpovídají. „Vyhledávače používají indexové soubory, ve kterých hledají odpovídající odkazy, mohou rovněž procházet plné texty dokumentů na webu“ [4].

WCM se tedy snaží získat znalost z obsahu webových stránek. Tyto znaloti se využijí především pro to, aby se tyto stránky vyskytovaly na prvních pozicích ve vyhledávačích. Tak budou moci přilákat co nejvíce návštěvníků. Návštěvníkům je pak možno nabídnout takovou strukturu webové stránky, kde se budou snadno orientovat a rychleji naleznou hledanou

[2] Tato informace byla uvedena ve výsledcích Křišťálové lupy v roce 2010, viz: http://kristalova.lupa.cz/2010/vysledky-hlasovani/

informaci. WCM se podle [6] dělí na přístup založený na vyhledávacích agentech (*Agent-Based Approach*) a databázový přístup (*Database Approach*).

### 1.2.1 Přístup založený na vyhledávacích agentech

Systémy, které jsou založeny na tomto přístupu se dále dle [6] dají rozdělit na tři kategorie: Inteligentní vyhledávací agenti (*Intelligent Search Agents*), Filtrování informací (*Information Filtering/ Categorization*), Personalizovaní weboví agenti (*Personalized Web Agents*).

Inteligentní vyhledávací agenti vyhledávají relevantní informace pomocí charakteristik domén a profilů uživatelů tak, aby organizovali a interpretovali nalezené informace. Agenti pro filtrování a kategorizaci informací používají techniky pro vyhledávání informací a charakteristiky hypertextových webových dokumentů. Personalizovaní weboví agenti studují uživatelovy preference a objevují informace právě na základě těchto preferencí. K vyhledávání používají i velmi příbuzné preference ostatních uživatelů [6].

### 1.2.2 Databázový přístup

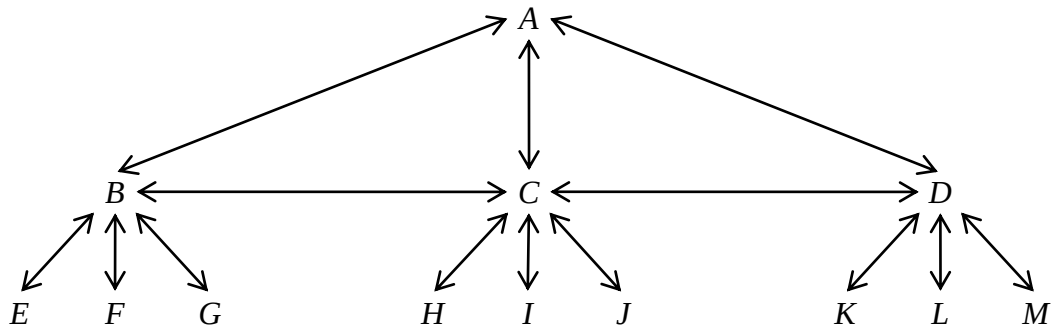
Podle [6] se databázový přístup Web miningu zaměřuje na techniky organizující polostrukturovaná data, která jsou dostupná na webu, do více strukturalizovaných zdrojů. K analýze dat používají standardní databázový dotazovací mechanismus a data miningové technologie. Tento přístup se dle [6] rozděluje dále na víceúrovňové databáze (*Multilevel Databases*) a tzv. webové dotazovací systémy (*Web Query Systems*).

Hlavní myšlenka víceúrovňových databází v kontextu Web miningu je založena na tom, že na nejnižších úrovních jsou polostrukturované informace, které představují hypertextové dokumenty. Na vyšších úrovních se nacházejí metadata nebo generalizovaná data, která jsou extrahována z nižších úrovní. Data ve vyšších úrovních jsou potom organizována do strukturální podoby, tj. do relačních nebo objektově orientovaných databází. Mnoho webových dotazovacích systémů a jazyků využívá znalostí ze standardních databázových dotazovacích jazyků jako je SQL, strukturální informace o webových dokumentech a dokonce také zpracování přirozeného jazyka pro dotazy, které jsou používány v internetových vyhledávačích [6].

## 1.3 Web Structure Mining

Web Structure Mining, neboli dobývání znalostí ze struktury webu (dále jen WSM) pracuje s daty, která popisují organizaci obsahu webových stránek. Podle [7] je možné si WWW prostor představit jako množinu webových stránek, které jsou mezi sebou propojeny odkazy. Webovou stránku je možné si představit jako orientovaný graf, kde uzly jsou webové dokumenty a hrany jsou odkazy mezi nimi. Odkazy jsou zde dvojího typu: vnitřní a vnější. Vnitřní odkazy jsou ty, které se odkazují z konkrétní webové stránky na mnoho dalších. Vnější jsou takové, které se odkazují z ostatních stránek na tuto konkrétní.

Příklad pohledu na webovou stránku jako na graf je možné si představit z takové struktury stránky, která má domovskou stránku, tzv. homepage, a ta je propojena odkazy na další obsah. Potom je možné se na tuto stránku dívat jako na strom, kde vrchol je domovská stránka a ten se větví podle obsahu na další části, které představují jednotlivé uzly. Hrany mezi uzly jsou potom jednotlivé odkazy. Jednoduchý příklad grafu je na zobrazen na obr. 1.2, kde vrchol A tvoří domovskou stránku:




**Obr. 1.2 – Webová stránka jako graf [8]**

Takovýto graf může být použit ke kategorizaci webových stránek na základě jejich struktury nebo získávání informací ohledně podobnosti či vztahu dvou stránek. Dalším využitím je fakt, že takovéto struktury v sobě obsahují implicitní informace, které mohou pomoci ve filtrování obsahu a stanovování pořadí ve vyhledávačích. Podle [9] a [10] jsou nejpoužívanějšími algoritmy pro stanovování pořadí ve vyhledávačích, které využívají informací ze struktury webových stránek, algoritmy HITS (*Hypertext Induced Topic Search*) a PageRank[3]. Druhý algoritmus, PageRank, využívá pro hodnocení pořadí vyhledávací nástroj Google.

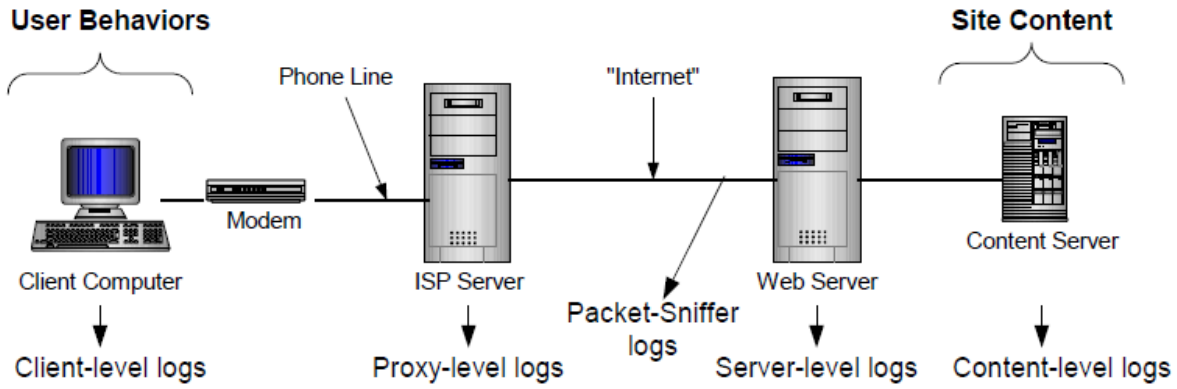
[3] Podrobnější popis a srovnání těchto algoritmů je popsán v [10]

Původně byl algoritmus PageRank založen na citační analýze. Ovšem pouze sledování počtu odkazování se na webovou stránku nestačilo, proto byl tento algoritmus vylepšen o vlastnost, která přiřazuje váhu ohodnocení na základě stránky, ze které je odkazováno.

## *1.4 Web Usage Mining*

Cílem Web Usage Miningu, neboli dobývání znalostí na základě používání webu (dále jen WUM), je objevení vzorů chování návštěvníků na webových stránkách. Podle [3] je možné WUM rozdělit na dvě části. První část zahrnuje transformaci a předzpracování dat do takové podoby, aby byly použitelné pro zpracování. Druhá část zahrnuje jednotlivé metody, použité na předzpracovaná data.

Data o chování návštěvníků jsou získávána z logovacích souborů, které zaznamenávají údaje o přístupech na webové stránky. Tyto logy mohou být získány sběrem dat z úrovně serveru, klienta a také z úrovně proxy serveru [1]. Úrovně sběru dat jsou graficky zobrazeny na obr. 1.3.




**Obr. 1.3 - Úrovně sběru dat pro WUM [5]**

### Sběr dat z úrovně serveru

Logové soubory z úrovně serveru explicitně zaznamenávají chování návštěvníka na stránkách ať už se jedná o jednoho návštěvníka nebo jich může být i více. „Původním standardem pro logové soubory je standard *Common Log Format* (CLF), definovaný internetovým konsorciem W3C, který obsahuje sedm základních datových položek. Jako jeho rozšíření byl posléze dodefinován takzvaný *Extended Common Log Format* (ECLF), kde k původním položkám přibyly další dvě“ [11]. Zmiňované položky logového souboru jsou stručně popsány v tab. 1.1.

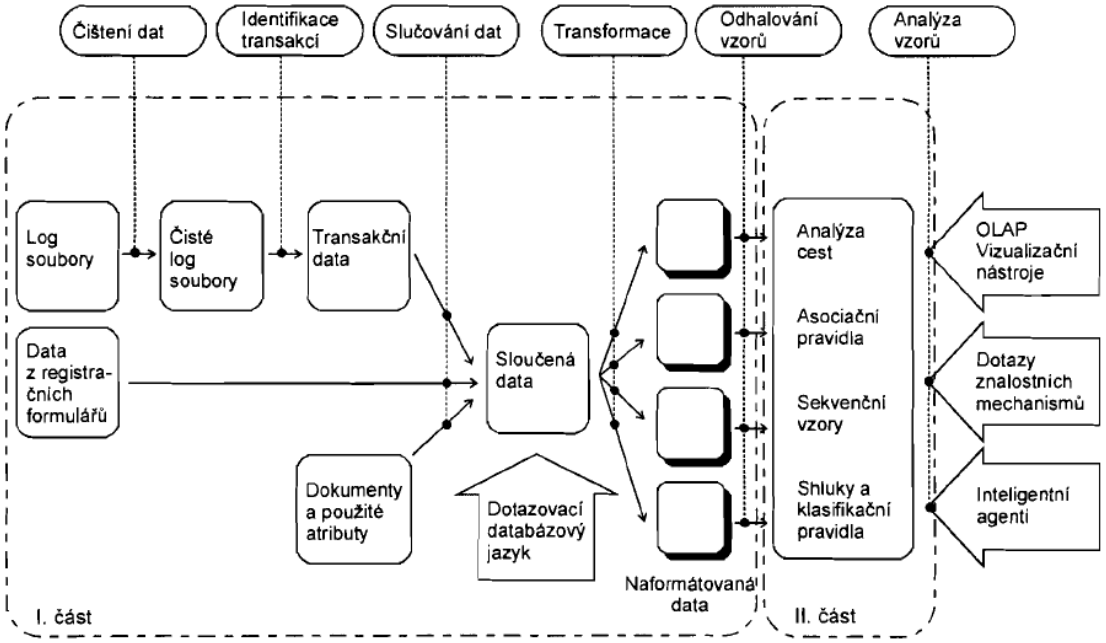
**Tab. 1.1 - Struktura logového souboru [12]**

| Název položky | Popis položky |
|---|---|
| remotehost | IP adresa nebo doménové jméno vzdáleného počítače, který zaslal požadavek |
| rfc931 | identifikační údaj uživatele (v současné době se již nepoužívá) |
| auth-username | login uživatele |
| timestamp | datum a čas požadavku přijeté požadavku web serverem |
| request-line | cesta a jméno objektu, který je požadován uživatelem, metoda a verze http protokolu |
| response-code | kód, který je vrácen uživateli na jeho požadavek |
| response-size | velikost obsahu vráceným serverem na základě požadavku |
| referrer | URL stránky, ze které byl zaslán požadavek uživatelem |
| user-agent | popis aplikace, která zaslala požadavek na server (internetový prohlížeč, indexovací robot vyhledávače) |

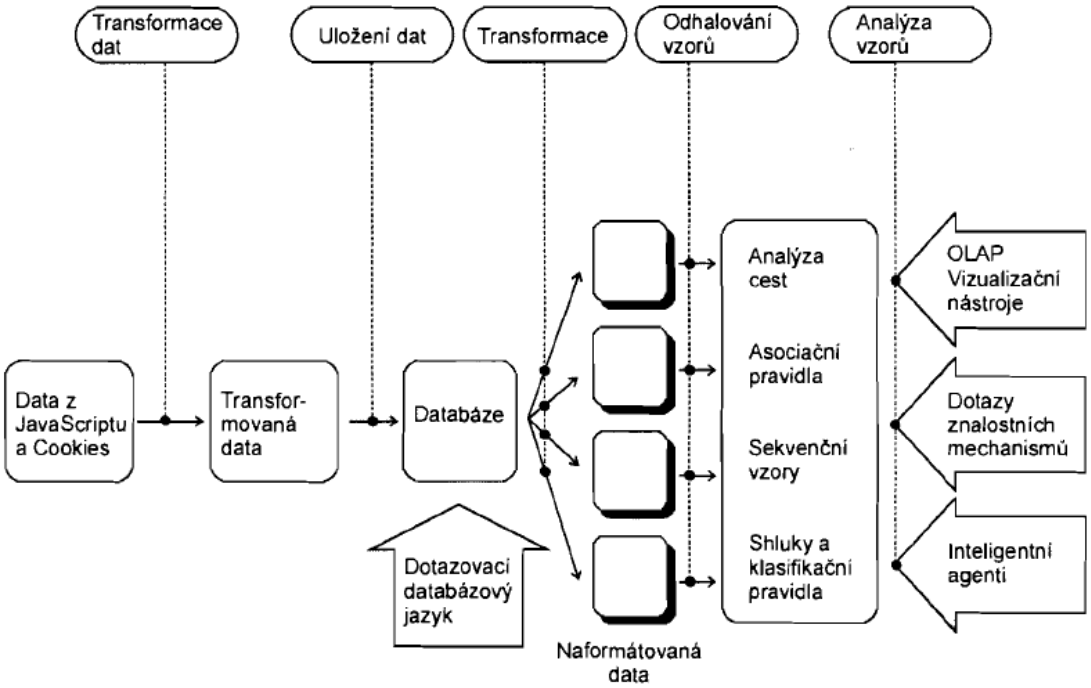
Ačkoliv serverové logy poskytují mnoho velmi užitečných informací, je zde i několik nevýhod. Jednou z nich může být i to, že standardní serverové logy neobsahují požadavky, které byly pokládány přes metodu POST [1]. Jelikož jsou v souborech ukládány informace o všech návštěvnících, kteří mohou přistupovat na server pokaždé s jinou adresou, je pracné vystopovat transakce jednoho konkrétního. Dalším problémem může být cachování stránek, které je využíváno ke zmenšení zatížení serveru. Výsledkem je potom situace, kdy pokud návštěvník klikne na tlačítko „*zpět*“, stránka je načtena z cache paměti a tento krok není v logu obsažen.

Další možností, jak získat data, je využití skriptovacího jazyku nebo cookies. Např. Javascriptem lze naprogramovat skripty, které se vloží na webové stránky a odtud získávají potřebné informace, které jsou odesílány na server.

Porovnání získaných dat přes logové soubory a Javascriptem je možno vidět na obr. 1.4 a obr. 1.5. Je zřejmé, že za použití Javascriptu není potřeba s daty při předzpracování provádět tolik operací, jako v případě logových souborů, které obsahují pro potřeby WUM také nepotřebná data.




**Obr. 1.4 - Architektura WUM se zdrojem dat z logovacího souboru [3]**




**Obr. 1.5 - Architektura WUM se zdrojem dat z Javascriptu a cookies [3]**

## Sběr dat z úrovně klienta

Serverové logy jsou generovány automaticky; naopak v případě sběru dat z úrovně klienta je již nutná spolupráce uživatele. Klientem je zde myšlen webový prohlížeč. Logy z této úrovně se získávají přes vzdálené agenty, jako můžou být Javascripty, Java applety nebo přes modifikovaný zdrojový kód existujících webových prohlížečů. Ať už se jedná o povolení Javascriptu nebo dobrovolné používání upraveného webového prohlížeče, vždy je potřeba spolupráce uživatele [1].

V dnešní době se dostává do podvědomí uživatelů nový webový prohlížeč vyvinutý Googlem – Google Chrome. Chrome zaznamenává kromě standardních informací i informace, které jsou uložené v cookies, vyhledávané fráze, navštívené stránky, špatně zadané adresy a mnoho dalšího. Díky těmto informacím vytváří uživatelův profil. Problém osobních údajů je zde vyřešen přijetím licenčních podmínek užívání prohlížeče a uživatel má možnost navštěvovat webové stránky v tzv. anonymním režimu [13].

I v případě získávání informací z klienta existují nevýhody a to zejména v nutnosti spolupráce uživatele. Javascripty zachycují pouze chování jednoho uživatele na jedné stránce. Modifikovaný webový prohlížeč je v tomto směru universálnější a umožňuje sbírat data o jednom uživateli, ale i na více stránkách. Ovšem zde nastává problém v tom, aby uživatelé tyto webové prohlížeče používali [1]. V případě cookies, které jsou uloženy na straně klienta je zde problém takový, že uživatel je může smazat nebo zamezit jejich ukládání. Nespornou výhodou je naopak fakt, že informace, které jsou získány tímto způsobem ukazují na chování konkrétního uživatele.

## Sběr dat z úrovně proxy serveru

Proxy server je používán jako prostředník mezi klientem a cílovým webovým serverem. Používání proxy serveru je možno mimo jiné použít k redukování času načítání stránek, snížení zatížení serveru nebo pro zvýšení bezpečnosti. Problémem zde je, že za proxy serverem může být několik uživatelů, ale na straně serveru se objeví pouze adresa tohoto proxy serveru. Tímto přístupem lze tedy získat data, která charakterizují chování skupiny anonymních uživatelů, kteří sdílejí společný proxy server. V tab. 1.2 je porovnáno, jaká data jsou sbírána jednotlivými přístupy.

**Tab. 1.2 - Porovnání přístupů sběru dat [vlastní]**

| Úroveň sběru dat | Rozsah dat |
|---|---|
| server | údaje o přístupech různých uživatelů ke konkrétnímu serveru |
| klient | údaje o uživateli prohlížeče vzhledem ke všem navštíveným serverům |
| proxy | údaje o všech uživatelích konkrétního proxy serveru a všech jimi navštívených serverech |

Všechny výše popsané způsoby získání dat mají své výhody a nevýhody. Na každé úrovni je možno získat jiný druh informací. Logové soubory obsahují veškeré informace, které server zaznamenal. Informace získané přes Javascript nebo z cookies obsahují vhodnější informace a zpracování takovýchto dat je jednodušší. Stejně jako v data miningu, i zde je potřeba data nejdříve vhodně upravit, aby byla vhodná pro jednotlivé metody jejich zpracování.

### 1.4.1 Předzpracování dat

Předzpracování dat je velmi důležité před aplikováním jednotlivých metod. Data, která jsou získána výše uvedenými způsoby, obsahují množství nepotřebných informací a nejsou v ucelené podobě. Je nutné tedy provést jejich čištění a strukturalizaci (viz obr 1.4).

Čištění dat v tomto smyslu znamená odstranění nepotřebných dat, jako mohou být obrázky, multimediální soubory nebo návštěvy indexovacích robotů vyhledávacích nástrojů. Indexovací roboti mají většinou naprogramované určité chování na stránkách a proto by mohli zkreslovat výsledky analýz [14]. Z textu vyplývá, že již v tomto případě je potřeba vědět, jaké metody budou nad daty prováděny z toho důvodu, že jiná data jsou vhodná např. pro predikci návštěvnosti a jiná data pro klasifikaci webových stránek. Výsledkem čištění dat jsou tedy data vhodná pro jednotlivé metody.

Další částí přípravy dat je strukturalizace dat. Během této fáze jsou data seskupena podle uživatelů, jejich sezení a jsou zjišťovány jednotlivé transakce.

Seskupování požadavků, kladených na server podle uživatelů neboli identifikace uživatele, záleží na politice stránek. Jestliže se jedná o stránky, kde je nutné přihlášení, je problém identifikace uživatelů vyřešen. V ostatních případech (např. pomocí IP adres) je identifikace jednotlivých uživatelů obtížnější. V tomto případě je pod každou IP adresou uvažován jeden uživatel [14]. V tomto případě je nutné počítat s tím, že v případě proxy serveru může být pod jednou adresou uživatelů více. V pracích [6] a [13] je zmíněno o heuristické metodě, která je

založena na kombinaci IP adresy a prohlížeči nebo operačního systému. Na základě tohoto přístupu je možné předpokládat, že pokud se v logu objeví ta samá IP adresa v kombinaci s různými typy prohlížeče, tak se jedná o dva různé návštěvníky. Ovšem ani tato metoda nemůže být neomylná.

Identifikací sezení návštěvníků jsou zde myšleny všechny požadavky uživatele na server, které provede během určitého času na jednom počítači stejným klientem [14]. Pokud není možno z logového souboru zjistit, kdy návštěvník stránky opustil, považuje se za čas strávený na těchto stránkách 30 minut. Identifikace sezení na základě časového limitu není přesná, nelze zaručit, že návštěvník po určitém čase nezměnil oblast zájmu. V tomto případě by díky časovému limitu bylo jedno sezení rozděleno do dvou [13].

Dalším problémem je identifikace posloupnosti transakcí ze sezení. Existuje více přístupů, jak lze transakce zjistit. V práci [14] je uvedena technika Maximální dopředné odkazování *(Maximal Forward Reference*). Je založena na domněnce, že uživatel zachází s webovou stránkou dvěma způsoby: buď jako navigační, jež mu umožní nalézt posloupností odkazů jím požadovanou informaci, nebo jako obsahovou. „Rozdělení stránek na jednotlivé typy je možné bud' ručně, nebo automaticky, podle způsobu, jakým je uživatelé procházejí. Při automatickém procházení se nejčastěji za obsahové stránky považují ty, na nichž uživatel stiskl tlačítko *„zpět"* pro návrat na předchozí stránku. Každá transakce je v tomto případě ukončena obsahovou stránkou. Nová transakce může začít ihned nebo až po posledním stisku tlačítka *„zpět"* (v případě, že se uživatel vrací o více stránek)" [16].

Předzpracováním dat jsou získána data, která již jsou vhodná pro jednotlivé metody. Nejčastější metody, které se v oblasti Web miningu používají, jsou předmětem následující podkapitoly.

### 1.4.2 Používané metody ve Web Miningu

V této podkapitole jsou stručně popsány metody, které se používají v souvislosti s Web miningem. Jedná se o známé algoritmy, které jsou používány běžně v dolování znalostí z databází, proto zde nebudou rozepisovány podrobnosti, ale spíše jejich přínos pro aplikaci ve Web miningu. Mezi nejčastější Web miningové metody podle [1] patří statistické analýzy, asociační pravidla, shluková analýza nebo analýza průchodu webem.

### 1.4.3 Statistické analýzy

Statistické analýzy se řadí mezi nejznámější metody k extrakci znalostí o návštěvnících webových stránek. Logové soubory produkují reporty v různých časových intervalech. Tyto reporty obsahují informace jako jsou počet návštěvníků stránek, průměrné a střední hodnoty, informace o nejnavštěvovanějších stránkách, délce setrvání návštěvníka na stránkách, průměrnou délku průchodu stránkami atd. Reporty také mohou obsahovat informace o neautorizovaných přístupech [1].

Ačkoliv se nejedná o hloubkové analýzy webových stránek, tyto informace jsou velmi důležité např. pro správce serveru, protože poskytují informace o tom, kdy je na stránkách největší návštěvnost, tudíž nejvíce zatížen server. Dále mají uplatnění pro marketingové účely, jelikož podávají informace v kterou denní, či roční dobu zákazníci nejvíce nakupují zboží – tím je možné plánovat zásobování zboží nebo plánování marketingových akcí. Na základě informací o neautorizovaných přístupech je možné posílit zabezpečení a mnoho dalšího [1].

### 1.4.4 Asociační pravidla

Metoda generování asociačních pravidel je založena na objevování takových asociací a korelací mezi daty, kde výskyt jedné množiny objektů v transakci implikuje s jistou hodnotou podpory a důvěry výskyt jiných objektů. V oblasti Web Miningu se používají nejčastěji data o chování návštěvníků. Dolováním znalostí z chování se návštěvníků je více uvedeno v podkapitole věnované Web Usage Miningu.

Každá transakce se dle [6] skládá z množiny odkazů, kterými návštěvník projde během jeho návštěvy na serveru. Z těchto transakcí, které splňují jistou hodnotu podpory, je potom možné odvodit různá pravidla typu:

- 40% návštěvníků, kteří navštívili stránku */výrobky/produktA*, také navštívili */výrobky/produktB* [6].

Na základě takovýchto pravidel je potom možné upravit stránky tak, aby návštěvníkovi, který se momentálně zajímá o produkt A, automaticky nabídly i produkt B.

### 1.4.5 Segmentace

Podstata segmentace je v roztřídění objektů, které mají podobné charakteristiky do homogenních skupin. Jedná se o metodu učení bez učitele, tj. kdy nejsou dopředu známy tyto skupiny. Z pohledu Web miningu jsou zajímavé dva druhy segmentace: segmentace na základě chování návštěvníků, tj. segmentace návštěvníků a segmentace stránek (respektive dokumentů a ostatních částí) [1].

Segmentace návštěvníků se zaměřuje na objevení nových skupin, které charakterizují podobné vzory v procházení stránek. Tyto znalosti jsou především užitečné pro odvozování demografických údajů za účelem provádění segmentace trhu v prostředí elektronického obchodování a také pro personalizace webových stránek. Na druhé straně, segmentace stránek objevuje skupiny stránek, které mají podobný obsah. Tyto informace jsou potom užitečné pro internetové vyhledávače [1].

### 1.4.6 Klasifikace

Pomocí klasifikace, podobně jako u segmentace, je možné roztřídit objekty na základě jejich vlastností do skupin. Rozdíl je zde v tom, že klasifikace patří mezi metody učení s učitelem. Zde jsou tedy dopředu známy skupiny, do kterých objekty patří. V oblasti Web miningu je důležité vhodně popsat a stanovit charakteristiky každé skupiny. Na základě těchto skupin je možné potom zatřídit nové návštěvníky podle jejich demografických údajů nebo chování.

Používají se zde různé algoritmy, jako jsou např. rozhodovací stromy, naivní Bayesovské klasifikátory, metoda nejbližšího souseda a další [1]. Díky klasifikaci serverového logu můžou být objevena zajímavá klasifikační pravidla typu:

- 40% návštěvníků, kteří si objednali zboží z */výrobky/Hudba* jsou ve skupině 18-25 let a žijí ve východních Čechách.

### 1.4.7 Analýza průchodu webem

Analýza průchodu webem (*Path Analysis*) je založena na tom, že webová stránka se dá považovat za graf. Takový graf může mít více podob, nejčastějším typem je takový, který reprezentuje fyzické uspořádání webové stránky. Uzly grafu v tomto případě tvoří webové dokumenty (jednotlivé stránky) a odkazy mezi těmito stránkami tvoří hrany grafu. Jiným typem grafu může být takový, kde hrany jsou ohodnoceny číslem. Toto číslo udává, kolik

návštěvníků přes tuto hranu prošlo, aby se dostali ke stránce, která je touto hranou spojená z jiné stránky [6]. Z analýzy průchodu webem je možné získat informace jako jsou:

- 40% návštěvníků, kteří zakoupili *produkt B*, přišli přes stránku *Úvod* a *Akční produkty*.
- 90% návštěvníků začíná na stránce *Úvod*.
- 60% návštěvníků opustilo stránky po zhlédnutí pěti a méně stránek.

Tyto informace jsou opět velmi dobře využitelné při úpravě struktury webu do takové podoby, aby zde návštěvníci vytrvali a nalezli rychle to, co hledají. Informace z analýzy průchodu webem mohou být využitelné v problému umisťování reklamy na stránkách.

### 1.4.8 Sekvenční vzory

Sekvenční vzory (*Sequential Patterns*) se snaží nalézt vzory mezi jednotlivými transakcemi prováděnými uživateli. Princip je založen na tom, že v logových souborech jsou zaznamenány jednotlivé transakce během časového období. Díky tomuto přístupu je možné predikovat vzory chování při další návštěvě stránek. Sekvenční vzory také napomáhají umístit reklamu cílenou na určitou skupinu návštěvníků [1], [6]. Touto metodou je možné zjistit zajímavé vzory chování v čase, kterými mohou být:

- 30% návštěvníků, kteří přišli z vyhledávače Google a hledali slovo *kolo*, si do pěti dnů objednali *produkt A*.
- 60% návštěvníků, kteří si objednali *produkt A*, si objednali *produkt B* do deseti dnů od nákupu *produktu A*.

Pokud jsou známy takovéto informace, je možné návštěvníkovi nabídnout produkt dříve, než by ho to napadlo samotného a mohl si ho koupit u konkurence.

## *1.5 Dílčí závěr kapitoly*

Úvodní kapitola se věnuje poměrně málo rozšířenému tématu oblasti data miningu – Web miningu. Tato metoda blíže zkoumá internetové stránky a především chování uživatelů na nich. Na základě získaných dat je potom možné aplikovat mnoho data miningových metod, které v konečném důsledku vedou k takové struktuře webových stránek, že je návštěvník ve vyhledávači nalezne, jsou pro něj co nejpříjemnějšími a rychle na nich nalezne to, kvůli čemu přichází. Data, která jsou získána touto metodou, se uplatní především v potřebách rozhodování a marketingu.

# 2 Neuronové sítě

## *2.1 Základní pojmy z oblasti neuronových sítí*

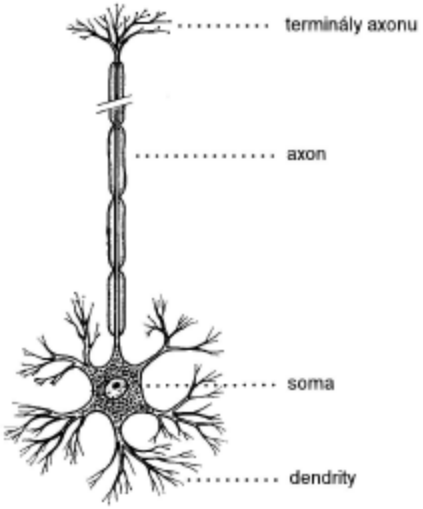
V této podkapitole jsou nejdříve stručně popsány základní pojmy z oblasti neuronových sítí, se kterými se dále pracuje. Dále je zde blíže popsána struktura neuronových sítí typu radiálně bazických funkcí, zkráceně RBF sítě.

### 2.1.1 Neuronová síť

Existuje velké množství definic, co je to neuronová síť. Například podle [17] je neuronová síť určená jako orientovaný graf $G = (V, E)$. Vrcholy grafu jsou neprázdnou množinou $V = \{v_1, v_2, \dots, v_N\}$. Vrcholy jsou tvořeny neurony. Spoje grafu jsou tvořeny neprázdnou množinou $E = \{e_1, e_2, \dots, e_M\}$ a tvoří je synapse. Každý spoj $e \in E$ je uspořádaná dvojice dvou neuronů z množiny $V$, $e = (v, v')$. Uspořádaná dvojice neuronů, protože každý spoj začíná v neuronu $v$ a končí v neuronu $v'$.

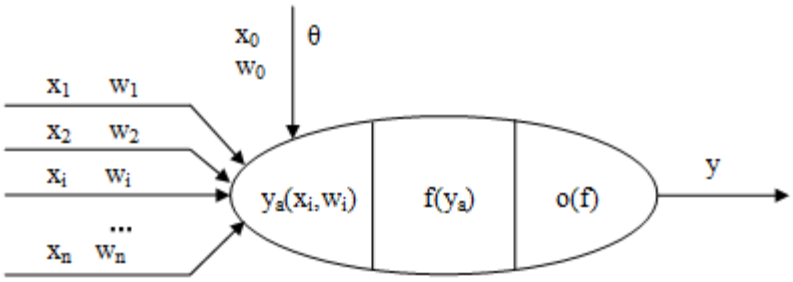
### 2.1.2 Neuron

Neuron je základní stavební jednotkou neuronové sítě. Je zjednodušením biologického neuronu, který se nachází v mozku, a lze vyjádřit matematicky. Biologický neuron je zobrazen na obr. 2.1 a jeho formální tvar, tzv. McCulloch-Pittsův model neuronu je zobrazen na obr. 2.2.




**Obr. 2.1 - Biologický neuron [18]**

Detailní popis a princip biologického neuronu je popsán v [17], [18], [19].




**Obr. 2.2 - McCulloch - Pittsův model neuronu [20]**

McCulloch – Pittsův model neuronu je dle [20] složen z následujících částí:

$x_i$ – vstupy neuronu (výstupy z předcházející vrstvy), i = 1, 2, …, n,

n – počet vstupů (počet neuronů v předcházející vrstvě),

$w_i$ – synaptické váhy,

$y_a$ – vstupní potenciál neuronu,

f – aktivační funkce neuronu,

o – výstupní funkce neuronu,

θ – práh neuronu,

y – výstup neuronu.

Matematický popis neuronu je dán vztahem

$$y = f\left(\sum_{i=1}^{n} x_i(t) \times w_i(t) + \theta\right). \tag{2.1}$$

Všechny vstupy do neuronu jsou dle [20] ohodnoceny určitou hodnotou příslušné synaptické váhy. Tato hodnota udává citlivost, s jakou příslušný neuron působí na výstup z neuronu. Hodnota prahu $\theta$ určuje, kdy je neuron aktivní. Jestliže neuron nedosáhne hodnoty prahu, není dále šířen. Jakmile dojde k překročení prahové hodnoty a neuron se zaktivuje, dochází k růstu výstupního signálu do určité maximální hodnoty, která je dána oborem hodnot příslušné aktivační funkce *f*.

Vstupní vektor hodnot $\boldsymbol{x}_i$, i=0,1,…,n je transformován na skalární signál $y_a$, který je dále vstupem do aktivační funkce neuronu. Proces agregace je možné popsat vztahem [20]

$$y_a(t) = \sum_{i=1}^{n} x_i(t) \times w_i(t) + \theta. \tag{2.2}$$

Jelikož lze práh $\theta$ považovat za speciální případ synaptické váhy, která vede od fiktivního neuronu, jehož výstup má trvale hodnotu +1, resp. -1, lze zavést substituci

$$\theta = x_0 \times w_0. \tag{2.3}$$

Potom lze zapsat vstupní potenciál $y_a(t)$ ve tvaru [20]

$$y_a(t) = \sum_{i=0}^{n} x_i(t) \times w_i(t). \tag{2.4}$$

Aktivační funkce má za úkol převést hodnotu vstupního potenciálu na výstupní hodnotu neuronu, což je dáno vztahem

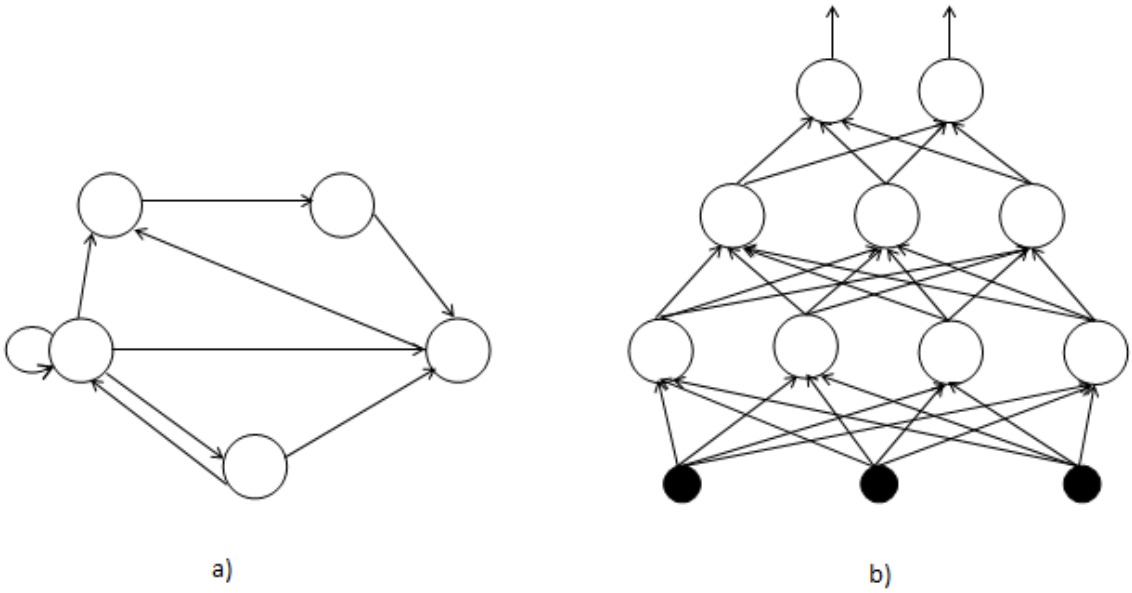
$$y = f\big(y_a(t)\big). \tag{2.5}$$

Výběr vhodné aktivační funkce je třeba volit s ohledem na konkrétní typ řešené úlohy, případně na konkrétní polohu neuronu v neuronové síti.

