

Univerzita Pardubice
Fakulta ekonomicko správní

Klasifikace ekonomických subjektů
Nikol Popelková

Bakalářská práce
2009

Univerzita Pardubice
Fakulta ekonomicko-správní
Ústav systémového inženýrství a informatiky
Akademický rok: 2008/2009

ZADÁNÍ BAKALÁŘSKÉ PRÁCE

(PROJEKTU, UMĚLECKÉHO DÍLA, UMĚLECKÉHO VÝKONU)

Jméno a příjmení: **Nikol POPELKOVA**
Studijní program: **B6209 Systémové inženýrství a informatika**
Studijní obor: **Regionální a informační management**

Název tématu: **Klasifikace ekonomických subjektů**

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

1. Charakteristika klasifikačního problému.
2. Charakteristika fuzzy logiky a neuronové sítě.
3. Návrh modelu na klasifikaci ekonomických subjektů.
4. V programovém prostředí Nefclass, popř. jiném prostředí, verifikace tohoto modelu na reálných ekonomických datech.
5. Realizace analýzy výsledků.

Rozsah grafických prací:


Rozsah pracovní zprávy:

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

Seznam odborné literatury:

- [1]OLEJ, V. Modelovanie ekonomických procesov na báze výpočtovej inteligencie. Hradec Králové : M&V, 2003. ISBN 80-903024-9-1.
- [2]HAYKIN, S.S. Neural Networks: A Comprehensive Foundation. Upper Saddle River : Prentice Hall, 1999. ISBN 0132733501.
- [3]KVASNIČKA, V. a kol. Úvod do teórie neurónových sietí. Bratislava : IRIS, 1997. ISBN 80-88778-30-1.
- [4]MAŘÍK, V. a kol. Umělá inteligence (1). Praha : Academia, 1993. ISBN 80-200-0496-3.
- [5]MAŘÍK, V. a kol. Umělá inteligence (4). Praha : Academia, 2003. ISBN: 80-200-1044-0.

Vedoucí bakalářské práce:


Ing. Petr Hájek, Ph.D.
Ústav systémového inženýrství a informatiky

Datum zadání bakalářské práce:

6. října 2008

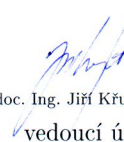
Termín odevzdání bakalářské práce:

1. května 2009



doc. Ing. Renáta Myšková, Ph.D.
děkanka

L.S.


doc. Ing. Jiří Křupka, Ph.D.
vedoucí ústavu

V Pardubicích dne 6. října 2008

Prohlášení autora

Prohlašuji:

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

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

Souhlasím s prezenčním zpřístupněním své práce v Univerzitní knihovně.

V Kolíně dne 27. 4. 2009

Nikol Popelková

Poděkování

Touto cestou bych ráda poděkovala vedoucímu mé bakalářské práce Ing. Petru Hájkovi, Ph.D. za cenné rady a pomoc, kterou mi poskytl při konzultacích k této práci.

ANOTACE

V práci jsou definované základní pojmy, mezi které patří fuzzy logika, fuzzy množina, neuronové sítě a klasifikační problém.

V další části práce je realizováno modelování bonity obcí, ratingu podniků a bonity zákazníků pomocí fuzzy logických neuronových sítí. Jsou navrženy a analyzovány modely fuzzy logických neuronových sítí na ohodnocování bonity obcí. Výsledky jsou porovnány s modely dopředných neuronových sítí. Navržené modely jsou verifikovány v programovém prostředí Nefclass a v programovém prostředí Weka. Vstupní data těchto modelů jsou předzpracována k dosažení vhodných výsledků. Výsledkem navržených modelů je klasifikace obcí do tříd.

KLÍČOVÁ SLOVA

fuzzy logika, fuzzy množina, neuronové sítě, klasifikace, perceptron

TITLE

Classification of economics subjects

ANNOTATION

There are defined basic concepts, including fuzzy logic, fuzzy set, neural networks and classification problems. In other parts of the work the modeling of municipal creditworthiness, business credit rating and creditworthiness of customers is realized using fuzzy logic neural networks. There are designed and analyzed models of fuzzy logic neural networks for the classification of the economic subjects according to their creditworthiness. The results are compared with models of feed-forward neural networks. The proposed models are verified in the programming environments Nefclass and WEKA. Input data of these models are appropriately preprocessed to achieve the results. The result of the proposed models is the classification of subjects into classes.

KEYWORDS

fuzzy logic, fuzzy sets, neural network, classification, perceptron

Obsah

Úvod.....	8
1 Charakteristika klasifikačního problému.....	10
2 Charakteristika fuzzy logiky a neuronové sítě.....	12
2.1 Fuzzy logika.....	12
2.2 Neuronové sítě	15
2.2.1 Biologická neuronová síť.....	15
2.2.2 Umělá neuronová síť	17
2.2.3 Dopředná neuronová síť.....	20
2.2.4 Další modely neuronových sítí.....	21
2.3 Fuzzy logická neuronová síť.....	22
3 Klasifikace ekonomických subjektů	26
3.1 Charakteristika použitých dat	26
3.1.1 Bonita obce.....	26
3.1.2 Rating podniku	29
3.1.3 Bonita zákazníků	32
3.2 Návrh modelu.....	33
3.3 Analýza výsledků dopředné neuronové sítě	35
3.4 Analýza výsledků fuzzy logické neuronové sítě.....	40
Závěr	50
Použitá literatura.....	52

Úvod

Klasické přístupy v rozhodování založené na metodách umělé inteligence jsou charakterizované reprezentací znalostí a sekvenčním zpracováním. Umělá inteligence je v současnosti nejrychlejší se rozvíjející oblast informatiky a výpočetní techniky.

Při analýze rozhodovacích procesů vystupuje do popředí složitost rozhodování. Rozhodování je tak složité, až je nepoužitelné. A rozhodování, kde při popisu procesů vystupuje do popředí neurčitost, která je způsobená neschopností přesně definovat základní pojmy. Přístup na řešení této problematiky můžeme uvést výpočetní inteligenci [9]. Obsahem výpočetní inteligence je zkoumání fuzzy množin, neuronových sítí, evolučních stochastických optimalizačních algoritmů a tvorba hybridních systémů. Hybridní systémy umožňují využití výhod například fuzzy množina neuronových sítí pomocí fuzzy logických neuronových sítí.

V této bakalářské práci se zabývám zkoumáním fuzzy logiky a neuronových sítí s cílem návrhu struktur fuzzy logických neuronových sítí. Fuzzy množiny v praxi členíme na fuzzy množiny s neurčitostí a fuzzy množiny s neúplností. Fuzzy množina s neurčitostí se nazývá stochastická a dá se modelovat pomocí teorie pravděpodobnosti. Fuzzy množina s neúplností se nazývá sémantická, týká se samotného významu jevů nebo výroků o nich.

Neuronové sítě jsou založeny na biologických neuronových sítích. Tyto modely neuronových sítí se používají v praktických aplikacích a jejich společným rysem je vrstevnatost sítě. Neuronové sítě jsou tvořeny několika nebo mnoha jednoduchými procesy, z nichž každý má několik vstupů a jeden výstup. Používány jsou na rozvoj aplikací v odvětví techniky, medicíny i společenských věd, na vývoj algoritmů učení, na minimalizaci paměti, atd. Nejčastěji jsou k těmto účelům využívány dopředné neuronové sítě.

Dopředné neuronové sítě [11] se používají pro řešení klasifikačních i regresních úloh. Vstupní data přitom nemusí být lineárně separabilní.

Bakalářská práce je zaměřena na možnosti využití fuzzy logických neuronových sítí při klasifikaci ekonomických subjektů. Pro porovnání jsou použity tři datové soubory, a to data charakterizující bonitu obcí, rating podniků a bonitu zákazníků. Rating podniků vyjadřuje třídu, která je podnikům přiřazena ratingovou agenturou. Rating vyjadřuje bonitu podniků. Bonita [1] je schopnost subjektu řádně plnit své závazky. Nejčastějším způsobem modelování bonity je klasifikace subjektů do tříd podle jejich bonity. Modelování bonity je

realizováno jako klasifikační problém. Klasifikaci můžeme realizovat dvěma metodami, první metoda je učení s učitelem a druhá metoda je učení bez učitele.

Cílem práce je návrh modelu na klasifikaci ekonomických subjektů do tříd podle jejich bonity. Dále dílčím cílem práce je porovnání výsledků fuzzy logické neuronové sítě s dopřednou neuronovou sítí. Porovnání bude realizováno správností klasifikace a jejich efektivností, tj. podle počtu podmíněných pravidel a podle počtu funkcí příslušnosti. Výsledky obou neuronových sítí budou v závěru analyzovány. Klasifikace bude realizována v programovém prostředí Weka a Nefclass.

1 Charakteristika klasifikačního problému

Cílem této bakalářské práce je klasifikovat ekonomické subjekty podle jejich bonity. Tento problém je proto realizován jako klasifikační.

Klasifikaci je možné uskutečnit dvojím způsobem v závislosti na tom, zda je příslušnost subjektů do tříd předem známá nebo neznámá. Pokud není bonita subjektu známá apriori, je možné klasifikaci realizovat metodami učení bez učitele [6]. Tyto metody zahrnují redukci dat a shlukování [2]. Vhodné jsou metody shlukování, jejichž cílem je rozdělení vstupních dat do disjunktních shluků tak, aby si subjekty ze stejného shluku byly vzájemně podobné, ale nebyly si příliš podobné se subjekty vyskytujícími se mimo tento shluk. Ve druhém případě je předem známá příslušnost subjektů do tříd. Tento problém se nazývá klasifikační a pro jeho realizaci jsou vhodné metody učení s učitelem [9]. Metody učení s učitelem vhodné pro klasifikaci jsou např.:

- Diskriminační analýza [6] - je technika klasifikace případů do dvou nebo více skupin pomocí číselných proměnných a z nich odvozených syntetických kritérií;
- Logistická regresní analýza [6] - označení metody matematické statistiky zabývající se problematikou odhadu pravděpodobnosti nějakého jevu na základě určitých známých skutečností, které mohou ovlivnit výskyt jevu;
- Rozhodovací stromy [10] - cílem je identifikovat subjekty popsané různými atributy do tříd tak, že jsou rozdělovány postupně do podprostorů;
- Neuronové sítě [11] – zejména dopředné neuronové sítě, více v následující kapitole 3;
- Fuzzy klasifikátory [9] - založené na fuzzy inferenčních systémech, jedná se o expertní systémy využívající produkční pravidla založená na fuzzy množinách.

Všeobecný klasifikační problém lze definovat následujícím způsobem [4]. Necht' $F(x)$ je funkce definovaná na množině O , která přiřadí každému prvku $x \in O$ obraz z množiny Ω , kde $\Omega = \{\omega_1, \omega_2, \dots, \omega_c\}$ je množina tříd. Klasifikátor (stroj, který provádí klasifikaci) $\hat{x} = F(x) \in \Omega$ je pak zobrazení

$$F : O \rightarrow \Omega. \quad (1)$$

Necht' $G(x, w)$ je funkce, jejíž argumenty jsou z konečné množiny $O_{\text{train}} = \{x_1, x_2, \dots, x_r\} \subset O$ (nazývané trénovací množina) a w je parametr (nebo parametry) zobrazení G .

Pak platí $\hat{x} = G(x, w) \in \Omega_{\text{train}} \subset \Omega$. Doplněk O_{train} vzhledem k množině O je testovací množina O_{test} . Necht' je znám pro každý $x_i \in O_{\text{train}}$ požadovaný obraz \hat{x}_i . Cílem klasifikace je pak nalezení parametrů w funkce $G(x, w)$, aby funkční hodnoty argumentů z trénovací množiny O_{train} byly co nejbližší obrazům funkce $F(x)$. Dále, necht' je daná účelová funkce v následujícím tvaru

$$E(w) = \frac{1}{2} \sum_{i=1}^r (G(x_i, w) - \hat{x}_i)^2 \quad (2)$$

Potom cílem je minimalizace této funkce. Globálního minima je dosaženo optimální hodnotou parametrů w . Pak je funkce $G(x, w)$ adaptovaná.

Necht' optimální hodnota parametru w je označena jako \bar{w} . Pak se adaptovaná funkce $G(x, \bar{w})$ používá pro klasifikaci trénovací i validační množiny. Předpokládá se, že adaptovaná funkce dobře aproximuje původní funkci $F(x)$ také mimo trénovací množiny.

Jako metody vhodné pro ohodnocování bonity subjektů jsou zvoleny neuronové sítě (Neural Networks) a fuzzy logické neuronové sítě (Fuzzy Logic Neural Networks). Neuronové sítě zejména vzhledem k jejich schopnosti učit se a zevšeobecňovat. Fuzzy logické neuronové sítě pak díky jejich schopnosti pracovat s neurčitostí a znalostmi experta. Umožňují rovněž realizovat uvedené výhody neuronových sítí.

2 Charakteristika fuzzy logiky a neuronové sítě

Tato kapitola popisuje základní pojmy a definice fuzzy logiky, neuronové sítě a fuzzy logické neuronové sítě.