Nejčastější aktivační funkce jsou skokové, lineární, po částech lineární, sigmoidální, hyperbolický tangens, Gaussova, atd [20].

### 2.1.3 Topologie neuronové sítě

Topologie sítě je určena na základě počtu neuronů a jejich propojení mezi sebou. Jestliže se v propojení neuronů vyskytuje alespoň jeden cyklus, hovoří se o rekurentní (cyklické) neuronové síti. Jestliže zde žádný cyklus neexistuje, jedná se o acyklickou neuronovou síť (dopřednou). V případě acyklické (dopředné) neuronové sítě je možné jednotlivé vrstvy uspořádat na vstupní, skrytou a výstupní vrstvu. Potom je možné označit v takové síti neurony jako vstupní, skryté a výstupní. Podle toho, zda síť obsahuje jeden nebo více výstupních neuronů, se dělí neuronové sítě na klasifikační a predikční. Pokud je na výstupu pouze jeden neuron, jedná se o síť predikční a pokud jich je zde více, jedná se o síť klasifikační. Vstupních neuronů je tolik, kolik je vstupních parametrů. Co se týče skryté vrstvy, zde není přesně stanoveno, kolik by měla obsahovat neuronů. Stanovení tohoto počtu je dáno experimentálně expertem a tato část se řadí k těm nejpracnějším při návrhu neuronové sítě. Příklad cyklické a acyklické neuronové sítě jsou zobrazeny na obr. 2.3.




**Obr. 2.3 - Příklad cyklické (a) a acyklické (b) neuronové sítě [18]**

### 2.1.4 Učení neuronové sítě

„Proces učení (adaptace) neuronové sítě je možno charakterizovat jako tu etapu její činnosti, kdy se podle daných požadavků mění vlastnosti jejích výkonných prvků a případně též její konfigurace“ [19]. Učení neuronové sítě je její základní vlastností a tato vlastnost je výrazně odlišuje od algoritmických systémů pro zpracování informací. Vlastní učení se uskutečňuje především modifikací hodnot synaptických vah (popř. pomocí prahů a parametrů aktivačních funkcí jednotlivých neuronů).

Podle [20] lze rozdělit proces učení nastavováním hodnot synaptických vah na dvě části – fázi učení a fázi života. Fáze učení (adaptace) je stav, kdy se znalosti ukládají do synaptických vah neuronové sítě. Jestliže $\boldsymbol{W}$ je matice všech synaptických vah neuronové sítě, pak učením sítě je stav, kdy platí

$$\frac{\partial \boldsymbol{W}}{\partial t} \neq 0. \tag{2.6}$$

Fáze života je stav, kdy se již synaptické váhy nemění. Naučené znalosti se využívají v řešení problému, jako může být například klasifikace nebo predikce. Tento stav lze tedy zapsat jako

$$\frac{\partial \boldsymbol{W}}{\partial t} = 0. \tag{2.7}$$

Pro učení a ověření správnosti učení neuronové sítě je třeba rozdělit množinu dat na tzv. trénovací a testovací množinu. Zde je zapotřebí dbát na výběr dat pro trénovací množinu, jelikož síť se na základě těchto hodnot učí. Na testovací množině je potom ověřováno, jak je síť naučena, respektivě jak správně jsou nastavené váhy mezi neurony.

Učení neuronové sítě rozdělit podle způsobu učení na učení s učitelem a učení bez učitele. Učení s učitelem znamená, že ke každé hodnotě vstupu je známa příslušná hodnota výstupu. Vlastní učení potom spočívá v nastavování vah a srovnávání aktuálního výstupu s požadovaným výstupem. Váhy se dále přenastavují tak, aby se snížil rozdíl mezi skutečným a požadovaným výstupem. Metodika snižování tohoto rozdílu je určena učícím algoritmem, kam patří například metody posilovaného učení, stochastického učení a zpětné šíření chyby (*Back-Propagation*). Učení bez učitele představuje situaci, kdy příslušný vektor výstupních hodnot není známý. V tomto případě učení spočívá v hledání určitých vzorů společných vlastností ve vstupních datech. Jedná se tedy o samoorganizaci. Metody, které lze uvést jako příklad učení bez učitele, jsou například Hebbovské učení nebo kompetice [21].

Neuronové sítě mají v dnešní době velké uplatnění v mnoha oborech. Využívají se především pro predikci, klasifikaci, aproximaci, kompresi dat a mnoho dalšího. Předmětem této práce je predikce časové řady.

Predikce je podle [21] předpovídání výstupní hodnoty jisté veličiny na základě jejího průběhu v minulosti. „Při predikci jde o to, abychom v průběhu nějaké známé číselné řady, jejíž hodnoty se mění v závislosti na některém nezávisle proměnném parametru sledovaného jevu (tím může být kterákoliv fyzikální veličina, ale i čas), nalezli co nejpravděpodobnější průběh nezávislé proměnné. Predikce je vlastně speciálním případem extrapolace; tou se rozumí způsob odvození, nebo závěr plynoucí z chování funkce uvnitř známého oboru pro její chování mimo tento obor.“ [21]

Kvalitu predikce je možné měřit tzv. predikční chybou, která vyjadřuje míru nepřesnosti mezi predikovaným a originálním výstupem. V praxi se nejčastěji používají tyto chyby [22]:
Střední kvadratická chyba MSE (*Mean Squared Error*), definovaná vztahem

$$MSE = \frac{1}{n}\sum_{i=0}^{n-1}(t_i - o_i)^2, \tag{2.8}$$

kde: $t_i$ je predikovaná hodnota,

$o_i$ je originální výstup.

Tento vztah je často nahrazován RMSE chybou (*Root Mean Squared Error*), která je odmocninou MSE chyby

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=0}^{n-1}(t_i - o_i)^2}\,. \tag{2.9}$$

Mezi další chyby, které jsou často používány, patří také Střední absolutní chyba MAE (*Mean Absolute Error*), která je definována vztahem

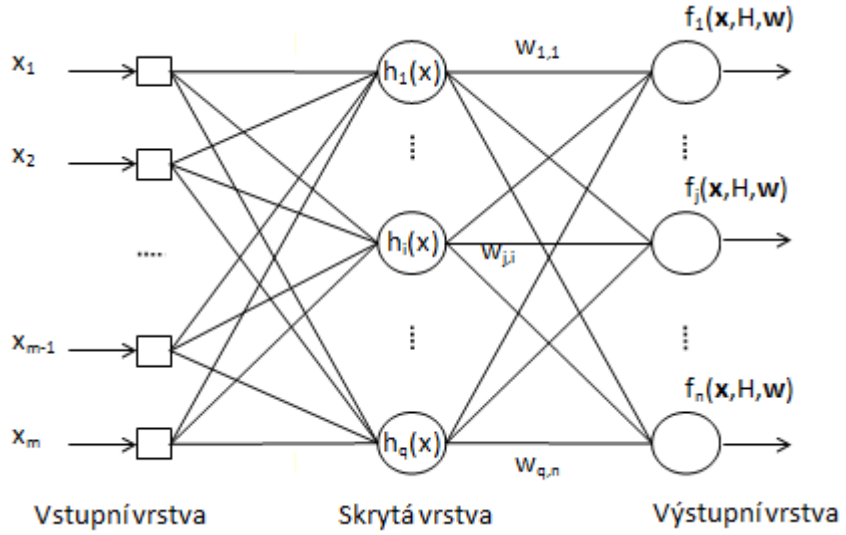
$$MAE = \frac{1}{n}\sum_{i=1}^{n}|t_i - o_i|. \tag{2.10}$$

## 2.2 Neuronové sítě typu RBF

Neuronové sítě typu radiálně bazických funkcí (dále jen RBF sítě) se začaly vyvíjet podle [18] začátkem osmdesátých let dvacátého století. V této době se hledaly nové metody, jakými funkcemi by se daly nejlépe aproximovat data. Vědci se zaměřili na tzv. radiální bazické funkce (*Radial Basis Functions*). „Radiální funkci si můžeme představit jako funkci určenou nějakým významným bodem – středem – která pro argumenty se stejnou vzdáleností od tohoto středu dává stejné funkční hodnoty. Měříme – li vzdálenost pomocí eukleidovské metriky a uvažujeme například dvojrozměrný vstupní prostor, pak množiny se stejnou funkční hodnotou tvoří kružnice, proto tedy hovoříme o radiálních funkcích“ [18]. RBF sítěmi je možno nazývat všechny dopředné neuronové sítě, které využívají RBF jako aktivační funkci [23].

### 2.2.1 Struktura RBF sítí

Neuronová síť typu RBF je zobrazena na obr. 2.4. Je složena ze třech vrstev [18], [23]. Vstupní vrstva slouží pouze k přenosu vstupních hodnot. Každý z $m$ vstupů vektoru $\boldsymbol{x} = (x_1, x_2, \dots, x_k, \dots, x_m)$ je vstupní hodnotou pro $q$ radiálních bazických funkcí. Jediná skrytá vrstva je složena z tzv. RBF neuronů, které realizují jednotlivé radiální bazické funkce. Poslední, výstupní vrstva je složena z neuronů perceptronového typu. Výstup je dán váženým součtem výstupů ze skryté vrstvy [23].




**Obr. 2.4 - Neuronová síť typu RBF [23]**

Jak již bylo zmíněno výše, vstupní vrstva slouží pouze k přenosu vektoru vstupních hodnot ***x***, které jsou dále parametry aktivačních funkcí v RBF neuronech ve skryté vrstvě. Množina těchto funkcí je označena jako $H = \{h_1(x), h_2(x), \dots, h_i(x), \dots, h_q(x)\}$, kde $q$ je počet neuronů ve skryté vrstvě. Aktivační funkce RBF neuronů ve skryté vrstvě je speciální třídou matematických funkcí. Její hlavní charakteristika je monotónní stoupání nebo klesání se zvyšující se vzdáleností od centra $c_i$ aktivačních funkcí $h_i(x)$ RBF neuronů. Jako aktivační funkce lze využít například Gaussovu aktivační funkci, rotační Gaussovu aktivační funkci, multikvadratickou a inverzní multikvadratickou aktivační funkci, Cauchyho funkci atd.

Gaussovu aktivační funkci *hi(x)* je možno zapsat jako [23]

$$h(\boldsymbol{x}, C, R) = \sum_{i=1}^{q} exp\left(-\frac{\|\boldsymbol{x} - \boldsymbol{c_i}\|^2}{r_i}\right), \tag{2.11}$$

kde: $\boldsymbol{x} = (x_1, x_2, \dots, x_k, \dots, x_m)$ reprezentuje vstupní vektor,

$C = \{c_1, c_2, \dots, c_i, \dots, c_q\}$ jsou centra aktivačních funkcí $h_i(x)$ RBF neuronů,

$R = \{r_1, r_2, \dots, r_i, \dots, r_q\}$ jsou poloměry (šířky) aktivačních funkcí $h_i(x)$,

$q$ je počet neuronů ve skryté vrstvě.

*J-tý* výstup $f_j(\boldsymbol{x}, H, \boldsymbol{w})$ je dán váženým součtem výstupů ze skryté vrstvy, což lze podle [23] zapsat jako

$$f_j(\boldsymbol{x}, H, \boldsymbol{w}) = \sum_{i=1}^{q} w_{j,i} \times h_i(\boldsymbol{x}), \tag{2.12}$$

kde: $\boldsymbol{x} = (x_1, x_2, \dots, x_k, \dots, x_m)$ reprezentuje vstupní vektor,

$H = \{h_1(x), h_2(x), \dots, h_i(x), \dots, h_q(x)\}$ je množina aktivačních funkcí RBF neuronů ve skryté vrstvě,

$w_{j,i}$ jsou synaptické váhy mezi skrytou a výstupní vrstvou,

$q$ je počet neuronů ve skryté vrstvě.

### 2.2.2 Učení RBF sítě

Učení RBF sítě sestává podle [18] ze třech kroků, které jsou složeny jak s učením bez učitele, tak s učením s učitelem. Předmětem učení je nalézt souřadnice středů RBF neuronů (první krok), jejich šířky (druhý krok) a dále koeficienty lineární kombinace výstupní hodnoty (třetí krok). První dva kroky představují učení bez učitele, poslední krok je potom klasické učení s učitelem. Vstupem do procesu učení je dle [18] tréninková množina, která je složena z párů vektorů $\{(\boldsymbol{x}^{(t)}, \boldsymbol{d}^{(t)}); t = 1, \dots, m\}$, sestávající ze vstupů $\boldsymbol{x}^{(t)} \in R^n$ a požadovaných výstupů $\boldsymbol{d}^{(t)} \in R^m$.

V prvním kroku se jedná o učení bez učitele. Je zapotřebí určit souřadnice $q$ středů RBF neuronů, které jsou reprezentovány vahami $\{c_{ji}; i = 1, \dots, m; j = 1, \dots, q\}$ mezi vstupní a skrytou vrstvou. K tomu je využíváno různých přístupů. [18] uvádí přístupy rovnoměrného rozložení středů, výběr náhodných vzorků, výběr optimálních vzorků nebo samoorganizaci.

V případě rovnoměrného rozložení, které patří mezi nejjednodušší postup, se středy jednotek pravidelně rozmístí rovnoměrně po vstupním prostoru. Tento přístup je vhodný v případě, že vstupní data jsou také rozmístěna rovnoměrně (typicky v případech, kdy jsou tréninková data výsledkem pravidelného měření). Přístup náhodných vzorků spočívá v tom, že se vybere $q$ náhodných vzorků z tréninkové množiny a na jejich vstupní části se umístí středy RBF jednotek. Výběr optimálních vzorků je podobný náhodným vzorkům, ale s tím rozdílem, že výběr zde není náhodný, nýbrž se používá metoda ortogonálních nejmenších čtverců k tomu, aby se minimalizovala chyba neuronové sítě. Při samoorganizaci se využívá k nastavení středů RBF neuronů aplikace znalostí ze shlukové analýzy. Nejčastěji se využívá metody K-Means, jejíž algoritmus je v tomto případě dle [18] následující:

(i) rozmísti $c_j$ náhodně po vstupním prostoru

(ii) v čase *t* dělej:

(a) najdi střed $\boldsymbol{c}_c$, který je nejblíže vstupu $\boldsymbol{x}^{(t)}$

(b) posuň $\boldsymbol{c}_c$ k $\boldsymbol{x}^{(t)}$ podle

$$\boldsymbol{c}_c := \boldsymbol{c}_c + \theta(t)\left\|\boldsymbol{x}^{(t)} - \boldsymbol{c}_c\right\|,$$

kde $0 < \theta(t) < 1$ je parametr učení.

(c) pokud dochází ke změnám, pokračuj od bodu (ii), jinak skonči.

V druhém kroku učení, které je také založené na učení bez učitele, je cílem nalézt parametry, které určují šířku radiální oblasti kolem středu $\boldsymbol{c}$, ve kterém má RBF neuron významné výstupní hodnoty. Parametry, které určují šířku, mají vliv na generalizační schopnosti neuronové sítě. Čím jsou menší, tím horší generalizaci lze očekávat, ale naopak, pokud bude okolí příliš široké, neurony pak ztrácejí svůj lokální charakter [18].

Tento parametr je možné určit jako střední kvadratickou vzdálenost vzorů od středu shluku [24]

$$r_k = \sqrt{\frac{1}{Q}\sum_{i=1}^{Q}\left\|\overline{c_k} - \overline{x_q}\right\|^2}, \tag{2.13}$$

kde: $x_q$ je $q$-tý vzor náležející ke shluku se středem $c_k$.

V třetím kroku učení, které je již založeno na učení s učitelem, zbývá nastavit váhy $w_{ji}$ lineární kombinace. Výsledkem učení by měla být síť s co nejmenší odchylkou mezi vstupním předkládaným vektorem hodnot a výstupním vektorem hodnot. Jedná se tedy o minimalizaci chybové funkce [18], [24]

$$\mathrm{E}(\mathbf{w}) = \frac{1}{2}\sum_{\mathrm{t}=1}^{\mathrm{m}}\left\|\mathbf{d}^{(\mathrm{t})} - \mathbf{y}^{(\mathrm{t})}\right\|^2 = \frac{1}{2}\sum_{\mathrm{t}=1}^{\mathrm{m}}\sum_{\mathrm{j}=1}^{\mathrm{n}}(\mathrm{d}_{\mathrm{j}}^{(\mathrm{t})} - \mathrm{y}_{\mathrm{j}}^{(\mathrm{t})})^2, \tag{2.14}$$

kde: $\boldsymbol{y}^{(t)}$ ( resp. $y_j^{(t)}$) je aktuální výstup sítě (výstup *j*-tého výstupního neuronu) po předložení vstupního vektoru $\boldsymbol{x}^{(t)}$,

$\boldsymbol{w} = (w)_{qr}, q = 1, \ldots, m$ a $r = 1, \ldots, q$,

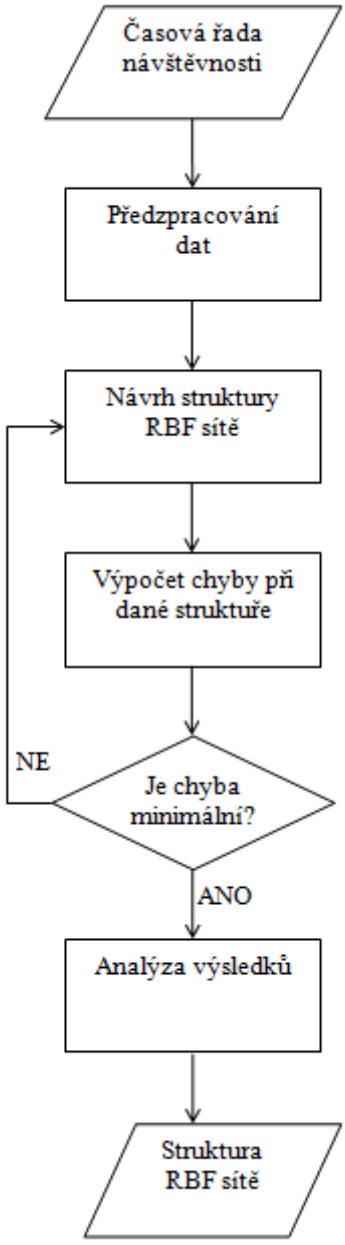
$d_j^{(t)}$ je požadovaný výstup sítě.

## 2.3 Dílčí závěr kapitoly

Druhá kapitola se zabývá neuronovými sítěmi. V dnešní době je toto téma velmi aktuální, neuronové sítě se využívají v mnoha oblastech a setkávají se s čím dál větší oblibou. Ačkoliv mají oproti klasickým výpočetním algoritmům nespornou výhodu takovou, že jsou schopné se učit, jejich hlavní nevýhodou je fakt, že doposud není zřejmé, jak k výsledku dospějí. Chovají se jako tzv. černá skříňka. Vznikly ve snaze napodobit myšlení lidského mozku, který se skládá z neuronů. Právě tyto neurony byly předlohou pro vytvoření tzv. formálního neuronu. Dále je zde stručně popsána architektura neuronových sítí a jejich učení. Druhá část kapitoly je věnována RBF neuronovým sítím, které byly dále použity v návrhu predikčního modelu návštěvnosti webové domény upce.cz.

# 3 Návrh modelu pro predikci časové řady

V této kapitole jsou popsány kroky spojené s modelem predikce návštěvnosti časové řady návštěvnosti domény upce.cz. Na obr. 3.1 je znázorněn celý proces, který se skládá z několika kroků.




**Obr. 3.1 - Návrh modelu [vlastní]**

## 3.1 Vstupní data

Vstupní data tvoří návštěvy webové domény upce.cz v období od 21. srpna do 31. května 2009. Tyto data byly pořízeny pomocí Google Analytics[4]. Google Analytics nabízí velkou řadu nástrojů, které se dají využít v oblasti Web miningu. Jeho největší předností je fakt, že je to služba, která je nabízená zdarma. Díky implementaci JavaScriptového kódu do webových stránek je možné získat široké spektrum jejich provozních charakteristik, kterým se říká webové metriky. Podle [25] lze tyto metriky rozdělit do čtyř skupin. Skupiny a základní přehled informací, které poskytují, jsou popsány v tab. 3.1.

**Tab. 3.1 - Služby Google Analytics [25] [26]**

| Skupina | Informace |
|---|---|
| návštěvy | počet návštěv zhlédnutých stránek; poměr nových návštěvníků vůči těm, kteří se na stránky vracejí; z jaké země návštěvníci přicházejí. |
| zdroje přístupů | odkud návštěvníci přicházejí. |
| obsah | zobrazované stránky, doba setrvání návštěvníka na stránkách, popularita jednotlivých stránek, přehledy pro nejlepší vstupní stránky, výstupní stránky a překryvná data stránek. |
| konverze | možnost definování cest na stránkách (cílů) a jejich statistiky. |

Pro účely této práce jsou potřeba data, která se týkají návštěv webové stránky Univerzity Pardubice www.upce.cz. Jak již bylo zmíněno na začátku kapitoly, počet návštěv byl měřen v období od 1. září 2008 do 31. května 2009. Návštěvou je zde rozuměna neopakovatelná kombinace IP adresy a cookies v tomto časovém období.

Informace, které lze získat pomocí Google Analytics o webu www.upce.cz za květen 2009 jsou následující [25]:

- Celková návštěvnost během měsíce klesá. Je zde zřejmý trend, kdy v pondělí je návštěvnost nejvyšší a klesá až po zbytek týdne. Sobota má nejnižší návštěvnost.
- Průměrný počet navštívených stránek je větší než tři.
- Návštěvníci zůstávají na stránce průměrně pět a půl minuty.
- Míra opuštění[5] je přibližně 60%
- Návštěvníci na stránky přicházejí napřímo, což je dobré.

[4] http://www.google.com/analytics/

[5] Míra opuštění (bounce rate) udává % návštěvníků, kteří webovou stránku opustí ihned po jejím vstupu.

- Mezi nejnavštěvovanější stránky patří hlavní stránka, dále pak stránky fakulty ekonomicko-správní a fakulty filozofické.

## *3.2 Předzpracování dat*

Data, získaná přes Google Analytics byla dále rozdělena na tři různě dlouhé časové řady: tzv. krátkodobou, střednědobou a dlouhodobou. Hodnoty těchto řad jsou vždy v intervalu jednoho dne. Jelikož se v každém intervalu vyskytují dny, kdy návštěvy byly výrazně vyšší (nižší), než v ostatní dny, byla data standardizována. Díky standardizaci, tedy převedení dat do určité škály, tyto výkyvy nebudou ovlivňovat výslednou hodnotu. Standardizovaná data mají potom střední hodnotu rovnou nule a směrodatnou odchylku rovnou jedné.

Standardizace dat byla provedena podle vztahu [27]

$$x_{ik}^{*} = \frac{x_{ik} - \bar{x}_k}{s_k}, \tag{3.1}$$

kde: $x_{ik}^{*}$ je standardizovaná hodnota původní hodnoty $x_{ik}, i = 1,2, \dots, n,$

$\bar{x}_k, k = 1,2, \dots, p$ je střední hodnota souboru hodnot,

$s_k$ je směrodatná odchylka.

Další předzpracování dat bylo realizované pomocí indikátorů technické analýzy. Použité indikátory a jejich charakteristiky jsou zobrazeny v tab. 3.2 [25].

**Tab. 3.2 - Metody technické analýzy [25]**

| Metody | Charakteristika |
|---|---|
| Jednoduchý klouzavý průměr (JKP) | 5, 7, 9 denní |
| Centrovaný klouzavý průměr (CKP) | 4, 6, 8 denní |
| Klouzavý medián (KM) | 5, 7, 9 denní |
| Jednoduché exponenciální vyrovnání (JEV) | pro α = 0.1 a α = 0.2 |
| Dvojité exponenciální vyrovnání (DEV) | pro α = 0.7 a α = 0.9 |

Jednoduchý klouzavý průměr (JKP) je obdobou klasického aritmetického průměru hodnot časové řady. Všem hodnotám je při výpočtu přiřazována stejná váha. JKP se tedy vypočítá jako součet hodnot v dané periodě, který se dále vydělí délkou této periody [28]:

$$JKP_{t,n} = \frac{1}{n}\sum_{t=1}^{n} Y_{t,n}, \tag{3.2}$$

kde: $JKP_{t,n}$ je hodnota průměru v čase *t* za periodu *n*,

*n* je délka periody,

$Y_{t,n}$ je empiricky zjištěná hodnota v čase *t*.

Centrovaný klouzavý průměr (CKP) je typem váženého klouzavého průměru se speciálními váhami, které jsou voleny tak, aby z časové řady eliminovaly sezónní složku. Délka CKP je vždy o jedničku větší než délka sezóny. Váhy jsou voleny tak, aby krajní hodnoty měly poloviční váhy oproti ostatním. Obecně lze CKP zapsat jako [28]:

$$CKP_{t,n} = \frac{1}{n-1}\left(\frac{Y_1}{2} + Y_2 + \cdots + Y_{n-1} + \frac{Y_n}{2}\right), \tag{3.3}$$

$$CKP_{t,n} = \frac{1}{2n-2}(Y_1 + 2Y_2 + \cdots + 2Y_{n-1} + Y_n), \tag{3.4}$$

kde: $CKP_{t,n}$ je hodnota průměru v čase *t* za periodu *n*,

*n* je délka periody,

$Y_n$ je empiricky zjištěná hodnota.

Klouzavý medián (KM) je představitelem tzv. robustních klouzavých průměrů, které slouží k potlačení odlehlých pozorování. Medián hodnot je prostřední hodnota pozorování, které jsou uspořádány podle velikosti [29].

Exponenciální vyrovnání (EV) patří v praxi k nejčastějším technikám vyrovnávání časových řad. Princip této metody spočívá ve využití historických pozorování. Váhy exponenciálně klesají směrem do minulosti. Čím starší jsou data, tím menší mají váhy. Jednoduché exponenciální vyrovnání je definováno vztahem [29]

$$\hat{Y}_t = \alpha Y_t + (1-\alpha)\hat{Y}_{t-1}, \tag{3.5}$$

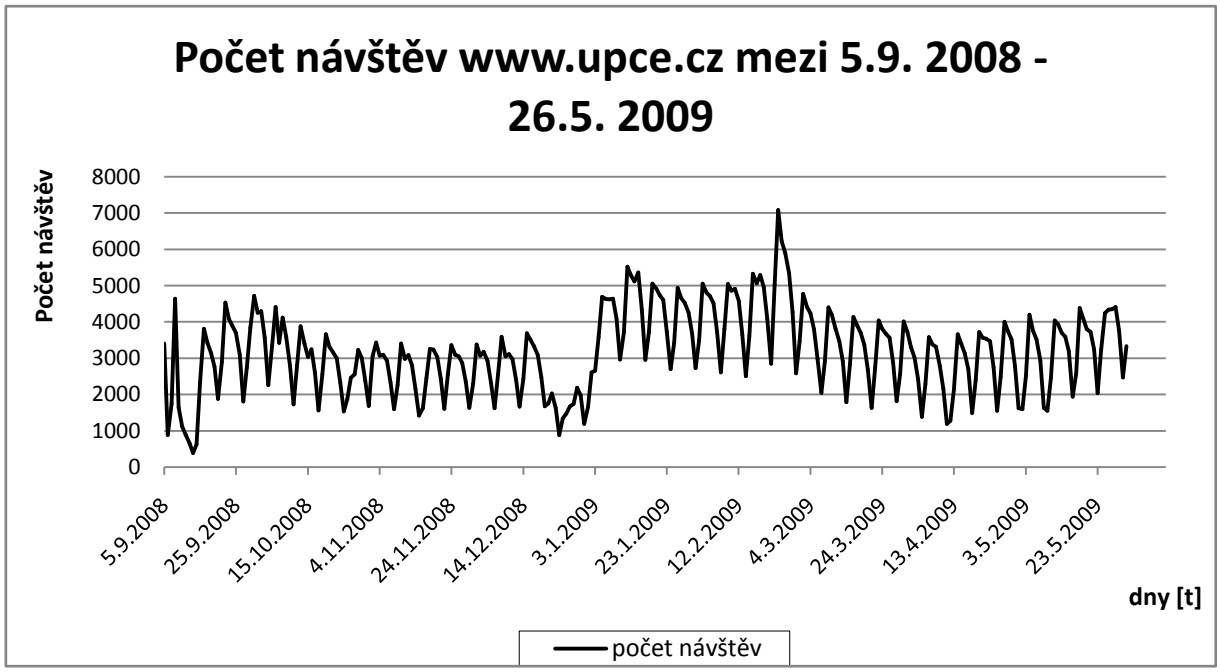
kde: $\hat{Y}_t$ je exponenciální průměr v čase t,

$\hat{Y}_{t-1}$ je exponenciální průměr v čase t-1,

$\alpha$ je vyrovnávací konstanta. Platí, že $\alpha \in \langle 0, 1\rangle$. Čím je hodnota tohoto parametru bližší k nule, tím se přiřazuje větší váha hodnotám z dávnější doby. Čím je bližší k hodnotě jedné, tím větší váha se přiřazuje hodnotám z nedávné doby [30].

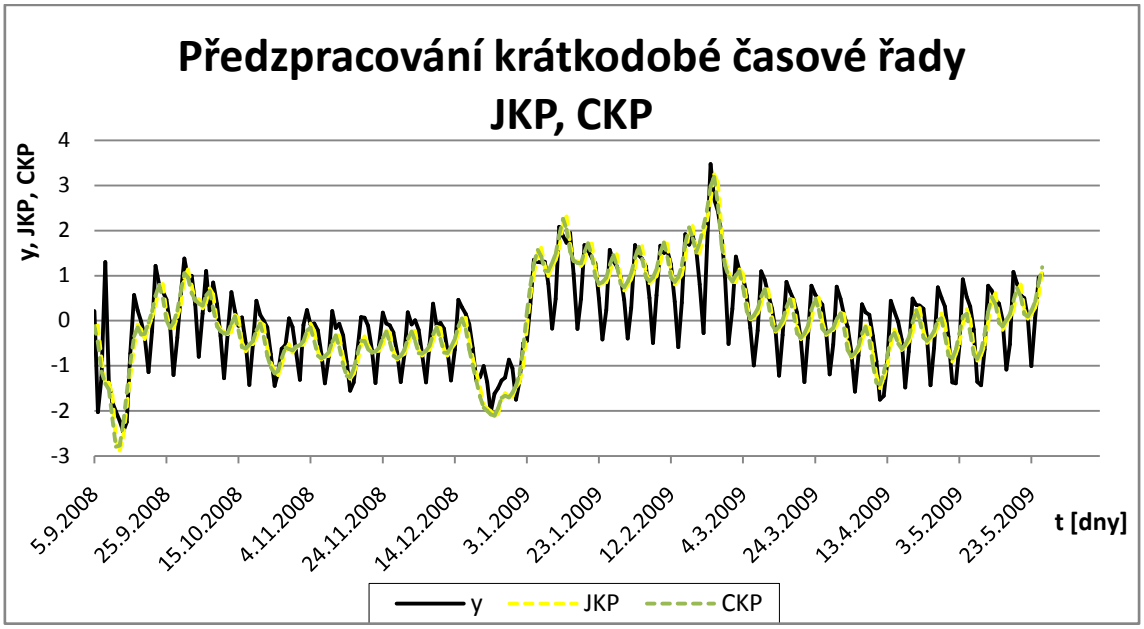
### 3.2.1 Krátkodobá časová řada

Data v této řadě tvoří návštěvy domény upce.cz v období od 5. září 2008 do 26. května 2009. Průběh návštěv je zobrazen na grafu 3.1. Z grafu je patrné, že největší návštěvnosti na www.upce.cz bylo během února 2009. Tento fakt je vysvětlován tím, že v této době se podávají přihlášky na vysoké školy a studenti se více zajímají o informace o univerzitě. Naopak, nejmenší počet návštěv byl zaznamenán na začátku měsíce září 2008, kdy jsou stále akademické prázdniny. Dále je nízká návštěvnost vykazována v období vánočních svátků.



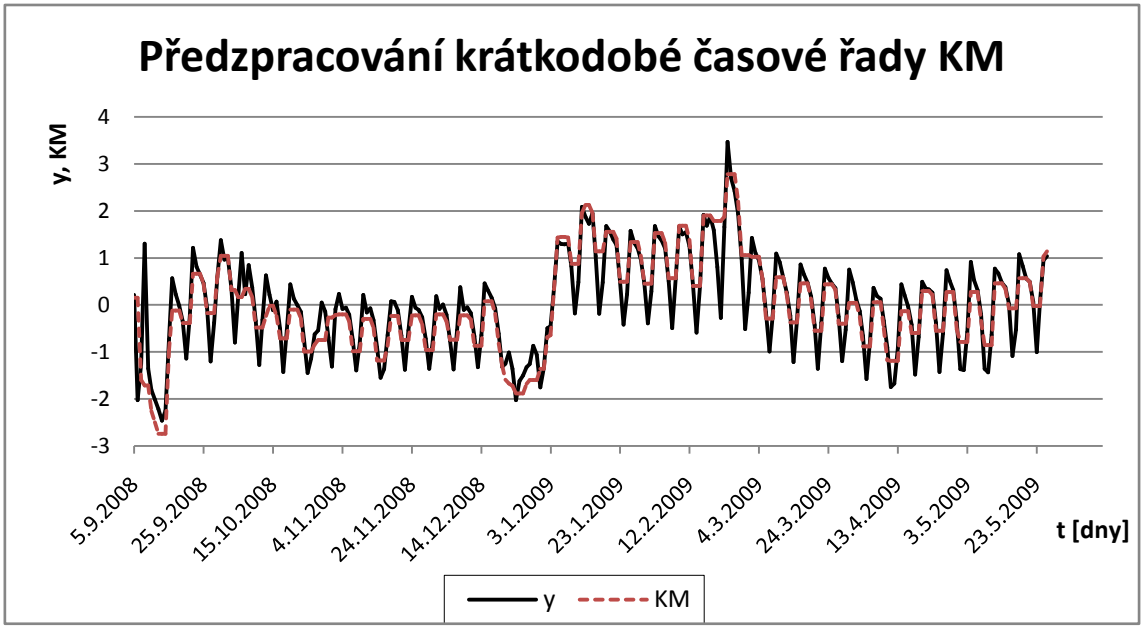

**Graf 3.1 - Krátkodobá řada: počet návštěv [vlastní]**

Předzpracování této řady spočívalo ve standardizaci dat a následné aplikaci výše zmíněných metod technické analýzy na tyto data. Na grafu 3.2 je zobrazen průběh původní standardizované časové řady *y*, jednoduchého klouzavého průměru (JKP) a centrovaného klouzavého průměru (CKP).



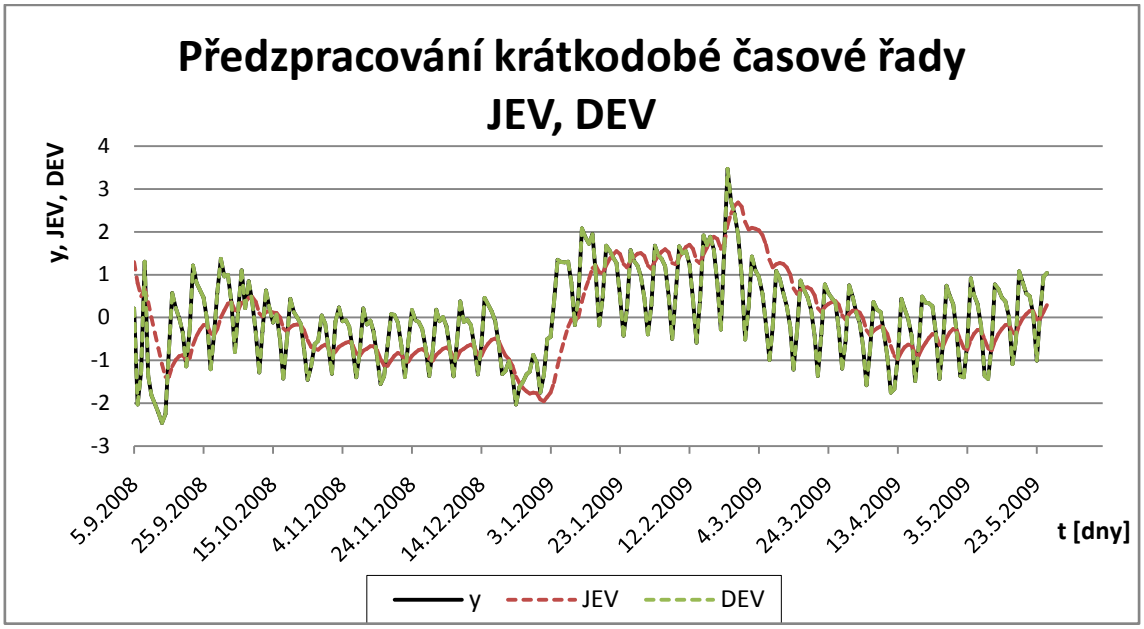

**Graf 3.2 - Předzpracování krátkodobé řady JKP, CKP [vlastní]**

Dalším indikátorem byl klouzavý medián (KM), jehož průběh ve srovnání s *y* je zobrazen na grafu 3.3.




**Graf 3.3 - Předzpracování krátkodobé časové řady KM [vlastní]**

Jako poslední indikátory byly zvoleny jednoduché exponenciální vyrovnání (JEV) a dvojité exponenciální vyrovnání (DEV), zobrazeny na grafu 3.4.

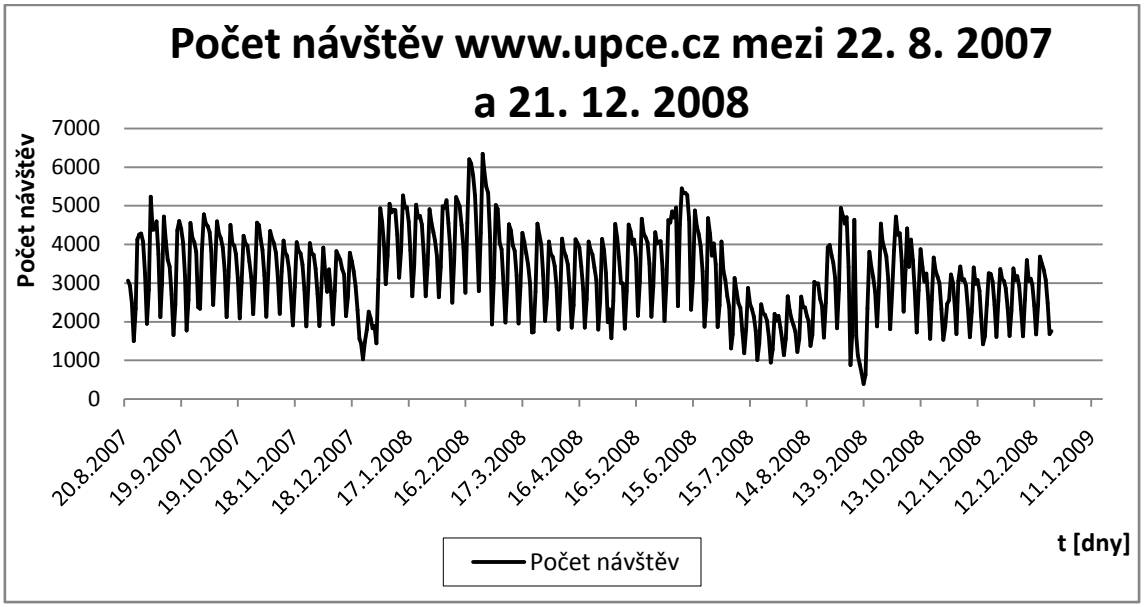



**Graf 3.4 - Předzpracování krátkodobé časové řady JEV, DEV [vlastní]**

Vypočítané hodnoty budou následně použity jako vstupní hodnoty pro predikci budoucí hodnoty v neuronové síti. Původní standardizovaná řada *y* je použita jako výstup.

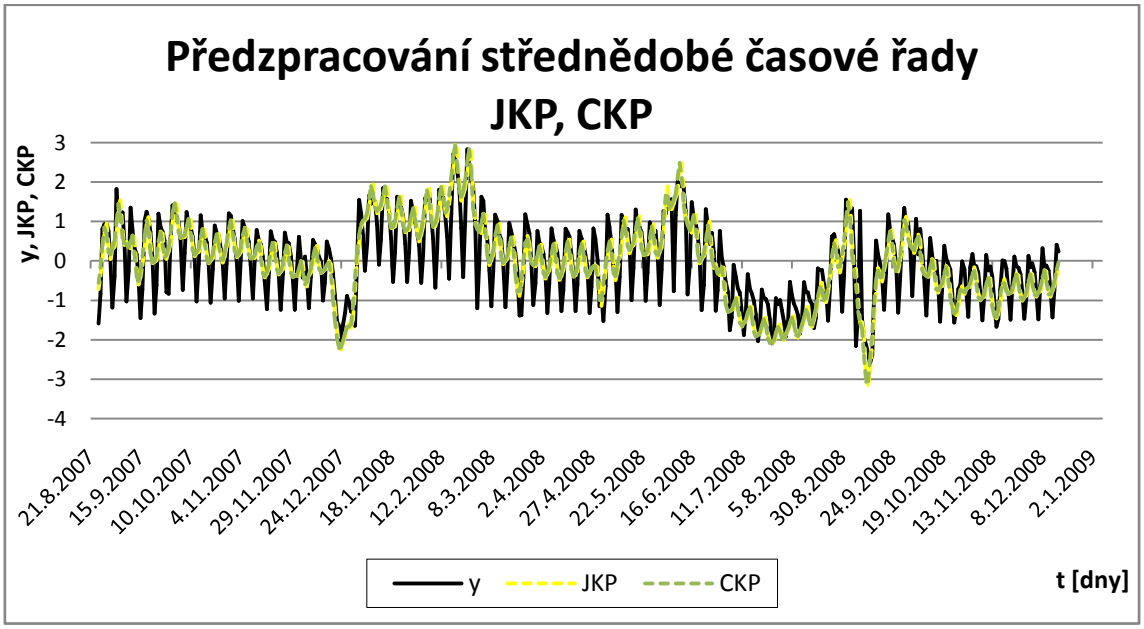
### 3.2.2 Střednědobá časová řada

Původní data pro střednědobou časovou řadu jsou v rozmezí od 22. srpna 2007 do 21. prosince 2008. Počet návštěv v této době je zobrazen na grafu 3.5. Na první pohled je z průběhu grafu patrné, že největší návštěvnost na www.upce.cz je opět během února. Naopak nejmenší návštěvnost je během vánočních svátků a v období akademických prázdnin.

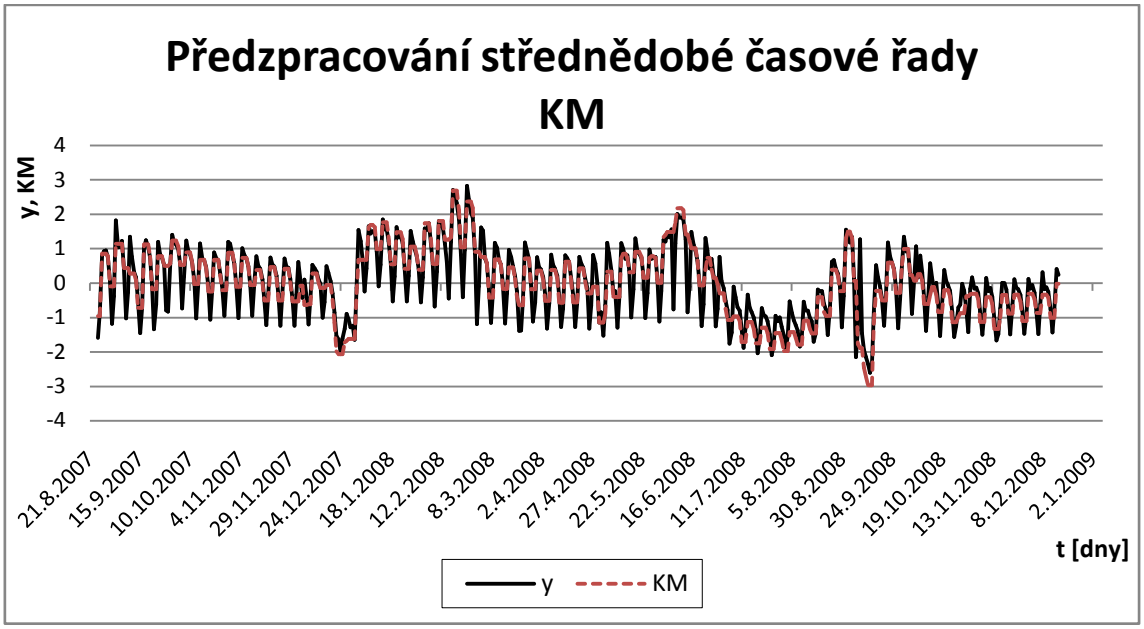



**Graf 3.5 - Střednědobá řada: počet návštěv [vlastní]**

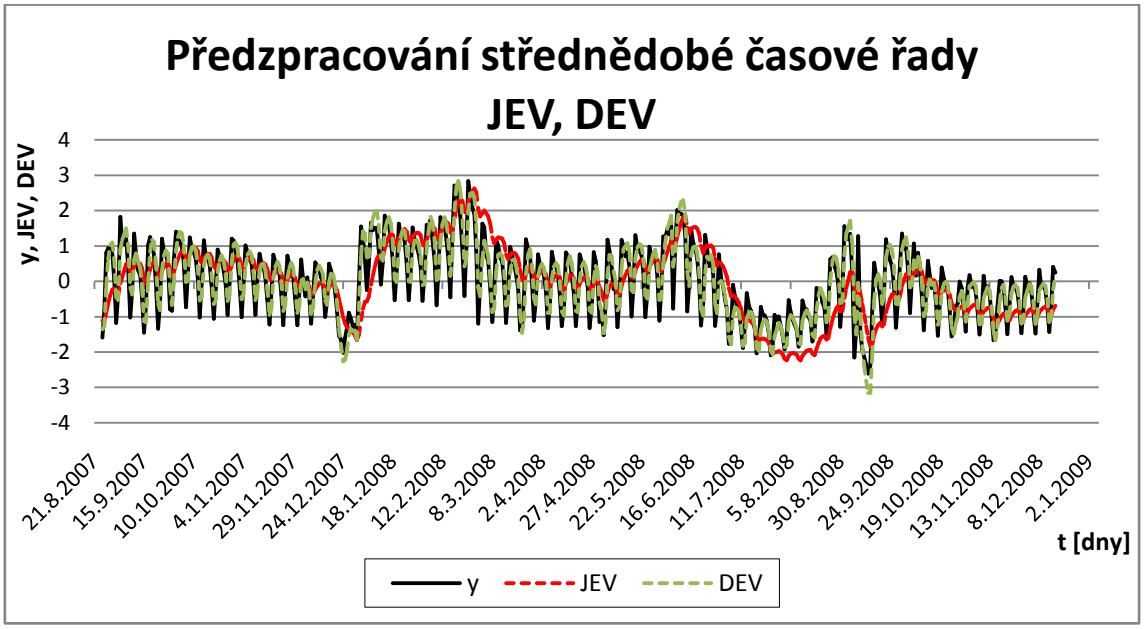
Stejně tak jako krátkodobá časová řada, i tato byla standardizována a dále na ni byly použity indikátory technické analýzy, které byly následně vstupy do predikčního modelu neuronové sítě. Průběh jednotlivých indikátorů ve srovnání s původními standardizovanými daty je zobrazen na grafech 3.6 až 3.8.




**Graf 3.6 - Předzpracování střednědobé časové řady JKP, CKP [vlastní]**




**Graf 3.7 - Předzpracování střednědobé časové řady KM [vlastní]**



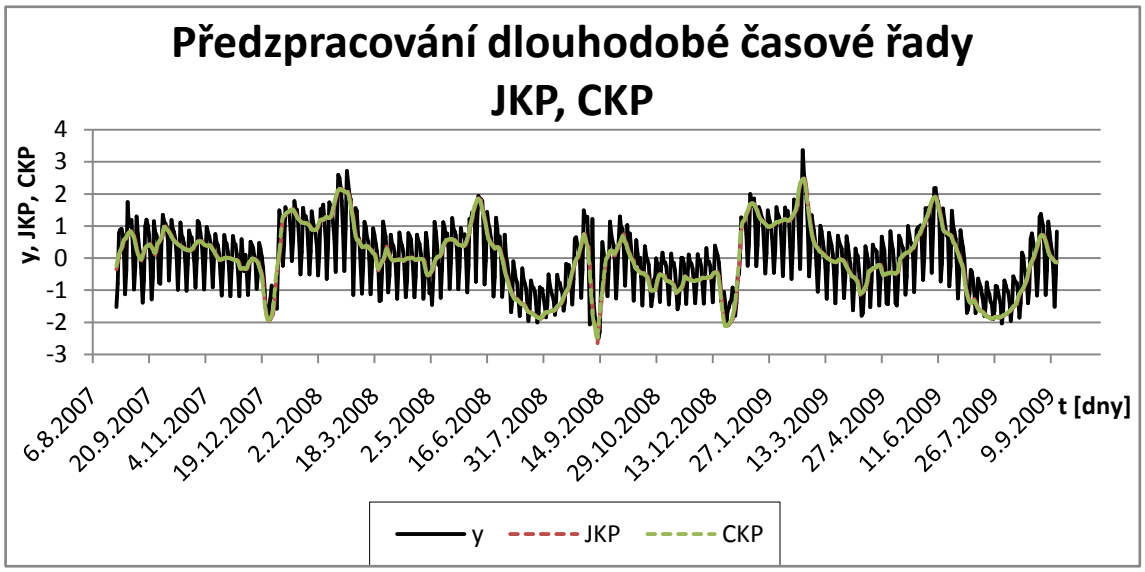

**Graf 3.8 - Předzpracování střednědobé časové řady JKP, CKP [vlastní]**

### 3.2.3 Dlouhodobá časová řada

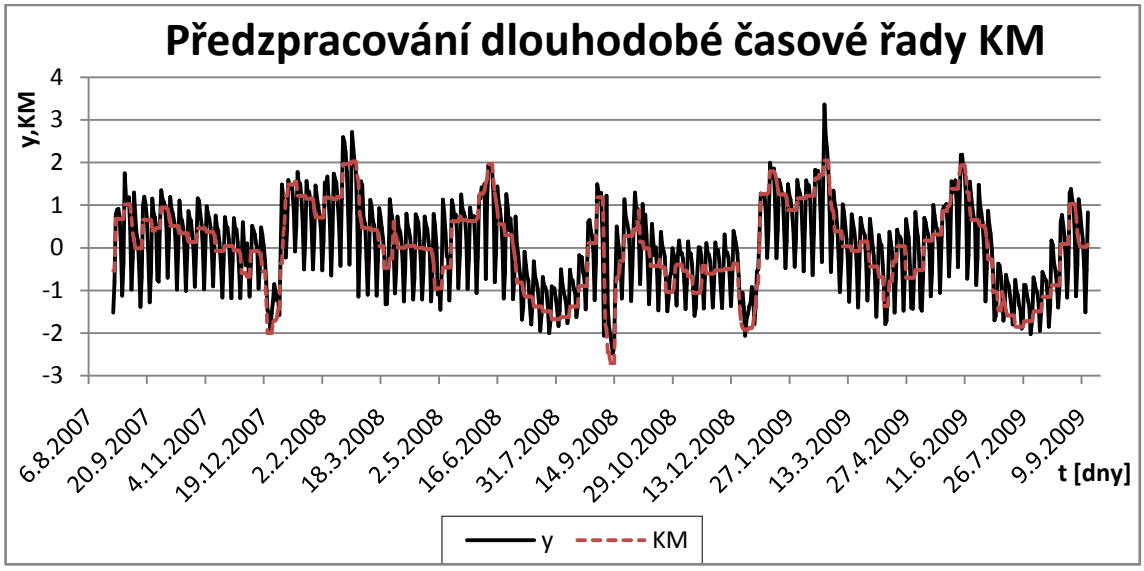
Dlouhodobá časová řada pozorování návštěvnosti domény upce.cz nabývá hodnot pozorování od 21. 8. 2007 do 20. 9. 2009. I na této časové řadě, která je vyobrazena na grafu 3.9, jsou patrné jisté výkyvy během období vánočních svátků, akademických prázdnin nebo času podávání přihlášek na vysoké školy.


**Graf 3.9 - Dlouhodobá řada: počet návštěv [vlastní]**

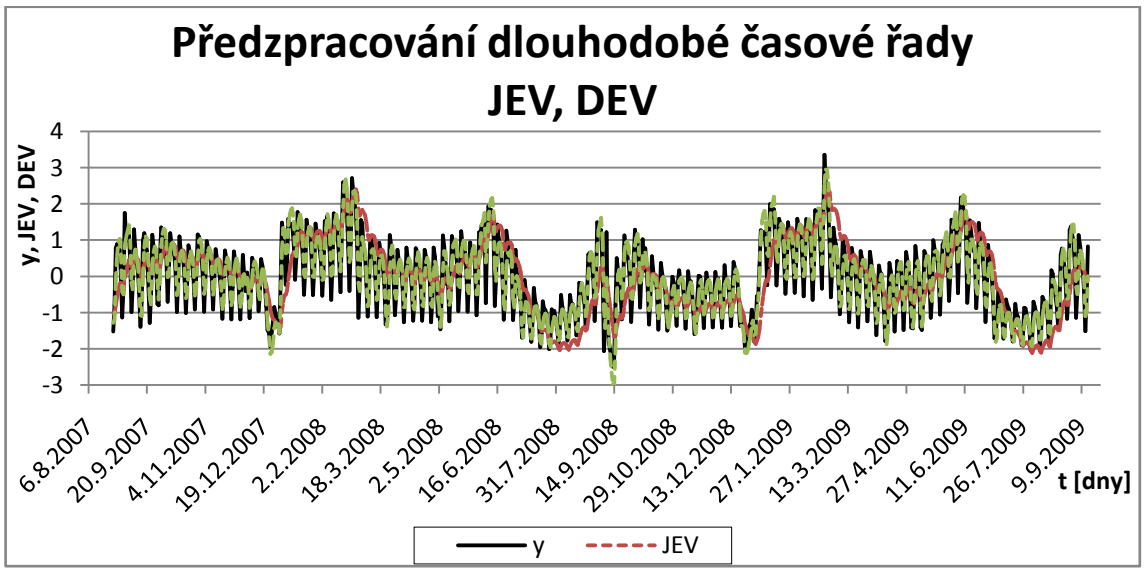
Průběh indikátorů technické analýzy je opět zobrazen na grafech 3.10 až 3.12.




**Graf 3.10 - Předzpracování dlouhodobé časové řady JKP, CKP [vlastní]**



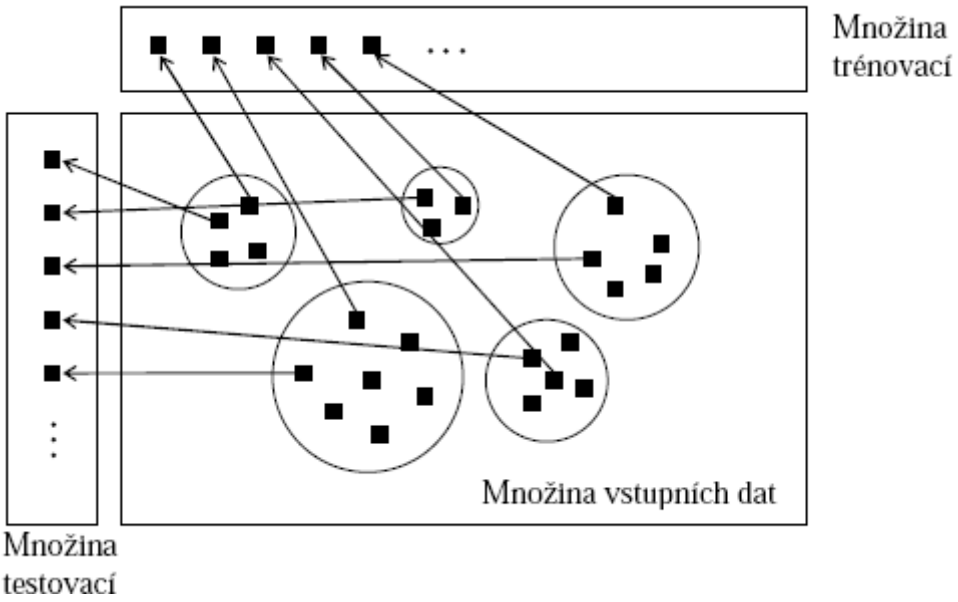

**Graf 3.11 - Předzpracování dlouhodobé časové řady KM [vlastní]**




**Graf 3.12 - Předzpracování dlouhodobé časové řady JEV, DEV [vlastní]**

### 3.2.4 Rozdělení dat na trénovací a testovací množiny

Rozdělení dat na trénovací a testovací množinu lze realizovat více způsoby. Rozdělení může proběhnout náhodným výběrem, výběrem každého n-tého prvku, výběrem na základě shlukové analýzy aj. Ať už se jedná o jakýkoliv způsob rozkladu, je vždy potřeba, aby testovací množina reprezentovala veškerá data, která jsou obsažena v trénovací množině [31]. Schematické rozdělení dat na trénovací a testovací množinu je zobrazeno na obr. 3.2.




**Obr. 3.2 - Rozdělení dat na trénovací a testovací množinu [31]**

Rozdělení dat pro účely této práce bylo realizováno v programovém prostředí SPSS Clementine 10.1, kde se proces rozdělení na trénovací a testovací množinu provádí přes uzel *Partition*. Zde je možné navolit procentuální poměr rozdělení dat. Rozdělení dat je realizováno náhodným výběrem, a pro každou řadu bylo voleno rozložení dat v poměrech trénovací ku testovací množině: 50:50, 60:40, 70:30, 80:20, 90:10.

## *3.3 Návrh struktury RBF sítě*

V této podkapitole budou dále popsány jednotlivé kroky, spojené s predikcí návštěvnosti web domény upce.cz, které byly realizovány v softwarovém prostředí SPSS Clementine 10.1. Výstupy byly dále zpracovány v MS Excel.

Nastavení parametrů RBF neuronové sítě v prostředí Clementine zahrnovalo tyto parametry:

- Počet cyklů, po kterých má být proces zastaven.
- Počet skrytých neuronů (*RBF clusters*).
- Persistence.
- Eta.
- Alpha.
- Překrytí RBF neuronů (*RBF overlapping*, v práci značeno jako *v*).

**Počet cyklů** je u všech rozdělení časových řad stanoven pevně na 600 cyklů.

**Počet skrytých RBF neuronů** se liší pro krátkodobou, střednědobou a dlouhodobou časovou řadu. Tab. 3.3 obsahuje počet skrytých neuronů pro jednotlivou časovou řadu.

**Tab. 3.3 - Počet RBF neuronů ve skryté vrstvě [32]**

| Typ řady | Počet skrytých RBF neuronů |
|---|---|
| Krátkodobá | 80 |
| Střednědobá | 120 |
| Dlouhodobá | 25 |

**Parametr *Persistence*** určuje počet cyklů, po které síť dále pracuje i když nedochází k žádným změnám. Tento parametr je nastaven na hodnotu 30 cyklů.

**Parametr *Eta*** je zde konstantou, která znamená rychlost učení neuronové sítě. V prostředí Clementine je možnost jejího automatického výpočtu, který se v průběhu učení sítě mění. Na začátku je *počáteční hodnota Ety* (*Initial Eta*), jež během učení klesne k hodnotě *Nízká Eta* (*Low Eta*), poté je resetována do hodnoty *Vysoká Eta* (*High Eta*) a opět klesá k hodnotě *Nízká Eta*. Poslední dva kroky se opakují, dokud není trénování sítě ukončeno.

**Parametr *alpha*** představuje momentum, které se používá v aktualizaci vah během procesu učení. Jeho hodnota je v rozmezí <0;1>.

**Parametr *Překrytí RBF neuronů (v)*** se týká skrytých RBF neuronů. Tyto neurony reprezentují radiální bazické funkce, které definují shluky nebo oblasti dat. Tento parametr umožňuje nastavit, jak moc se tyto shluky nebo oblasti překrývají.

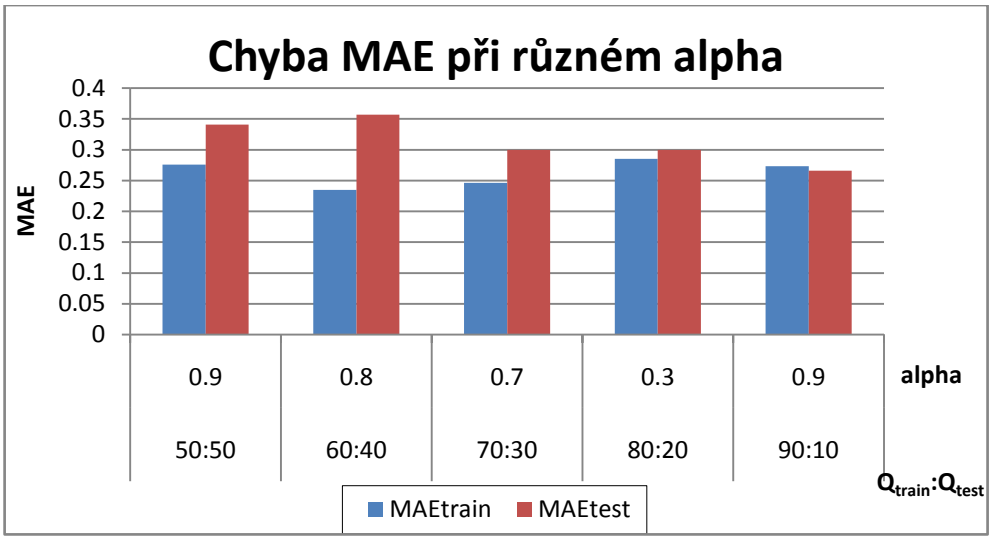
Jako vstupní hodnoty do neuronové sítě jsou vždy voleny indikátory technické analýzy popsané výše. Jedná se tedy o pět vstupních proměnných: JKP, CKP, KM, JEV, DEV. Naměřená standardizovaná hodnota *y* je použita jako výstupní hodnota. Jedná se tedy o učení s učitelem. Cílem navržení struktury RBF sítě pro predikci návštěvnosti domény upce.cz je nalézt takovou strukturu neuronové sítě, při které bude dosaženo nejmenší chyby. V prostředí Clementine je počítáno se střední absolutní chybou MAE (*Mean Absolute Error*).

### 3.3.1 Krátkodobá časová řada

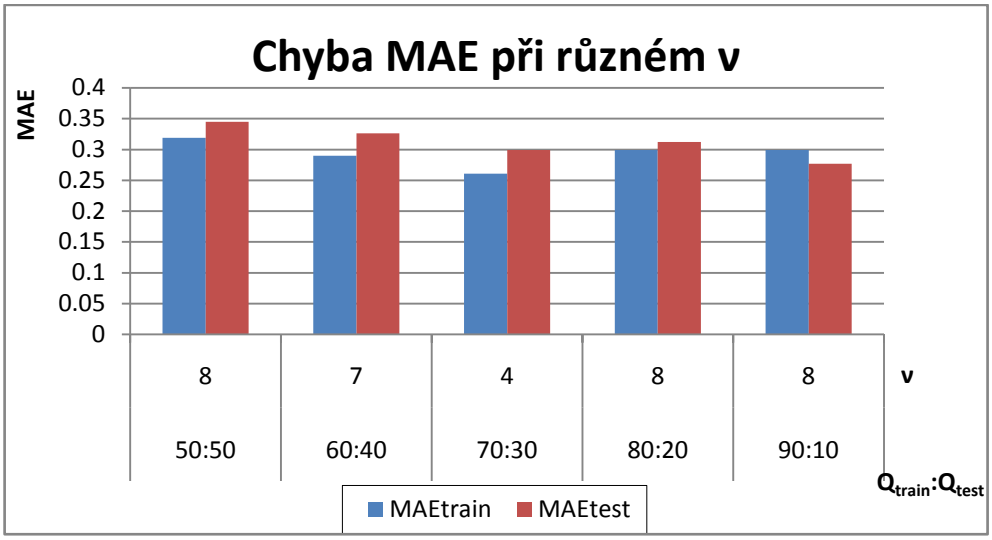
Jak již bylo zmíněno dříve, data byla rozdělena v různých poměrech na množinu trénovací a testovací. Základní statistické charakteristiky rozdělení dat ve všech poměrech trénovací a testovací množiny jsou v příloze 1 a 2 tohoto dokumentu. Po rozdělení dat následovalo učení neuronové sítě. Nejprve bylo třeba nalézt hodnotu parametru *alpha*, při které bude

dosaženo nejmenší chyby na testovacích datech. Proto byly voleny hodnoty *alpha* od 0.1 do 1 a ostatní proměnné byly konstantní. Na základě nejmenší chyby, dosažené na testovacích datech, byla dále zvolena hodnota parametru *alpha* pro hledání takové velikosti parametru *v*, při které bude síť vykazovat nejmenší chybu na testovacích datech. Parametr *v* byl volen v rozmezí hodnot od 1 do 9, ostatní proměnné byly konstantní.

Porovnání jednotlivých nalezených hodnot, při kterých bylo dosaženo nejmenší chyby na testovacích datech pro každé rozdělení dat, je zobrazeno na grafu 3.13. Na grafu 3.14 je potom zobrazeno jaké chyby bylo dosaženo při nejlepších výsledcích hledání parametru *v* v jednotlivých rozděleních. Průběhy chyb při všech parametrech *alpha* jsou v příloze 3 a pro parametr *v* jsou příloze 4.




**Graf 3.13 - Chyba MAE při různém *alpha* v jednotlivých rozdělení dat krátkodobé ČŘ [vlastní]**

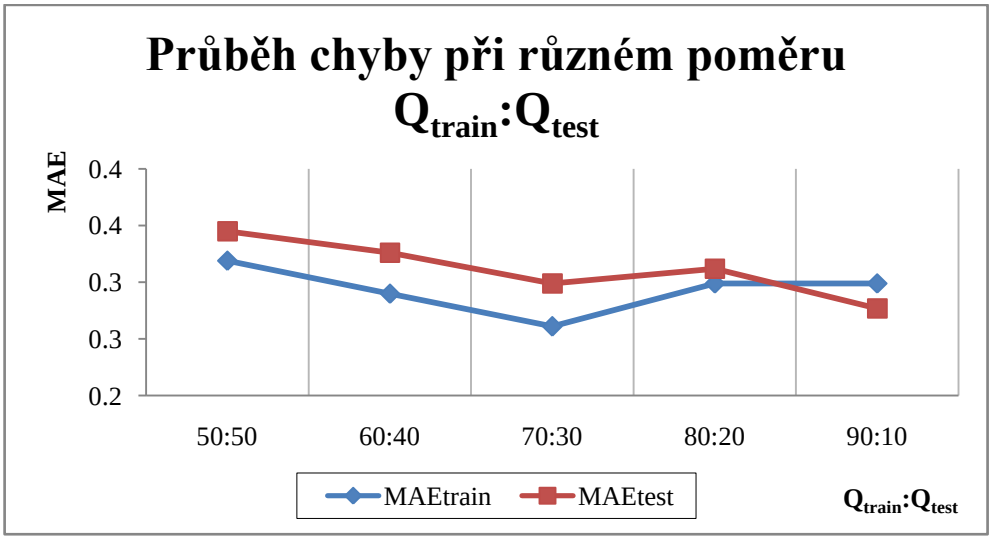



**Graf 3.14 - Chyba MAE při různém *v* v jednotlivých rozděleních dat krátkodobé ČŘ [vlastní]**

Takto nastavené hodnoty vykazovaly různé chyby, které jsou zobrazeny v tab. 3.4 a jejich průběh v grafu 3.15. S rostoucím počtem trénovacích dat nejprve klesala chyba na testovacích datech, která dosáhla lokálního minima při rozdělení dat v poměru 70:30 a globálního minima v případě rozdělení dat 90:10.

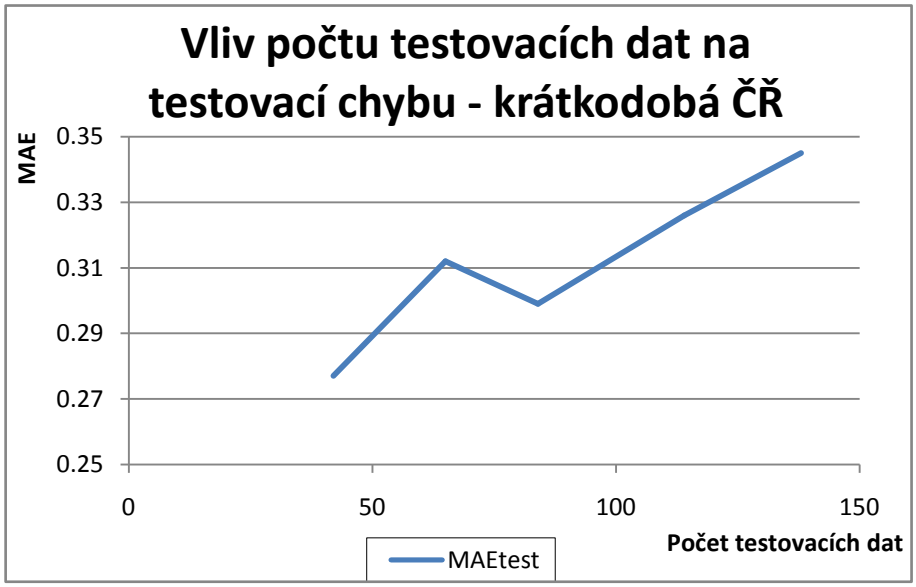
**Tab. 3.4 - Nejmenší dosažené chyby pro jednotlivá rozdělení dat pro krátkodobou ČŘ [vlastní]**

| počet neuronů | alpha | v | počet cyklů | persistence | Qtrain:Qtest | MAEtrain | MAEtest |
|---|---|---|---|---|---|---|---|
| 80 | 0.9 | 8 | 600 | 30 | 50:50 | 0.319 | 0.345 |
| 80 | 0.8 | 7 | 600 | 30 | 60:40 | 0.29 | 0.326 |
| 80 | 0.7 | 4 | 600 | 30 | 70:30 | 0.261 | 0.299 |
| 80 | 0.3 | 8 | 600 | 30 | 80:20 | 0.299 | 0.312 |
| 80 | 0.9 | 8 | 600 | 30 | 90:10 | 0.299 | 0.277 |




**Graf 3.15 - Průběh nejmenších testovacích chyb jednotlivých rozdělení dat krátkodobé ČŘ [vlastní]**

Na grafu 3.16 je zobrazen vliv počtu testovacích dat na průběh testovací chyby. Z grafu 3.16 je jasně vidět, že s rostoucím počtem dat testovací chyba nejprve stoupala, následovně začala klesat (což odpovídá předešlému grafu 3.15 a rozložení dat v poměru 70:30). Následně chyba opět rostla. Nejvíce testovacích dat bylo dat síti k dispozici při rozdělení 50:50.



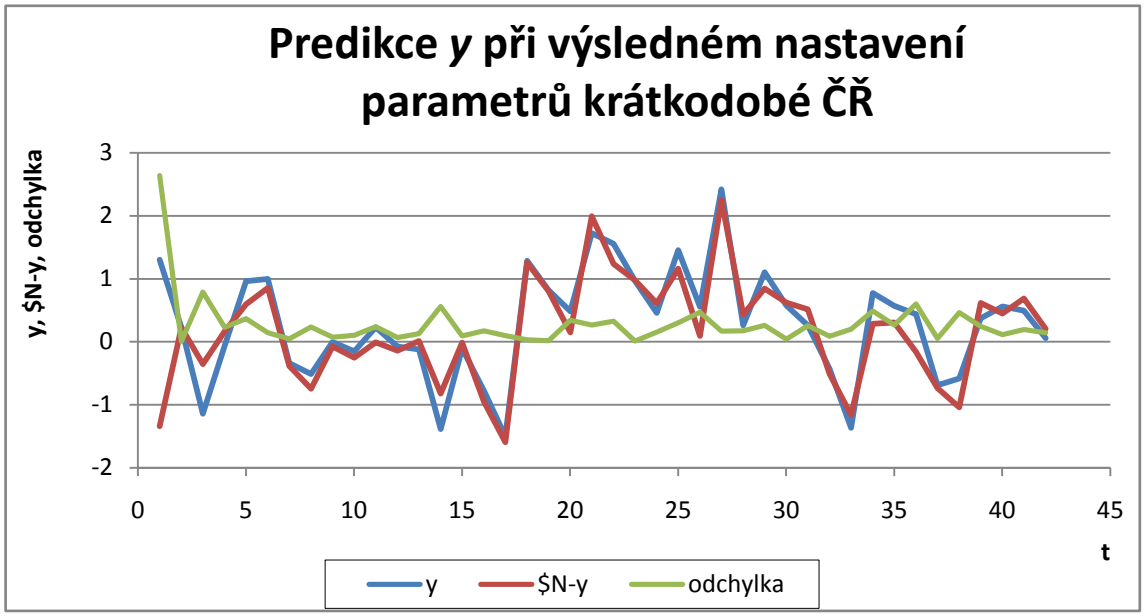

**3.16 - Vliv počtu testovacích dat na testovací chybu - krátkodobá ČŘ [vlastní]**

Výsledným nastavením parametrů *alpha* a *v* a poměru trénovací ku testovací množině dat, kdy bylo dosaženo nejmenší MAE chyby, je shrnuto v tab. 3.5.

**Tab. 3.5 - Výsledné nastavení parametrů pro krátkodobou ČŘ [vlastní]**

| počet neuronů | alpha | ν | počet cyklů | persistence | Qtrain:Qtest | MAEtrain | MAEtest |
|---|---|---|---|---|---|---|---|
| 80 | 0.9 | 8 | 600 | 30 | 90:10 | 0.299 | 0.277 |

V grafu 3.17 jsou zobrazeny měřené původní hodnoty *y*, predikované hodnoty *$N-y* a odchylka mezi těmito hodnotami. Tyto hodnoty se týkají výsledného nastavení parametrů neuronové sítě, která je uvedena v tab. 3.5.