2.1 Fuzzy logika

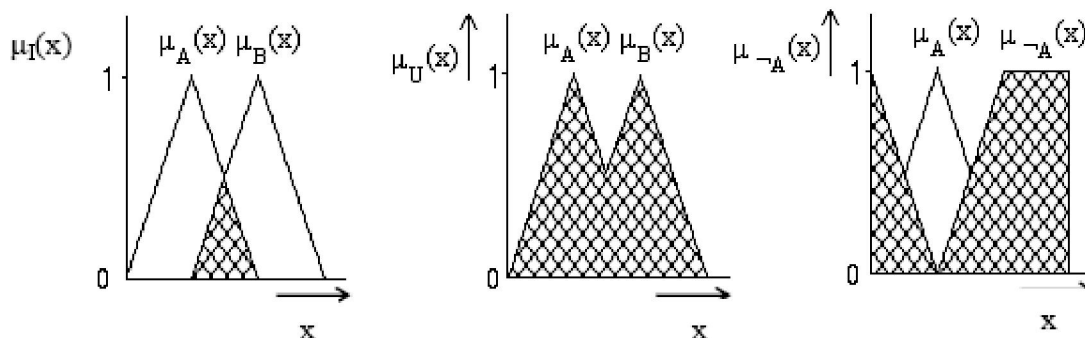
Fuzzy logika [13] byla zavedena roku 1965 Lotfím Zadehem z Kalifornské univerzity v Berkeley. Tehdy byl definován základní pojem fuzzy logiky a to fuzzy¹ množina.

Pod pojmem fuzzy logika [7] se rozumí dvě různé logiky ovlivněné pojmem fuzzy množiny. Vícehodnotová logika, to je logika s více než dvěma pravdivostními hodnotami zpravidla z intervalu $\langle 0, 1 \rangle$, tedy každé dvě pravdivostní hodnoty jsou srovnatelné. Je přímým zobecněním klasické dvouhodnotové logiky a po matematické stránce je hluboce propracovaná. Lingvistická logika to je logika, jejíž pravdivostní hodnoty jsou jazykově vyjádřeny (např. slovy pravda, více méně pravda apod.). Význam těchto pravdivostních hodnot je modelován pomocí fuzzy množin v intervalu $\langle 0, 1 \rangle$. Je dosti odvážným pokusem vytvořit teorii, která dalekosáhle zobecňuje logiku. Po matematické stránce není hluboce propracovaná a je z velké části postavena na intuici a sémantické interpretaci zaváděných pojmů.

Základní pojmy fuzzy logiky jsou fuzzy množiny, jazyková proměnná, fuzzy výroky [12]. Fuzzy množina [13] je množina, která kromě úplného nebo žádného členství připouští i členství částečné. To znamená, že prvek patří do množiny s jistou pravděpodobností (stupeň příslušnosti). Funkce, která každému prvku universa přiřadí stupeň příslušnosti se nazývá funkce příslušnosti.

¹ Slovo fuzzy znamená neostrý, matný, mlhavý, neurčitý, vágní

Dále jsou uváděny základní pojmy výrokové fuzzy logiky [9]. Necht' A, B, C jsou fuzzy množiny a \emptyset je prázdná množina. Potom můžeme definovat funkce příslušnosti μ , které jsou znázorněny na obrázku 1.



obrázek 1: Funkce příslušnosti $\mu_I(x)$, $\mu_U(x)$, $\mu_{\neg A}(x)$. Zdroj [9]

Funkci příslušnosti [9] $\mu_I(x)$ průniku $I = A \cap B$ ve tvaru

$$\mu_I(x) = \text{MIN} \{ \mu_A(x), \mu_B(x) \}. \quad (3)$$

Funkci příslušnosti [9] $\mu_U(x)$ sjednocení $U = A \cup B$ ve tvaru

$$\mu_U(x) = \text{MAX} \{ \mu_A(x), \mu_B(x) \}. \quad (4)$$

Funkci příslušnosti [9] $\mu_{\neg A}(x)$ doplňku ve tvaru

$$\mu_{\neg A}(x) = 1 - \mu_A(x). \quad (5)$$

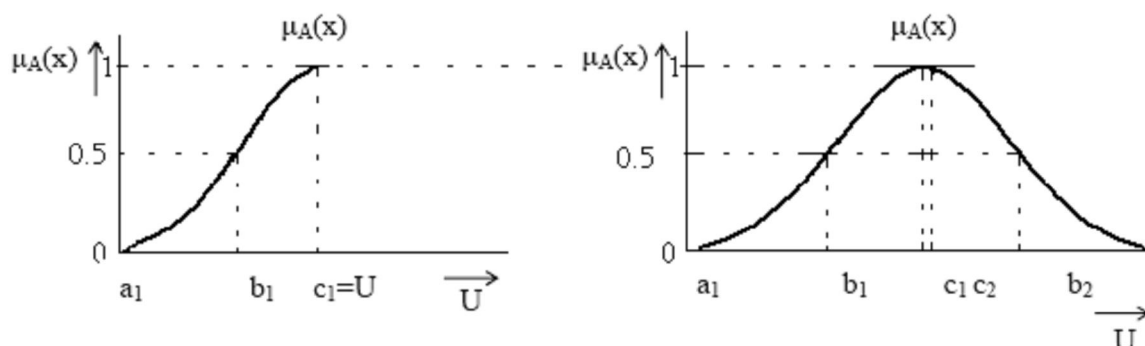
Třída funkcí [9], která vyhovuje vztahu (3) se nazývá trojúhelníková norma (t-norma). Může být vyjádřena následujícím způsobem

$$I = A \cap B \Leftrightarrow \forall x \in X: \mu_I(x) = \mu_A(x) \text{ t } \mu_B(x) \leq \text{MIN} \{ \mu_A(x), \mu_B(x) \}. \quad (6)$$

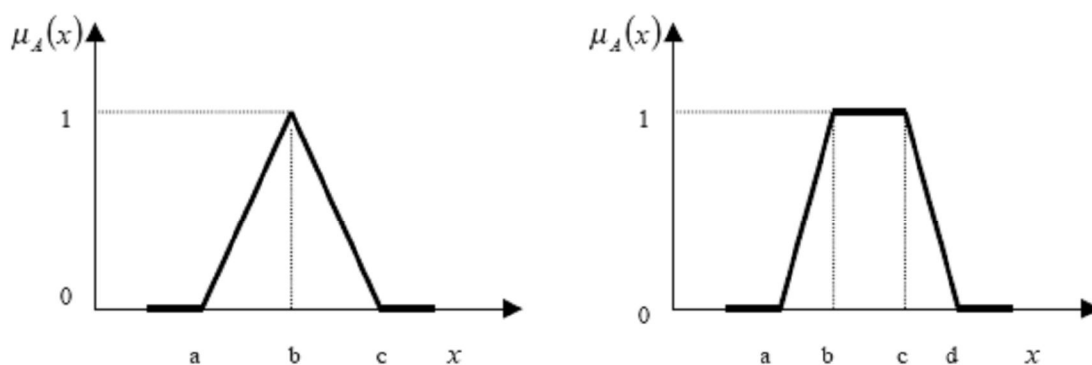
Třída funkcí [9], která vyhovuje vztahu (4) se nazývá s – norma. Může být vyjádřena následujícím způsobem

$$U = A \cup B \Leftrightarrow \forall x \in X: \mu_U(x) = \text{MAX}\{\mu_A(x), \mu_B(x)\} \leq \mu_A(x) \text{ s } \mu_B(x). \quad (7)$$

Funkce příslušnosti (3), (4) a (5) definované pomocí operací konjunkce, disjunkce a negace nad fuzzy množinami jsou používané nejčastěji [9]. Na obrázku 2 a obrázku 3 jsou znázorněny základní tvary funkcí příslušnosti.



obrázek 2: Funkce příslušnosti S a Π fuzzy množiny. Zdroj [9]



obrázek 3: Trojúhelníková a lichoběžníková funkce příslušnosti. Zdroj [9]

Trojúhelníkovou funkci [9] příslušnosti lze popsat pomocnými parametry a, b, c , kde parametry a, c jsou meze nosiče fuzzy množiny a parametr b je její jádro. Funkce příslušnosti se získá spojením bodů $[a, 0]$, $[b, 1]$ a $[b, 1]$, $[c, 0]$ přímkou. Výhody tohoto

postupu spočívají v jeho jednoduchosti, protože určení parametrů a, b, c je jednoduché a poměrně přirozené.

Lichoběžníkovou funkci [9] příslušnosti znázorněnou také na obrázku 3 charakterizujeme pomocí parametrů a, b, c, d , kde parametry a, d jsou meze nosiče fuzzy množiny a parametry b, c jsou její jádro. Funkce příslušnosti se získá spojením bodů $[a, 0]$, $[b, 1]$ a $[b, 1]$, $[c, 1]$ a $[c, 1]$, $[d, 0]$ přímkou.

2.2 Neuronové sítě

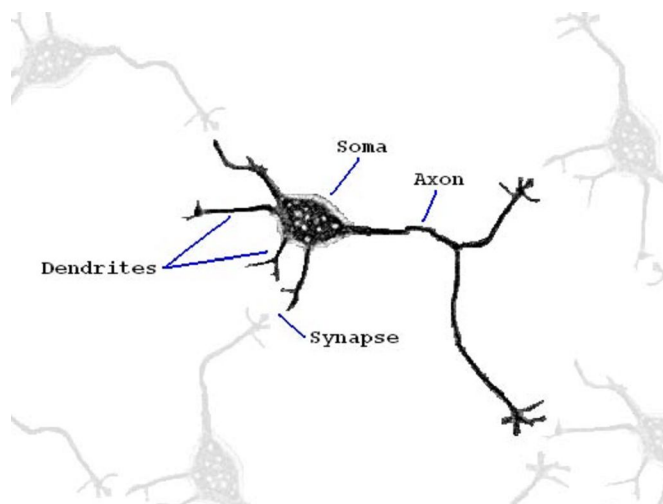
Neuronové sítě [11] jsou tvořeny několika nebo mnoha jednoduchými procesory, z nichž každý má několik vstupů a jeden výstup. Tyto procesory jsou propojeny jednosměrnými informačními kanály.

Základními stavebními prvky neuronových sítí [9] jsou neurony a synapse. Všeobecně můžeme neuronové sítě popsat jako dvojici množin (U, H) , kde U je množina uzlů neuronové sítě, to je neuron, vstup a výstup a H je podmnožina množiny $N \times N$. Určuje synapse mezi jednotlivými uzly. Synapse jsou přiřazeny váhy $w \in R$. Neurony představují nelineárně, typicky analogové jednotky, které jsou málo porovnatelné s moderními číslicovými obvody. Úkolem neuronů je agregace signálu ze vstupních synapsí a realizace předem určené nelineární funkce.

Rozlišujeme neuronové sítě jednoduché a vícevrstvé neuronové sítě, dopředné a rekurentní neuronové sítě, atd.

2.2.1 Biologická neuronová síť

Neuronové sítě [11] tvoří základ všech částí biologického informačního systému. Základní stavební jednotkou je nervová buňka – neuron. Neurony jsou vysoce specializované buňky, schopné přijmout, vést, zpracovat a odpovědět na speciální signály. Přenášejí a zpracovávají informace z vnitřního i vnějšího prostředí a tím podmiňují schopnost organismu na ně reagovat, viz obrázek 4.

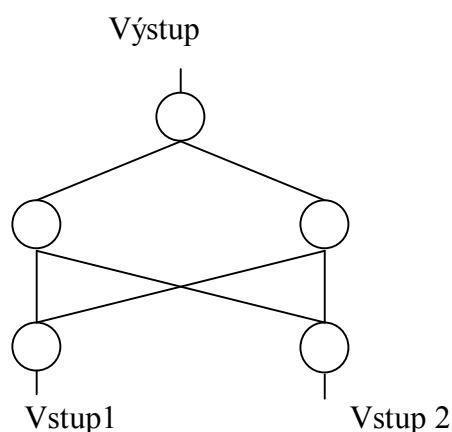


obrázek 4: Neuron

Biologický neuron [11] se skládá z těla s buněčným jádrem nazývaným soma. Z těla neuronu vystupují nervové výběžky – dendrity a axony. Dendrity jsou krátké a tenké, vedou vzruch směrem k buňce, tvoří vstupy do neuronu. Jejich počet je kolem 10^4 . Axony jsou poměrně dlouhé a silnější. Každý neuron obsahuje pouze 1 axon, který tvoří výstup z neuronu. Axony jsou ukončeny synapsami, které tvoří informační rozhraní.

Jednoduchá neuronová síť

Nechť je daná jednoduchá neuronová síť [9], znázorněná na obrázku 5. Skládá se z pěti neuronů, které jsou spojené šesti synapsami. Každá synapse reprezentuje jeden gen. Alelou tohoto genu je váha zodpovědné synapse.



obrázek 5: Jednoduchá neuronová síť. Zdroj [9]

Nechť je potřebné určit správné váhy synapsí neuronových sítí z obrázku 5 tak, aby realizovala Booleovu tabulku zadanou tabulkou 1.

tabulka 1: Booleova funkce. Zdroj [9]

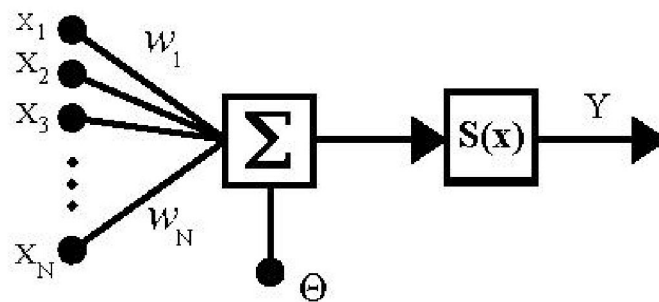
Vstup 1	Vstup 2	Výstup
0	0	0
0	1	1
1	0	0
1	1	0

Potom vhodnost η bude rovna počtu správných výstupů. Možnými vstupy do neuronových sítí jsou dvojice 00, 01, 10 a 11. Vhodnost η se bude rovnat počtu správných výsledků k počtu správných výstupů neuronových sítí za předpokladu, že správné výstupy jsou 0, 1, 1 a 0.