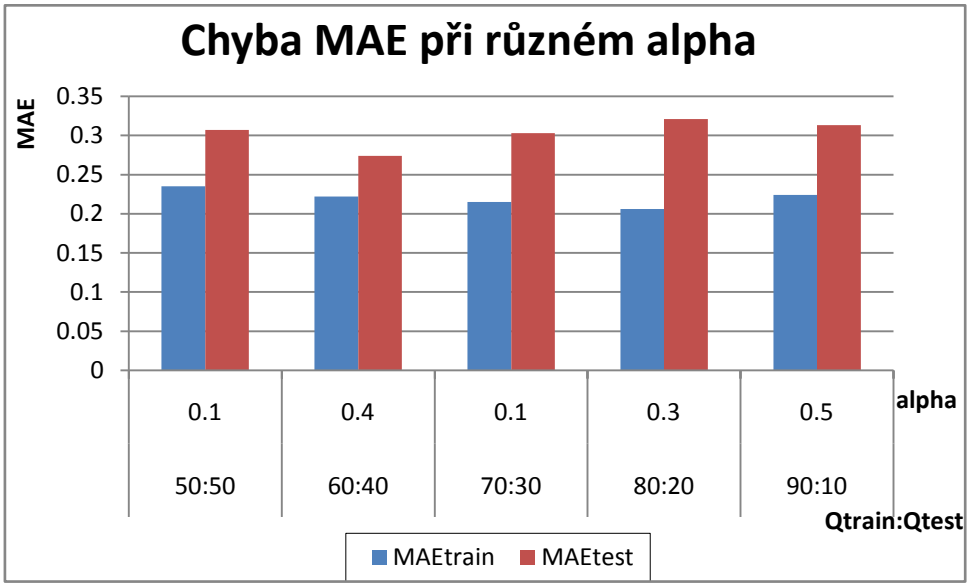

**Graf 3.17 - Predikce *y* při výsledném nastavení parametrů krátkodobé ČŘ [vlastní]**

### 3.3.2 Střednědobá časová řada

V případě střednědobé časové řady návštěvnosti domény upce.cz bylo k dispozici pro naučení sítě více dat než v případě krátkodobé časové řady. Statistiky rozdělení dat pro tuto řadu a jednotlivé výsledky průběhu chyby na testovacích datech při změně jednotlivých parametrů a rozdělení dat jsou kompletně zobrazeny v přílohách tohoto dokumentu.

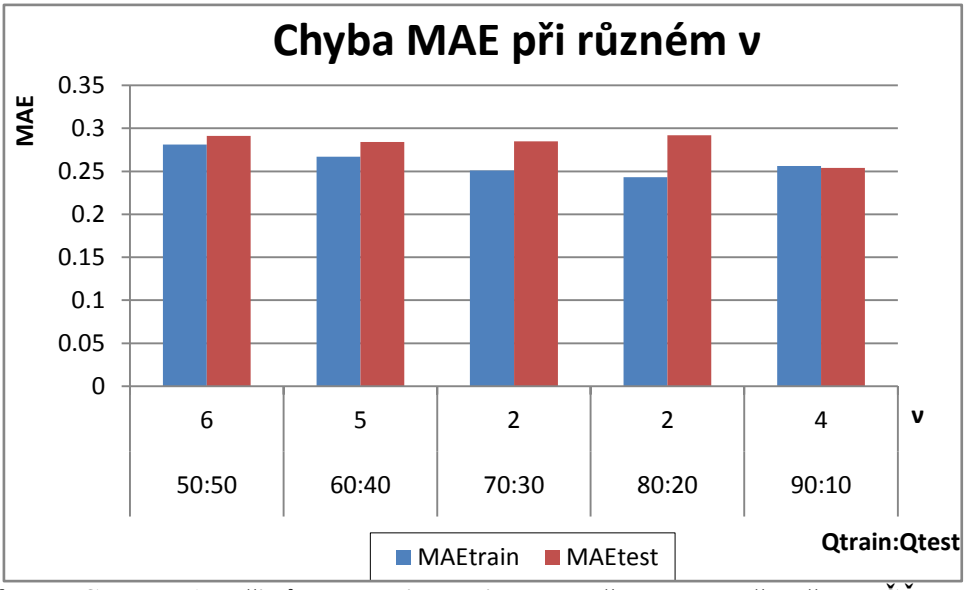
Co se týče parametru *alpha*, opět byl měněn v hodnotách od 0.1 do 1. Nejlepších výsledků, tj. nejmenší chyby na testovacích datech, kterých bylo dosaženo v jednotlivých rozděleních dat, je zobrazeno na grafu 3.18. V příloze 8 jsou k dispozici všechny tabulky a grafy, které zobrazují jednotlivé průběhy chyb při hledání parametru *alpha*. Z jednotlivých průběhů vlivu velikosti parametru *alpha* na testovací chybu je jasně vidět, že chyba s rostoucí velikostí tohoto parametru kolísavě rostla a klesala. V případě nejvyšší hodnoty parametru *alpha* byla

chyba na testovacích datech nejvyšší ve všech případech rozdělení dat na trénovací a testovací množiny.




**Graf 3.18- Chyba MAE při různém *alpha* v jednotlivých rozděleních dat střednědobé ČŘ [vlastní]**

Na základě získaných hodnot parametru *alpha* nastalo opět hledání velikosti parametru *v*. V případě této časové řady bylo zjištěno, že s rostoucí hodnotou parametru *v* rapidně rostla chyba na testovacích datech. V některých případech ani tato chyba nebyla ve výsledcích uvedena. Podle průběhu předchozích hodnot chyby se předpokládá, že tato chyba byla velmi vysoká. Průběhy chyby ve všech rozděleních pro jednotlivé hodnoty parametru *v* jsou v příloze 9. Na grafu 3.19 je již zobrazena taková hodnota parametru *v*, při které bylo dosaženo nejmenší chyby na testovacích datech v jednotlivých rozděleních.



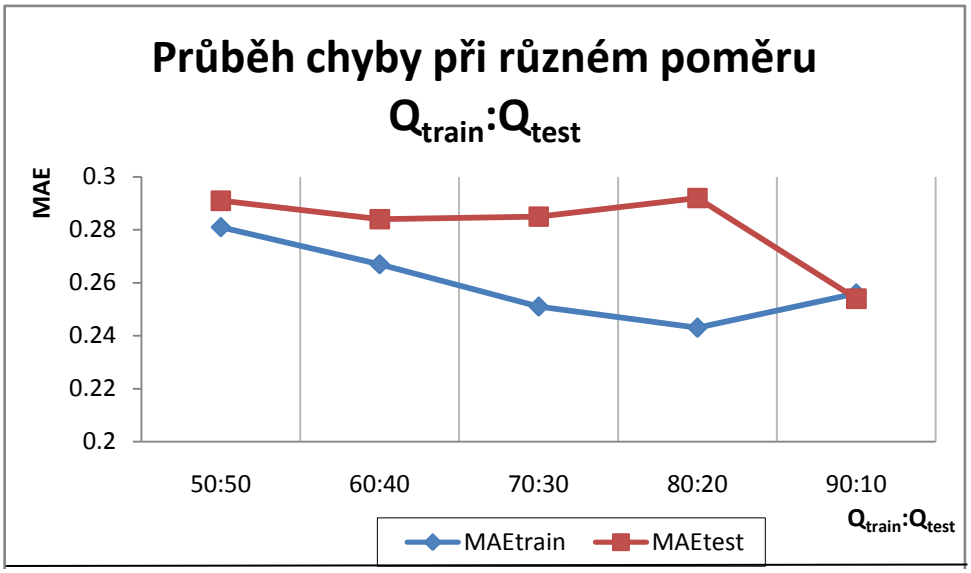

**Graf 3.19 - Chyba MAE při různém *v* v jednotlivých rozděleních dat střednědobé ČŘ [vlastní]**

V tab. 3.6 jsou zobrazeny nejlepší výsledky, kterých bylo dosaženo v rámci jednotlivých rozdělení dat na trénovací a testovací množinu pro střednědobou časovou řadu návštěvnosti domény upce.cz.

**Tab. 3.6 - Nejmenší dosažené chyby pro jednotlivá rozdělení dat pro střednědobou ČŘ [vlastní]**

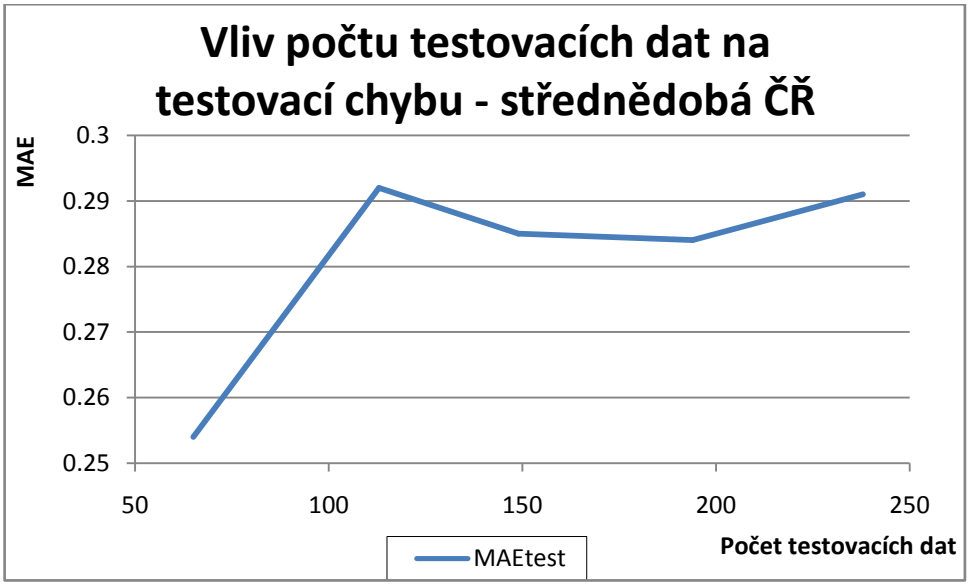
| počet neuronů | alpha | v | počet cyklů | persistence | $Q_{train}$:$Q_{test}$ | $MAE_{train}$ | $MAE_{test}$ |
|---|---|---|---|---|---|---|---|
| 125 | 0.1 | 6 | 600 | 30 | 50:50 | 0.281 | 0.291 |
| 125 | 0.4 | 5 | 600 | 30 | 60:40 | 0.267 | 0.284 |
| 125 | 0.1 | 2 | 600 | 30 | 70:30 | 0.251 | 0.285 |
| 125 | 0.3 | 2 | 600 | 30 | 80:20 | 0.243 | 0.292 |
| 125 | 0.5 | 4 | 600 | 30 | 90:10 | 0.256 | 0.254 |

Tab. 3.6 je pro lepší zobrazení průběhu chyby převedena do grafu 3.20, ze kterého je nejlépe vidět, jak chyba na testovacích datech nejprve mírně klesala až opět do rozdělení dat v poměru 70:30, poté začala růst a v rozdělení dat v poměru 90:10 klesla do svého minima.




**Graf 3.20 - Průběh nejmenších testovacích chyb jednotlivých rozdělení dat střednědobé ČŘ [vlastní]**

Stejně jako v případě krátkodobé časové řady i zde platí, že čím méně testovacích dat má síť k dispozici, tím menší je testovací chyba, což je patrné z grafu 3.21.



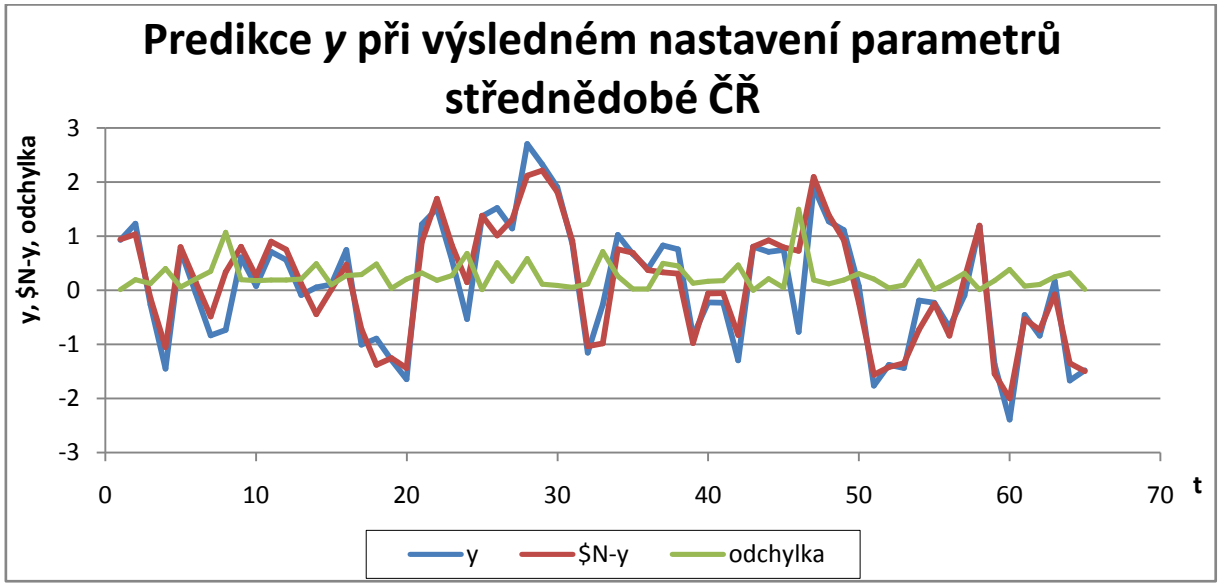

**Graf 3.21 - Vliv počtu testovacích dat na testovací chybu - střednědobá ČŘ [vlastní]**

Výsledné nastavené parametrů, při kterých bylo dosaženo nejmenší chyby na testovacích datech, je zobrazeno v tab. 3.7.

**Tab. 3.7 - Výsledné nastavení parametrů pro střednědobou ČŘ [vlastní]**

| počet neuronů | alpha | ν | počet cyklů | persistence | $Q_{train}$:$Q_{test}$ | $MAE_{train}$ | $MAE_{test}$ |
|---|---|---|---|---|---|---|---|
| 125 | 0.5 | 4 | 600 | 30 | 90:10 | 0.256 | 0.254 |

Na grafu 3.22 jsou zobrazeny původní hodnoty časové řady *y*, predikované hodnoty *$N-y* a odchylka mezi těmito hodnotami při výsledném nastavení parametrů pro střednědobou časovou řadu. V porovnání s předchozí krátkodobou časovou řadou je zde patrnější menší odchylka mezi původními a predikovanými daty. Střednědobá časová řada obsahovala více dat než krátkodobá, což dopomohlo k lepšímu naučení sítě.



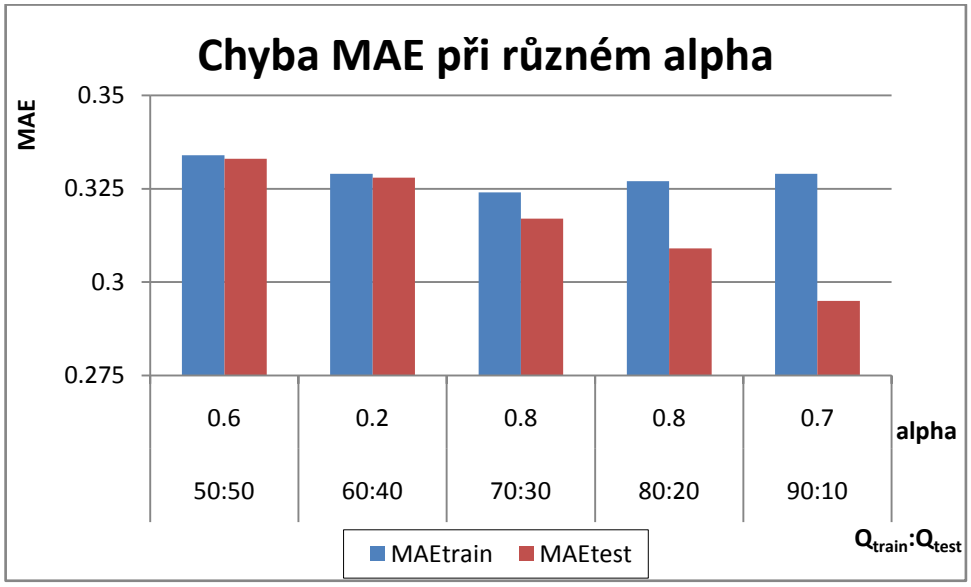

**3.22 - Predikce y při výsledném nastavení parametrů střednědobé ČŘ [vlastní]**

### 3.3.3 Dlouhodobá časová řada

Pro dlouhodobou časovou řadu návštěvnosti domény upce.cz bylo použito nejvíce dat. Učení neuronové sítě probíhalo stejným způsobem jako v případě dvou předcházejících řad. V případě hledání parametru *alpha* učení sítě vykazovalo podobné rysy jako v předcházející časové řadě. Největší chyby na testovacích datech bylo dosaženo v největší možné hodnotě tohoto parametru, tedy v hodnotě 1. Průběhy jednotlivých hodnot jsou součástí přílohy 13.

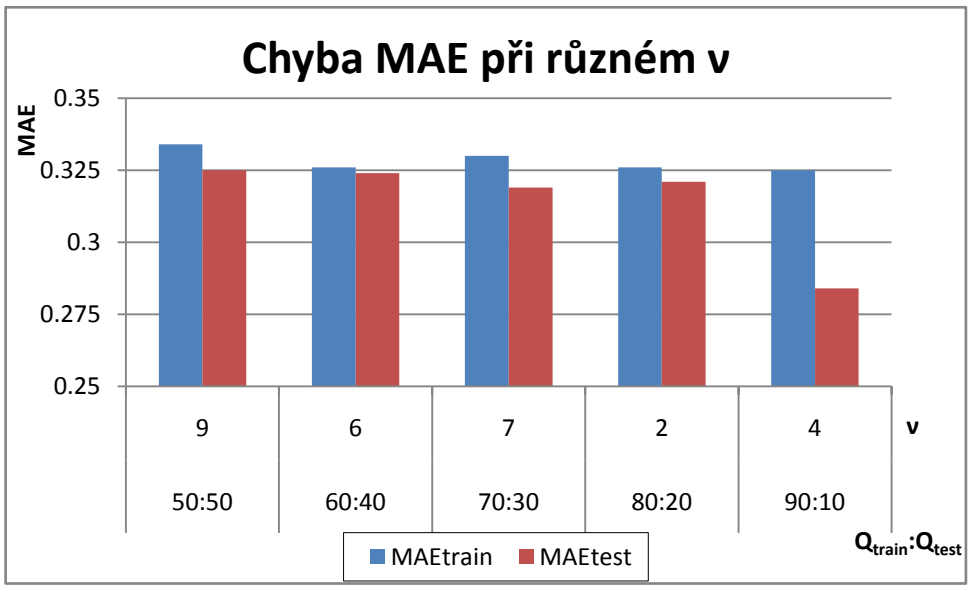
Na grafu 3.23 je opět přehled nejmenších dosažených chyb na trénovacích a testovacích datech v jednotlivých rozděleních dat při hledání nastavení parametru *alpha*. Významným

rozdílem oproti předcházejícím časovým řadám je zde situace, kdy ve všech případech chyba na testovacích datech byla nižší než chyba na trénovacích datech.




**Graf 3.23 - Chyba MAE při různém *alpha* jednotlivých rozděleních dlouhodobé ČŘ [vlastní]**

Pro každé rozdělení dat bylo dále hledáno nastavení parametru *v*, kde bude vykazována nejmenší chyba. Průběhy těchto chyb jsou součástí přílohy 14. Na grafu 3.24 jsou zobrazené nejmenší chyby na trénovacích a testovacích datech při různém nastavení parametru v v jednotlivých rozděleních. Opět zde testovací chyba byla ve všech případech nižší než chyba na trénovacích datech.



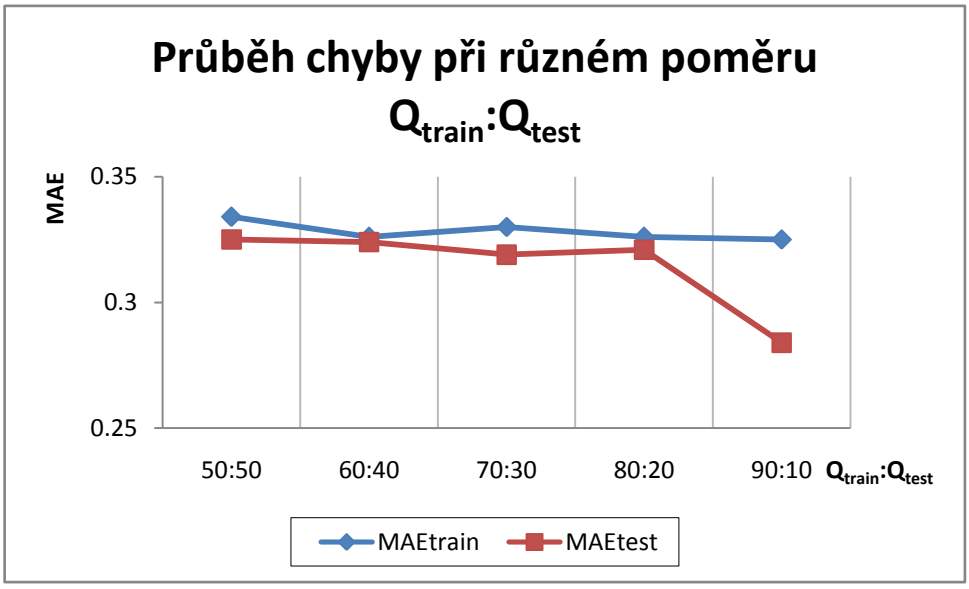

**Graf 3.24 - Chyba MAE při různém *v* v jednotlivých rozděleních dat dlouhodobé ČŘ [vlastní]**

Shrnutí výše popsaných poznatků obsahuje tab. 3.8, která zobrazuje jednotlivé nastavení parametrů pro rozdělení dat s nejmenší chybou na testovacích datech.

**Tab. 3.8 - Nejmenší dosažené chyby pro jednotlivá rozdělení dat pro dlouhodobou ČŘ [vlastní]**

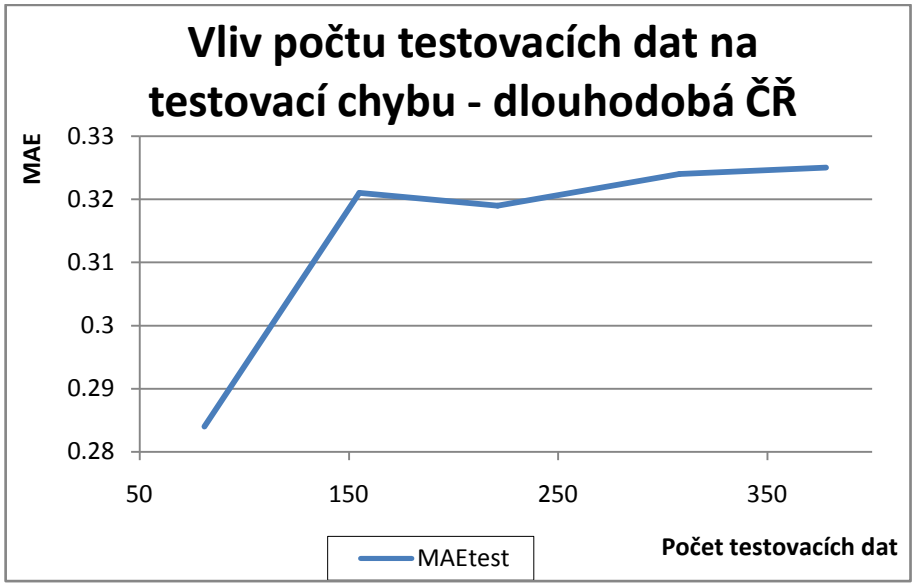
| počet neuronů | alpha | ν | Počet cyklů | Persistence | Qtrain:Qtest | MAEtrain | MAEtest |
|---|---|---|---|---|---|---|---|
| 25 | 0.6 | 9 | 600 | 30 | 50:50 | 0.334 | 0.325 |
| 25 | 0.2 | 6 | 600 | 30 | 60:40 | 0.326 | 0.324 |
| 25 | 0.8 | 7 | 600 | 30 | 70:30 | 0.33 | 0.319 |
| 25 | 0.8 | 2 | 600 | 30 | 80:20 | 0.326 | 0.321 |
| 25 | 0.7 | 4 | 600 | 30 | 90:10 | 0.325 | 0.284 |

Tyto výsledky jsou opět převedeny do grafické podoby, kde je opět nejlépe vidět, že testovací chyba nejprve klesala až do rozdělení dat v poměru 70:30, poté mírně stoupla a nejnižší byla v rozdělení dat v poměru 90:10.




**Graf 3.25 - Průběh nejmenších testovacích chyb jednotlivých rozdělení dlouhodobé ČŘ [vlastní]**

Tento průběh je možné vysvětlit tím, že v případě výsledného rozdělení dat v poměru 90:10 bylo k testování nejméně hodnot, což zobrazuje i graf 3.26, kde je zobrazen vliv počtu testovacích dat na testovací chybu. S počtem těchto dat chyba, stejně jako v ostatních případech, rostla.



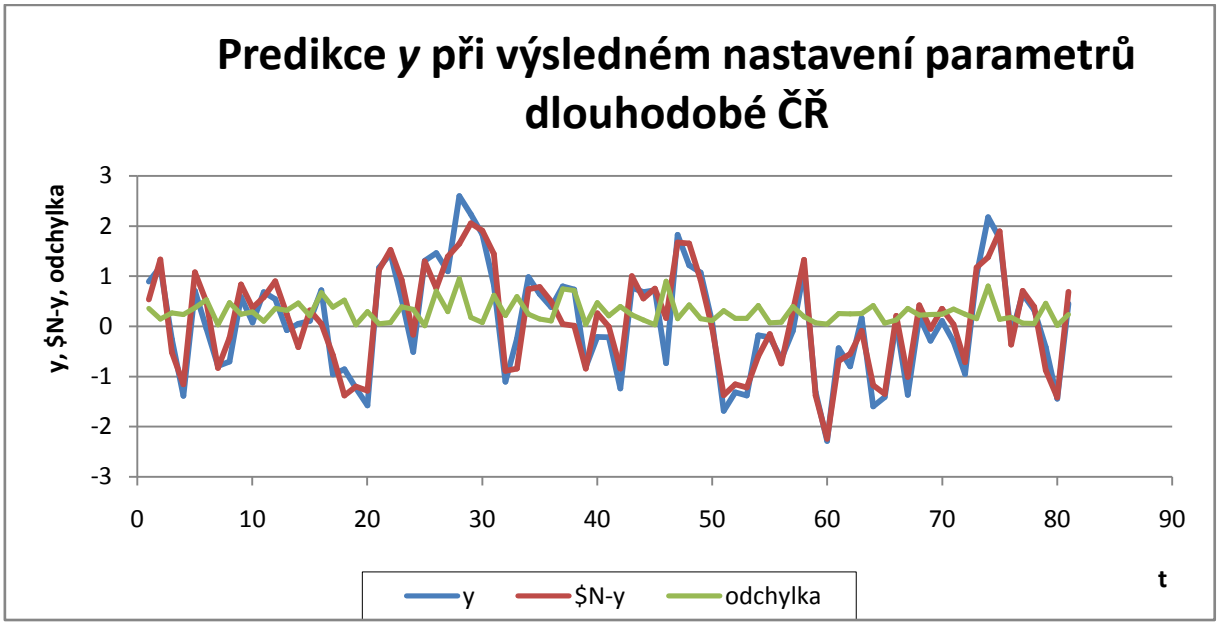

**Graf 3.26 - Vliv počtu testovacích dat na testovací chybu [vlastní]**

Výsledné nastavení neuronové sítě, při které bylo dosaženo nejmenší chyby na testovacích datech, je v tab. 3.9. Na grafu 3.27 je potom opět zobrazen průběh původních hodnot, predikovaných hodnot a jejich odchylka.

**Tab. 3.9 - Výsledné nastavení parametrů pro dlouhodobou ČŘ [vlastní]**

| počet neuronů | alpha | ν | počet cyklů | persistence | $Q_{train}$:$Q_{test}$ | $MAE_{train}$ | $MAE_{test}$ |
|---|---|---|---|---|---|---|---|
| 25 | 0.7 | 4 | 600 | 30 | 90:10 | 0.325 | 0.284 |

Ačkoliv by se mohlo zdát, že v případě této časové řady, kde bylo k dispozici nejvíce dat, bude výsledek ve srovnání s krátkodobou a střednědobou časovou řadou nejlepší, nebylo tomu tak. V případě dlouhodobé časové řady síť vykazovala největší chybu. Z těchto výsledku plyne, že pro učení RBF sítě v případě návštěvnosti upce.cz je lepší použít méně dat.




**Graf 3.27 - Predikce y při výsledném nastavení parametrů dlouhodobé ČŘ [vlastní]**

### 3.3.4 Srovnání časových řad

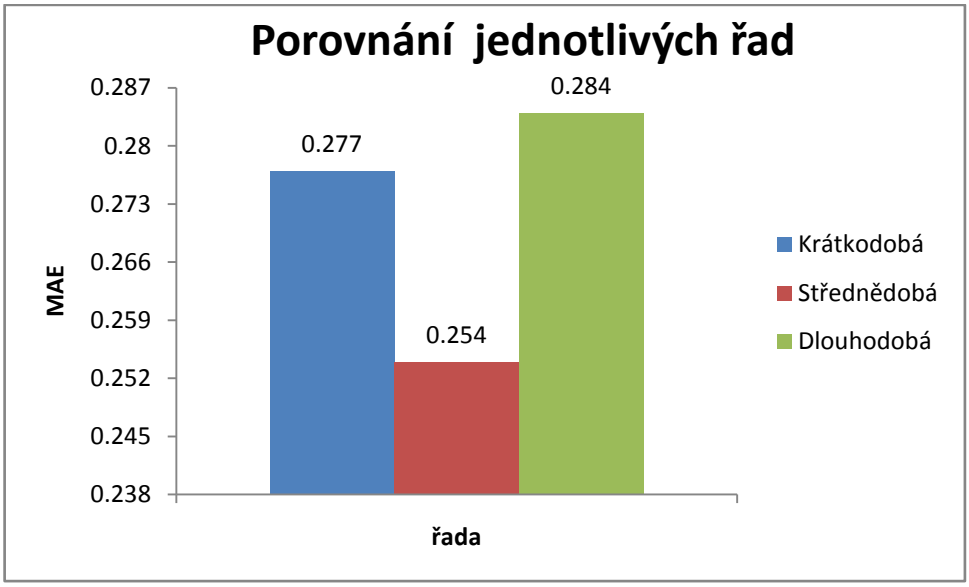
Jednotlivé časové řady se od sebe lišily především počtem neuronů ve skryté vrstvě a počtem dat, která měla síť k dispozici pro učení a testování.

Srovnání MAE chyb, kterých bylo dosaženo při výše zmíněných nejlepších nastaveních parametrů *alpha*, *v* a poměru rozdělení dat na trénovací a testovací množinu, je zobrazeno dohromady v tab. 3.10.

**Tab. 3.10 - Porovnání časových řad [vlastní]**

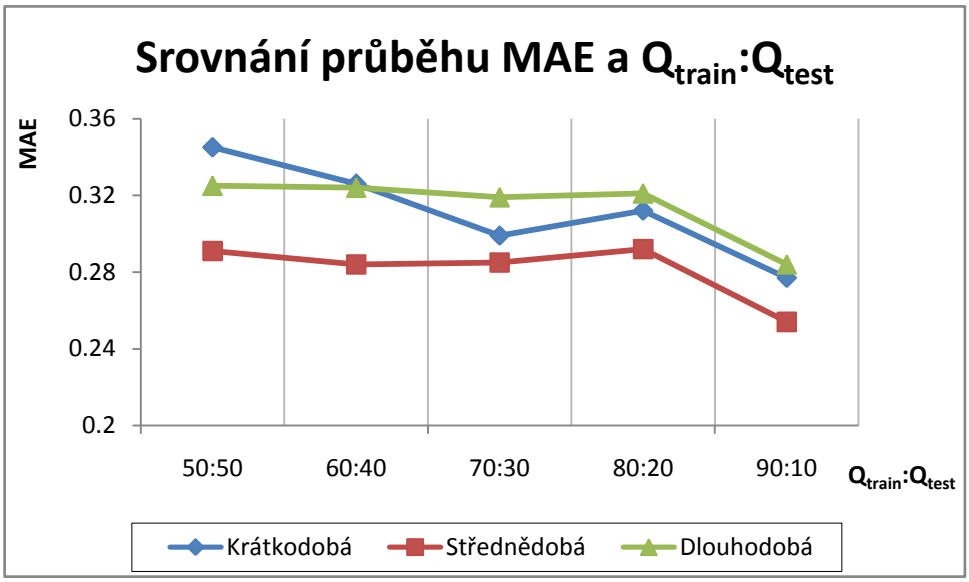
| řada | počet neuronů | alpha | v | počet cyklů | persistence | $Q_{train}$:$Q_{test}$ | $MAE_{train}$ | $MAE_{test}$ |
|---|---|---|---|---|---|---|---|---|
| Krátkodobá | 80 | 0.9 | 8 | 600 | 30 | 90:10 | 0.299 | 0.277 |
| Střednědobá | 125 | 0.5 | 4 | 600 | 30 | 90:10 | 0.256 | 0.254 |
| Dlouhodobá | 25 | 0.7 | 4 | 600 | 30 | 90:10 | 0.325 | 0.284 |

Na grafu 3.28 je zobrazené srovnání jednotlivých řad na základě MAE chyby testovacích dat. Z grafu je jasně vidět, že nejmenší MAE chyby na testovacích datech 0.254 bylo dosaženo u střednědobé časové řady návštěvnosti domény upce.cz.




**Graf 3.28 - Porovnání časových řad [vlastní]**

Na grafu 3.29 je přehledně zobrazen průběh testovací MAE chyby jednotlivých řad a jednotlivých rozdělení trénovací a testovací množiny. Všechny řady vykazovaly nejmenší chybu při rozdělení dat v poměru 90:10. To znamená, že síť měla k dispozici 90% všech dat k trénování a naučení sítě bylo testováno na 10% datech.




**Graf 3.29 – Srovnání řad v jednotlivém rozdělení dat [vlastní]**

Z těchto výsledků plyne, že čím menší počet testovacích dat měla každá časová řada k dispozici, tím menší chyba na těchto datech byla.

Z výše popsaného lze říci, že pro predikci návštěvnosti časové řady návštěvnosti domény upce.cz je lepší použít střednědobou časovou řadu. Tato časová řada obsahovala 480 hodnot sledování návštěvnosti. V případě dlouhodobé časové řady, která obsahovala celkově 752 pozorování, byla chyba v navrhovaných modelech nejvyšší, proto není pro predikci těchto dat vhodná.

## *3.4 Dílčí závěr kapitoly*

V této závěrečné kapitole jsou popsány jednotlivé kroky vedoucí k predikci návštěvnosti webové domény upce.cz. Začátek kapitoly je věnován předzpracování dat. Dále následuje popis parametrů RBF sítě, které používá prostředí SPSS Clementine 10.1, kde byly prováděny návrhy modelů neuronových sítí typu RBF k predikci časové řady. Další část kapitoly je věnována jednotlivým modelům, které byly navrženy a výběru vždy nejlepšího modelu, kde síť vykazovala nejmenší chybu pro danou časovou řadu.

Na závěr kapitoly je uvedené porovnání jednotlivých časových řad, ze kterého vyplynulo, že pro predikci návštěvnosti domény upce.cz RBF neuronovými sítěmi by byla nejvhodnější střednědobá časová řada.

# Závěr

Diplomová práce se zabývá predikcí návštěvnosti webové domény upce.cz neuronovými sítěmi typu RBF. První část práce je věnována problematice Web miningu, do kterého predikce návštěvnosti časové řady spadá. Druhá část práce je věnována základním poznatkům z oblasti neuronových sítí, dále pak neuronovým sítím typu RBF. Třetí část práce je věnovaná predikci návštěvnosti webové domény upce.cz.

Pro návrh modelu predikce neuronové sítě bylo využito třech časových řad různé délky. Cílem této práce bylo nalézt takové parametry modelu neuronové sítě typu RBF, při kterých je vykazována nejmenší chyba. Návrh jednotlivých modelů byl realizován v programovém prostředí SPSS Clementine 10.1. V tomto prostředí bylo nutné nalézt velikosti jednotlivých parametrů, které ovlivňují učení neuronové sítě. To bylo provedeno pro každou časovou řadu zvlášť v různých rozděleních dat na trénovací a testovací množiny. Kvalita navržených modelů byla měřena velikostí testovací chyby těchto modelů.

Každé časové řadě – pojmenované jako krátkodobá, střednědobá a dlouhodobá – byla věnována jedna podkapitola. V těchto podkapitolách jsou popsány výsledky navrhovaných modelů pro jednotlivá rozdělení dat na množinu trénovací a testovací. Závěrem kapitoly je porovnání všech třech časových řad z hlediska vykazovaných chyb. Bylo zjištěno, že pro predikci návštěvnosti je vhodné použít střednědobou časovou řadu. Zde byla testovací chyba na datech nejmenší. Naopak největší chyba byla naměřena v případě dlouhodobé časové řady.

# Seznam literatury


[1] SRIVASTAVA, J. at al. *Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data*. ACM SIGKDD Explorations Newsletter [online]. 2000, Vol. 1, [cit. 2011-02-28]. Dostupný z WWW: <http://www.sigkdd.org/explorations/ issues/1-2-2000-01/srivastava.pdf>.