2.2.2 Umělá neuronová síť

Stejně jako u biologických, tak i umělých neuronových sítí [11] je jejich základní stavební jednotkou neuron. V případě umělé neuronové sítě je to jeho matematický model. Rozdíl mezi jednotlivými modely jsou vytvářeny jednak používanými matematickými funkcemi, jednak různou topologií vlastního modelu.

Neuron [11] si můžeme představit jako matematický procesor, kde vstupem je n -rozměrný vektor parametrů, výstupem je skalár na obrázku 6. Každý model neuronu má dvě části – obvodovou funkci a aktivační funkci. Obvodové funkce určují, jakým způsobem budou vstupní parametry kombinovány uvnitř neuronu. Aktivační funkce lze přirovnat k přenosovým funkcím.



obrázek 6: Model umělého neuronu. Zdroj [15]

Výstupy neuronu je dáno jako

$$y = f\left(\sum_{i=1}^N w_i x_i - \Theta\right), \quad (8)$$

kde y je výstup neuronu, u je vnitřní potenciál neuronu, w_i je váhy neuronu, x_i je vstupy neuronu, Θ je práh neuronu, f je neuronová aktivační funkce.

Pomocí tabulky 2 lze znázornit základní analogii mezi biologickými a umělými neuronovými sítěmi.

tabulka 2: Analogie mezi biologickými a umělými neuronovými sítěmi. Zdroj [11]

Biologické neuronové síť	Umělé neuronové síť
dendrit	vstup informací do neuronu
axon	výstup informací z neuronu
synapse, průchodnost synapsí	paměťové elementy, synaptické váhy, aktuální vstupy, předchozí zkušenosti
soma	příjem signálů a jejich slučování, prahování, nelineární zobrazení
neuron	matematický procesor

Proces učení lze definovat jako modifikaci synaptických vah a prahu podle zvoleného algoritmu učení. Podstatou učení je

- výběr charakteristických rysů a zkušeností ze vstupních signálů
- nastavení parametrů umělé neuronové sítě tak, aby odchylka mezi požadovaným a skutečným výstupem při odezvě na soubor trénovacích vzorů byla minimální

Způsoby učení se rozlišují podle několik hledisek [11].

Podle prvního hlediska je rozlišováno učení neasociativní a asociativní:

1. neasociativní – cílem je uchování vzorů v paměti a jejich následné vybavování,
2. asociativní – cílem je extrakce vzájemných vztahů mezi jednotlivými vzory nebo skupinami vzorů.

Podle druhého hlediska se rozlišuje učení s učitelem a bez učitele:

1. učení s učitelem – jsou známy požadované výsledky, které se během učení porovnávají s výstupem sítě,
2. učení bez učitele – je učení, které je založeno na schopnosti neuronových sítí rozeznat ve vstupních vzorech stejné nebo blízké vlastnosti a třídit přicházející vektory podle nich.

Podle třetího hlediska existuje učení jednorázové a opakované:

1. jednorázové učení – v jednom učícím cyklu je dosaženo nejlepšího výsledku, nedochází k žádným dalším úpravám,
2. opakované učení – opakovaně jsou překládány síti vzory a snahou je dosažení minimální odchylky od požadovaného výstupu.

Mezi výhody [11] umělých neuronových sítí patří hlavně dvě charakteristické vlastnosti:

- Získávání znalostí učením pomocí množiny vzorů bez nutnosti explicitní znalosti algoritmu řešení. Schopnost učit se je spolu se schopností generalizace tou nejdůležitější vlastností neuronových sítí.
- Schopnost generalizace, kdy do správných tříd jsou klasifikovány i vstupní vzory, které nejsou součástí tréninkové neboli učící množiny.

Mezi nevýhody lze řadit:

- obtížnou volbu optimální struktury sítě,
- velikost a složitost sítí,
- dobu potřebnou k natrénování,
- obtížné zjištění, zda síť správně generalizuje.

2.2.3 Dopředná neuronová síť

Perceptrony se používají pro řešení jednoduchých klasifikačních úloh. Důležitou podmínkou je, že vstupní data musí být lineárně separabilní. Pouze jedna vrstva trénuje, a to pomocí perceptronového trénovacího algoritmu. Tento algoritmus patří mezi chybové učení. Učící algoritmus Perceptronu je následující.

- Pro obvodovou funkci $u(\underline{x})$ a výstupní funkci [11] platí

$$u(\underline{x}) = w_0 + \sum_{i=1}^n w_i x_i \quad (9)$$

$$y(\underline{x}) = 1 \quad u(\underline{x}) \geq 0, \quad (10)$$

$$y(\underline{x}) = 0 \quad u(\underline{x}) < 0. \quad (11)$$

- Výstup je porovnán s cílovou hodnotou a je spočítána chyba e [11]

$$e = c - y. \quad (12)$$

- je-li $y = c$ a chyba $e = 0$, váhy se nezmění
- je-li $y = 0$ a $c = 1$, chyba $e = 1$ a k vahám se přičte vstupní hodnota, váhový vektor se přiblíží ke vstupnímu vektoru a v následující iteraci se zvýší pravděpodobnost klasifikace 1
- je-li $y = 1$ a $c = 0$, chyba $e = -1$ a od vah se odečte vstupní hodnota, váhový vektor se vzdálí od vstupního vektoru a v následující iteraci se zvýší pravděpodobnost klasifikace 0

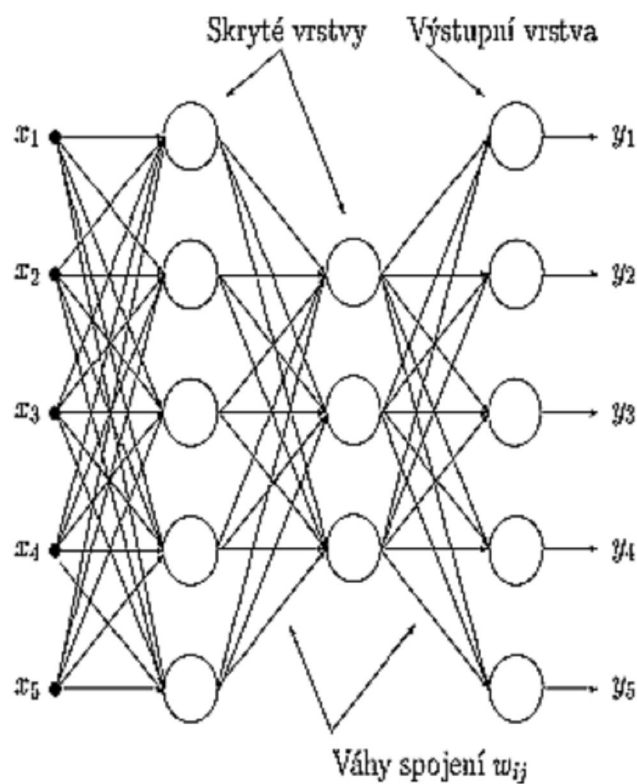
- pro adaptaci vah a prahů [11] platí rovnice

$$w_{ij}(t+1) = w_{ij}(t) + e_j x_i, \quad (13)$$

$$\Theta_j(t+1) = \Theta_j + e_j, \quad (14)$$

kde t označuje čas. Z rovnice (12) a (13) plyne, že velikost chyby je závislá na velikosti vstupního vektoru. Pomocí jednoho perceptronu je možné klasifikovat pouze do dvou tříd. Pro klasifikaci do více tříd je třeba použít více perceptronů nebo vícevrstvou síť s učením.

V této neuronové síti je základ tvořen modelem neuronu s lineárně váženou obvodovou funkcí a aktivační funkcí, kterou je skoková funkce. Vstup je tvořen vektorem vzorků $\underline{x} = (x_1, x_2, \dots, x_n)$ a vektorem požadovaných hodnot $\underline{c} = (c_1, c_2, \dots, c_n)$ v n -dimensionálním prostoru. Jedná se o učení s učitelem. Síť obsahuje vstupní vrstvu, jednu nebo dvě skryté vrstvy a výstupní vrstvu viz obrázek 9. Neurony jsou propojeny synapsemi tak, že každý neuron v jedné vrstvě je propojen s každým neuronem v následující vrstvě.



obrázek 7: Vícevrstvý perceptron. Zdroj [14]

2.2.4 Další modely neuronových sítí

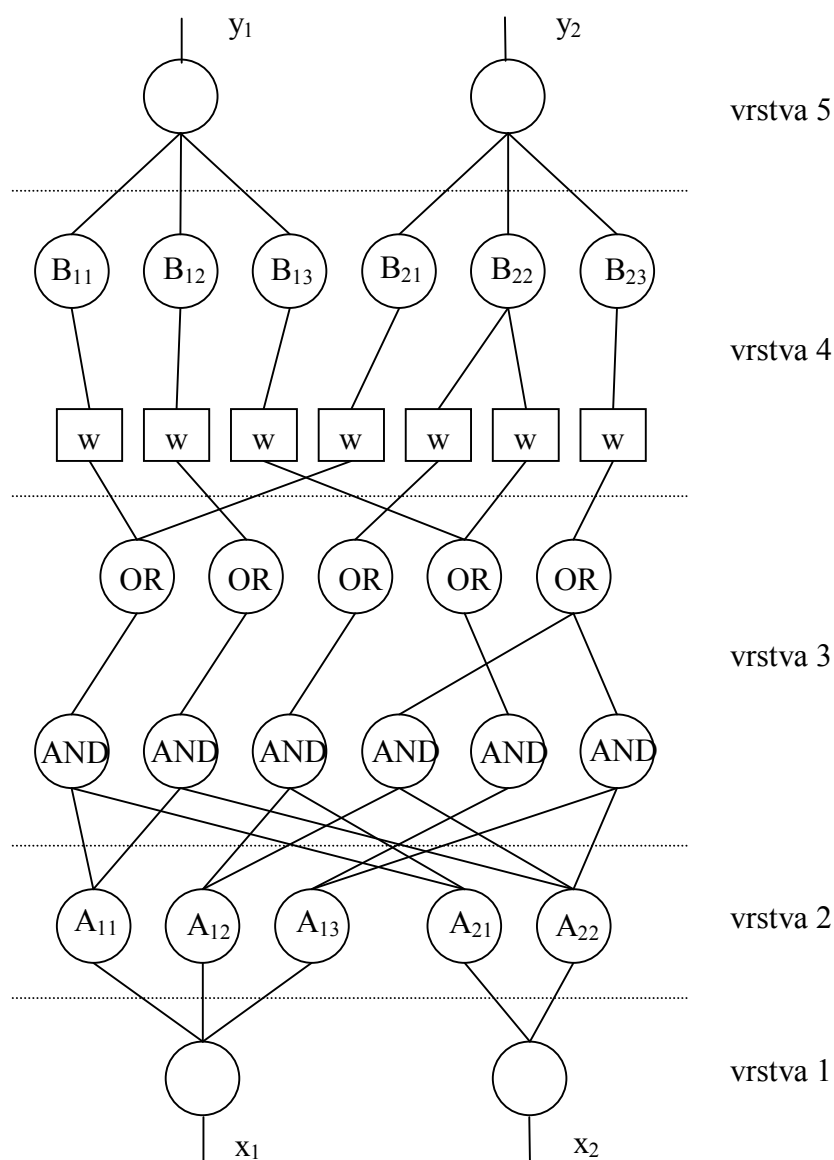
Mezi další modely patří například asociativní paměti, které jsou založeny na asociativním učení. Parametry sítě se nastavují explicitně, pomocí jednorázového nastavení.

Hopfieldova síť je propojená symetrická síť. Taková síť je řazena do skupiny zpětnovazebních (rekurentních) sítí. Během učení dochází k iterativnímu procesu. V daném čase se mění hodnota synaptických vah pouze jediného neuronu. Dělí se na binární a na spojité. Grossbergovo učení bylo vyvinuto jako zpětnovazební síť pro vysvětlení biologických aspektů učení. Pro učení bez učitele byly navrženy modely neuronových sítí typu ART a samoorganizující se mapy. Pro učení s učitelem jsou pak vhodné jejich varianty tj. ARTMAP a LVQ neuronové sítě.

2.3 Fuzzy logická neuronová síť

Fuzzy logická neuronová síť (FLNS) [12] je neuronová síť, která vykonává fuzzy inferenci, viz obrázek 10. Standardní neurony ve FLNS jsou nahrazeny fuzzy neurony. Na rozdíl od standardního neuronu, váhy synapsí mezi fuzzy neurony buď přiřazují vstupní hodnotě hodnotu funkce příslušnosti k příslušné fuzzy množině, nebo reprezentují argument fuzzy operace. Výstup fuzzy neuronu je obvykle hodnota z intervalu $\langle 0,1 \rangle$. Fuzzy logická neuronová síť patří mezi modely, které spojují neuronové sítě a fuzzy inferenční systémy.

Výhodou struktury FLNS je dobrá interpretovatelnost vnitřních parametrů. Ta umožňuje uživateli specifikovat podmnožinu parametrů, jež nemohly být správně nastaveny na základě dostupné apriorní znalosti expertů. Pro klasifikaci je vhodná topologie FLNS typu perceptron se třemi vrstvami. Je odvozena z obecného fuzzy perceptronu. Vstupní vrstva reprezentuje vstupní proměnné $x_1, x_2, \dots, x_i, \dots, x_n$, skrytá vrstva podmíněná pravidla $R_1, R_2, \dots, R_k, \dots, R_m$ a výstupní vrstva reprezentuje výstupní proměnné, jimiž jsou třídy $\omega_1, \omega_2, \dots, \omega_j, \dots, \omega_c$. Neurony používají jako aktivační funkce t-normy a t-conormy. Hodnoty jazykové proměnné $A_j(i)$ příslušné k odpovídajícím fuzzy množinám jsou reprezentovány pomocí synaptických vah $w(x_i, R_k)$. Synaptické váhy $w(R_k, \omega_j)$ reprezentují váhy podmíněných pravidel R_k . Každý neuron ve skryté vrstvě je spojen právě s jedním neuronem výstupní vrstvy. Klasifikátor je vytvořen ve dvou krocích. Nejprve je určena topologie této FLNS, poté jsou určeny parametry systému v procesu učení. Topologie FLNS pro realizaci klasifikačních problémů je uveden na obrázku 11. Těmito parametry jsou podmíněná pravidla a fuzzy množiny. Určení podmíněných pravidel se uskutečňuje následujícím algoritmem.



obrázek 8: Model FLNS. Zdroj [13]

Nechť existuje FLNS s n vstupy x_1, x_2, \dots, x_n , $m < m_{\max}$ neurony, reprezentující podmíněná pravidla R_1, R_2, \dots, R_m a c výstupními neurony, reprezentující třídy $\omega_1, \omega_2, \dots, \omega_c$. Dále, necht' je dána množina s vzorů $O = \{(\mathbf{p}_1, \mathbf{t}_1), (\mathbf{p}_2, \mathbf{t}_2), \dots, (\mathbf{p}_s, \mathbf{t}_s)\}$, kde $\mathbf{p} \in \mathbb{R}^n$ je vektor hodnot proměnných a $\mathbf{t} \in \{0,1\}^c$ je vektor příslušností vzorů z množiny O do jednotlivých tříd. Pak učící algoritmus podmíněných pravidel R_k má následující tvar:

1. Vyber objekt $(\mathbf{p}_i, \mathbf{t}_i) \in O$.
2. Pro každou vstupní proměnnou x_i najdi funkci příslušnosti $\mu_{j_i}^{(i)}(\mathbf{p}_i)$ takovou, že platí

$$\mu_{j_i}^{(i)}(p_i) = \max_{j=\{1,2,\dots,q_i\}} \{\mu_j^{(i)}(x_i)\}, \quad (15)$$

kde q_i je počet funkcí příslušnosti odpovídající vstupní proměnné x_i .

1. Jestliže neexistuje podmíněné pravidlo R_k s $w(x_1, R_k)=\mu_{j_1}^{(1)}, w(x_2, R_k)=\mu_{j_2}^{(2)}, \dots, w(x_n, R_k)=\mu_{j_n}^{(n)}$, pak tento neuron R_k vytvoř a spoj jej s výstupním neuronem ω_j , pro který $t_j=1$.
2. Jestliže jsou v množině O objekty, které neprošly trénovacím algoritmem, pokračuj bodem 1. Jinak urči nejlepších m podmíněných pravidel a všechna ostatní podmíněná pravidla odstraň.

Fuzzy množiny jsou určeny pomocí tohoto algoritmu:

1. Vyber objekt $(p_i, t_j) \in O$, nech jej projít FLNS a urči výstupní vektor c .
2. Pro každý výstupní neuron ω_j urči

$$\Delta_{\omega_j} = t_j - o_{\omega_j}, \text{ kde } y_{\omega_j} \text{ je výstup neuronu } \omega_j. \quad (16)$$

3. Pro každé podmíněné pravidlo R_k s $y_{R_k} > 0$

- a) Urči $\Delta_{R_k} = y_{R_k} (1 - y_{R_k}) \sum_{c_i \in U_\Delta} w(R_k, \omega_j) \Delta_{\omega_j}$, kde y_{R_k} je výstup neuronu R_k . (17)

- b) Najdi x' takové, pro které platí

$$w(x', R_k) y_{x'} = \min_{x_i \in U_1} \{w(x_i, R_k) y_{x_i}\}, \text{ kde } y_{x_i} \text{ je výstup neuronu } x_i. \quad (18)$$

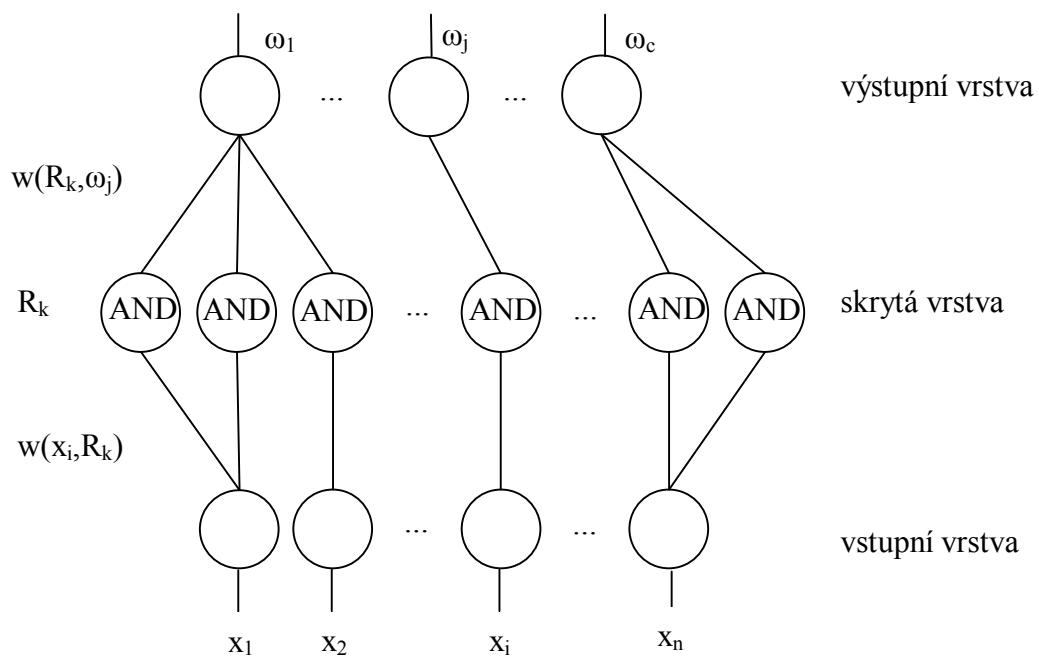
- c) Pro fuzzy množinu reprezentovanou synaptickou váhou $w(x', R_k)$ urči parametry jejich funkcí příslušnosti (v tomto případě trojúhelníkových) takto:

$$\Delta_b = \eta \Delta_{R_k} (c - a) \operatorname{sgn}(y_{x'} - b), \quad (19)$$

$$\Delta_a = -\eta \Delta_{R_k} (c - a) + \Delta_b, \quad (20)$$

$$\Delta_c = \eta \Delta_{R_k} (c - a) + \Delta_b, \text{ kde } \eta > 0 \text{ je rychlost učení.} \quad (21)$$

4. Jestliže bylo splněno cílové kritérium (požadovaná chyba E_r nebo počet iterací), ukonči proces učení, jinak pokračuj bodem 1.



obrázek 9: Model FLNS vhodný pro klasifikaci. Zdroj [13]

3 Klasifikace ekonomických subjektů

V této kapitole budou popsána vstupní data a navržen model pro klasifikaci ekonomických subjektů podle bonity. Klasifikace ekonomických subjektů je realizována pomocí FLNS a dopředné neuronové sítě typu perceptron. Budou použity různé parametry učení. Modely budou verifikovány ve dvou programových prostředí Weka a Nefclass a výsledky budou porovnány a analyzovány.

3.1 Charakteristika použitých dat

V bakalářské práci jsou použity tři datové soubory. Mezi ně patří data o bonitě obce, ratingu podniku a bonitě zákazníků.

3.1.1 Bonita obce

Bonita [1] je schopnost subjektu řádně plnit své závazky. Zahrnuje solventnost a úroveň podnikatelských aktivit daného subjektu. Podle bonity hodnotí obchodní partneři a banky úvěruschopnost zákazníků. Rozlišujeme vysokou a nízkou bonitu subjektu. Vysoká bonita subjektu značí nízké úvěrové riziko, nízká bonita vysoké úvěrové riziko.

Ohodnocování bonity subjektů je v současnosti realizováno metodami, které kombinují matematicko-statistické metody a ohodnocování expertem. Patří mezi ně bodovací modely (Scoring Models), modely selhání (Default Models) a modely ratingu [1].

V datech bonita obce jsou uvedeny společné kategorie parametrů pro ohodnocování bonity obcí, a to ekonomické, dluhové, finanční a administrativní. Klíčové jsou parametry ekonomické, dluhové a finanční.

Ekonomické parametry ovlivňují dlouhodobé úvěrové riziko obcí. Ekonomický růst však může vést k nutnosti rozšíření veřejných služeb a tím k růstu zadluženosti. Ekonomické parametry pro ohodnocování bonity obcí lze navrhnout takto:

$$\text{Parametr } x_1 = PO_r, \quad (22)$$

kde PO_r je počet obyvatel v r -tém roce.

$$\text{Parametr } x_2 = PO_r/PO_{r-s}, \quad (23)$$

kde: - PO_{r-s} je počet obyvatel v roce r-s,
- s je zvolený časový interval.

$$\text{Parametr } x_3 = u, \quad (24)$$

kde u je míra nezaměstnanosti v obci.

$$\text{Parametr } x_4 = \sum_{i=1}^k (PZO_i/PZ)^2, \quad (25)$$

kde: - PZO_i je počet obyvatel zaměstnaných v i-tém odvětví ekonomiky, $i=1,2, \dots, k$,
- PZ je celkový počet zaměstnaných obyvatel,
- k je počet odvětví ekonomiky

Dluhové parametry zahrnují velikost a strukturu dluhu. Lze navrhnout následující dluhové parametry:

$$\text{Parametr } x_5 = DS/OP, \quad (26)$$

kde: - $x_5 \in \langle 0,1 \rangle$,
- DS je dluhová služba,
- OP jsou opakující se příjmy.

$$\text{Parametr } x_6 = CD/PO, [\text{Kč}], \quad (27)$$

kde CD je celkový dluh.

$$\text{Parametr } x_7 = KD/CD, \quad (28)$$

kde: - $x_7 \in \langle 0,1 \rangle$,
- KD je krátkodobý dluh

Finanční parametry informují o rozpočtovém hospodaření obce. Jejich hodnoty jsou získávány z rozpočtu obce. Finanční parametry pro ohodnocování bonity obcí lze navrhnout takto:

$$\text{Parametr } x_8 = OP/BV, \quad (29)$$

kde: - $x_8 \in \mathbb{R}^+$,

- BV jsou běžné výdaje rozpočtu.

$$\text{Parametr } x_9 = \text{VP/CP}, \quad (30)$$

- kde
- $x_9 \in \langle 0,1 \rangle$,
 - VP jsou vlastní příjmy,
 - CP jsou celkové příjmy.

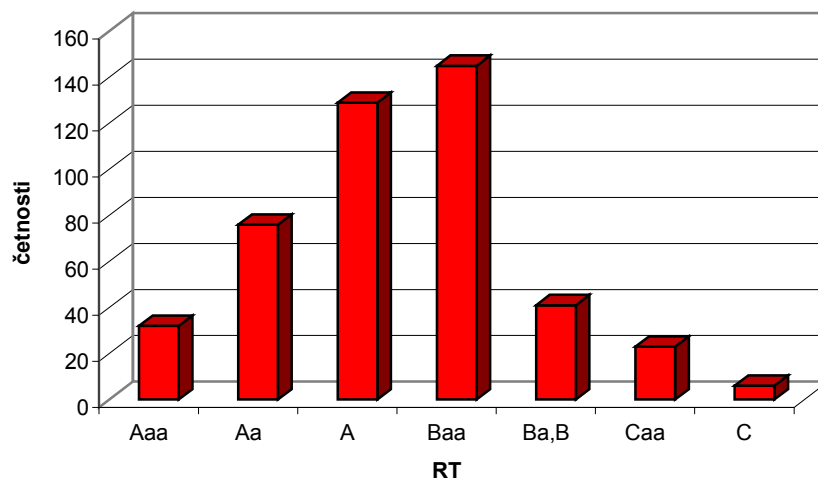
$$\text{Parametr } p_{10} = \text{KV/CV}, \quad (31)$$

- kde:
- $x_{10} \in \langle 0,1 \rangle$,
 - KV jsou kapitálové výdaje,
 - CV jsou celkové výdaje.