[2] *Web Data Mining* [online]. [cit. 2011-02-27]. Dostupné z WWW: <http://www.web-datamining.net/>.

[3] FILIPOVÁ, J. *Metody výpočetní inteligence a web mining ve veřejné správě*. Pojednání ke státní závěrečné dokorské zkoušce, FES, Univerzita Pardubice, 2009.

[4] BERKA, P. *Dobývání znalostí z databází*. Praha: Academica, 2003. 366 s.

[5] SRIVASTAVA, J., DESIKAN, P., KUMAR, V. *Web Mining - Accomplishments & Future Directions* [online], 2009 [cit. 2011-03-01]. Dostupné z WWW: <http://www.ieee.org.ar/downloads/Srivastava-tut-paper.pdf>.

[6] COOLEY, R., MOBASHER, B., SRIVASTAVA, J. *Web Mining: Information and Pattern Discovery on the World Wide Web*. In Proc. of the 9th IEEE International Conference on Tools With Artifical Intelligence (ICTAI '97) [online]. Newport Beach : CA, 1997 [cit. 2011-03-01]. Dostupné z WWW: <http://maya.cs.depaul.edu/~mobasher/papers/webminer-tai97.pdf>.

[7] LEE, W. *Hierarchical Web Structure Mining* [online]. Faculty of Computer Science, Sungkyul University, 147-2 Anyang, Kyongkido, Korea. 2006 [cit. 2011-03-04]. Dostupné z WWW: <http://www.ieice.org/~de/DEWS/DEWS2006/doc/2A-v1.pdf>.

[8] GAUL, W. SCHMIDT-THIEME, L. *Mining Web Navigation Path Fragments* [online]. Institut für Entscheidungstheorie und Unternehmensforschung, Universität Karlsruhe, D-76128 Karlsruhe, Germany. 2000 [cit. 2011-04-04]. Dostupné z WWW: <http://robotics.stanford.edu/~ronnyk/WEBKDD2000/papers/ thieme.pdf>.

[9] DA COSTA, M. G., GONG, Z. *Web Structure Mining: An Introduction* [online]. Information Acquisition, 2005 IEEE International Conference, 2005 [cit. 2011-03-01]. Dostupné z WWW: <http://www.ceng.metu.edu.tr/~nihan/ceng553/StudentPapers/01635156.pdf>.

[10] KUMAR, P., SINGH, A. *Web Structure Mining: Exploring Hyperlinks and Algorithms*

*for Information Retrieval*. American Journal of Applied Sciences [online]. 2010, 7(6), [cit. 2011-04-03]. Dostupný z WWW: <http://www.scipub.org/fulltext/ajas/ajas76840-845.pdf>. ISSN 1546-923.

[11] REHBERGER, I. *Obsahují webové logy bohatství?* [online]. 2002 [cit. 2011-03-05]. Dostupné z WWW: <http://www.lupa.cz/clanky/obsahuji-webove-logy-bohatstvi/>.

[12] TESAŘ, M. *Získávání dat pomocí log souboru webového serveru* [online]. 2007 [cit. 2011-03-04]. Dostupné z WWW: <http://www.milantesar.net/analyza-navstevnosti/ziskavani-dat-pomoci-log-souboru-weboveho-serveru>.

[13] BENKOVSKÁ, P. *Web Usage Mining*. Jindřichův Hradec, 2008. 76 s. Diplomová práce. Vysoká škola ekonomická v Praze.

[14] TANASA, D. *Web Usage Mining: Contributions to Intersites Logs Preprocessing and Sequential Pattern Extraction with Low Support*. [online]. 2005, [cit. 2011-03-07]. Dostupný z WWW: <http://www-sop.inria.fr/axis/personnel/Doru.Tanasa/these_TANASA.pdf>.

[15] COOLEY, R., MOBASHER, B., SRIVASTAVA, J. *Grouping Web Page References into Transactions for Mining World Wide Web Browsing Patterns*. In Technical Report TR 97-021 [online]. Minneapolis: University of Minnesota, Dept. of Computer Science, 1997 [cit. 2011-03-07]. Dostupné z WWW: <http://maya.cs.depaul.edu/~mobasher/papers/webminer-kdex97.pdf>.

[16] VEČEŘA, M. *Dobývání znalostí z webu*. Brno, 2007. 59 s. Diplomová práce. Masarykova univerzita.

[17] KVASNIČKA, V. at al. *Úvod do teóorie neurónových sietí*. Bratislava: IRIS, 1997. 285 s. ISBN 80-88778-30-1.

[18] ŠÍMA, J., NERUDA, R. *Teoretické otázky neuronových sítí.* [online]. 1996 [cit. 2011-02-23]. Dostupné z WWW: <http://www2.cs.cas.cz/~sima/kniha.pdf>.

[19] NOVÁK, M. at al. *Umělé neuronové sítě: teorie a aplikace*. 1. Praha : C.H. Beck, 1998. 382 s. ISBN 80-7179-132-6.

[20] OLEJ, V., HÁJEK, P. *Úvod do umělé inteligence: Moderní přístupy: distanční opora*. 1. Pardubice: Univerzita Pardubice, 2010. 98 s. ISBN 978-80-7395-307-2.

[21] BERÁNEK, L. *Umělé neuronové sítě*. [online]. 2010, [cit. 2011-03-17]. Dostupný z WWW: <http://www.eamos.cz%2Famos%2Fkat_inf%2Fexterni%2Fkat_inf_76600%2F4.ppt>.

[22] ZELINKA, I. *Umělá inteligence I : Neuronové sítě a genetické algoritmy*. 1. Brno : VUTIUM, 1998. 126 s. ISBN 80-214-1163-5.

[23] OLEJ, V., HÁJEK, P. *Municipal Creditworthiness Modelling by Radial Basis Function Neural Networks and Sensitive Analysis their Inputs Parameters.* 19th International Conference on Artifical Neural Network, ICANN 2009, 14-17 September, Limassol, Cyprus, Alippi, C., Polycarpou, M., Panayiotou, Ch., Ellinas, G., Eds., Springer Berlin Heidelberg New York, pp. 505-514, 2009.

[24] KORDÍK, P. *Course Ware* [online]. České vysoké učení technické v Praze : 2010 [cit. 2011-03-14]. Vytěžování dat. Dostupné z WWW: <http://cw.felk.cvut.cz/lib/exe /fetch.php/courses/y336vd/prednasky/p10-doprednens.pdf>.

[25] Olej, V., Hájek, P., Filipová, J. *Modelling of Web Domain Visits by IF-Inference System*. WSEAS Transaction on Computers, WSEAS Press, Issue 10, Vol.9. October 2010, pp. 1170-1180. ISSN 1790-5079.

[26] *Google Analytics* [online]. 2011 [cit. 2011-03-17]. Nápověda Analytics. Dostupné z WWW: <http://www.google.com/support/analytics/>.

[27] KUBANOVÁ, J. *Statistické metody pro ekonomickou a technickou praxi*. Bratislava: Statis, 2003. 247 s. ISBN 80-85659-31-X.

[28] LERCH, J. *Porovnání klasifikačních a rozhodovacích modelů v oblasti životního prostředí*. Pardubice, 2008. 79 s. Diplomová práce. Univerzita Pardubice. Dostupné z WWW: <http://hdl.handle.net/10195/29189>.

[29] CIPRA, T. *Praktický průvodce finanční a pojistnou matematikou*. Praha: HZ Editio, 1995. 320 s. ISBN 80-901918-0-0.

[30] OLEJ, V. *Modelovanie ekonomických procesov na báze výpočtovej inteligencie*. Hradec Králové : Miloš Vognar - M&V, 2003. 159 s. ISBN 80-90324-9-1.

[31] KOKEŠ, R. *Modelování bonity obcí pomocí RBF neuronových sítí (predikční modely)*. Pardubice, 2008. 63 s. Diplomová práce. Univerzita Pardubice. Dostupné z WWW: <http://hdl.handle.net/10195/29185>.

[32] OLEJ, V., HÁJEK, P. *Municipal Creditworthiness Modelling by Radial Basis Function Neural Networks and Sensitive Analysis their Inputs Parameters*. 19th International Conference on Artificial Neural Network, ICANN 2009, Limassol, Cyprus, Alippi, C., Polycarpou, M., Panayiotou,Ch., Ellinas, G., Eds., Springer Berlin Heidelberg New York, 2009, pp.505-514, ISSN 0302-9743.

# Seznam obrázků

# Seznam tabulek

# Seznam grafů

# Seznam příloh

## PŘÍLOHA 1: STATISTIKY TRÉNOVACÍCH DAT KRÁTKODOBÉ ČR

| Poměr | Parametr | Střední hodnota | Odchylka | Minimum | Maximum | Počet |
|---|---|---|---|---|---|---|
| 50:50 | JKP | -0.05 | 1.04 | -2.77 | 2.47 | 126 |
| 50:50 | CKP | -0.05 | 1.02 | -2.88 | 2.22 | 126 |
| 50:50 | KM | -0.07 | 1.04 | -2.74 | 2.2 | 126 |
| 50:50 | JEV | -0.004 | 0.94 | -1.85 | 2.69 | 126 |
| 50:50 | DEV | -0.12 | 1.02 | -2.96 | 2.63 | 126 |
| 50:50 | y | -0.14 | 1.05 | -2.47 | 2.08 | 126 |
| 60:40 | JKP | -0.06 | 1.01 | -2.77 | 2.47 | 150 |
| 60:40 | CKP | -0.06 | 1 | -2.88 | 2.22 | 150 |
| 60:40 | KM | -0.09 | 1.03 | -2.74 | 2.2 | 150 |
| 60:40 | JEV | -0.04 | 0.96 | -1.96 | 2.69 | 150 |
| 60:40 | DEV | -0.13 | 1 | -2.96 | 2.63 | 150 |
| 60:40 | y | -0.14 | 1.03 | -2.47 | 2.08 | 150 |
| 70:30 | JKP | -0.07 | 1.01 | -2.8 | 2.98 | 180 |
| 70:30 | CKP | -0,07 | 0,99 | -2,88 | 2,66 | 180 |
| 70:30 | KM | -0,08 | 1,02 | -2,74 | 2,78 | 180 |
| 70:30 | JEV | -0,06 | 0,96 | -1,96 | 2,69 | 180 |
| 70:30 | DEV | -0,11 | 1,02 | -2,96 | 2,63 | 180 |
| 70:30 | y | -0,11 | 1,03 | -2,47 | 3,47 | 180 |
| 80:20 | JKP | -0.05 | 1 | -2.8 | 2.98 | 199 |
| 80:20 | CKP | -0.06 | 0.99 | -2.88 | 2.66 | 199 |
| 80:20 | KM | -0.06 | 1.01 | -2.74 | 2.78 | 199 |
| 80:20 | JEV | -0.05 | 0.98 | -1.96 | 2.69 | 199 |
| 80:20 | DEV | -0.09 | 1 | -2.96 | 2.63 | 199 |
| 80:20 | y | -0.08 | 1.02 | -2.47 | 3.47 | 199 |
| 90:10 | JKP | -0.04 | 1.01 | -2.8 | 3.19 | 222 |
| 90:10 | CKP | -0.05 | 1 | -2.88 | 3.25 | 222 |
| 90:10 | KM | -0.04 | 1.01 | -2.74 | 2.78 | 222 |
| 90:10 | JEV | -0.04 | 1 | -1.96 | 2.69 | 222 |
| 90:10 | DEV | -0.06 | 1 | -2.96 | 2.99 | 222 |
| 90:10 | y | -0.05 | 1.02 | -2.47 | 3.47 | 222 |

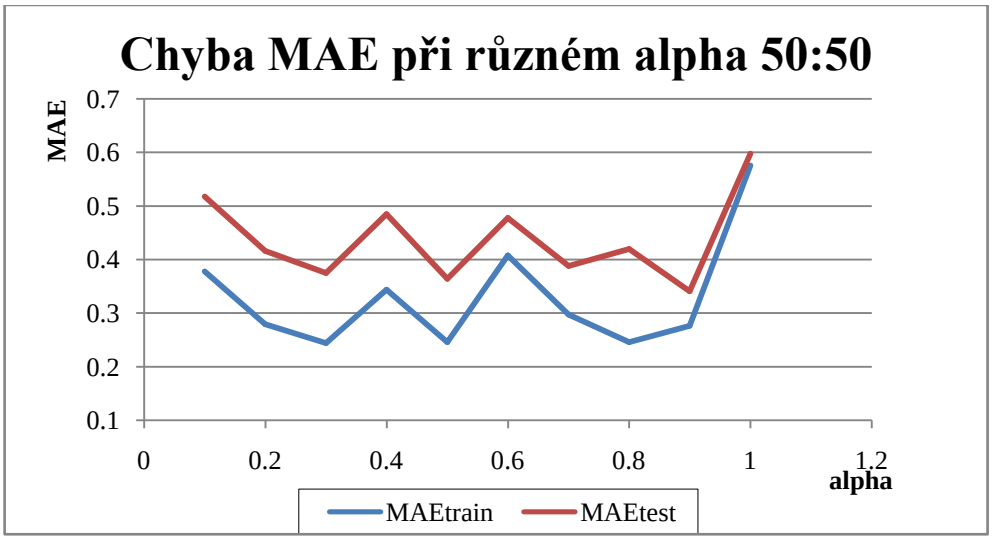
## PŘÍLOHA 2: STATISTIKY TESTOVACÍCH DAT KRÁTKODOBÉ ČŘ

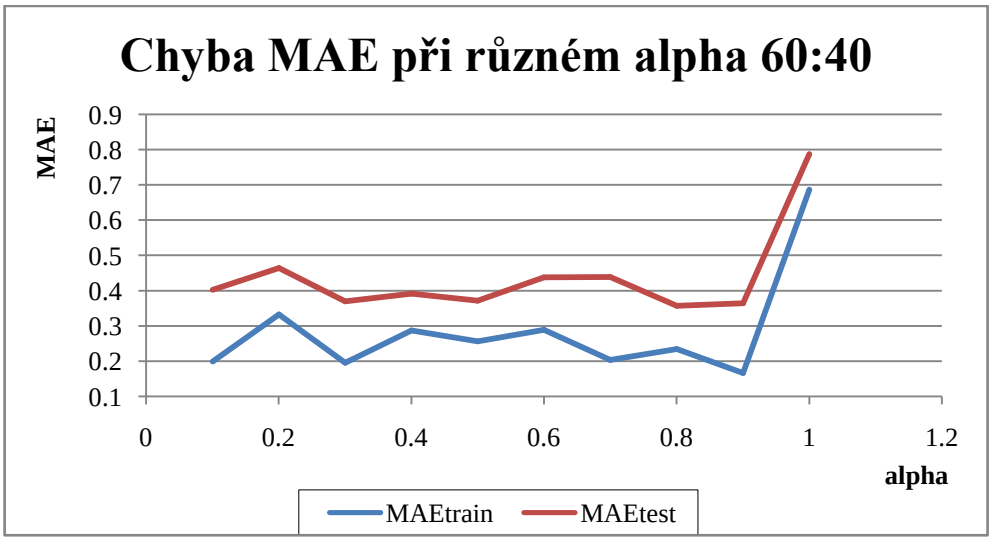
| Poměr | Parametr | Střední hodnota | Odchylka | Minimum | Maximum | Počet |
|---|---|---|---|---|---|---|
| 50:50 | JKP | 0.04 | 0.97 | -2.8 | 3.19 | 138 |
| 50:50 | CKP | 0.04 | 0.99 | -2.64 | 3.25 | 138 |
| 50:50 | KM | 0.07 | 0.96 | -2.5 | 2.78 | 138 |
| 50:50 | JEV | 0.004 | 1.06 | -1.96 | 2.61 | 138 |
| 50:50 | DEV | 0.11 | 0.98 | -2.31 | 2.99 | 138 |
| 50:50 | y | 0.12 | 0.94 | -2.02 | 3.47 | 138 |
| 60:40 | JKP | 0.08 | 0.99 | -2.8 | 3.19 | 114 |
| 60:40 | CKP | 0.08 | 0.99 | -2.64 | 3.25 | 114 |
| 60:40 | KM | 0.12 | 0.96 | -2.5 | 2.78 | 114 |
| 60:40 | JEV | 0.06 | 1.06 | -1.91 | 2.61 | 114 |
| 60:40 | DEV | 0.17 | 0.97 | -2.31 | 2.99 | 114 |
| 60:40 | y | 0.18 | 0.94 | -2.02 | 3.47 | 114 |
| 70:30 | JKP | 0,14 | 0,98 | -2,01 | 3,2 | 84 |
| 70:30 | CKP | 0,15 | 1,00 | -2,07 | 3,25 | 84 |
| 70:30 | KM | 0,18 | 0,95 | -1,89 | 2,78 | 84 |
| 70:30 | JEV | 0,13 | 1,07 | -1,91 | 2,61 | 84 |
| 70:30 | DEV | 0,24 | 0,93 | -1,92 | 2,99 | 84 |
| 70:30 | y | 0,23 | 0,91 | -1,75 | 2,7 | 84 |
| 80:20 | JKP | 0.15 | 1.00 | -2.01 | 3.19 | 65 |
| 80:20 | CKP | 0.18 | 1.03 | -2.07 | 3.25 | 65 |
| 80:20 | KM | 0.19 | 0.97 | -1.89 | 2.78 | 65 |
| 80:20 | JEV | 0.15 | 1.08 | -1.91 | 2.61 | 65 |
| 80:20 | DEV | 0.29 | 0.95 | -1.92 | 2.99 | 65 |
| 80:20 | y | 0.23 | 0.92 | -1.75 | 2.70 | 65 |
| 90:10 | JKP | 0.22 | 0.93 | -1.94 | 2.56 | 42 |
| 90:10 | CKP | 0.26 | 0.99 | -2.07 | 3.06 | 42 |
| 90:10 | KM | 0.22 | 0.95 | -1.89 | 2.78 | 42 |
| 90:10 | JEV | 0.19 | 0.99 | -1.65 | 2.61 | 42 |
| 90:10 | DEV | 0.32 | 0.94 | -1.92 | 2.98 | 42 |
| 90:10 | y | 0.28 | 0.87 | -1.50 | 2.42 | 42 |

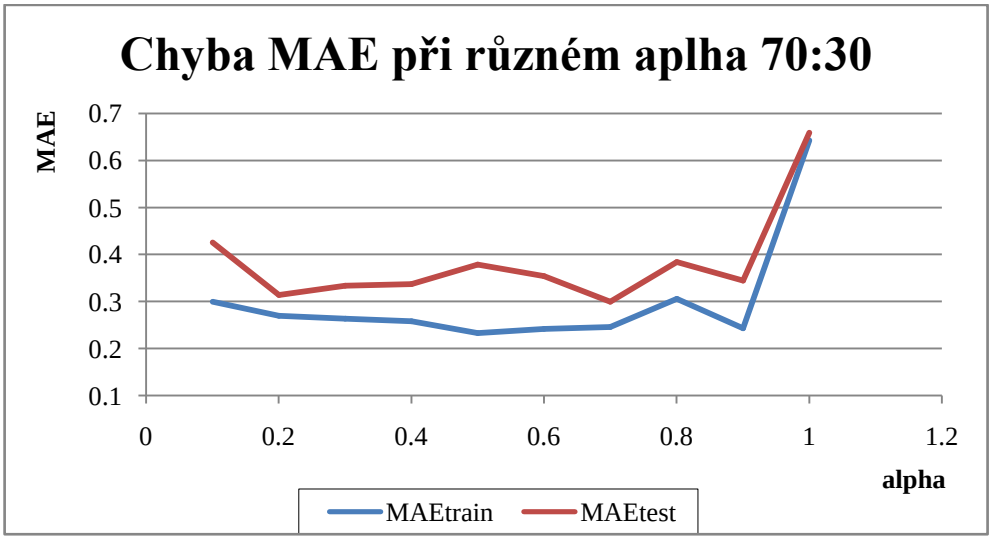
## PŘÍLOHA 3: PRŮBĚH CHYBY PŘI RŮZNÉM ALPHA KRÁTKODOBÉ ČŘ

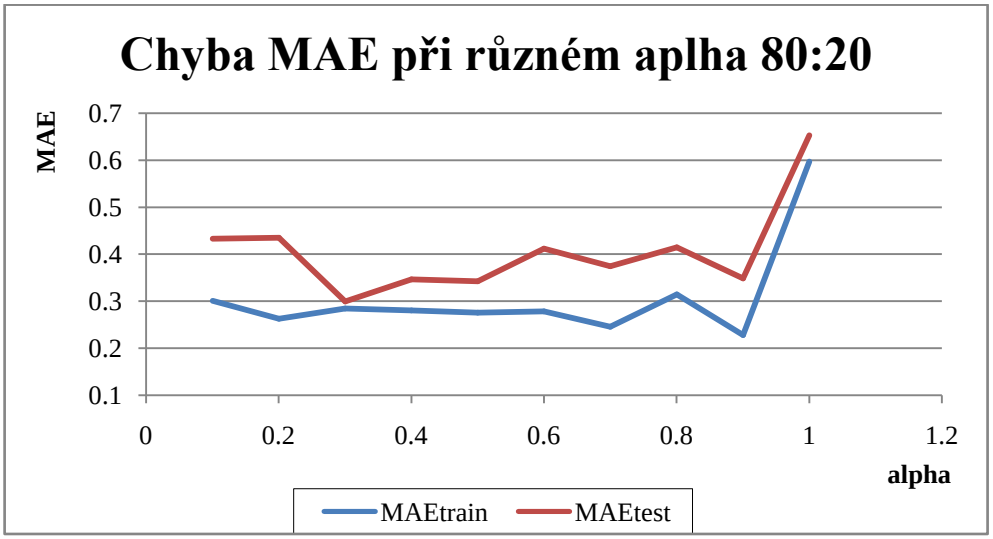
| počet neuronů | alpha | ν | počet cyklů | persistence | Qtrain:Qtest | MAEtrain | MAEtest |
|---|---|---|---|---|---|---|---|
| 80 | 0.1 | 1 | 600 | 30 | 50:50 | 0.378 | 0.518 |
| 80 | 0.2 | 1 | 600 | 30 | 50:50 | 0.279 | 0.416 |
| 80 | 0.3 | 1 | 600 | 30 | 50:50 | 0.244 | 0.375 |
| 80 | 0.4 | 1 | 600 | 30 | 50:50 | 0.344 | 0.485 |
| 80 | 0.5 | 1 | 600 | 30 | 50:50 | 0.246 | 0.364 |
| 80 | 0.6 | 1 | 600 | 30 | 50:50 | 0.408 | 0.478 |
| 80 | 0.7 | 1 | 600 | 30 | 50:50 | 0.297 | 0.388 |
| 80 | 0.8 | 1 | 600 | 30 | 50:50 | 0.246 | 0.42 |
| 80 | 0.9 | 1 | 600 | 30 | 50:50 | 0.276 | 0.341 |
| 80 | 1 | 1 | 600 | 30 | 50:50 | 0.576 | 0.598 |
| 80 | 0.1 | 1 | 600 | 30 | 60:40 | 0.199 | 0.403 |
| 80 | 0.2 | 1 | 600 | 30 | 60:40 | 0.333 | 0.464 |
| 80 | 0.3 | 1 | 600 | 30 | 60:40 | 0.196 | 0.37 |
| 80 | 0.4 | 1 | 600 | 30 | 60:40 | 0.287 | 0.392 |
| 80 | 0.5 | 1 | 600 | 30 | 60:40 | 0.257 | 0.372 |
| 80 | 0.6 | 1 | 600 | 30 | 60:40 | 0.289 | 0.438 |
| 80 | 0.7 | 1 | 600 | 30 | 60:40 | 0.204 | 0.439 |
| 80 | 0.8 | 1 | 600 | 30 | 60:40 | 0.235 | 0.357 |
| 80 | 0.9 | 1 | 600 | 30 | 60:40 | 0.167 | 0.365 |
| 80 | 1 | 1 | 600 | 30 | 60:40 | 0.687 | 0.788 |
| 80 | 0.1 | 1 | 600 | 30 | 70:30 | 0.3 | 0.426 |
| 80 | 0.2 | 1 | 600 | 30 | 70:30 | 0.27 | 0.314 |
| 80 | 0.3 | 1 | 600 | 30 | 70:30 | 0.264 | 0.334 |
| 80 | 0.4 | 1 | 600 | 30 | 70:30 | 0.258 | 0.337 |
| 80 | 0.5 | 1 | 600 | 30 | 70:30 | 0.233 | 0.379 |
| 80 | 0.6 | 1 | 600 | 30 | 70:30 | 0.242 | 0.354 |
| 80 | 0.7 | 1 | 600 | 30 | 70:30 | 0.246 | 0.3 |
| 80 | 0.8 | 1 | 600 | 30 | 70:30 | 0.306 | 0.384 |
| 80 | 0.9 | 1 | 600 | 30 | 70:30 | 0.243 | 0.345 |
| 80 | 1 | 1 | 600 | 30 | 70:30 | 0.643 | 0.659 |
| 80 | 0.1 | 1 | 600 | 30 | 80:20 | 0.301 | 0.433 |
| 80 | 0.2 | 1 | 600 | 30 | 80:20 | 0.263 | 0.435 |
| 80 | 0.3 | 1 | 600 | 30 | 80:20 | 0.285 | 0.3 |
| 80 | 0.4 | 1 | 600 | 30 | 80:20 | 0.281 | 0.347 |
| 80 | 0.5 | 1 | 600 | 30 | 80:20 | 0.276 | 0.343 |
| 80 | 0.6 | 1 | 600 | 30 | 80:20 | 0.279 | 0.412 |
| 80 | 0.7 | 1 | 600 | 30 | 80:20 | 0.246 | 0.375 |
| 80 | 0.8 | 1 | 600 | 30 | 80:20 | 0.315 | 0.415 |
| 80 | 0.9 | 1 | 600 | 30 | 80:20 | 0.228 | 0.349 |
| 80 | 1 | 1 | 600 | 30 | 80:20 | 0.597 | 0.653 |

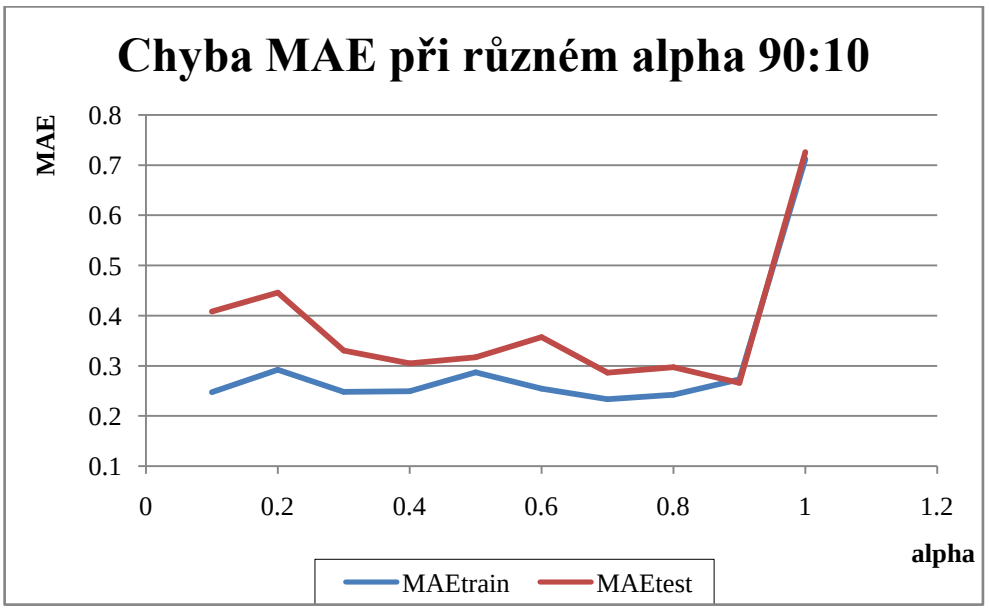
| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 80 | 0.1 | 1 | 600 | 30 | 90:10 | 0.247 | 0.408 |
| 80 | 0.2 | 1 | 600 | 30 | 90:10 | 0.292 | 0.446 |
| 80 | 0.3 | 1 | 600 | 30 | 90:10 | 0.248 | 0.33 |
| 80 | 0.4 | 1 | 600 | 30 | 90:10 | 0.249 | 0.305 |
| 80 | 0.5 | 1 | 600 | 30 | 90:10 | 0.287 | 0.317 |
| 80 | 0.6 | 1 | 600 | 30 | 90:10 | 0.254 | 0.357 |
| 80 | 0.7 | 1 | 600 | 30 | 90:10 | 0.233 | 0.286 |
| 80 | 0.8 | 1 | 600 | 30 | 90:10 | 0.242 | 0.297 |
| 80 | 0.9 | 1 | 600 | 30 | 90:10 | 0.273 | 0.266 |
| 80 | 1 | 1 | 600 | 30 | 90:10 | 0.712 | 0.726 |




Chyba MAE při různém aplha 70:30
MAE
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0.2
0.4
0.6
0.8
1
1.2
alpha
MAEtrain
MAEtest

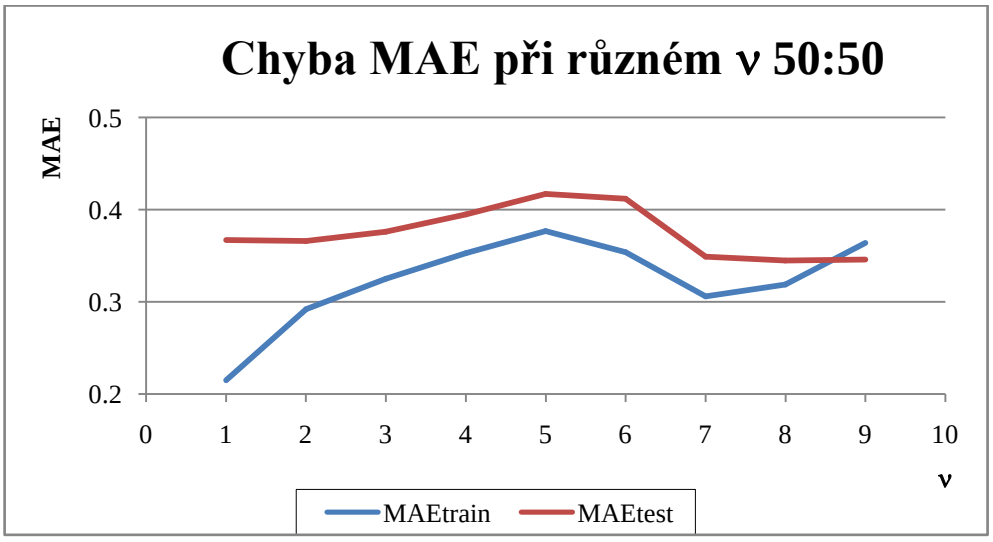

Chyba MAE při různém aplha 80:20
MAE
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0.2
0.4
0.6
0.8
1
1.2
alpha
MAEtrain
MAEtest

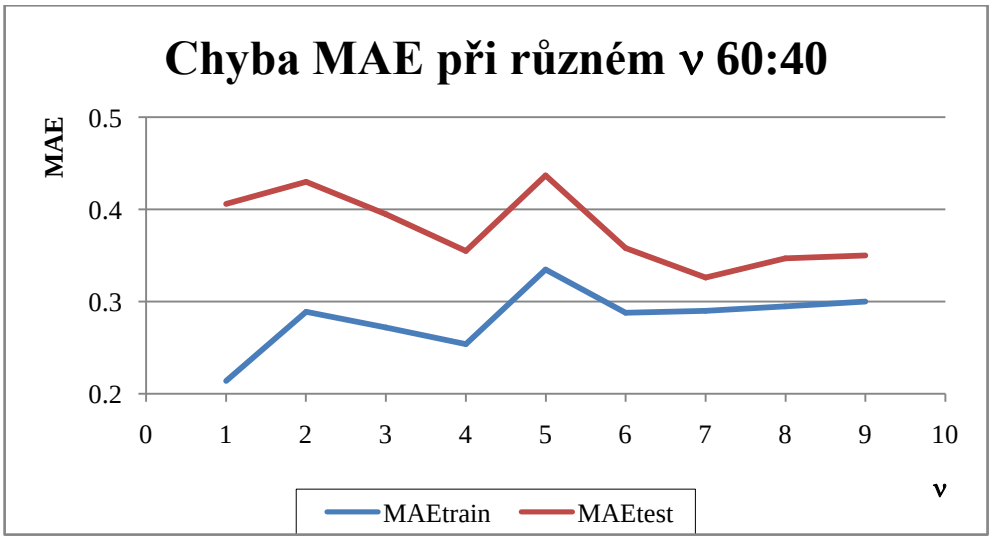

Chyba MAE při různém alpha 90:10
MAE
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0.2
0.4
0.6
0.8
1
1.2
alpha
MAEtrain
MAEtest

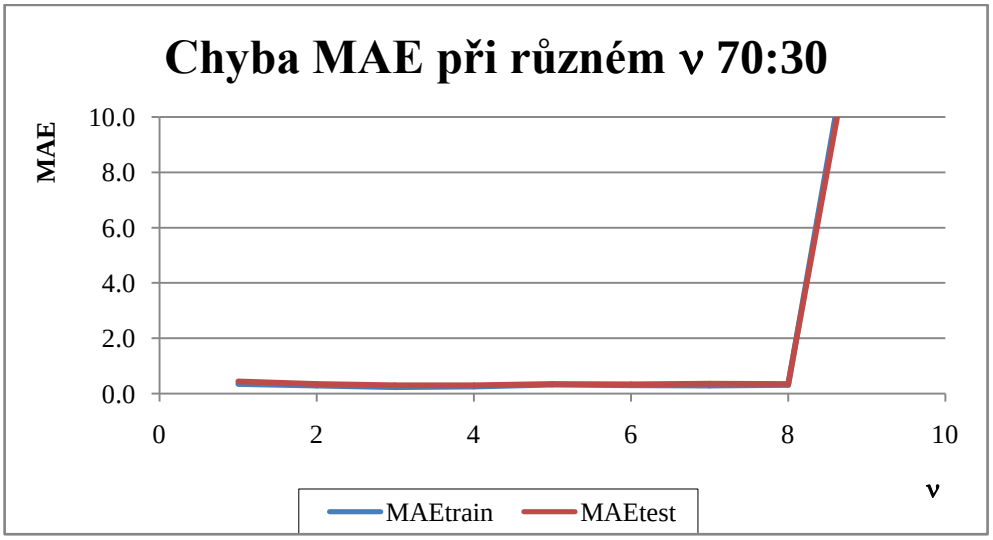
## PŘÍLOHA 4: PRŮBĚH CHYBY PŘI RŮZNÉM ν KRÁTKODOBÉ ČŘ