Bonita obce je rozdělena do sedmi tříd. Třídy byly obcím přiřazeny experty na danou problematiku v Pardubickém kraji.

tabulka 3: Třídy bonity obce. Zdroj [1]

RT	Popis
Aaa	Rozvíjející se obce s dobrým rozpočtovým hospodařením, bez problémů se zadlužeností.
Aa	Obce s vynikajícím ekonomickým prostředím, bez investičního rozvoje.
A	Obce průměrně zadlužené, s dobrým ekonomickým prostředím, investičně se rozvíjející.
Baa	Obce se špatným ekonomickým prostředím, bez zadluženosti, s průměrným rozpočtovým hospodařením.
Ba,B	Obce s velkým majetkem, středně zadlužené, se signály ekonomického poklesu.
Caa	Obce s vysokou zadlužeností, investičně se rozvíjející.
C	Obce s vysokou dluhovou službou, špatným ekonomickým prostředím, s investičním rozvojem.



graf 1: Četnosti subjektů v třídách - Bonita obce

Základní statistiky vstupních proměnných Bonity obcí jsou uvedeny v Příloze A.

3.1.2 Rating podniku

Rating je nezávislé ohodnocování, jehož cílem je zjistit (na základě komplexního rozboru veškerých známých rizik ohodnocovaného subjektu) jak je tento subjekt schopen a ochoten dostát včas a v plné výši svým splatným závazkům. Rating je ohodnocování založené na kvantitativních i kvalitativních parametrech.

Modelování ratingu je klasifikační problém. Klasifikaci je možné realizovat metodami učení s učitelem (pokud jsou známy ratingové třídy zkoumaných objektů) nebo metodami učení bez učitele (pokud tyto třídy známy nejsou). Vzhledem k tomu, že ve vstupních datech jsou obsaženy třídy, udávající zařazení podniků do ratingových tříd, jedná se o klasifikaci obcí do ratingových tříd pomocí metod učení s učitelem. Označení ratingových tříd je ve shodě s ratingovými agenturami.

Sada dat zahrnuje finanční i nefinanční informace o 1024 amerických společnostech. Tato data byla získána z databáze Standard and Poor. Kategorie jednotlivých tříd jsou zobrazeny v tabulce 4. Celkem tabulka 4 obsahuje 34 vstupních proměnných ($x_1, x_2, x_3, \dots, x_{34}$).

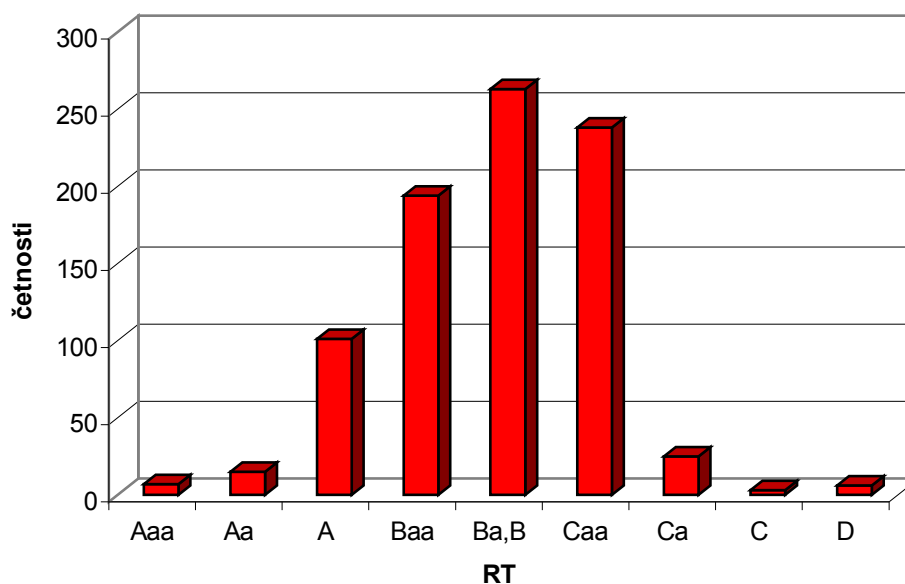
Návrh vstupních proměnných vychází z předchozích prací [3]. Proměnné reprezentující finanční ukazatele podniků, hlavně ukazatele ziskovosti, zadluženosti a ukazatele kapitálových trhů.

tabulka 4: Vstupy proměnných pro společný rating modelování. Zdroj [1]

x_n	Název	x_n	Název
x ₁	Velikost podniku	x ₁₈	Účetní hodnota vlastního kapitálu
x ₂	Cena akcie	x ₁₉	Nepeněžní pracovní kapitál
x ₃	Tržní hodnota podniku	x ₂₀	Cash Flow
x ₄	Hodnota podniku	x ₂₁	Efektivní sazba daně
x ₅	Cena akcie/Zisk na akcii	x ₂₂	Pokles
x ₆	Zisk na akcii	x ₂₃	Tržby
x ₇	Tržní hodnota/Účetní hodnota podniku	x ₂₄	Čistý příjem
x ₈	Cena podniku /Tržby	x ₂₅	Dividendy
x ₉	Tržní hodnota podniku/Zisk před úroky a zdaněním	x ₂₆	Tržní hodnota dluhu/Celkový kapitál
x ₁₀	Růst zisku na akcii	x ₂₇	Účetní hodnota dluhu/Celkový kapitál
x ₁₁	Beta koeficient	x ₂₈	Dividendový výnos
x ₁₂	Největší cena akcie/Nejnižší cena akcie	x ₂₉	Podíl akcií ve vlastnictví institucionálních investorů
x ₁₃	Odchylka z ceny akcie	x ₃₀	Čistý převod
x ₁₄	Korelace ceny akcie	x ₃₁	Tržby/čisté jmění podniku
x ₁₅	Výplatní poměr	x ₃₂	Tržní hodnota podniku/Celkový dluh
x ₁₆	Rentabilita celkového kapitálu	x ₃₃	Tržba/celková aktiva
x ₁₇	Čistý zisk	x ₃₄	Cena akcie/ Cash Flow

tabulka 5: Třídy ratingu. Zdroj [1]

RT	Popis
Aaa	prvotřídní subjekt s nízkým stupněm rizika
Aa	velmi kvalitní subjekt s mírným rizikem v delším časovém horizontu
A	nadprůměrný subjekt se skutečnostmi, které by v budoucnu mohly ohrozit nízkou míru rizika
Baa	středně rizikový subjekt se stabilní současností a s možnými problémy v budoucnu
Ba,B	již spekulativní subjekt s nejistou až problematickou úrovní rizika
Caa	již spekulativní subjekt s nejistou až problematickou úrovní rizika
Ca	podprůměrný až vysoce rizikový subjekt s problémy při dodržování svých závazků/s vysokou mírou nedodržování svých závazků
C	podprůměrný až vysoce rizikový subjekt s problémy při dodržování svých závazků/s vysokou mírou nedodržování svých závazků
D	vysoce spekulativní subjekt neschopný splácet své závazky



graf 2: Četnosti subjektů v třídách - Rating podniků

Základní statistiky vstupních proměnných Ratingu podniků jsou uvedeny v Příloze B.

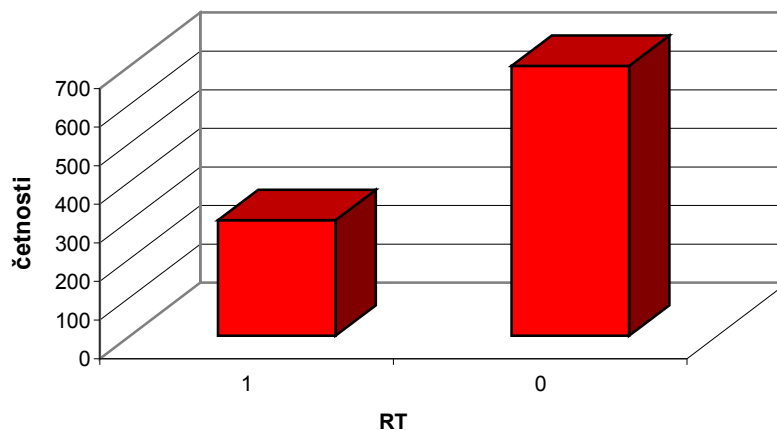
3.1.3 Bonita zákazníků

Úvěr je formou dočasného postoupení zboží nebo peněžních prostředků (půjčka) věřitelem, na principu návratnosti, dlužníkovi, který je ochoten za tuto půjčku po uplynutí nebo v průběhu doby splatnosti zaplatit určitý úrok ve formě peněžité prémie. Úvěrová bonita zákazníků je popsána vstupními proměnnými uvedenými v tabulce 5. Jsou reprezentovány 24 proměnnými, o čtyřech z nich se nepodařilo nalézt informace. Data pocházejí z UCI raclusive Leasing Repository [1]. Představují proto benchmarková data, která jsou používána pro porovnání nově navržených klasifikátorů.

tabulka 6: Popis bonity zákazníků. Zdroj [1]

x_n	Název	x_n	Název
x_1	Stav existujícího účtu v bance	x_{11}	Délka bydlení v současném bydlišti
x_2	Délka v měsících	x_{12}	Vlastnictví
x_3	Historický úvěr	x_{13}	Věk
x_4	Účel, cíl	x_{14}	Ostatní plánovaná splátka
x_5	Úvěrová částka	x_{15}	Typy bydlení
x_6	Spořitelní účet	x_{16}	Počet existujících úvěrů v bance
x_7	Délka zaměstnání	x_{17}	Práce
x_8	Podíl na disponibilním důchodu	x_{18}	Počet lidí poskytující zajištění
x_9	Osobní stav	x_{19}	Telefon
x_{10}	Ostatní dlužníci	x_{20}	Zahraniční pracovník

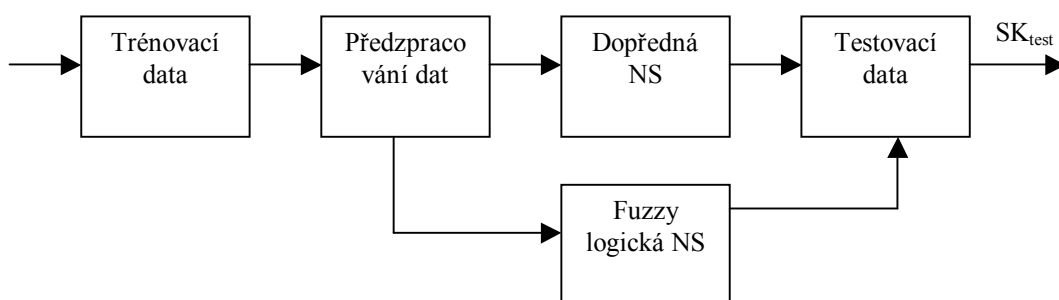
Subjektům jsou přiřazeny třídy 1 a 0, podle toho zda zákazník je schopen dostát svým závazkům (0) nebo ne (1). Základní statistiky vstupních proměnných jsou uvedeny v Příloze C.



graf 3: Četnosti subjektů v třídách - Bonita zákazníků

3.2 Návrh modelu

Modelování bonity ekonomických subjektů je možné uskutečnit následujícím způsobem (obrázek 12). Na vstupu modelu jsou vstupní data, z kterých byly navrženy vstupní proměnné klasifikace subjektů podle bonity. Trénovací data pro bonitu obcí obsahují finanční, dluhové a ekonomické proměnné. Dále jsou zde použita data o ratingu podniků, kde jsou x_1, x_2, \dots, x_{34} vstupní proměnné (viz tabulka 4) a pro bonitu zákazníků, kde jsou x_1, x_2, \dots, x_{20} vstupní proměnné (viz tabulka 5).



obrázek 10: Blokové schéma modelu

Dále byla data předzpracována. Předzpracování bylo provedeno pomocí standardizace, pomocí níž byly odstraněny závislosti na proměnných jednotkách. Standardizace je provedena tak, že jsou od původních hodnot odečteny průměry a výsledek je vydělen směrodatnou odchylkou proměnné. Data jsou přivedena na vstupy dvou navržených struktur NS a FLNS pomocí nichž je provedena klasifikace ekonomických subjektů.

Dopředné neuronové sítě byly realizovány v programovém prostředí Weka a fuzzy logické neuronové sítě v programovém prostředí Nefclass.

Vstupní parametry učení pro dopředné neuronové sítě jsou:

- rychlost učení: 0,05; 0,12 a 0,2;
- momentum, aby nedocházelo k přelučení sítě: 0,2; 0,3; 0,4 a 0,5;
- počet cyklů: 10.

Vstupní parametry učení pro fuzzy logické neuronové sítě jsou nastaveny takto :

- počet funkcí příslušnosti: 2, 3, 4 a 5;
- typ funkcí příslušnosti: trojúhelníková nesymetrická;
- agregační funkce: maximum;
- velikost báze podmíněných pravidel: automaticky;
- učení podmíněných pravidel: nejlepší pro každou třídu;
- rychlost učení: 0,05; 0,12 a 0,2;
- počet cyklů: 10;
- ukončení procesu:
 - maximální počet epoch = 200,
 - minimální počet epoch = 0,
 - počet epoch po ukončení procesu = 20,
 - požadovaná chyba klasifikace = 0.