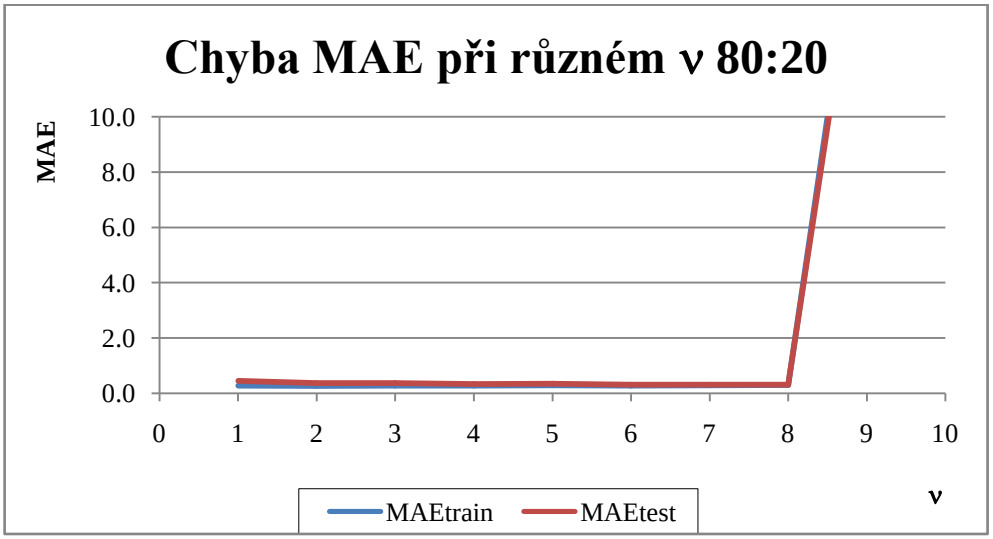
| počet neuronů | alpha | ν | počet cyklů | persistence | Qtrain:Qtest | MAEtrain | MAEtest |
|---|---|---|---|---|---|---|---|
| 80 | 0.9 | 1 | 600 | 30 | 50:50 | 0.215 | 0.367 |
| 80 | 0.9 | 2 | 600 | 30 | 50:50 | 0.292 | 0.366 |
| 80 | 0.9 | 3 | 600 | 30 | 50:50 | 0.325 | 0.376 |
| 80 | 0.9 | 4 | 600 | 30 | 50:50 | 0.353 | 0.395 |
| 80 | 0.9 | 5 | 600 | 30 | 50:50 | 0.377 | 0.417 |
| 80 | 0.9 | 6 | 600 | 30 | 50:50 | 0.354 | 0.412 |
| 80 | 0.9 | 7 | 600 | 30 | 50:50 | 0.306 | 0.349 |
| 80 | 0.9 | 8 | 600 | 30 | 50:50 | 0.319 | 0.345 |
| 80 | 0.9 | 9 | 600 | 30 | 50:50 | 0.364 | 0.346 |
| 80 | 0.8 | 1 | 600 | 30 | 60:40 | 0.214 | 0.406 |
| 80 | 0.8 | 2 | 600 | 30 | 60:40 | 0.289 | 0.43 |
| 80 | 0.8 | 3 | 600 | 30 | 60:40 | 0.272 | 0.395 |
| 80 | 0.8 | 4 | 600 | 30 | 60:40 | 0.254 | 0.355 |
| 80 | 0.8 | 5 | 600 | 30 | 60:40 | 0.335 | 0.437 |
| 80 | 0.8 | 6 | 600 | 30 | 60:40 | 0.288 | 0.358 |
| 80 | 0.8 | 7 | 600 | 30 | 60:40 | 0.29 | 0.326 |
| 80 | 0.8 | 8 | 600 | 30 | 60:40 | 0.295 | 0.347 |
| 80 | 0.8 | 9 | 600 | 30 | 60:40 | 0.3 | 0.35 |
| 80 | 0.7 | 1 | 600 | 30 | 70:30 | 0.344 | 0.444 |
| 80 | 0.7 | 2 | 600 | 30 | 70:30 | 0.297 | 0.353 |
| 80 | 0.7 | 3 | 600 | 30 | 70:30 | 0.23 | 0.299 |
| 80 | 0.7 | 4 | 600 | 30 | 70:30 | 0.261 | 0.299 |
| 80 | 0.7 | 5 | 600 | 30 | 70:30 | 0.328 | 0.346 |
| 80 | 0.7 | 6 | 600 | 30 | 70:30 | 0.303 | 0.342 |
| 80 | 0.7 | 7 | 600 | 30 | 70:30 | 0.287 | 0.366 |
| 80 | 0.7 | 8 | 600 | 30 | 70:30 | 0.317 | 0.351 |
| 80 | 0.7 | 9 | 600 | 30 | 70:30 | 16.453 | 15.845 |
| 80 | 0.3 | 1 | 600 | 30 | 80:20 | 0.285 | 0.451 |
| 80 | 0.3 | 2 | 600 | 30 | 80:20 | 0.267 | 0.375 |
| 80 | 0.3 | 3 | 600 | 30 | 80:20 | 0.281 | 0.369 |
| 80 | 0.3 | 4 | 600 | 30 | 80:20 | 0.277 | 0.336 |
| 80 | 0.3 | 5 | 600 | 30 | 80:20 | 0.289 | 0.352 |
| 80 | 0.3 | 6 | 600 | 30 | 80:20 | 0.284 | 0.317 |
| 80 | 0.3 | 7 | 600 | 30 | 80:20 | 0.294 | 0.315 |
| 80 | 0.3 | 8 | 600 | 30 | 80:20 | 0.299 | 0.312 |
| 80 | 0.3 | 9 | 600 | 30 | 80:20 | 19.651 | 18.959 |
| 80 | 0.9 | 1 | 600 | 30 | 90:10 | 0.264 | 0.362 |
| 80 | 0.9 | 2 | 600 | 30 | 90:10 | 0.27 | 0.329 |
| 80 | 0.9 | 3 | 600 | 30 | 90:10 | 0.249 | 0.316 |
| 80 | 0.9 | 4 | 600 | 30 | 90:10 | 0.283 | 0.29 |

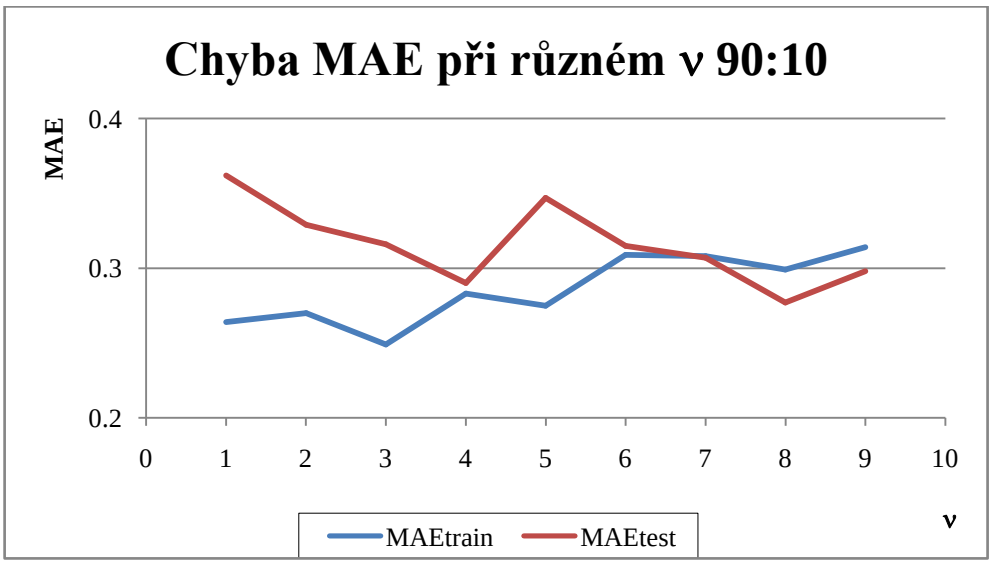
| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 80 | 0.9 | 5 | 600 | 30 | 90:10 | 0.275 | 0.347 |
| 80 | 0.9 | 6 | 600 | 30 | 90:10 | 0.309 | 0.315 |
| 80 | 0.9 | 7 | 600 | 30 | 90:10 | 0.308 | 0.307 |
| 80 | 0.9 | 8 | 600 | 30 | 90:10 | 0.299 | 0.277 |
| 80 | 0.9 | 9 | 600 | 30 | 90:10 | 0.314 | 0.298 |



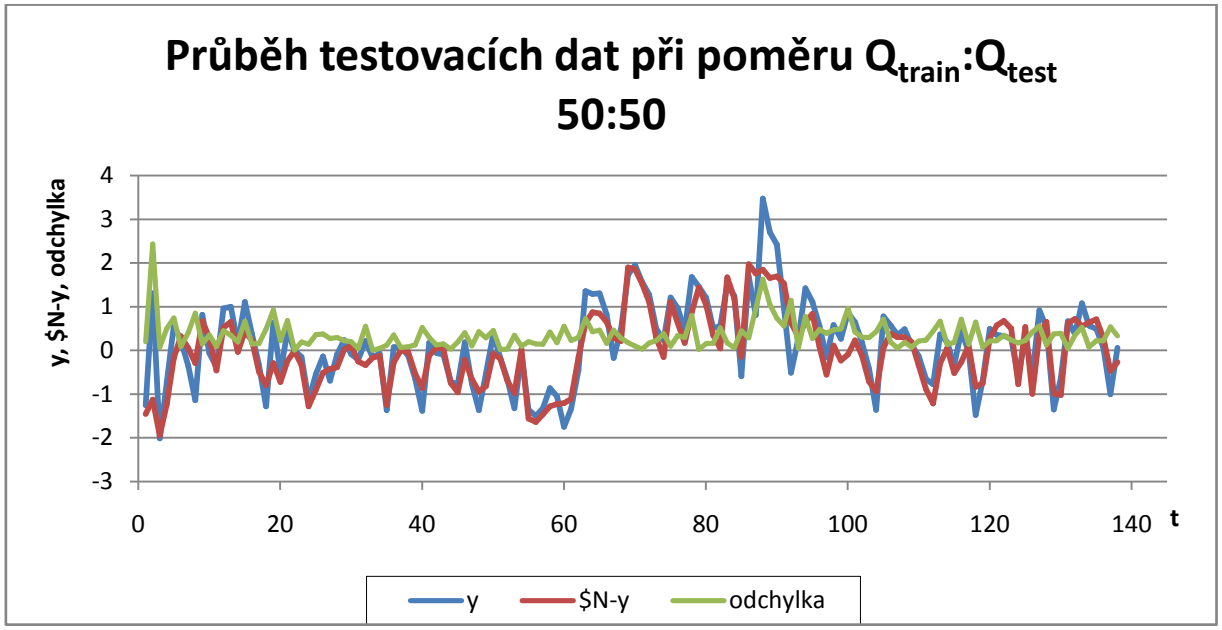



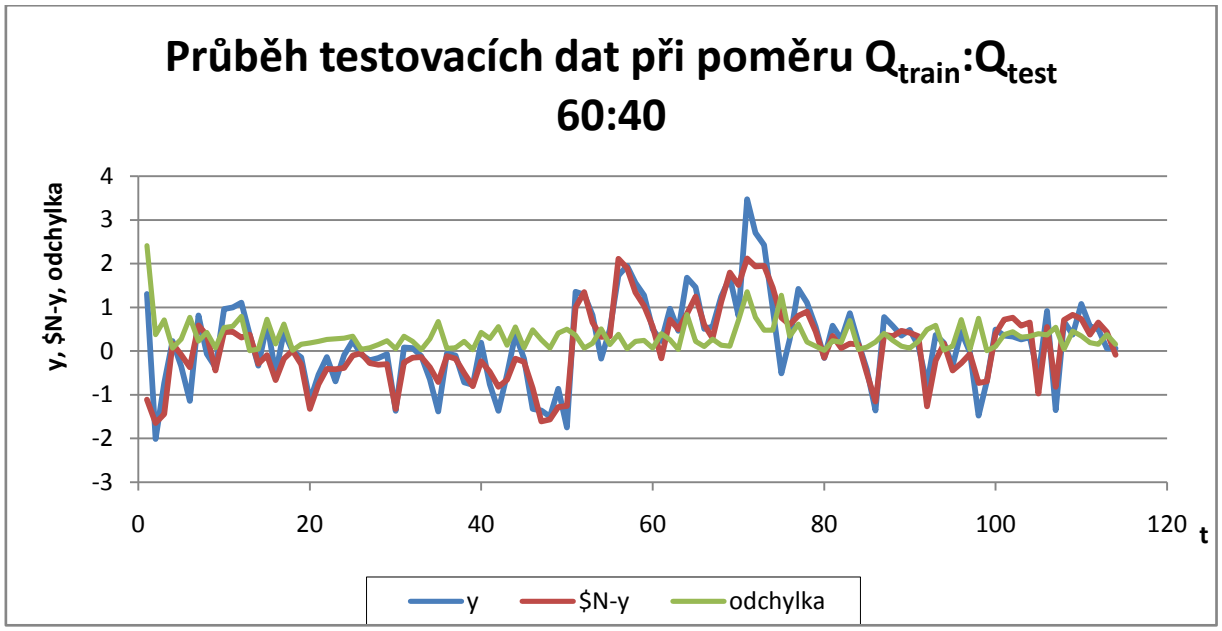

Chyba MAE při různém ν 70:30
MAE
10.0
8.0
6.0
4.0
2.0
0.0
0
2
4
6
8
10
ν
MAEtrain
MAEtest

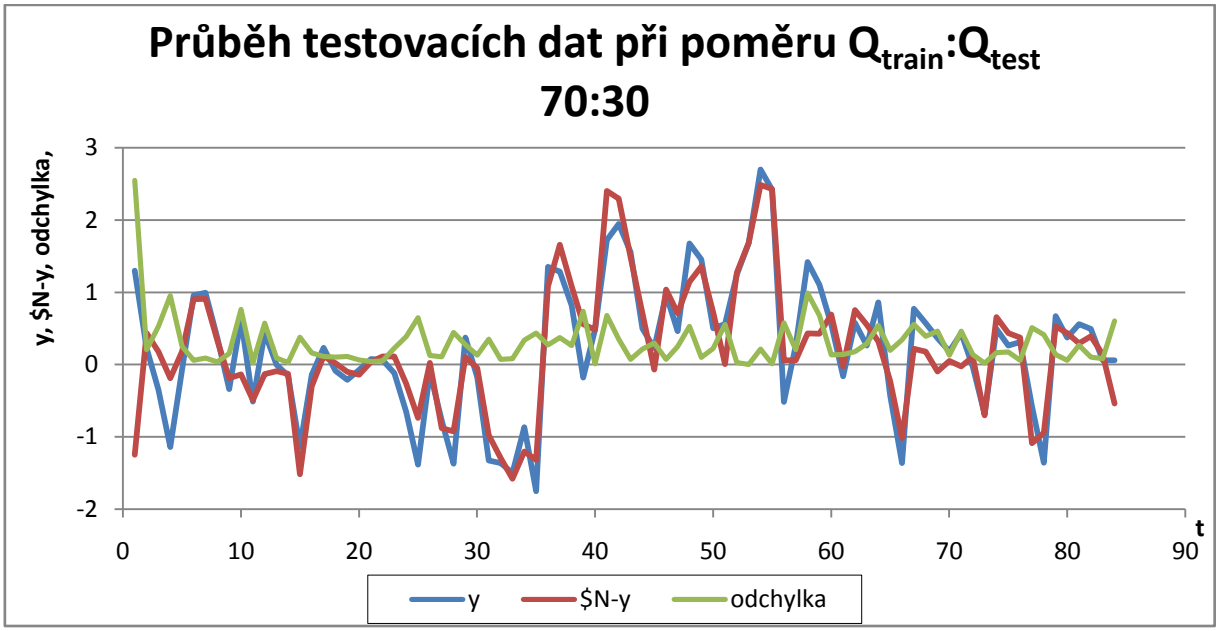

Chyba MAE při různém ν 80:20
MAE
10.0
8.0
6.0
4.0
2.0
0.0
0
1
2
3
4
5
6
7
8
9
10
ν
MAEtrain
MAEtest

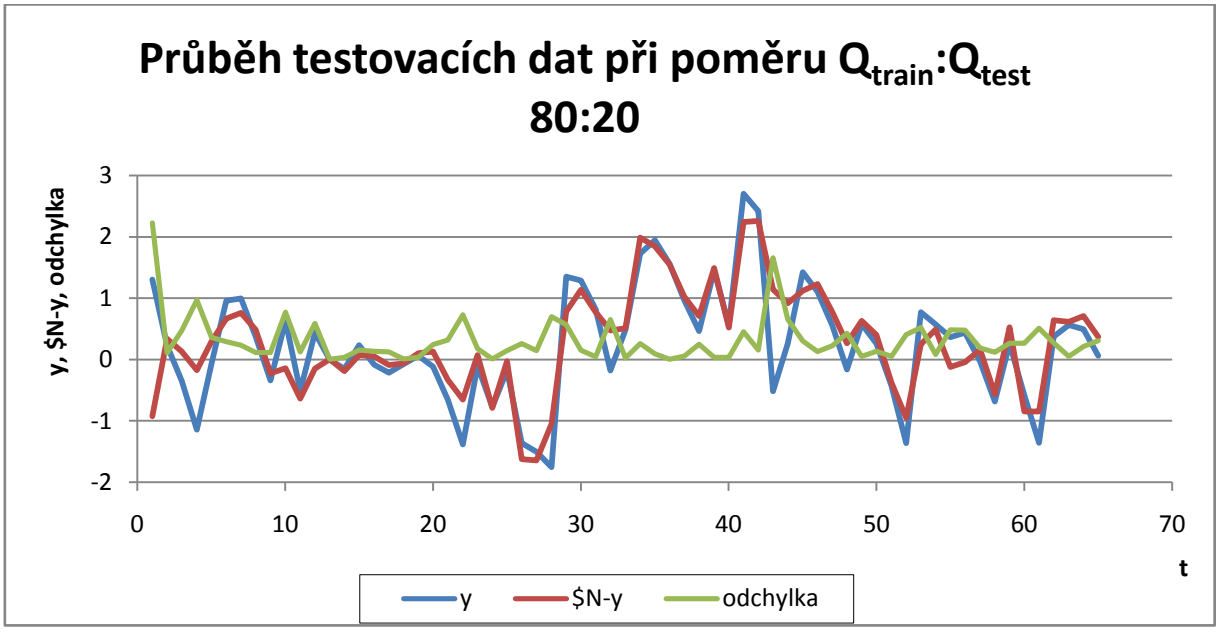

Chyba MAE při různém ν 90:10
MAE
0.4
0.3
0.2
0
1
2
3
4
5
6
7
8
9
10
ν
MAEtrain
MAEtest

## PŘÍLOHA 5: PREDIKCE y V JEDNOTLIVÝCH ROZDĚLENÍCH DAT KRÁTKODOBÉ ČŘ




Průběh testovacích dat při poměru $Q_{train}$:$Q_{test}$ 70:30
y, $N-y, odchylka,
3
2
1
0
-1
-2
0
10
20
30
40
50
60
70
80
90
t
y
$N-y
odchylka


Průběh testovacích dat při poměru $Q_{train}$:$Q_{test}$ 80:20
y, $N-y, odchylka
3
2
1
0
-1
-2
0
10
20
30
40
50
60
70
t
y
$N-y
odchylka

## PŘÍLOHA 6: STATISTIKY TRÉNOVACÍCH DAT STŘEDNĚDOBÉ ČR

| Poměr | Parametr | Střední hodnota | Odchylka | Minimum | Maximum | Počet |
|---|---|---|---|---|---|---|
| 50:50 | JKP | -0.01 | 0.99 | -3.04 | 2.84 | 242 |
| 50:50 | CKP | -0.02 | 0.98 | -3.15 | 2.81 | 242 |
| 50:50 | KM | -0.02 | 0.99 | -2.97 | 2.37 | 242 |
| 50:50 | JEV | -0.03 | 0.99 | -2.19 | 2.54 | 242 |
| 50:50 | DEV | -0.03 | 0.99 | -3.16 | 2.57 | 242 |
| 50:50 | y | 0.00 | 1.01 | -2.61 | 2.83 | 242 |
| 60:40 | JKP | -0.05 | 0.99 | -3.04 | 2.84 | 286 |
| 60:40 | CKP | -0.05 | 0.98 | -3.15 | 2.81 | 286 |
| 60:40 | KM | -0.05 | 0.99 | -2.97 | 2.37 | 286 |
| 60:40 | JEV | -0.05 | 0.98 | -2.24 | 2.54 | 286 |
| 60:40 | DEV | -0.05 | 0.97 | -3.16 | 2.57 | 286 |
| 60:40 | y | -0.02 | 1.00 | -2.61 | 2.83 | 286 |
| 70:30 | JKP | -0.02 | 0.99 | -3.07 | 2.84 | 331 |
| 70:30 | CKP | -0.02 | 0.98 | -3.15 | 2.81 | 331 |
| 70:30 | KM | -0.01 | 0.99 | -2.97 | 2.37 | 331 |
| 70:30 | JEV | -0.02 | 0.95 | -2.24 | 2.54 | 331 |
| 70:30 | DEV | -0.01 | 0.97 | -3.16 | 2.57 | 331 |
| 70:30 | y | 0.00 | 0.99 | -2.61 | 2.83 | 331 |
| 80:20 | JKP | -0.02 | 0.99 | -3.07 | 2.84 | 367 |
| 80:20 | CKP | -0.02 | 0.98 | -3.15 | 2.81 | 367 |
| 80:20 | KM | -0.01 | 0.98 | -2.97 | 2.37 | 367 |
| 80:20 | JEV | -0.02 | 0.96 | -2.24 | 2.54 | 367 |
| 80:20 | DEV | -0.01 | 0.96 | -3.16 | 2.57 | 367 |
| 80:20 | y | 0.00 | 0.99 | -2.61 | 2.83 | 367 |
| 90:10 | JKP | -0.03 | 1.00 | -3.07 | 2.94 | 415 |
| 90:10 | CKP | -0.03 | 0.99 | -3.15 | 2.94 | 415 |
| 90:10 | KM | -0.02 | 0.98 | -2.97 | 2.69 | 415 |
| 90:10 | JEV | -0.03 | 0.98 | -2.24 | 2.54 | 415 |
| 90:10 | DEV | -0.02 | 0.97 | -3.16 | 2.67 | 415 |
| 90:10 | y | -0.01 | 0.98 | -2.61 | 2.83 | 415 |

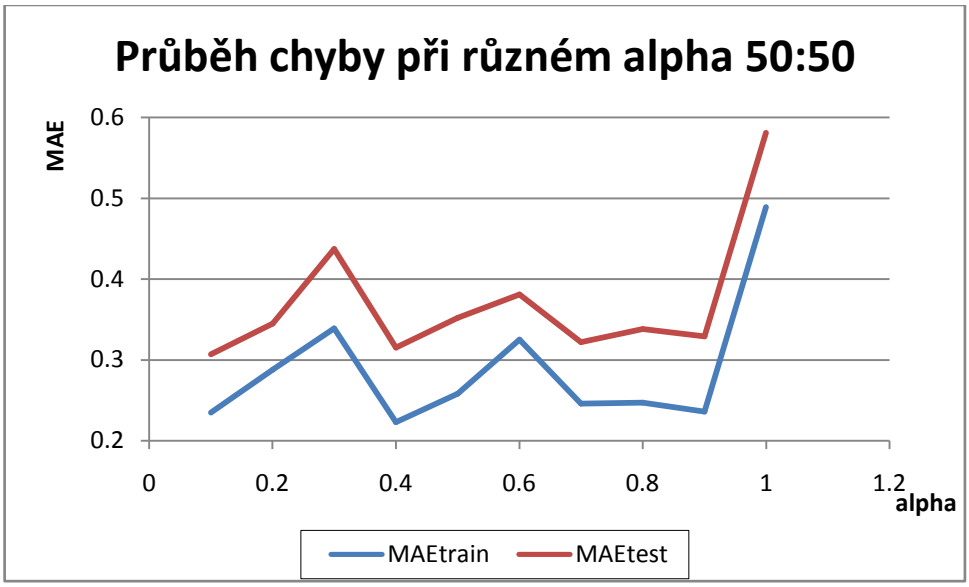
## PŘÍLOHA 7*:* STATISTIKY TESTOVACÍCH DAT STŘEDNĚDOBÉ ČR

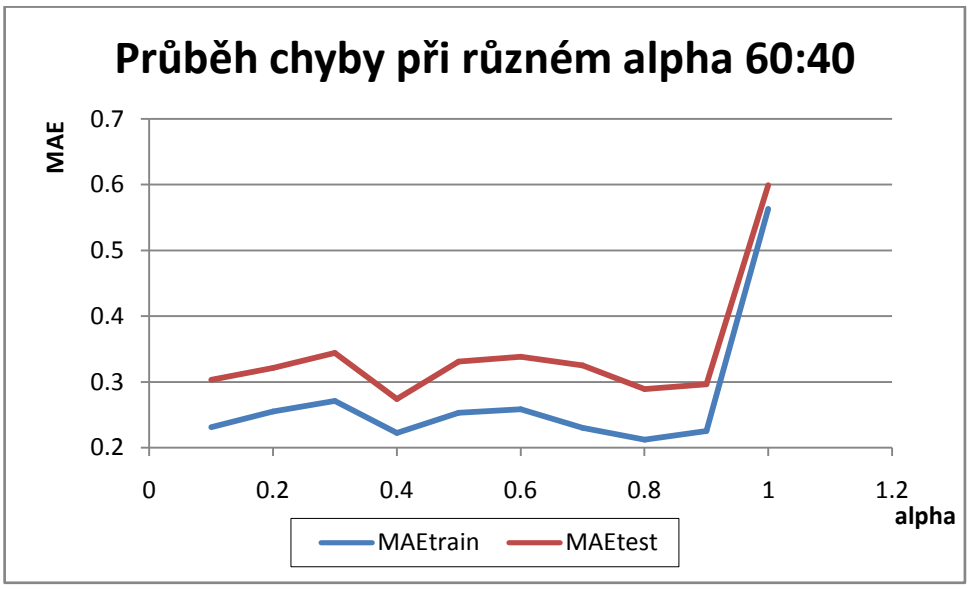
| Poměr | Parametr | Střední hodnota | Odchylka | Minimum | Maximum | Počet |
|---|---|---|---|---|---|---|
| 50:50 | JKP | 0.01 | 1.02 | -3.07 | 2.94 | 238 |
| 50:50 | CKP | 0.02 | 1.02 | -2.90 | 2.94 | 238 |
| 50:50 | KM | 0.02 | 1.01 | -2.97 | 2.69 | 238 |
| 50:50 | JEV | 0.03 | 1.02 | -2.24 | 2.62 | 238 |
| 50:50 | DEV | 0.03 | 1.02 | -3.15 | 2.83 | 238 |
| 50:50 | y | 0.00 | 1.00 | -2.39 | 2.70 | 238 |
| 60:40 | JKP | 0.07 | 1.01 | -3.07 | 2.94 | 194 |
| 60:40 | CKP | 0.08 | 1.03 | -2.90 | 2.94 | 194 |
| 60:40 | KM | 0.07 | 1.01 | -2.97 | 2.69 | 194 |
| 60:40 | JEV | 0.08 | 1.03 | -2.24 | 2.62 | 194 |
| 60:40 | DEV | 0.07 | 1.04 | -3.15 | 2.83 | 194 |
| 60:40 | y | 0.03 | 1.00 | -2.39 | 2.70 | 194 |
| 70:30 | JKP | 0.04 | 1.03 | -2.02 | 2.94 | 149 |
| 70:30 | CKP | 0.04 | 1.05 | -2.11 | 2.94 | 149 |
| 70:30 | KM | 0.03 | 1.04 | -2.97 | 2.69 | 149 |
| 70:30 | JEV | 0.03 | 1.10 | -2.24 | 2.62 | 149 |
| 70:30 | DEV | 0.03 | 1.08 | -3.15 | 2.83 | 149 |
| 70:30 | y | -0.01 | 1.02 | -2.39 | 2.70 | 149 |
| 80:20 | JKP | 0.07 | 1.04 | -2.02 | 2.94 | 113 |
| 80:20 | CKP | 0.05 | 1.07 | -2.11 | 2.94 | 113 |
| 80:20 | KM | 0.04 | 1.08 | -2.97 | 2.69 | 113 |
| 80:20 | JEV | 0.05 | 1.14 | -2.24 | 2.62 | 113 |
| 80:20 | DEV | 0.04 | 1.12 | -3.15 | 2.83 | 113 |
| 80:20 | y | -0.01 | 1.03 | -2.39 | 2.70 | 113 |
| 90:10 | JKP | 0.18 | 1.02 | -1.91 | 2.67 | 65 |
| 90:10 | CKP | 0.18 | 1.07 | -2.11 | 2.84 | 65 |
| 90:10 | KM | 0.15 | 1.14 | -2.97 | 2.69 | 65 |
| 90:10 | JEV | 0.20 | 1.13 | -2.09 | 2.62 | 65 |
| 90:10 | DEV | 0.15 | 1.19 | -3.15 | 2.83 | 65 |
| 90:10 | y | 0.08 | 1.11 | -2.39 | 2.70 | 65 |

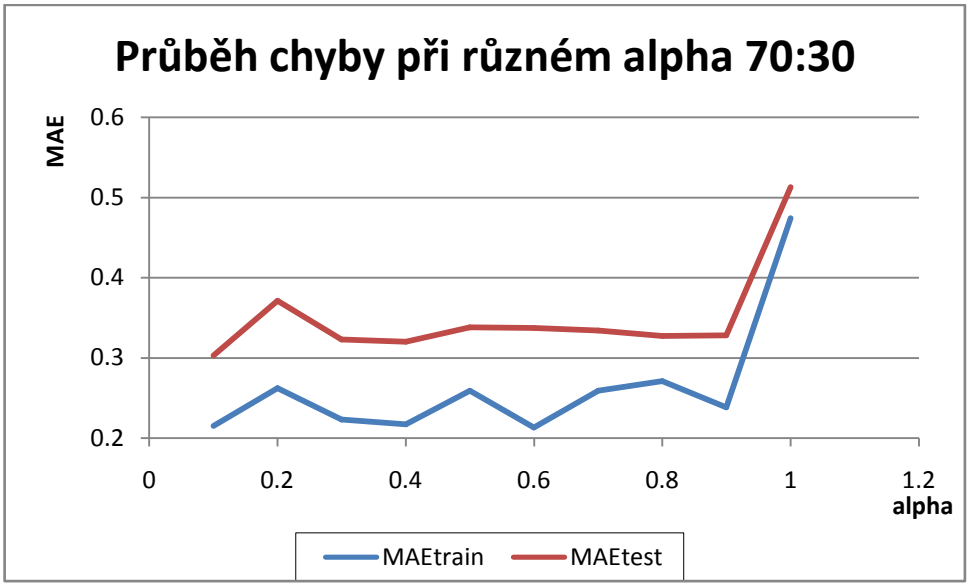
## PŘÍLOHA 8: PRŮBĚH CHYBY PŘI RŮZNÉM ALPHA STŘEDNĚDOBÉ ČR

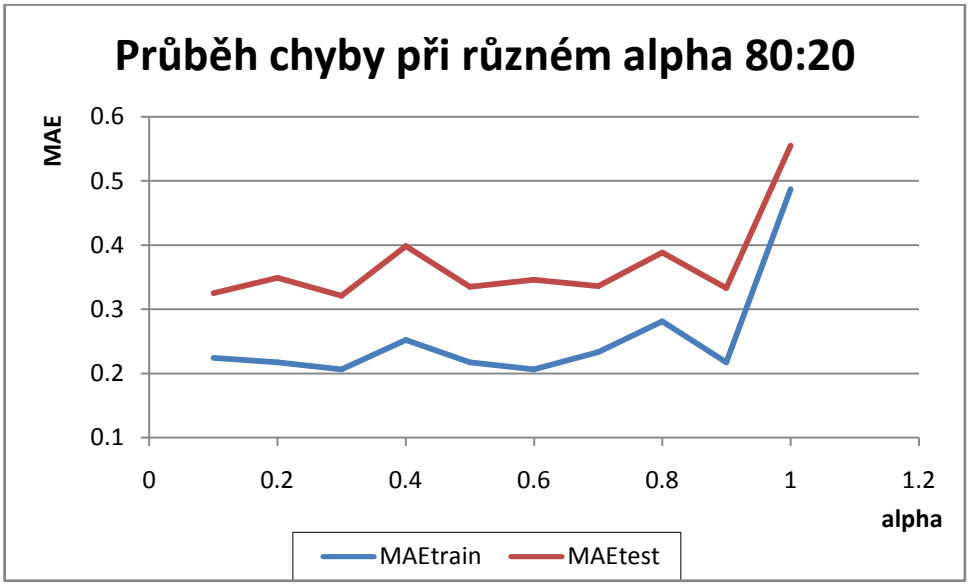
| počet neuronů | alpha | ν | Počet cyklů | Persistence | Qtrain:Qtest | MAEtrain | MAEtest |
|---|---|---|---|---|---|---|---|
| 125 | 1 | 1 | 600 | 30 | 50:50 | 0.489 | 0.581 |
| 125 | 0.9 | 1 | 600 | 30 | 50:50 | 0.236 | 0.329 |
| 125 | 0.8 | 1 | 600 | 30 | 50:50 | 0.247 | 0.338 |
| 125 | 0.7 | 1 | 600 | 30 | 50:50 | 0.246 | 0.322 |
| 125 | 0.6 | 1 | 600 | 30 | 50:50 | 0.325 | 0.381 |
| 125 | 0.5 | 1 | 600 | 30 | 50:50 | 0.258 | 0.352 |
| 125 | 0.4 | 1 | 600 | 30 | 50:50 | 0.223 | 0.315 |
| 125 | 0.3 | 1 | 600 | 30 | 50:50 | 0.339 | 0.437 |
| 125 | 0.2 | 1 | 600 | 30 | 50:50 | 0.288 | 0.345 |
| 125 | 0.1 | 1 | 600 | 30 | 50:50 | 0.235 | 0.307 |
| 125 | 1 | 1 | 600 | 30 | 60:40 | 0.563 | 0.599 |
| 125 | 0.9 | 1 | 600 | 30 | 60:40 | 0.225 | 0.296 |
| 125 | 0.8 | 1 | 600 | 30 | 60:40 | 0.212 | 0.289 |
| 125 | 0.7 | 1 | 600 | 30 | 60:40 | 0.23 | 0.325 |
| 125 | 0.6 | 1 | 600 | 30 | 60:40 | 0.258 | 0.338 |
| 125 | 0.5 | 1 | 600 | 30 | 60:40 | 0.253 | 0.331 |
| 125 | 0.4 | 1 | 600 | 30 | 60:40 | 0.222 | 0.274 |
| 125 | 0.3 | 1 | 600 | 30 | 60:40 | 0.271 | 0.344 |
| 125 | 0.2 | 1 | 600 | 30 | 60:40 | 0.255 | 0.321 |
| 125 | 0.1 | 1 | 600 | 30 | 60:40 | 0.231 | 0.303 |
| 125 | 1 | 1 | 600 | 30 | 70:30 | 0.474 | 0.513 |
| 125 | 0.9 | 1 | 600 | 30 | 70:30 | 0.238 | 0.328 |
| 125 | 0.8 | 1 | 600 | 30 | 70:30 | 0.271 | 0.327 |
| 125 | 0.7 | 1 | 600 | 30 | 70:30 | 0.259 | 0.334 |
| 125 | 0.6 | 1 | 600 | 30 | 70:30 | 0.213 | 0.337 |
| 125 | 0.5 | 1 | 600 | 30 | 70:30 | 0.259 | 0.338 |
| 125 | 0.4 | 1 | 600 | 30 | 70:30 | 0.217 | 0.32 |
| 125 | 0.3 | 1 | 600 | 30 | 70:30 | 0.223 | 0.323 |
| 125 | 0.2 | 1 | 600 | 30 | 70:30 | 0.262 | 0.371 |
| 125 | 0.1 | 1 | 600 | 30 | 70:30 | 0.215 | 0.303 |
| 125 | 1 | 1 | 600 | 30 | 80:20 | 0.487 | 0.555 |
| 125 | 0.9 | 1 | 600 | 30 | 80:20 | 0.217 | 0.333 |
| 125 | 0.8 | 1 | 600 | 30 | 80:20 | 0.281 | 0.388 |
| 125 | 0.7 | 1 | 600 | 30 | 80:20 | 0.233 | 0.336 |
| 125 | 0.6 | 1 | 600 | 30 | 80:20 | 0.206 | 0.346 |
| 125 | 0.5 | 1 | 600 | 30 | 80:20 | 0.217 | 0.335 |
| 125 | 0.4 | 1 | 600 | 30 | 80:20 | 0.252 | 0.398 |
| 125 | 0.3 | 1 | 600 | 30 | 80:20 | 0.206 | 0.321 |
| 125 | 0.2 | 1 | 600 | 30 | 80:20 | 0.217 | 0.349 |
| 125 | 0.1 | 1 | 600 | 30 | 80:20 | 0.224 | 0.325 |

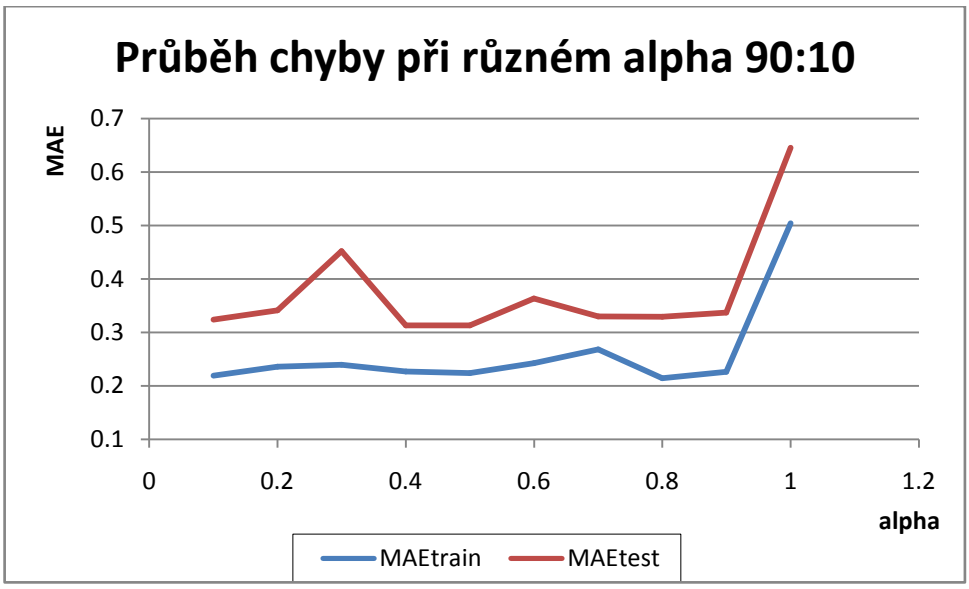
| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 125 | 1 | 1 | 600 | 30 | 90:10 | 0.504 | 0.645 |
| 125 | 0.9 | 1 | 600 | 30 | 90:10 | 0.226 | 0.337 |
| 125 | 0.8 | 1 | 600 | 30 | 90:10 | 0.214 | 0.329 |
| 125 | 0.7 | 1 | 600 | 30 | 90:10 | 0.268 | 0.33 |
| 125 | 0.6 | 1 | 600 | 30 | 90:10 | 0.242 | 0.363 |
| 125 | 0.5 | 1 | 600 | 30 | 90:10 | 0.224 | 0.313 |
| 125 | 0.4 | 1 | 600 | 30 | 90:10 | 0.227 | 0.313 |
| 125 | 0.3 | 1 | 600 | 30 | 90:10 | 0.239 | 0.452 |
| 125 | 0.2 | 1 | 600 | 30 | 90:10 | 0.236 | 0.341 |
| 125 | 0.1 | 1 | 600 | 30 | 90:10 | 0.219 | 0.324 |