Rozdělení na testovací a trénovací data bylo provedeno křížovou validací s 10 podmnožinami, kde K – násobná křížová validace postupuje takto:

- vstupní data jsou rozdělena na k stejně velkých podmnožin
- každá z těchto podmnožin je jednou použita pro testování hypotézy vytvořené trénováním na zbytku dat

Výsledkem modelu je správnost klasifikace SK_{test} na testovacích datech + střední absolutní chyba na testovacích datech.

3.3 Analýza výsledků dopředné neuronové sítě

V této části práce je provedena klasifikace dat pomocí dopředné neuronové sítě. Vstupy neuronových sítí jsou data o Bonitě obce, Ratingu podniku a Bonitě zákazníků. Parametry učení neuronové sítě jsou rychlosti učení (0,05; 0,1 a 0,2) a parametry momentum (0,2; 0,3; 0,4 a 0,5).

Následující tabulky 7-11 znázorňují výsledky experimentů dopředné neuronové sítě realizované v prostředí Weka. To nabízí různé možnosti testování těchto modelů: pomocí určité části dat (testovací data) a křížovou validaci. Parametr rychlost učení byl zvolen 0,05; 0,1 a 0,2 a parametr momentum 0,2; 0,3; 0,4 a 0,5.

Tabulky 7,9,11 znázorňují správnost klasifikace SK_{test} při různé rychlosti učení a momentu.

tabulka 7: Správnost klasifikace - Bonita obce

Momentum	Rychlost učení	Nesprávně klasifikované	Správně klasifikované
0,2	0,05	13,05%	86,95%
	0,1	14,38%	85,62%
	0,2	15,71%	84,29%
0,3	0,05	13,72%	86,28%
	0,1	13,72%	86,28%
	0,2	13,49%	86,51%
0,4	0,05	15,06%	84,94%
	0,1	14,38%	85,62%
	0,2	14,38%	85,62%
0,5	0,05	14,82%	85,18%
	0,1	15,49%	84,51%
	0,2	15,49%	84,51%

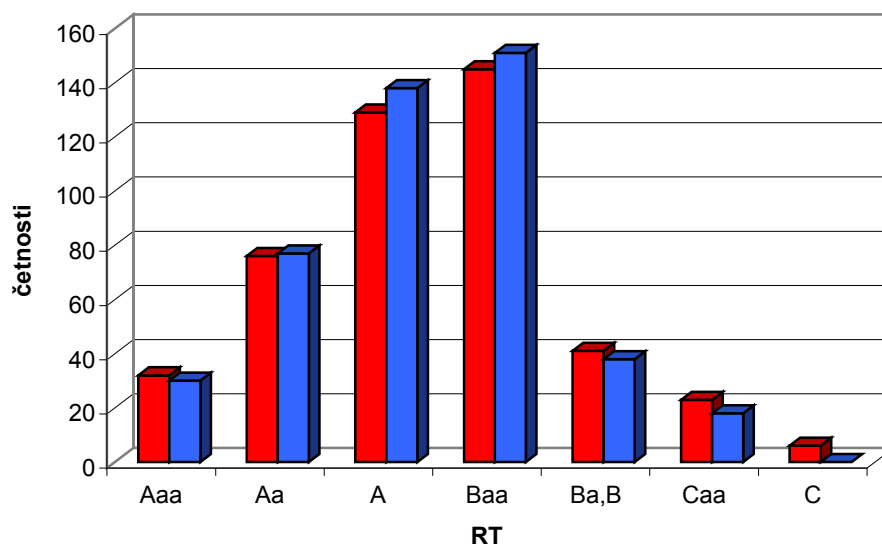
Z tabulky 7 lze vidět, že nejvyšší počet správně klasifikovaných subjektů je u momenta 0,2 s rychlosti učení 0,05, kde správnost klasifikace dosáhla 86,95%. Detailněji nám správnost klasifikace u dat Bonita obce ukáže tabulka 8, která znázorňuje matici záměn nejvyšší správně klasifikované hodnoty a to 86,95% u momenta 0,2 s rychlosti učení

0,05. Správně klasifikované subjekty se pohybují v rozmezí 84,51% - 86,95%. Nejvyšší hodnota nesprávně klasifikovaných subjektů je 15,49%. Nesprávně klasifikované subjekty se pohybují v rozmezí 13,05% - 15,49%.

tabulka 8: Matice záměn - Bonita obce

	Aaa	Aa	A	Baa	Ba,B	Caa	C	celkem
Aaa	28	3	0	1	0	0	0	32
Aa	0	63	7	6	0	0	0	76
A	0	2	119	3	2	3	0	129
Baa	0	3	3	137	2	0	0	145
Ba,B	0	0	4	4	33	0	0	41
Caa	0	4	5	0	1	13	0	23
C	2	2	0	0	0	2	0	6
celkem	30	77	138	151	38	18	0	452

Tabulka 8 znázorňuje matici záměn u dat Bonita obce. Tabulka obsahuje 7 tříd s celkovým počtem 452 obcí. V jednotlivých třídách je určitý počet obcí. Poslední sloupec v tabulce ukazuje původní rozdělení obcí do jednotlivých tříd. Poslední řádek v tabulce ukazuje rozdělení obcí do jednotlivých tříd uskutečněné pomocí dopředné neuronové sítě. Třídy Caa a C nebylo možné pomocí dopředné neuronové sítě správně klasifikovat.



graf 4: Porovnání četností klasifikace - Bonita obce

Legenda ke grafu 4, grafu 5 i grafu 6:

Červená barva – původní hodnoty

Modrá barva – hodnoty dopředné neuronové sítě

Z grafu 4 lze vyčíst, že ve třídě A vzrostl počet obcí z původních 129 na 138. Třída A se skládá z obcí, které jsou průměrně bonitní. Ve třídě Ba klesl počet obcí z původních 41 na 31. Třída Ba obsahuje obce, které jsou středně bonitní.

tabulka 9:Správnost klasifikace - Rating podniků

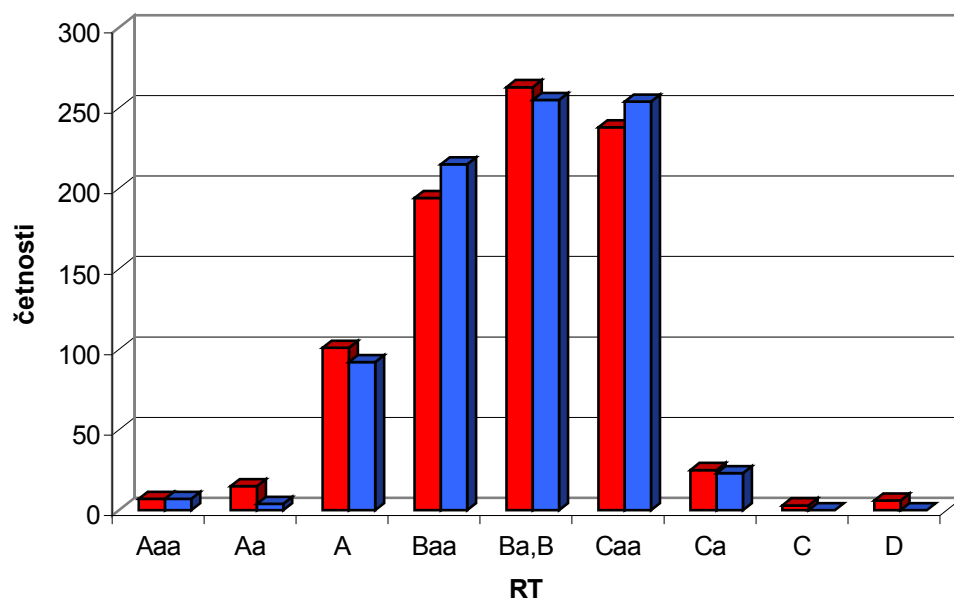
Momentum	Rychlost učení	Nesprávně klasifikované	Správně klasifikované
0,2	0,05	42,37%	57,63%
	0,1	45,65%	54,35%
	0,2	46,71%	53,29%
0,3	0,05	41,90%	58,10%
	0,1	45,11%	54,89%
	0,2	47,88%	52,12%
0,4	0,05	43,19%	56,81%
	0,1	46,36%	53,64%
	0,2	47,77%	52,23%
0,5	0,05	43,66%	56,34%
	0,1	44,48%	55,52%
	0,2	45,30%	54,70%

Z tabulky 9 lze vidět, že nejvyšší počet správně klasifikovaných subjektů je u momenta 0,2 s rychlostí učení 0,05, kde správnost klasifikace dosáhla 57,63%. Detailněji správnost klasifikace u dat Rating podniků ukázaných v tabulce 10, která znázorňuje matici záměn nejvyšší správně klasifikované subjekty a to 57,63% u momenta 0,2 s rychlostí učení 0,05. Správně klasifikované subjekty SK_{test} u dat Rating podniků se pohybují v rozmezí 52,23% - 57,63%. Špatně klasifikované subjekty jsou nejvyšší u momenta 0,4 s rychlostí učení 0,2 a to 47,77%.

tabulka 10: Matice záměn – Rating podniků

	Aaa	Aa	A	Baa	Ba,B	Caa	Ca	C	D	celkem
Aaa	4	0	3	0	0	0	0	0	0	7
Aa	2	1	10	2	0	0	0	0	0	15
A	1	2	49	42	6	0	1	0	0	101
Baa	0	1	23	119	44	7	0	0	0	194
Ba,B	0	0	6	45	145	65	2	0	0	263
Caa	0	0	1	6	56	165	10	0	0	238
Ca	0	0	0	0	3	14	8	0	0	25
C	0	0	0	1	0	0	2	0	0	3
D	0	0	0	0	1	3	2	0	0	6
celkem	7	4	92	215	255	254	23	0	0	852

Tabulka 10 obsahuje 9 tříd s počtem 852 podniků. Poslední sloupec zobrazuje původní počet podniků v jednotlivých třídách. Poslední řádek zobrazuje získané hodnoty klasifikace pomocí dopředné neuronové sítě, které jsou rozděleny do jednotlivých tříd. Dopředná neuronová síť nebyla schopna správně klasifikovat třídy Ca a D.



graf 5: Porovnání četností klasifikace - Rating podniků

Z grafu 5 lze vyčíst, že ve třídě B vzrostl počet podniků z původních 238 na 255 a naopak ve třídě Ba,B klesl počet podniků z původních 263 na 255. Třída Ba,B obsahuje subjekty s nejistou a problematickou úrovní rizika.

tabulka 11: Správnost klasifikace - Bonita zákazníků

Momentum	Rychlost učení	Nesprávně klasifikované	Správně klasifikované
0,2	0,05	28,30%	71,70%
	0,12	28,00%	72,00%
	0,2	28,30%	71,70%
0,3	0,05	27,90%	72,10%
	0,12	27,60%	72,40%
	0,2	27,50%	72,50%
0,4	0,05	27,90%	72,10%
	0,12	28,30%	71,70%
	0,2	27,60%	72,40%
0,5	0,05	28,60%	71,40%
	0,12	27,50%	72,50%
	0,2	28,40%	71,60%

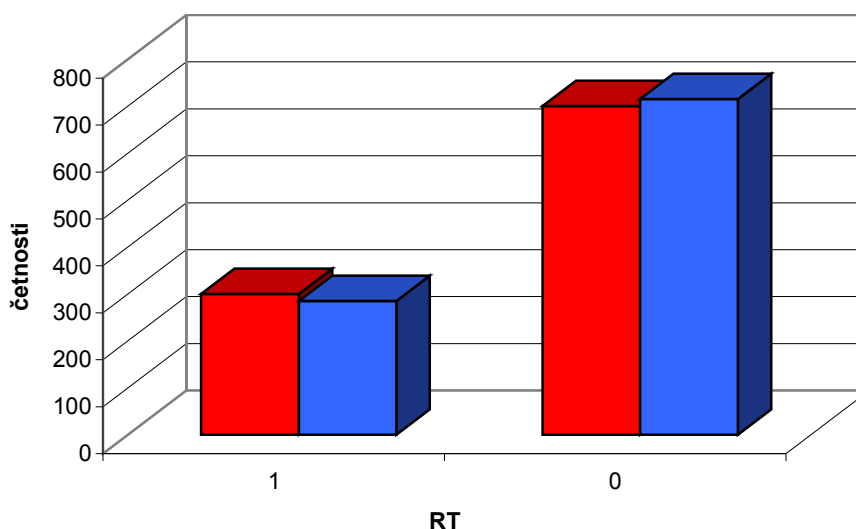
Z tabulky 9 lze vidět, že nejvyšší počet správně klasifikovaných subjektů je u momenta 0,3 s rychlostí učení 0,2, kde správnost klasifikace dosáhla 72,50%. Detailněji nám správnost klasifikace u dat Bonita zákazníků ukáže tabulka 11, která znázorňuje matici záměn nejvyšší správně klasifikované hodnoty a to 72,50% u momenta 0,3 s rychlostí učení 0,2. Správně klasifikované hodnoty u dat Bonita zákazníků se pohybují v rozmezí 71,40% - 72,50%. Nejvyšší hodnota nesprávně klasifikovaných subjektů u dat Bonita zákazníků byla dosažena u momenta 0,5 s rychlostí učení 0,2 a to 28,40%.

tabulka 12: Matice záměn - Bonita zákazníků

	0	1	<i>celkem</i>
0	570	130	<i>700</i>
1	145	155	<i>300</i>
<i>celkem</i>	<i>715</i>	<i>285</i>	<i>1000</i>

Tabulka 12 znázorňuje matici záměn, kde data Bonita zákazníků mají dvě třídy a to třídu 0 a 1 podle toho, zda zákazník je schopen dostát svým závazkům (0) nebo ne (1).