Průběh chyby při různém alpha 70:30
MAE
0.6
0.5
0.4
0.3
0.2
0
0.2
0.4
0.6
0.8
1
1.2
alpha
MAEtrain
MAEtest

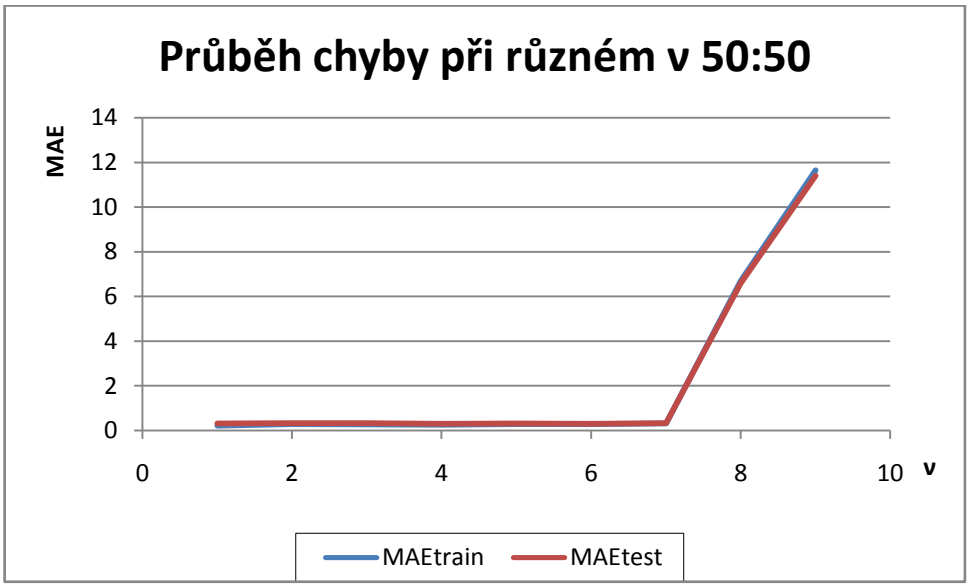

Průběh chyby při různém alpha 80:20
MAE
0.6
0.5
0.4
0.3
0.2
0.1
0
0.2
0.4
0.6
0.8
1
1.2
alpha
MAEtrain
MAEtest

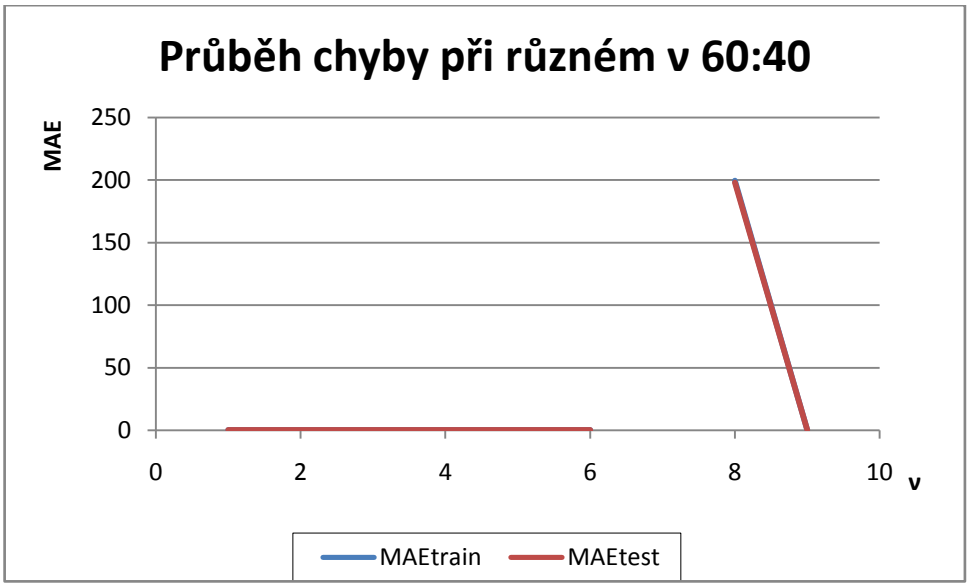

Průběh chyby při různém alpha 90:10
MAE
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0.2
0.4
0.6
0.8
1
1.2
alpha
MAEtrain
MAEtest

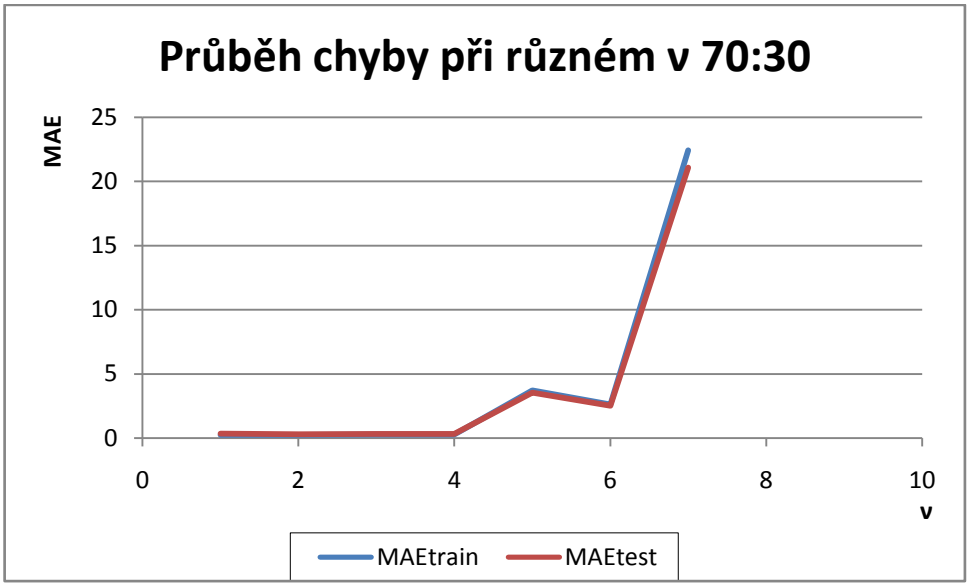
## PŘÍLOHA 9: PRŮBĚH CHYBY PŘI RŮZNÉM ν KRÁTKODOBÉ ČŘ

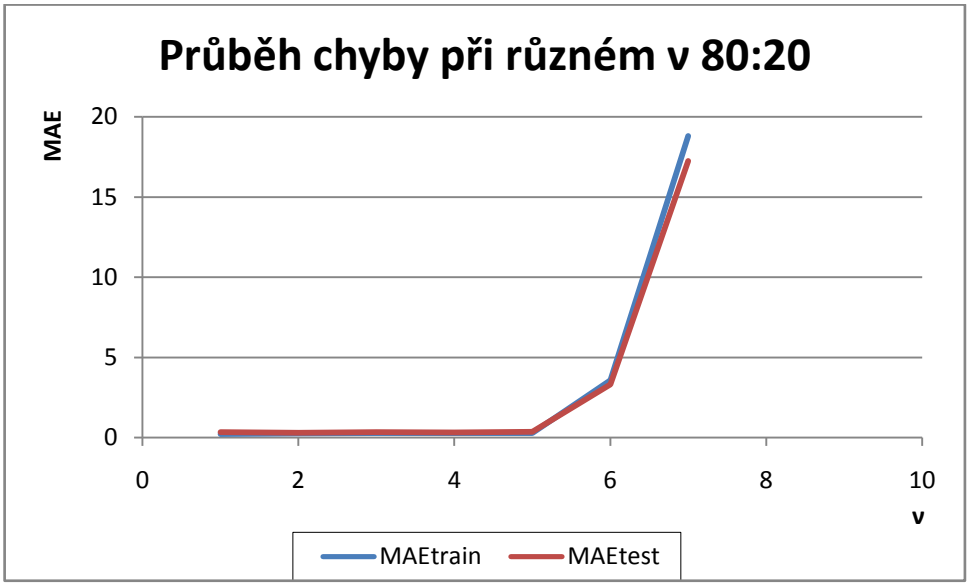
| počet neuronů | alpha | ν | počet cyklů | persistence | Qtrain:Qtest | MAEtrain | MAEtest |
|---|---|---|---|---|---|---|---|
| 125 | 0.1 | 1 | 600 | 30 | 50:50 | 0.214 | 0.308 |
| 125 | 0.1 | 2 | 600 | 30 | 50:50 | 0.28 | 0.316 |
| 125 | 0.1 | 3 | 600 | 30 | 50:50 | 0.276 | 0.313 |
| 125 | 0.1 | 4 | 600 | 30 | 50:50 | 0.261 | 0.292 |
| 125 | 0.1 | 5 | 600 | 30 | 50:50 | 0.284 | 0.304 |
| 125 | 0.1 | 6 | 600 | 30 | 50:50 | 0.281 | 0.291 |
| 125 | 0.1 | 7 | 600 | 30 | 50:50 | 0.315 | 0.32 |
| 125 | 0.1 | 8 | 600 | 30 | 50:50 | 6.694 | 6.585 |
| 125 | 0.1 | 9 | 600 | 30 | 50:50 | 11.648 | 11.4 |
| 125 | 0.4 | 1 | 600 | 30 | 60:40 | 0.241 | 0.304 |
| 125 | 0.4 | 2 | 600 | 30 | 60:40 | 0.297 | 0.331 |
| 125 | 0.4 | 3 | 600 | 30 | 60:40 | 0.245 | 0.289 |
| 125 | 0.4 | 4 | 600 | 30 | 60:40 | 0.268 | 0.31 |
| 125 | 0.4 | 5 | 600 | 30 | 60:40 | 0.267 | 0.284 |
| 125 | 0.4 | 6 | 600 | 30 | 60:40 | 0.3 | 0.321 |
| 125 | 0.4 | 7 | 600 | 30 | 60:40 | | |
| 125 | 0.4 | 8 | 600 | 30 | 60:40 | 199.538 | 197.89 |
| 125 | 0.4 | 9 | 600 | 30 | 60:40 | 0.27 | 0.302 |
| 125 | 0.1 | 1 | 600 | 30 | 70:30 | 0.246 | 0.345 |
| 125 | 0.1 | 2 | 600 | 30 | 70:30 | 0.251 | 0.285 |
| 125 | 0.1 | 3 | 600 | 30 | 70:30 | 0.271 | 0.315 |
| 125 | 0.1 | 4 | 600 | 30 | 70:30 | 0.26 | 0.311 |
| 125 | 0.1 | 5 | 600 | 30 | 70:30 | 3.726 | 3.521 |
| 125 | 0.1 | 6 | 600 | 30 | 70:30 | 2.61 | 2.496 |
| 125 | 0.1 | 7 | 600 | 30 | 70:30 | 22.427 | 21.068 |
| 125 | 0.1 | 8 | 600 | 30 | 70:30 | | |
| 125 | 0.1 | 9 | 600 | 30 | 70:30 | | |
| 125 | 0.3 | 1 | 600 | 30 | 80:20 | 0.208 | 0.322 |
| 125 | 0.3 | 2 | 600 | 30 | 80:20 | 0.243 | 0.292 |
| 125 | 0.3 | 3 | 600 | 30 | 80:20 | 0.259 | 0.328 |
| 125 | 0.3 | 4 | 600 | 30 | 80:20 | 0.258 | 0.318 |
| 125 | 0.3 | 5 | 600 | 30 | 80:20 | 0.276 | 0.347 |
| 125 | 0.3 | 6 | 600 | 30 | 80:20 | 3.553 | 3.31 |
| 125 | 0.3 | 7 | 600 | 30 | 80:20 | 18.789 | 17.233 |
| 125 | 0.3 | 8 | 600 | 30 | 80:20 | | |
| 125 | 0.3 | 9 | 600 | 30 | 80:20 | | |
| 125 | 0.5 | 1 | 600 | 30 | 90:10 | 0.239 | 0.31 |
| 125 | 0.5 | 2 | 600 | 30 | 90:10 | 0.264 | 0.347 |
| 125 | 0.5 | 3 | 600 | 30 | 90:10 | 0.255 | 0.29 |
| 125 | 0.5 | 4 | 600 | 30 | 90:10 | 0.256 | 0.254 |

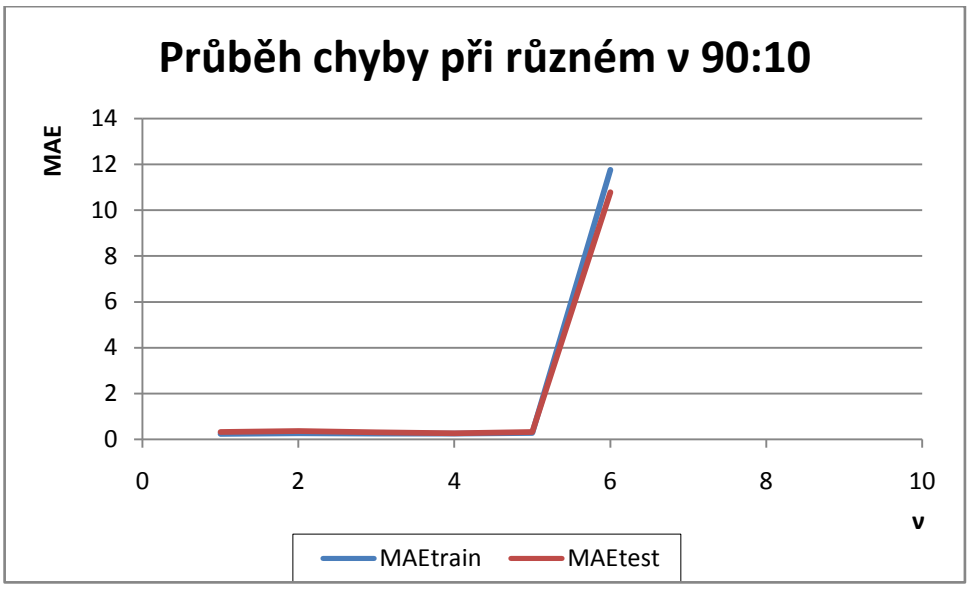
| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 125 | 0.5 | 5 | 600 | 30 | 90:10 | 0.276 | 0.304 |
| 125 | 0.5 | 6 | 600 | 30 | 90:10 | 11.766 | 10.787 |
| 125 | 0.5 | 7 | 600 | 30 | 90:10 | | |
| 125 | 0.5 | 8 | 600 | 30 | 90:10 | | |
| 125 | 0.5 | 9 | 600 | 30 | 90:10 | | |



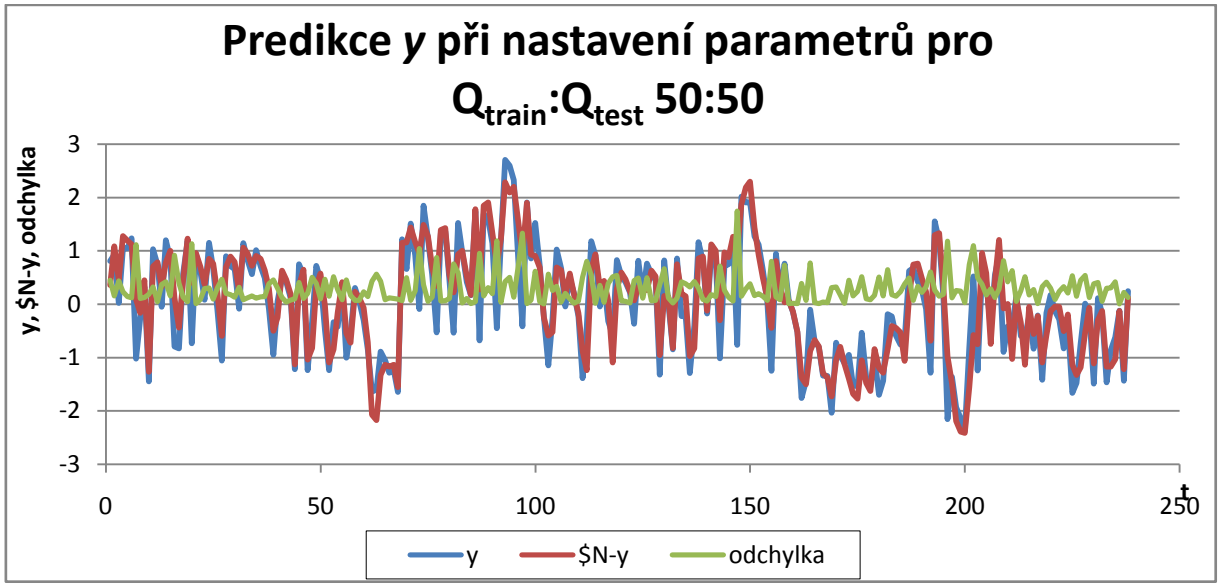



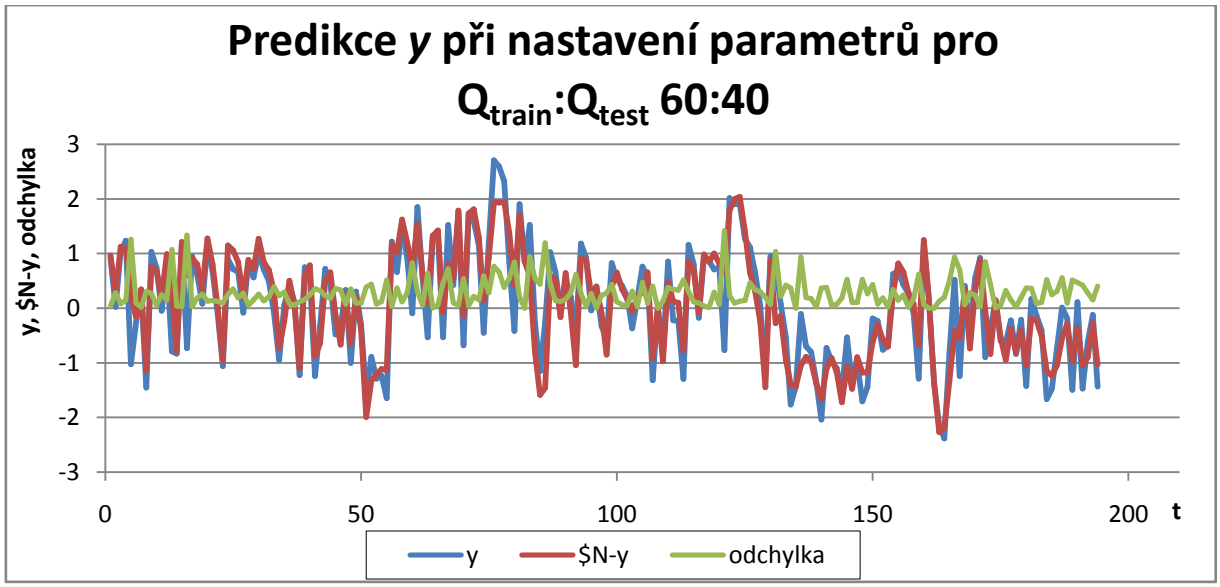

Průběh chyby při různém v 70:30
MAE
25
20
15
10
5
0
0
2
4
6
8
10
v
MAEtrain
MAEtest

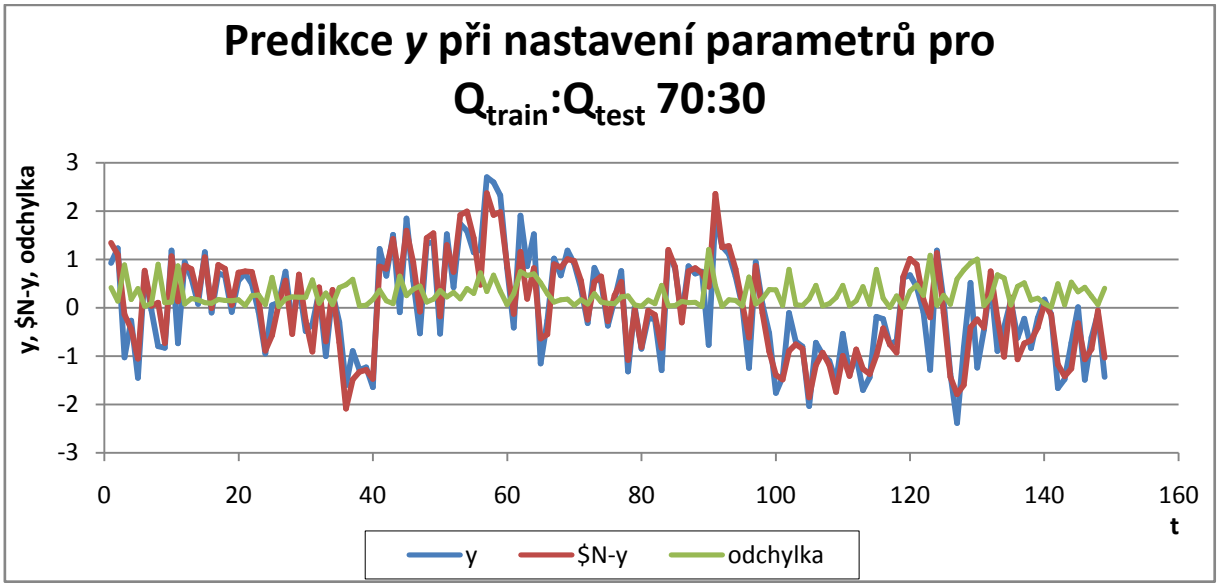

Průběh chyby při různém v 80:20
MAE
20
15
10
5
0
0
2
4
6
8
10
v
MAEtrain
MAEtest

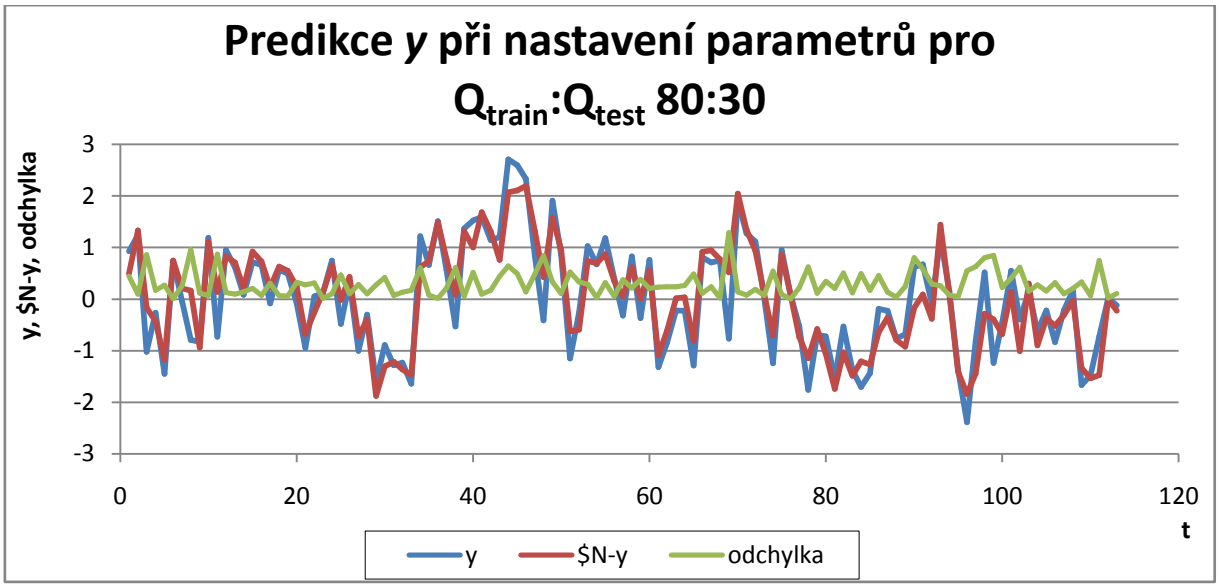

Průběh chyby při různém v 90:10
MAE
14
12
10
8
6
4
2
0
0
2
4
6
8
10
v
MAEtrain
MAEtest

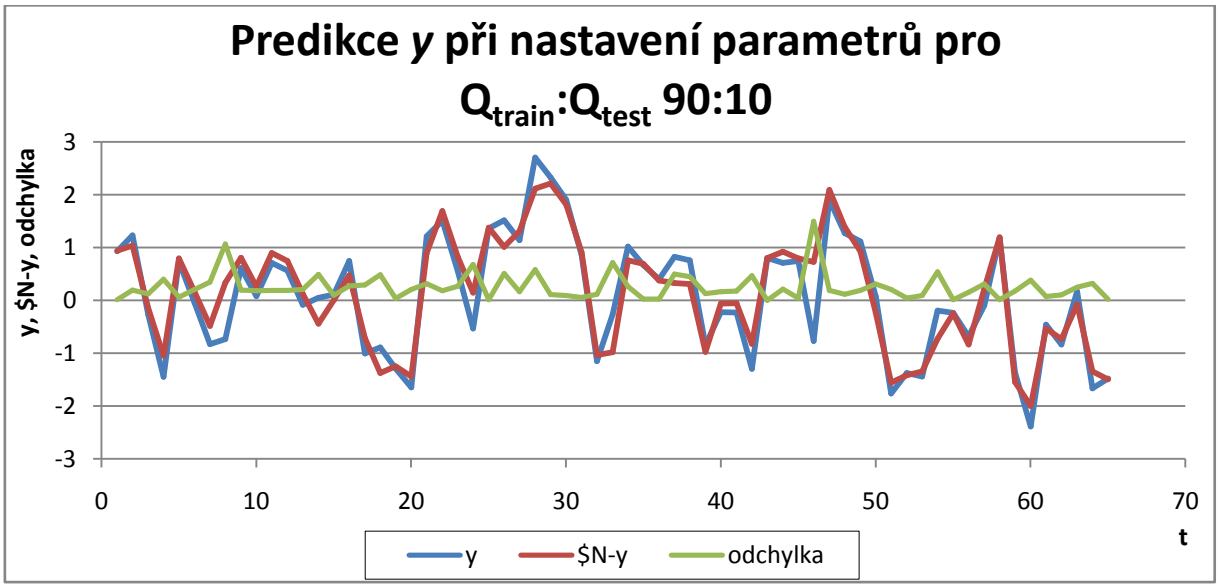
## PŘÍLOHA 10: PREDIKCE y V JEDNOTLIVÝCH ROZDĚLENÍCH DAT STŘEDNĚDOBÉ ČR




Predikce y při nastavení parametrů pro
Qtrain:Qtest 70:30
y, $N-y, odchylka
3
2
1
0
-1
-2
-3
0
20
40
60
80
100
120
140
160
t
y
$N-y
odchylka


Predikce y při nastavení parametrů pro
Qtrain:Qtest 80:30
y, $N-y, odchylka
3
2
1
0
-1
-2
-3
0
20
40
60
80
100
120
t
y
$N-y
odchylka


Predikce y při nastavení parametrů pro
Qtrain:Qtest 90:10
y, $N-y, odchylka
3
2
1
0
-1
-2
-3
0
10
20
30
40
50
60
70
t
y
$N-y
odchylka

## PŘÍLOHA 11: STATISTIKY TRÉNOVACÍCH DAT DLOUHODOBÉ ČŘ

| Poměr | Parametr | Střední hodnota | Odchylka | Minimum | Maximum | Počet |
|---|---|---|---|---|---|---|
| 50:50 | JKP | 0.04 | 1.01 | -2.47 | 2.41 | 374 |
| 50:50 | CKP | 0.04 | 1.01 | -2.65 | 2.43 | 374 |
| 50:50 | KM | 0.04 | 1.01 | -2.71 | 2.04 | 374 |
| 50:50 | JEV | 0.00 | 0.99 | -2.10 | 2.42 | 374 |
| 50:50 | DEV | 0.02 | 0.99 | -2.98 | 2.93 | 374 |
| 50:50 | y | 0.04 | 1.00 | -2.50 | 2.72 | 374 |
| 60:40 | JKP | -0.01 | 1.01 | -2.47 | 2.41 | 444 |
| 60:40 | CKP | -0.01 | 1.01 | -2.65 | 2.43 | 444 |
| 60:40 | KM | 0.00 | 1.01 | -2.71 | 2.04 | 444 |
| 60:40 | JEV | -0.04 | 0.99 | -2.12 | 2.42 | 444 |
| 60:40 | DEV | -0.02 | 0.99 | -2.98 | 2.93 | 444 |
| 60:40 | y | 0.00 | 1.00 | -2.50 | 2.72 | 444 |
| 70:30 | JKP | -0.02 | 1.01 | -2.47 | 2.41 | 531 |
| 70:30 | CKP | -0.02 | 1.01 | -2.65 | 2.43 | 531 |
| 70:30 | KM | -0.01 | 1.01 | -2.71 | 2.04 | 531 |
| 70:30 | JEV | -0.03 | 0.99 | -2.12 | 2.42 | 531 |
| 70:30 | DEV | -0.02 | 0.99 | -2.98 | 2.93 | 531 |
| 70:30 | y | -0.01 | 1.00 | -2.50 | 2.72 | 531 |
| 80:20 | JKP | -0.02 | 1.01 | -2.47 | 2.47 | 597 |
| 80:20 | CKP | -0.02 | 1.01 | -2.65 | 2.48 | 597 |
| 80:20 | KM | -0.01 | 1.00 | -2.71 | 2.04 | 597 |
| 80:20 | JEV | -0.03 | 0.99 | -2.12 | 2.42 | 597 |
| 80:20 | DEV | -0.02 | 0.99 | -2.98 | 2.95 | 597 |
| 80:20 | y | 0.00 | 1.00 | -2.50 | 3.36 | 597 |
| 90:10 | JKP | -0.02 | 1.01 | -2.47 | 2.47 | 671 |
| 90:10 | CKP | -0.02 | 1.01 | -2.65 | 2.48 | 671 |
| 90:10 | KM | -0.02 | 0.99 | -2.71 | 2.04 | 671 |
| 90:10 | JEV | -0.03 | 1.00 | -2.12 | 2.42 | 671 |
| 90:10 | DEV | -0.02 | 0.99 | -2.98 | 2.95 | 671 |
| 90:10 | y | -0.01 | 1.00 | -2.50 | 3.36 | 671 |

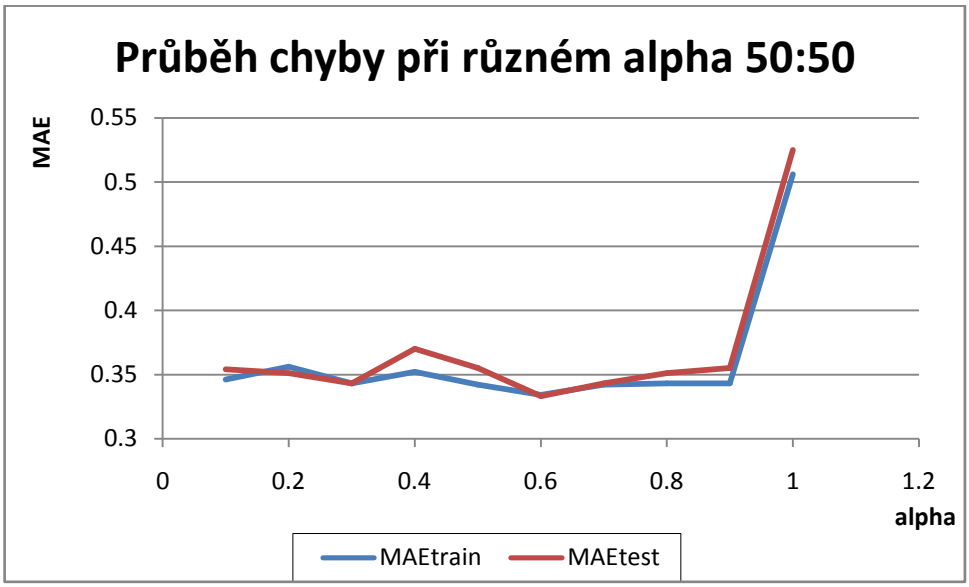
## PŘÍLOHA 12: STATISTIKY TESTOVACÍCH DAT DLOUHODOBÉ ČŘ

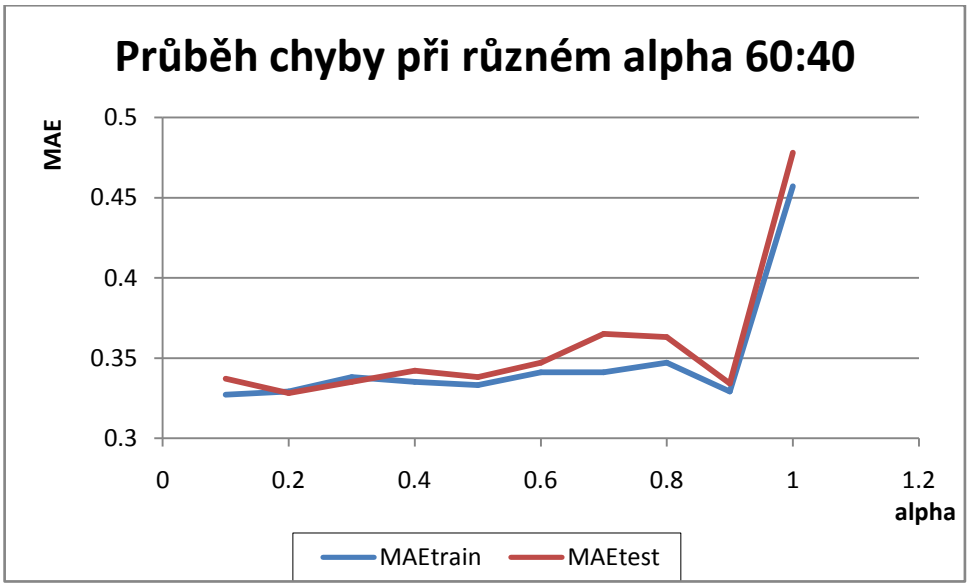
| Poměr | Parametr | Střední hodnota | Odchylka | Minimum | Maximum | Počet |
|---|---|---|---|---|---|---|
| 50:50 | JKP | -0.04 | 0.99 | -2.46 | 2.47 | 378 |
| 50:50 | CKP | -0.04 | 0.99 | -2.25 | 2.48 | 378 |
| 50:50 | KM | -0.04 | 0.99 | -2.71 | 2.04 | 378 |
| 50:50 | JEV | 0.00 | 1.01 | -2.12 | 2.40 | 378 |
| 50:50 | DEV | -0.02 | 1.01 | -2.98 | 2.95 | 378 |
| 50:50 | y | -0.04 | 1.00 | -2.28 | 3.36 | 378 |
| 60:40 | JKP | 0.01 | 0.99 | -2.46 | 2.47 | 308 |
| 60:40 | CKP | 0.01 | 0.99 | -2.25 | 2.48 | 308 |
| 60:40 | KM | 0.00 | 0.99 | -2.71 | 2.04 | 308 |
| 60:40 | JEV | 0.05 | 1.02 | -2.06 | 2.40 | 308 |
| 60:40 | DEV | 0.03 | 1.02 | -2.98 | 2.95 | 308 |
| 60:40 | y | -0.01 | 1.01 | -2.28 | 3.36 | 308 |
| 70:30 | JKP | 0.04 | 0.98 | -2.02 | 2.47 | 221 |
| 70:30 | CKP | 0.04 | 0.98 | -1.95 | 2.48 | 221 |
| 70:30 | KM | 0.02 | 0.99 | -2.71 | 2.04 | 221 |
| 70:30 | JEV | 0.08 | 1.03 | -2.06 | 2.40 | 221 |
| 70:30 | DEV | 0.06 | 1.03 | -2.98 | 2.95 | 221 |
| 70:30 | y | 0.03 | 1.01 | -2.28 | 3.36 | 221 |
| 80:20 | JKP | 0.08 | 0.97 | -1.93 | 2.33 | 155 |
| 80:20 | CKP | 0.08 | 0.97 | -1.95 | 2.27 | 155 |
| 80:20 | KM | 0.05 | 1.01 | -2.71 | 2.02 | 155 |
| 80:20 | JEV | 0.11 | 1.04 | -2.06 | 2.40 | 155 |
| 80:20 | DEV | 0.06 | 1.04 | -2.98 | 2.68 | 155 |
| 80:20 | y | 0.00 | 0.99 | -2.28 | 2.59 | 155 |
| 90:10 | JKP | 0.17 | 0.95 | -1.88 | 2.13 | 81 |
| 90:10 | CKP | 0.17 | 0.95 | -1.89 | 2.15 | 81 |
| 90:10 | KM | 0.12 | 1.06 | -2.71 | 2.02 | 81 |
| 90:10 | JEV | 0.23 | 1.01 | -1.89 | 2.40 | 81 |
| 90:10 | DEV | 0.16 | 1.09 | -2.98 | 2.68 | 81 |
| 90:10 | y | 0.08 | 1.04 | -2.28 | 2.59 | 81 |

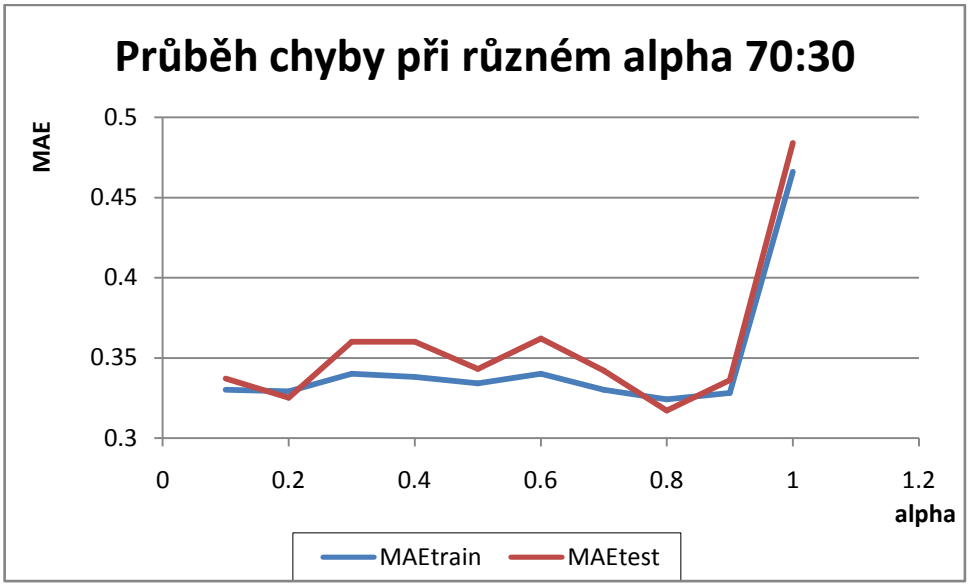
## PŘÍLOHA 13: PRŮBĚH CHYBY PŘI RŮZNÉM ALPHA DLOUHODOBÉ ČŘ

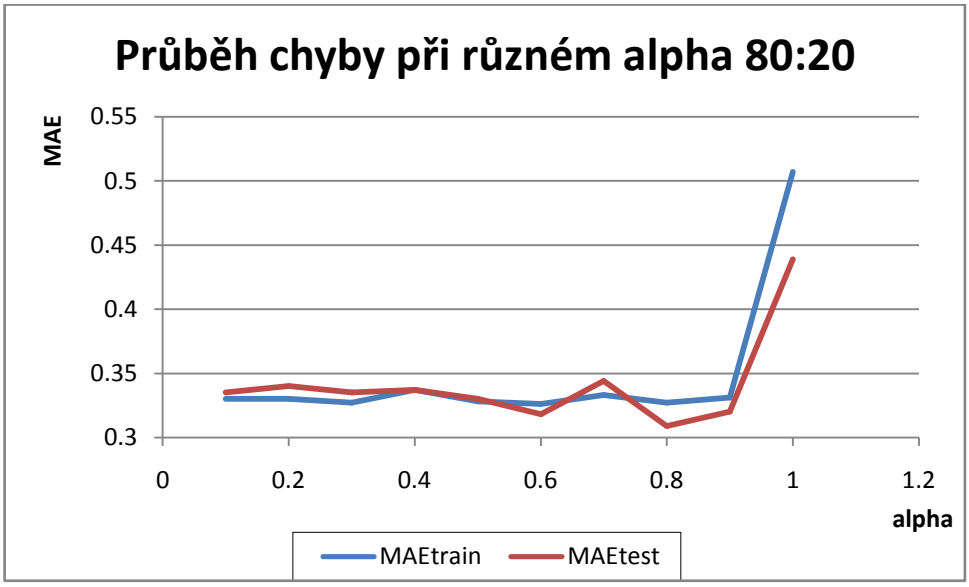
| počet neuronů | alpha | ν | Počet cyklů | Persistence | Qtrain:Qtest | MAEtrain | MAEtest |
|---|---|---|---|---|---|---|---|
| 25 | 1 | 1 | 600 | 30 | 50:50 | 0.506 | 0.525 |
| 25 | 0.9 | 1 | 600 | 30 | 50:50 | 0.343 | 0.355 |
| 25 | 0.8 | 1 | 600 | 30 | 50:50 | 0.343 | 0.351 |
| 25 | 0.7 | 1 | 600 | 30 | 50:50 | 0.342 | 0.343 |
| 25 | 0.6 | 1 | 600 | 30 | 50:50 | 0.334 | 0.333 |
| 25 | 0.5 | 1 | 600 | 30 | 50:50 | 0.342 | 0.355 |
| 25 | 0.4 | 1 | 600 | 30 | 50:50 | 0.352 | 0.37 |
| 25 | 0.3 | 1 | 600 | 30 | 50:50 | 0.343 | 0.343 |
| 25 | 0.2 | 1 | 600 | 30 | 50:50 | 0.356 | 0.351 |
| 25 | 0.1 | 1 | 600 | 30 | 50:50 | 0.346 | 0.354 |
| 25 | 1 | 1 | 600 | 30 | 60:40 | 0.457 | 0.478 |
| 25 | 0.9 | 1 | 600 | 30 | 60:40 | 0.329 | 0.334 |
| 25 | 0.8 | 1 | 600 | 30 | 60:40 | 0.347 | 0.363 |
| 25 | 0.7 | 1 | 600 | 30 | 60:40 | 0.341 | 0.365 |
| 25 | 0.6 | 1 | 600 | 30 | 60:40 | 0.341 | 0.347 |
| 25 | 0.5 | 1 | 600 | 30 | 60:40 | 0.333 | 0.338 |
| 25 | 0.4 | 1 | 600 | 30 | 60:40 | 0.335 | 0.342 |
| 25 | 0.3 | 1 | 600 | 30 | 60:40 | 0.338 | 0.335 |
| 25 | 0.2 | 1 | 600 | 30 | 60:40 | 0.329 | 0.328 |
| 25 | 0.1 | 1 | 600 | 30 | 60:40 | 0.327 | 0.337 |
| 25 | 1 | 1 | 600 | 30 | 70:30 | 0.466 | 0.484 |
| 25 | 0.9 | 1 | 600 | 30 | 70:30 | 0.328 | 0.336 |
| 25 | 0.8 | 1 | 600 | 30 | 70:30 | 0.324 | 0.317 |
| 25 | 0.7 | 1 | 600 | 30 | 70:30 | 0.33 | 0.342 |
| 25 | 0.6 | 1 | 600 | 30 | 70:30 | 0.34 | 0.362 |
| 25 | 0.5 | 1 | 600 | 30 | 70:30 | 0.334 | 0.343 |
| 25 | 0.4 | 1 | 600 | 30 | 70:30 | 0.338 | 0.36 |
| 25 | 0.3 | 1 | 600 | 30 | 70:30 | 0.34 | 0.36 |
| 25 | 0.2 | 1 | 600 | 30 | 70:30 | 0.329 | 0.325 |
| 25 | 0.1 | 1 | 600 | 30 | 70:30 | 0.33 | 0.337 |
| 25 | 1 | 1 | 600 | 30 | 80:20 | 0.507 | 0.439 |
| 25 | 0.9 | 1 | 600 | 30 | 80:20 | 0.331 | 0.32 |
| 25 | 0.8 | 1 | 600 | 30 | 80:20 | 0.327 | 0.309 |
| 25 | 0.7 | 1 | 600 | 30 | 80:20 | 0.333 | 0.344 |
| 25 | 0.6 | 1 | 600 | 30 | 80:20 | 0.326 | 0.318 |
| 25 | 0.5 | 1 | 600 | 30 | 80:20 | 0.328 | 0.33 |
| 25 | 0.4 | 1 | 600 | 30 | 80:20 | 0.337 | 0.337 |
| 25 | 0.3 | 1 | 600 | 30 | 80:20 | 0.327 | 0.335 |
| 25 | 0.2 | 1 | 600 | 30 | 80:20 | 0.33 | 0.34 |
| 25 | 0.1 | 1 | 600 | 30 | 80:20 | 0.33 | 0.335 |

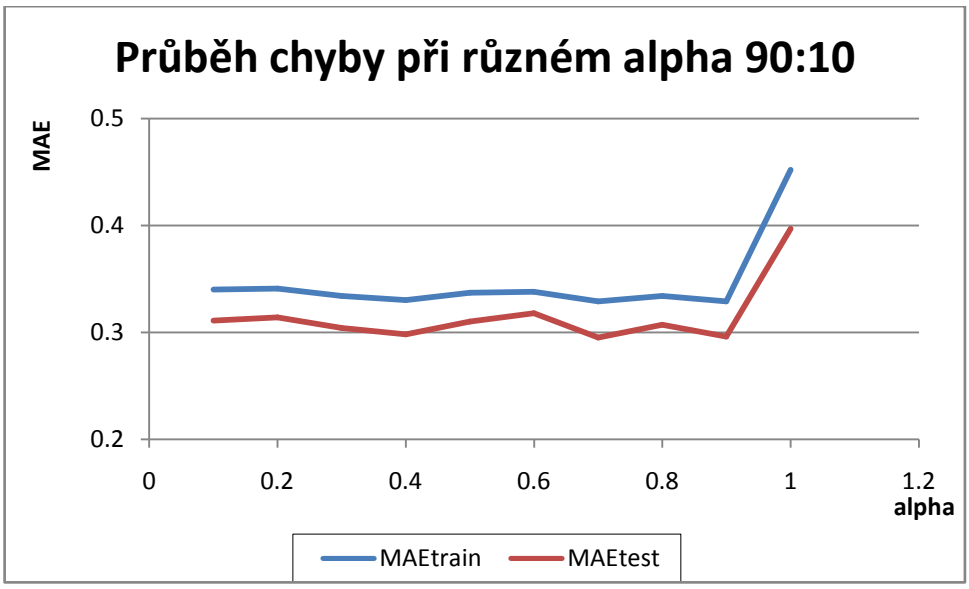
| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 25 | 1 | 1 | 600 | 30 | 90:10 | 0.452 | 0.397 |
| 25 | 0.9 | 1 | 600 | 30 | 90:10 | 0.329 | 0.296 |
| 25 | 0.8 | 1 | 600 | 30 | 90:10 | 0.334 | 0.307 |
| 25 | 0.7 | 1 | 600 | 30 | 90:10 | 0.329 | 0.295 |
| 25 | 0.6 | 1 | 600 | 30 | 90:10 | 0.338 | 0.318 |
| 25 | 0.5 | 1 | 600 | 30 | 90:10 | 0.337 | 0.31 |
| 25 | 0.4 | 1 | 600 | 30 | 90:10 | 0.33 | 0.298 |
| 25 | 0.3 | 1 | 600 | 30 | 90:10 | 0.334 | 0.304 |
| 25 | 0.2 | 1 | 600 | 30 | 90:10 | 0.341 | 0.314 |
| 25 | 0.1 | 1 | 600 | 30 | 90:10 | 0.34 | 0.311 |




Průběh chyby při různém alpha 70:30
MAE
0.5
0.45
0.4
0.35
0.3
0
0.2
0.4
0.6
0.8
1
1.2
alpha
MAEtrain
MAEtest

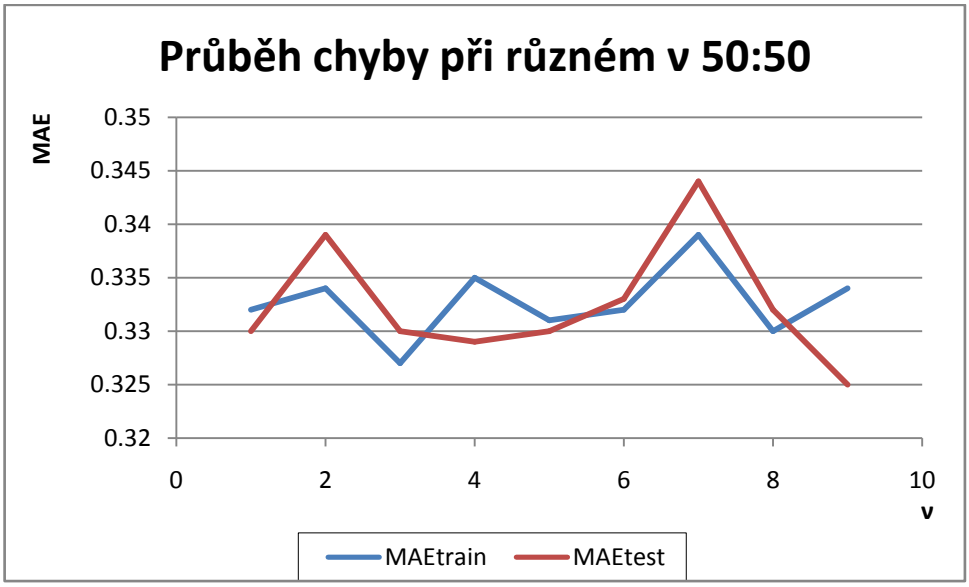

Průběh chyby při různém alpha 80:20
MAE
0.55
0.5
0.45
0.4
0.35
0.3
0
0.2
0.4
0.6
0.8
1
1.2
alpha
MAEtrain
MAEtest

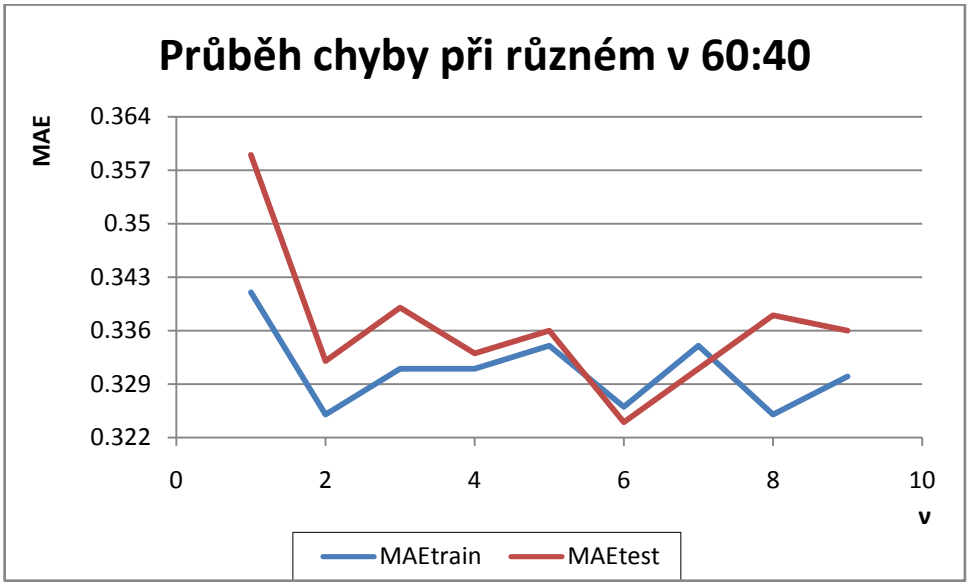

Průběh chyby při různém alpha 90:10
MAE
0.5
0.4
0.3
0.2
0
0.2
0.4
0.6
0.8
1
1.2
alpha
MAEtrain
MAEtest

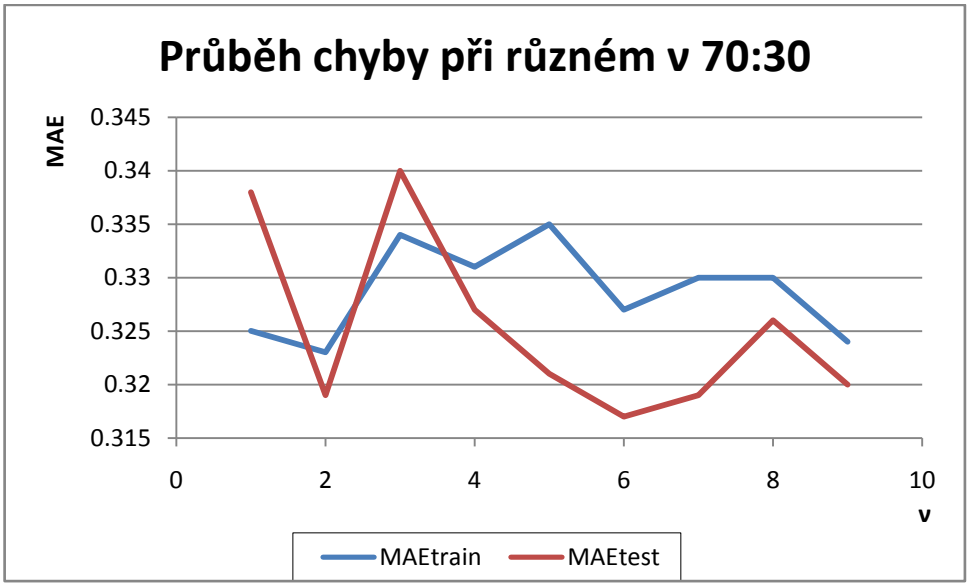
## PŘÍLOHA 14: PRŮBĚH CHYBY PŘI RŮZNÉM ν DLOUHODOBÉ ČŘ

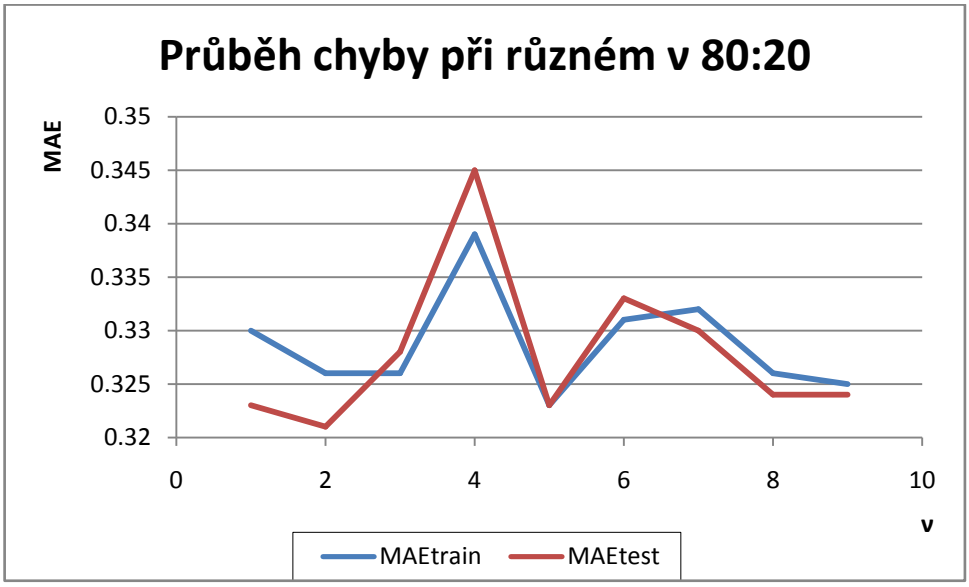
| počet neuronů | alpha | ν | Počet cyklů | Persistence | Qtrain:Qtest | MAEtrain | MAEtest |
|---|---|---|---|---|---|---|---|
| 25 | 0.6 | 1 | 600 | 30 | 50:50 | 0.332 | 0.33 |
| 25 | 0.6 | 2 | 600 | 30 | 50:50 | 0.334 | 0.339 |
| 25 | 0.6 | 3 | 600 | 30 | 50:50 | 0.327 | 0.33 |
| 25 | 0.6 | 4 | 600 | 30 | 50:50 | 0.335 | 0.329 |
| 25 | 0.6 | 5 | 600 | 30 | 50:50 | 0.331 | 0.33 |
| 25 | 0.6 | 6 | 600 | 30 | 50:50 | 0.332 | 0.333 |
| 25 | 0.6 | 7 | 600 | 30 | 50:50 | 0.339 | 0.344 |
| 25 | 0.6 | 8 | 600 | 30 | 50:50 | 0.33 | 0.332 |
| 25 | 0.6 | 9 | 600 | 30 | 50:50 | 0.334 | 0.325 |
| 25 | 0.2 | 1 | 600 | 30 | 60:40 | 0.341 | 0.359 |
| 25 | 0.2 | 2 | 600 | 30 | 60:40 | 0.325 | 0.332 |
| 25 | 0.2 | 3 | 600 | 30 | 60:40 | 0.331 | 0.339 |
| 25 | 0.2 | 4 | 600 | 30 | 60:40 | 0.331 | 0.333 |
| 25 | 0.2 | 5 | 600 | 30 | 60:40 | 0.334 | 0.336 |
| 25 | 0.2 | 6 | 600 | 30 | 60:40 | 0.326 | 0.324 |
| 25 | 0.2 | 7 | 600 | 30 | 60:40 | 0.334 | 0.331 |
| 25 | 0.2 | 8 | 600 | 30 | 60:40 | 0.325 | 0.338 |
| 25 | 0.2 | 9 | 600 | 30 | 60:40 | 0.33 | 0.336 |
| 25 | 0.8 | 1 | 600 | 30 | 70:30 | 0.325 | 0.338 |
| 25 | 0.8 | 2 | 600 | 30 | 70:30 | 0.323 | 0.319 |
| 25 | 0.8 | 3 | 600 | 30 | 70:30 | 0.334 | 0.34 |
| 25 | 0.8 | 4 | 600 | 30 | 70:30 | 0.331 | 0.327 |
| 25 | 0.8 | 5 | 600 | 30 | 70:30 | 0.335 | 0.321 |
| 25 | 0.8 | 6 | 600 | 30 | 70:30 | 0.327 | 0.317 |
| 25 | 0.8 | 7 | 600 | 30 | 70:30 | 0.33 | 0.319 |
| 25 | 0.8 | 8 | 600 | 30 | 70:30 | 0.33 | 0.326 |
| 25 | 0.8 | 9 | 600 | 30 | 70:30 | 0.324 | 0.32 |
| 25 | 0.8 | 1 | 600 | 30 | 80:20 | 0.33 | 0.323 |
| 25 | 0.8 | 2 | 600 | 30 | 80:20 | 0.326 | 0.321 |
| 25 | 0.8 | 3 | 600 | 30 | 80:20 | 0.326 | 0.328 |
| 25 | 0.8 | 4 | 600 | 30 | 80:20 | 0.339 | 0.345 |
| 25 | 0.8 | 5 | 600 | 30 | 80:20 | 0.323 | 0.323 |
| 25 | 0.8 | 6 | 600 | 30 | 80:20 | 0.331 | 0.333 |
| 25 | 0.8 | 7 | 600 | 30 | 80:20 | 0.332 | 0.33 |
| 25 | 0.8 | 8 | 600 | 30 | 80:20 | 0.326 | 0.324 |
| 25 | 0.8 | 9 | 600 | 30 | 80:20 | 0.325 | 0.324 |
| 25 | 0.7 | 1 | 600 | 30 | 90:10 | 0.344 | 0.321 |
| 25 | 0.7 | 2 | 600 | 30 | 90:10 | 0.329 | 0.307 |
| 25 | 0.7 | 3 | 600 | 30 | 90:10 | 0.333 | 0.298 |
| 25 | 0.7 | 4 | 600 | 30 | 90:10 | 0.325 | 0.284 |

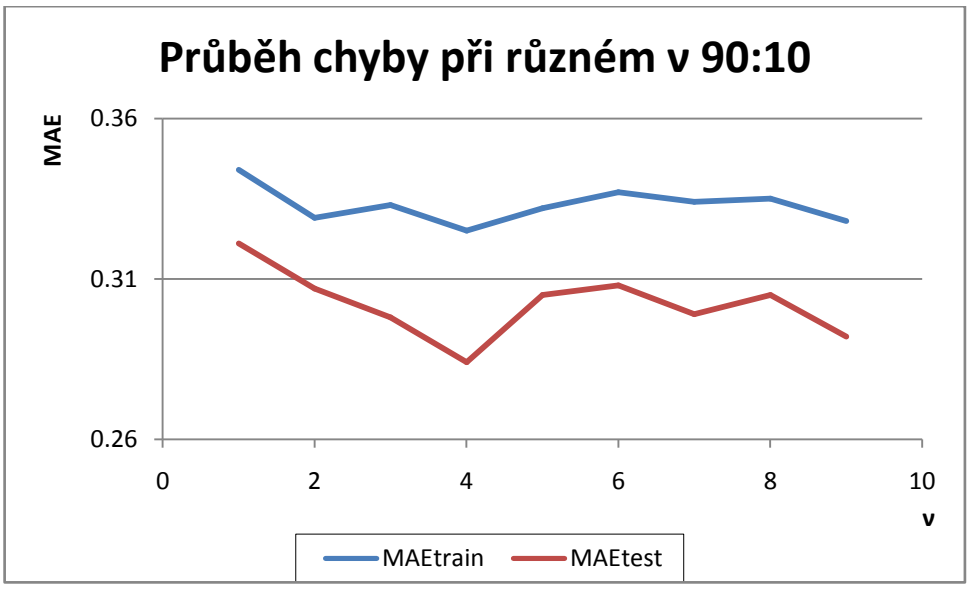
| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 25 | 0.7 | 5 | 600 | 30 | 90:10 | 0.332 | 0.305 |
| 25 | 0.7 | 6 | 600 | 30 | 90:10 | 0.337 | 0.308 |
| 25 | 0.7 | 7 | 600 | 30 | 90:10 | 0.334 | 0.299 |
| 25 | 0.7 | 8 | 600 | 30 | 90:10 | 0.335 | 0.305 |
| 25 | 0.7 | 9 | 600 | 30 | 90:10 | 0.328 | 0.292 |



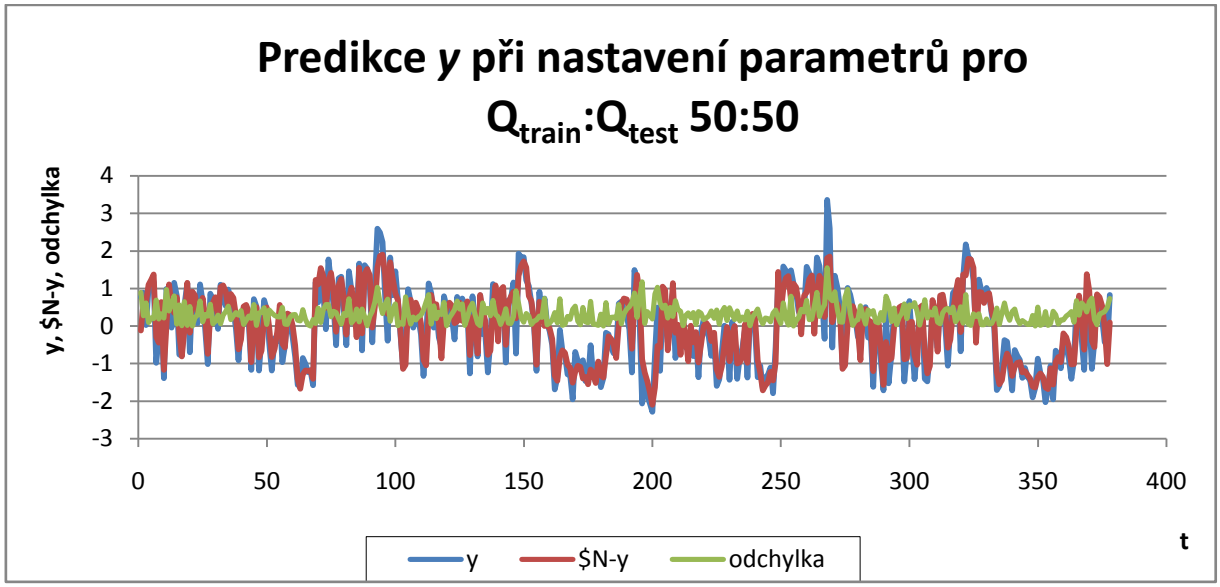



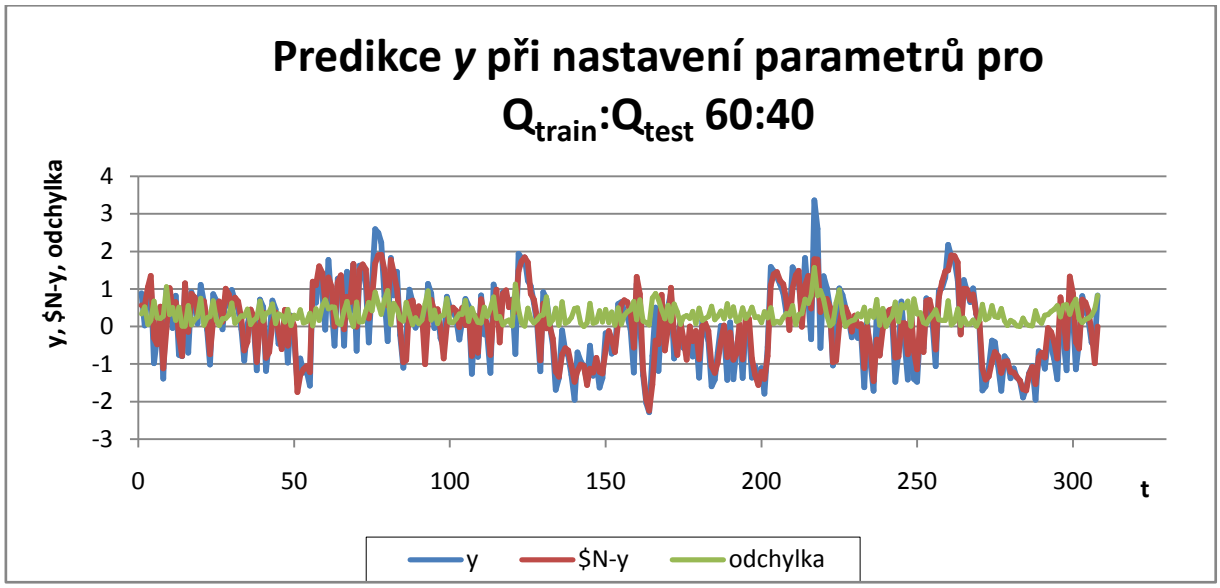

Průběh chyby při různém v 70:30
MAE
0.345
0.34
0.335
0.33
0.325
0.32
0.315
0
2
4
6
8
10
v
MAEtrain
MAEtest

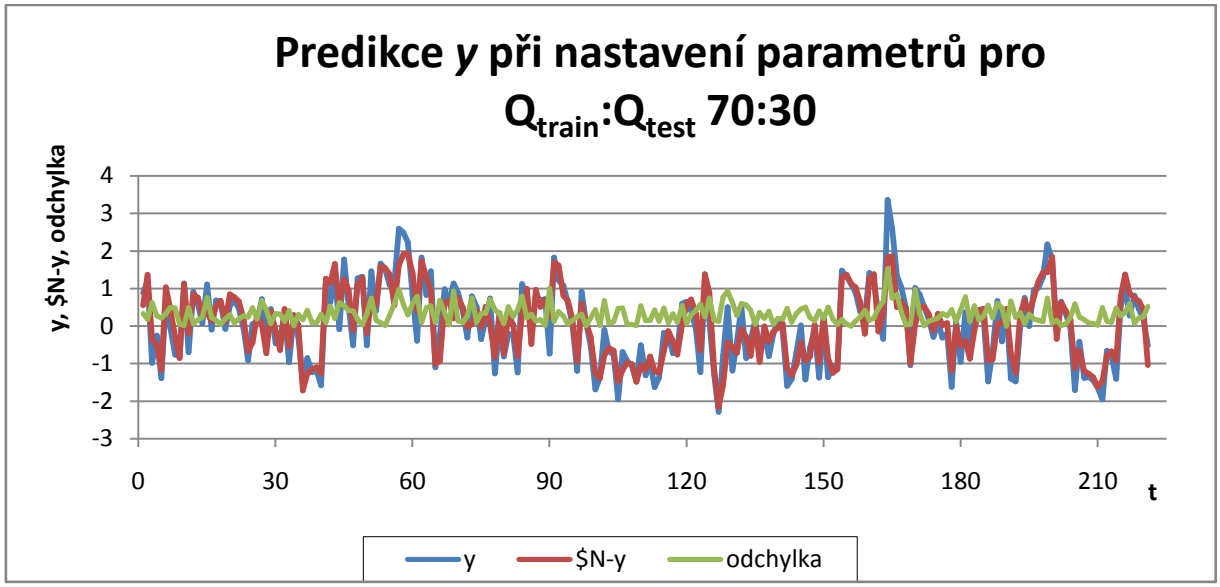

Průběh chyby při různém v 80:20
MAE
0.35
0.345
0.34
0.335
0.33
0.325
0.32
0
2
4
6
8
10
v
MAEtrain
MAEtest

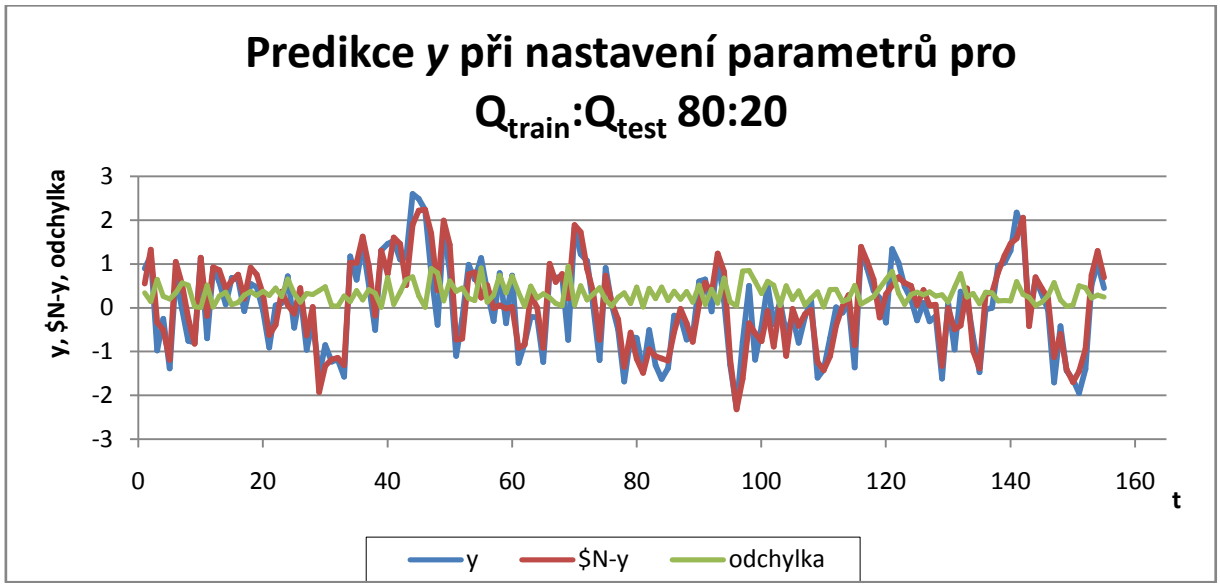

Průběh chyby při různém v 90:10
MAE
0.36
0.31
0.26
0
2
4
6
8
10
v
MAEtrain
MAEtest

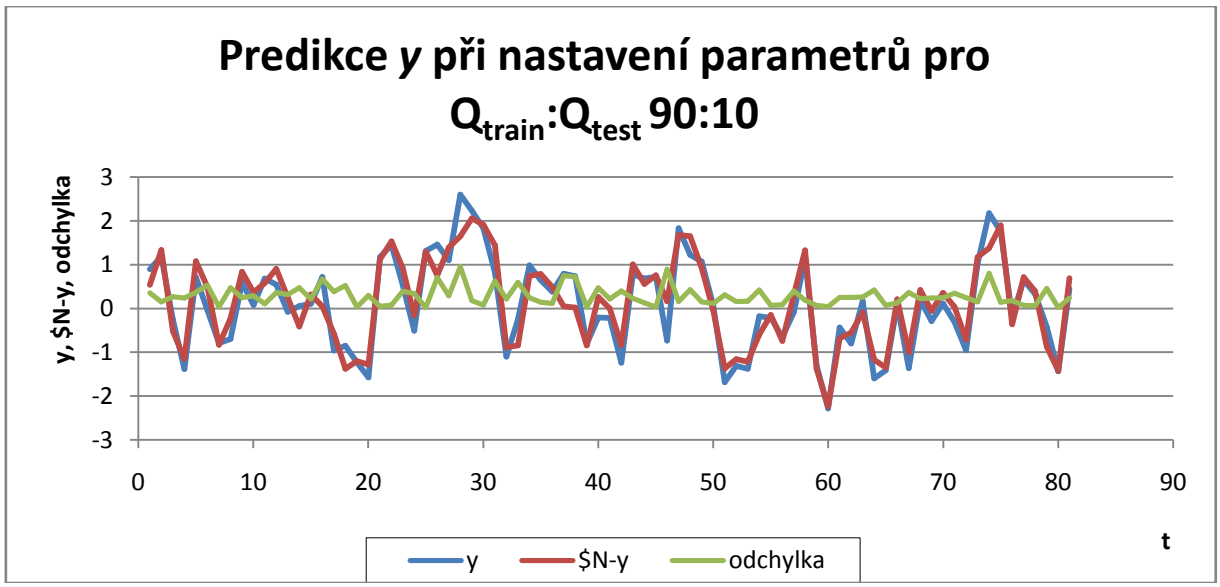
## PŘÍLOHA 15: PREDIKCE y V JEDNOTLIVÝCH ROZDĚLENÍCH DAT DLOUHODOBÉ ČŘ




Predikce y při nastavení parametrů pro
Qtrain:Qtest 70:30
y, $N-y, odchylka
4
3
2
1
0
-1
-2
-3
0
30
60
90
120
150
180
210
t
y
$N-y
odchylka


Predikce y při nastavení parametrů pro
Qtrain:Qtest 80:20
y, $N-y, odchylka
3
2
1
0
-1
-2
-3
0
20
40
60
80
100
120
140
160
t
y
$N-y
odchylka


Predikce y při nastavení parametrů pro
Qtrain:Qtest 90:10
y, $N-y, odchylka
3
2
1
0
-1
-2
-3
0
10
20
30
40
50
60
70
80
90
t
y
$N-y
odchylka