Dopředná neuronová síť klasifikovala správně většinu bonitních zákazníků a asi 50% nebonitních zákazníků.



graf 6: Porovnání četností klasifikace - Bonita zákazníků

Z grafu 8 a 9 je vidět, že se původní hodnoty ve třídě 1 změnilly z 300 na 285 a ve třídě 0 ze 700 na 715.

3.4 Analýza výsledků fuzzy logické neuronové sítě

Tato část práce obsahuje správnost klasifikace dat pomocí fuzzy logické neuronové sítě. Vstupy neuronových sítí jsou data o Bonitě obce, Ratingu podniku a Bonitě zákazníků. Parametry učení fuzzy logické neuronové sítě jsou rychlosti učení (0,05; 0,1 a 0,2) a parametry funkce příslušnosti (2, 3, 4 a 5).

Následující tabulky 13-17 znázorňují výsledky fuzzy logické neuronové sítě zpracované v prostředí Nefclass.

Tabulky 13,15,17 znázorňují správnost klasifikace SK_{test} při různém počtu funkcí příslušnosti a rychlosti učení. Počet funkce příslušnosti byl zvolen 2, 3, 4 a 5 a parametr rychlosti učení 0,2; 0,3; 0,4 a 0,5.

tabulka 13: Správnost klasifikace - Bonita obce

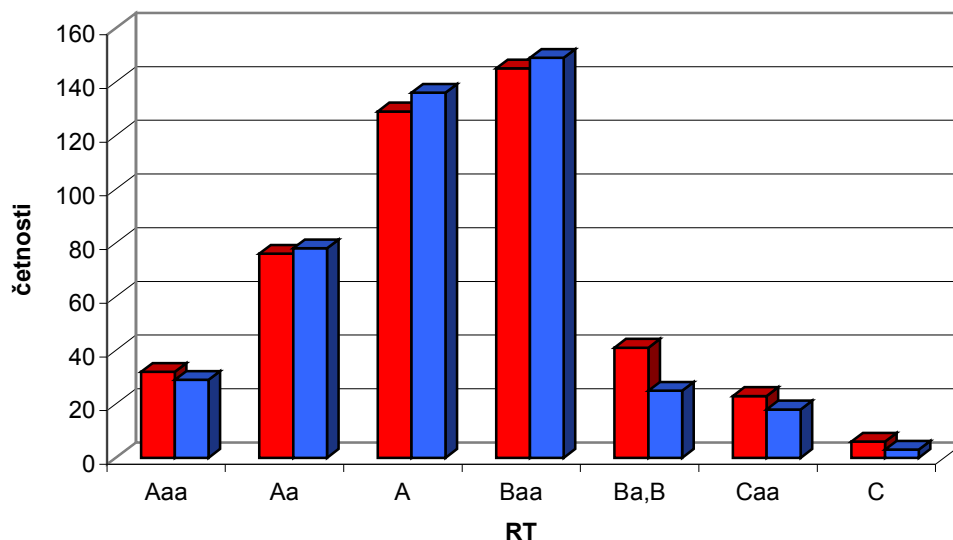
Počet funkcí příslušnosti	Rychlost učení	Nesprávně klasifikované	Správně klasifikované	Počet podmíněných pravidel
2	0,05	18,70%	81,30%	23
	0,1	18,59%	81,41%	23
	0,2	18,37%	81,63%	23
3	0,05	14,37%	85,63%	65
	0,1	15,49%	84,51%	65
	0,2	14,57%	85,43%	65
4	0,05	13,49%	86,51%	110
	0,1	13,08%	86,92%	110
	0,2	13,49%	86,51%	110
5	0,05	23,00%	77,00%	162
	0,1	22,30%	77,70%	162
	0,2	25,99%	74,01%	162

Z tabulky 13 lze vidět, že nejvyšší počet správně klasifikovaných subjektů je u funkce příslušnosti 4 s rychlostí učení 0,1 a to 86,92%. Počet nejvyšších správně klasifikovaných subjektů je u počtu funkcí příslušnosti 4 s počtem podmíněných pravidel 110. Počet správně klasifikovaných subjektů u dat Bonita zákazníků se pohybuje v rozmezí 74,01% až 86,92%. Nejvíce špatně klasifikovaných subjektů je u funkce příslušnosti 5 s rychlostí učení 0,2 a to 25,99%. Počet podmíněných pravidel u nejvyšší hodnoty špatně klasifikovaných subjektů je 162. Nejméně podmíněných pravidel program určil u nejnižšího počtu funkce příslušnosti. Matici záměn znázorňuje tabulka 14.

tabulka 14: Matice záměn - Bonita obce

	Aaa	Aa	A	Baa	Ba,B	Caa	C	celkem
Aaa	29	3	0	0	0	0	0	32
Aa	0	66	5	3	0	0	0	76
A	0	3	120	1	0	1	0	129
Baa	0	1	1	138	1	0	0	145
Ba,B	0	2	6	6	24	1	0	41
Caa	0	2	4	0	0	15	0	23
C	0	1	0	1	0	1	3	6
celkem	29	78	136	149	25	18	3	452

Z tabulky 14 lze určit, jak se počet obcí ve třídách změnil. Tabulka zobrazuje 7 tříd a udává kolik obcí je v každé třídě. Celkem data obsahují 452 obcí. Poslední sloupec zobrazuje původní počet obcí v jednotlivých třídách u dat bonita obce. Poslední řádek v tabulce zobrazuje získaný počet obcí uskutečněné pomocí fuzzy logické neuronové sítě. U třídy Caa fuzzy logická neuronová síť určila 50% správně klasifikovaných subjektů.



graf 7: Porovnání četností klasifikace - Bonita obce

Legenda ke grafu 7, grafu 8 i grafu 9:

Červená barva – původní hodnoty

Modrá barva – hodnoty fuzzy logické neuronové sítě

Z grafu 9 lze vyčíst, že počet obcí vzrostl ve třídě Aa z původních 76 na 78 a naopak počet obcí klesl u tříd Ba, B; Caa a C s průměrnou až vysokou zadlužeností.

Podmíněná pravidla pro správně klasifikované hodnoty s počtem funkce příslušnosti 4 a rychlostí učení 0,1 jsou následující:

R0: IF x_1 je velmi nízký AND x_2 je velmi nízký AND x_3 je velmi nízký AND x_4 je nízký AND x_5 je velmi nízký AND x_6 je velmi nízký AND x_7 je velmi nízký AND x_8 je velmi nízký AND x_9 je velmi nízký AND x_{10} je velmi nízký AND x_{11} je vysoký AND x_{12} je velmi vysoký THEN třída = 0

R2: IF x_1 je velmi nízký AND x_2 je nízký AND x_3 je velmi nízký AND x_4 je nízký AND x_5 je velmi nízký AND x_6 je velmi nízký AND x_7 střední AND x_8 je velmi nízký AND x_9

je velmi nízký AND x_{10} je velmi nízký AND x_{11} je velmi vysoký AND x_{12} je velmi nízký THEN třída = 0

·
·
·
·

R86: IF x_1 je velmi nízký AND x_2 je velmi nízký AND x_3 je nízký AND x_4 je střední AND x_5 je velmi nízký AND x_6 je velmi nízký AND x_7 velmi nízký AND x_8 je nízký AND x_9 je velmi nízký AND x_{10} je velmi nízký AND x_{11} je velmi nízký AND x_{12} je velmi nízký THEN třída = 1

R90: IF x_1 je velmi nízký AND x_2 je velmi nízký AND x_3 je nízký AND x_4 je nízký AND x_5 je velmi nízký AND x_6 je velmi nízký AND x_7 velmi nízký AND x_8 je nízký AND x_9 je velmi nízký AND x_{10} je velmi nízký AND x_{11} je velmi nízký AND x_{12} je velmi nízký THEN třída = 2

·
·
·
·

R110: IF x_1 je velmi nízký AND x_2 je velmi vysoký AND x_3 je velmi nízký AND x_4 je velmi nízký AND x_5 je velmi vysoký AND x_6 je velmi nízký AND x_7 velmi nízký AND x_8 je velmi nízký AND x_9 je velmi nízký AND x_{10} je velmi nízký AND x_{11} je střední AND x_{12} je velmi nízký THEN třída = 0

tabulka 15: Správnost klasifikace - Rating podniků

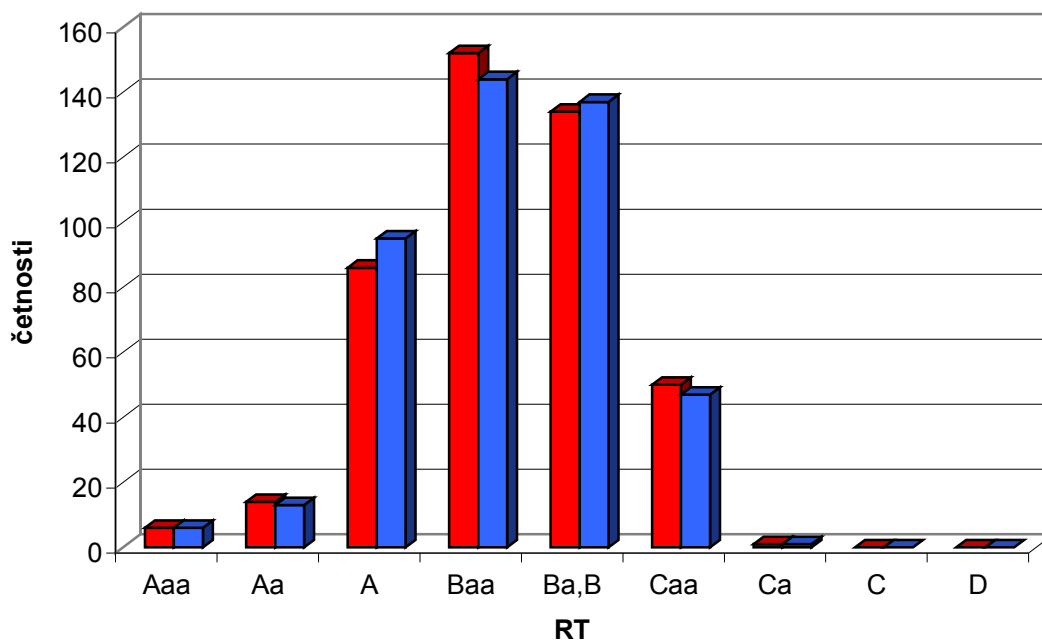
Počet funkcí příslušnosti	Rychlost učení	Nesprávně klasifikované	Správně klasifikované	Počet podmíněných pravidel
2	0,05	42,20%	57,80 %	172
	0,1	41,77%	58,23%	172
	0,2	42,20%	57,80%	172
3	0,05	-	-	-
	0,1	-	-	-
	0,2	-	-	-
4	0,05	31,61%	68,39%	264
	0,1	32,06%	67,94%	264
	0,2	32,74%	67,26%	264
5	0,05	23,06%	76,94%	333
	0,1	25,25%	74,75%	333
	0,2	27,27%	72,73%	333

Z tabulky 15 lze vidět, že nejvyšší správnost klasifikace je dosažena u počtu funkcí příslušnosti 5 s rychlostí učení 0,05 a to 76,94%. Počet podmíněných pravidel u správně klasifikovaných subjektů je 333. Počet správně klasifikovaných subjektů u dat Rating podniků se pohybuje v rozmezí 57,80% až 76,94%. Nejvíce špatně klasifikovaných subjektů je u funkce příslušnosti 2 s rychlostí učení 0,2 a to 42,20%. Počet podmíněných pravidel u nejvyšší hodnoty špatně klasifikovaných subjektů je 172. Hodnoty u počtu funkcí příslušnosti 3 s rychlostí učení 0,05; 0,1; a 0,2 program nezpracoval. Matici záměn zobrazuje tabulka 16.

tabulka 16: Matice záměn - Rating podniků

	Aaa	Aa	A	Baa	Ba,B	Caa	Ca	C	D	celkem
Aaa	6	0	0	0	0	0	0	0	0	6
Aa	0	13	0	1	0	0	0	0	0	14
A	0	0	70	0	7	9	0	0	0	86
Baa	0	0	18	122	12	0	0	0	0	152
Ba,B	0	0	7	12	103	12	0	0	0	134
Caa	0	0	0	9	15	26	0	0	0	50
Ca	0	0	0	0	0	0	1	0	0	1
C	0	0	0	0	0	0	0	0	0	0
D	0	0	0	0	0	0	0	0	0	0
celkem	6	13	95	144	137	47	1	0	0	443

Tabulka 16 znázorňující matici záměn u dat Rating podniků obsahuje 9 tříd s počtem podniků 443. Poslední sloupec zobrazuje původní hodnoty rozdělené do jednotlivých tříd. Poslední řádek v tabulce zobrazuje získané hodnoty, uskutečněné pomocí fuzzy logické neuronové sítě. Třídy Ca a D nebylo možné pomocí fuzzy logické neuronové sítě správně klasifikovat.



graf 8: Porovnání četností klasifikace - Rating podniků

Z grafu 8 lze vyčíst, že u třídy A vzrostl počet podniků od původních 86 na 95. A naopak u třídy Caa počet podniků klesl z původních 50 na 47.

Podmíněná pravidla pro správně klasifikované hodnoty s počtem funkce příslušnosti 5 a rychlostí učení 0,05 jsou následující:

R0: IF x_1 je velmi vysoký AND x_2 je velmi nízký AND x_3 je velmi vysoký AND x_4 je velmi vysoký AND x_5 je velmi nízký AND x_6 je velmi nízký AND x_7 je velmi nízký AND x_8 je velmi nízký AND x_9 je velmi nízký AND x_{10} je nízký AND x_{11} je nízký AND x_{12} je velmi nízký AND x_{13} je velmi nízký AND x_{14} je velmi vysoký AND x_{15} je velmi nízký AND x_{16} je nízký AND x_{17} je velmi nízký AND x_{18} je velmi vysoký AND x_{19} je velmi nízký AND x_{20} je nízký AND x_{21} je vysoký AND x_{22} je velmi vysoký AND x_{23} je velmi vysoký AND x_{24} je velmi vysoký AND x_{25} je vysoký AND x_{26} je velmi nízký AND x_{27} je velmi nízký AND x_{28} je velmi nízký AND x_{29} je střední AND x_{30} je

velmi nízký AND x_{31} je velmi nízký AND x_{32} je nízký AND x_{33} je nízký AND x_{34} je velmi nízký THEN třída = 0

.
. .
. .
. .

R266: IF x_1 je vysoký AND x_2 je velmi nízký AND x_3 je velmi nízký AND x_4 je velmi nízký AND x_5 je velmi nízký AND x_6 je velmi nízký AND x_7 je velmi nízký AND x_8 je velmi nízký AND x_9 je velmi nízký AND x_{10} je velmi nízký AND x_{11} je velmi nízký AND x_{12} je nízký AND x_{13} je velmi nízký AND x_{14} je velmi vysoký AND x_{15} je velmi nízký AND x_{16} je velmi nízký AND x_{17} je velmi nízký AND x_{18} je velmi nízký AND x_{19} je velmi nízký AND x_{20} je velmi nízký AND x_{21} je velmi nízký AND x_{22} je velmi nízký AND x_{23} je velmi nízký AND x_{24} je velmi nízký AND x_{25} je velmi nízký AND x_{26} je nízký AND x_{27} je velmi vysoký AND x_{28} je velmi nízký AND x_{29} je velmi vysoký AND x_{30} je velmi nízký AND x_{31} je velmi nízký AND x_{32} je velmi nízký AND x_{33} je velmi nízký AND x_{34} je velmi nízký THEN třída = 0

.
. .
. .
. .

R333: IF x_1 je velmi nízký AND x_2 je velmi nízký AND x_3 je velmi nízký AND x_4 je velmi nízký AND x_5 je velmi nízký AND x_6 je velmi nízký AND x_7 je velmi nízký AND x_8 je velmi nízký AND x_9 je velmi nízký AND x_{10} je střední AND x_{11} je střední AND x_{12} je nízký AND x_{13} je vysoký AND x_{14} je velmi nízký AND x_{15} je velmi nízký AND x_{16} je nízký AND x_{17} je velmi nízký AND x_{18} je velmi nízký AND x_{19} je velmi nízký AND x_{20} je velmi nízký AND x_{21} je vysoký AND x_{22} je velmi nízký AND x_{23} je velmi nízký AND x_{24} je velmi nízký AND x_{25} je velmi nízký AND x_{26} je vysoký AND x_{27} je střední AND x_{28} je velmi nízký AND x_{29} je nízký AND x_{30} je velmi nízký AND x_{31} je velmi nízký AND x_{32} je velmi nízký AND x_{33} je velmi nízký AND x_{34} je velmi nízký THEN třída = 0

tabulka 17: Správnost klasifikace - Bonita zákazníků

Počet funkcí příslušnosti	Rychlost učení	Nesprávně klasifikované	Správně klasifikované	Počet podmíněných pravidel
2	0,05	32,70%	67,30%	738
	0,1	32,50%	67,50%	738
	0,2	33,10%	66,90%	738
3	0,05	30,70%	69,30%	920
	0,1	35,40%	64,60%	920
	0,2	37,20%	62,80%	920
4	0,05	25,20%	74,80%	941
	0,1	29,80%	70,20%	941
	0,2	31,40%	68,60%	941
5	0,05	13,80%	86,20%	989
	0,1	15,00%	85,00%	989
	0,2	16,30%	83,70%	989

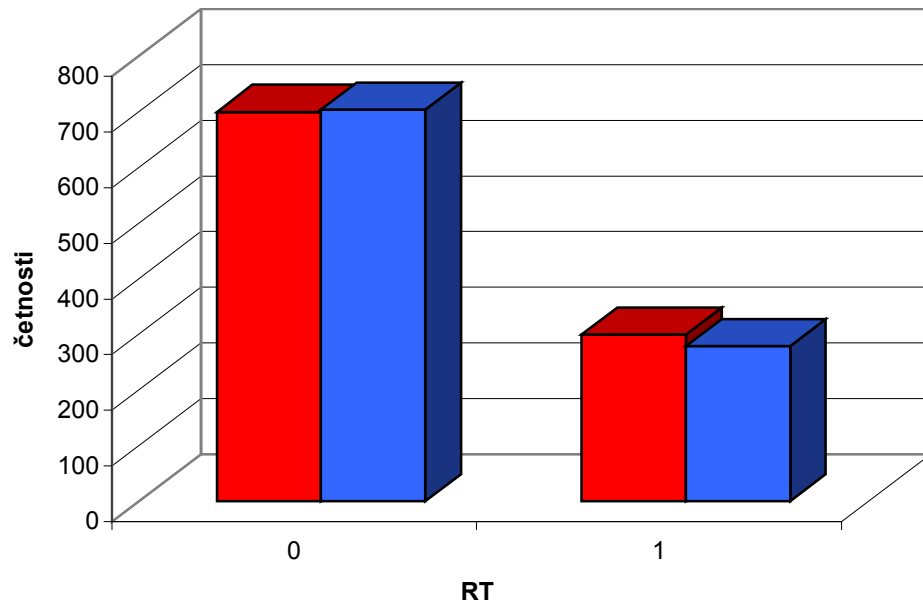
Z tabulky lze vidět, že nejvyšších správně klasifikovaných subjektů je dosaženo u počtu funkcí příslušnosti 5 s rychlostí učení 0,05 a to 86,20%. Počet nejvyšších správně klasifikovaných subjektů je u počtu funkcí příslušnosti 5 s počtem podmíněných pravidel 989. Počet správně klasifikovaných hodnot u dat Bonita zákazníků se pohybuje v rozmezí 67,30% až 86,20%. Nejvíce špatně klasifikovaných hodnot je u funkce příslušnosti 3 s rychlostí učení 0,2 a to 37,20%. Počet podmíněných pravidel u nejvyšší hodnoty špatně klasifikovaných subjektů je 920. Matici záměn znázorňuje tabulka 18.

tabulka 18: Matice záměn - Bonita zákazníků

	0	1	<i>celkem</i>
0	632	68	700
1	72	228	300
<i>celkem</i>	704	279	1000

Matice záměn, která je zobrazena v tabulce 18, obsahuje dvě třídy u dat Bonita zákazníků. Třídy 0 a 1 podle toho, zda zákazník je schopen dostát svým závazkům (0) nebo ne (1). V posledním sloupci je znázorněna původní hodnota a v posledním řádku získaná hodnota, která je uskutečněná pomocí fuzzy logické neuronové sítě. Fuzzy logická

neuronová síť klasifikovala správně většinu bonitních zákazníků a kolem 70% nebonitních zákazníků.



graf 9: Porovnání četností klasifikace - Bonita zákazníků

Z grafu 9 lze vyčíst, že ve třídě 1 vzrostl počet hodnot ze 700 na 704 a ve třídě 0 klesl počet hodnot z 300 na 279.

Podmíněná pravidla pro správně klasifikované hodnoty s počtem funkce příslušnosti 5 a rychlostí učení 0,05 jsou následující:

R0: IF x_1 je nízký AND x_2 je nízký AND x_3 je vysoký AND x_4 je nízký AND x_5 je vysoký AND x_6 je vysoký AND x_7 je vysoký AND x_8 je vysoký AND x_9 je nízký AND x_{10} je vysoký AND x_{11} je vysoký AND x_{12} je nízký AND x_{13} je nízký AND x_{14} je vysoký AND x_{15} je nízký AND x_{16} je nízký AND x_{17} je nízký AND x_{18} je vysoký AND x_{19} je nízký AND x_{20} je nízký AND x_{21} je vysoký AND x_{22} nízký AND x_{23} je nízký AND x_{24} je vysoký THEN = 0

·
·
·

R553: IF x_1 je vysoký AND x_2 je nízký AND x_3 je střední AND x_4 je nízký AND x_5 je nízký AND x_6 je nízký AND x_7 je vysoký AND x_8 je vysoký AND x_9 je vysoký AND x_{10} je nízký AND x_{11} je vysoký AND x_{12} je nízký AND x_{13} je nízký AND x_{14} je nízký AND x_{15} je nízký AND x_{16} je nízký AND x_{17} je nízký AND x_{18} je vysoký AND x_{19} je nízký AND x_{20} je nízký AND x_{21} je vysoký AND x_{22} nízký AND x_{23} je vysoký AND x_{24} je nízký THEN = 0

·
·
·
·
·

R920: IF x_1 je vysoký AND x_2 je nízký AND x_3 je střední AND x_4 je nízký AND x_5 je střední AND x_6 je střední AND x_7 je nízký AND x_8 je vysoký AND x_9 je vysoký AND x_{10} je vysoký AND x_{11} je vysoký AND x_{12} je nízký AND x_{13} je nízký AND x_{14} je nízký AND x_{15} je nízký AND x_{16} je nízký AND x_{17} je nízký AND x_{18} je vysoký AND x_{19} je nízký AND x_{20} je nízký AND x_{21} je vysoký AND x_{22} nízký AND x_{23} je vysoký AND x_{24} je nízký THEN = 0

Závěr

V práci je analyzován současný stav řešení v oblasti dopředných neuronových sítí a fuzzy logických neuronových sítí. Jsou uvedeny základní pojmy řešené problematiky a současné metody ohodnocování bonity obcí, ratingu podniků a bonity zákazníků. Ty jsou založeny na kombinaci znalosti expertů a matematicko-statistických metod. Metody ohodnocování bonity ekonomických subjektů lze rozdělit na bodovací systémy, rating a modely selhání. Rating je ohodnocování bonity, které provádí expert. Umožňuje tedy pracovat s přirozeným jazykem. Jeho nevýhodou jsou velké náklady a subjektivita ohodnocování. S touto skutečností je spojena nemožnost ověření přesnosti ohodnocování. Tyto nedostatky se snaží odstranit modely na bázi ratingu. Jejich cílem je klasifikovat objekty do tříd, například pomocí statistických metod. Výsledky těchto modelů se vyznačují nízkou správností klasifikace, která je dána omezeným počtem dat, použitím nevhodných metod, chybnou volbou parametrů ratingu, a také uvedenou subjektivitou ratingu. Selháním se rozumí neschopnost subjektu dostát svým závazkům.

Je potřebné navrhnout takový model, který by tyto nedostatky odstranil. K tomuto účelu je v práci navržen vektor parametrů bonity obcí, ratingu podniků a bonity zákazníků. Tyto vektory obsahují ekonomické, finanční a další parametry s cílem určit klíčové činitele ohodnocování bonity ekonomických subjektů. Pro klasifikaci ekonomických subjektů jsou vhodné neuronové sítě a fuzzy logické neuronové sítě. Na základě uvedených skutečností jsou v práci uvedeny základní pojmy a metody z oblasti neuronových sítí, pojmy z teorie fuzzy množin, fuzzy logiky a fuzzy logických neuronových sítí.

Dále byla provedena aplikace uvedených metod na ekonomických datech. Zpracovaná data se týkají bonity obcí, ratingu podniku a bonity zákazníků. Cílem práce bylo vytvoření takových klasifikátorů, které budou po naučení schopny klasifikovat správně jak trénovací, tak testovací data (tj. budou schopny generalizace) a zároveň budou uživateli srozumitelné rozhodovací procesy, které klasifikátory realizují.

Z procesu klasifikace byla získána správnost klasifikace pomocí rozdělení dat křížovou validací. Model byl verifikován v programovém prostředí Nefclass a v programovém prostředí Weka. Původní data byla klasifikována do sedmi (Bonita obcí), devíti (Rating podniků) a dvou tříd (Bonita zákazníků).

Výsledky ukazují správnost klasifikace testovacích dat. Po porovnání výsledků jsem zjistila následující výsledky správnosti klasifikace.

Pomocí dopředné neuronové sítě bylo nejlepších výsledků dosaženo u dat Bonita obce a to 86,95% s momentem 0,2 a rychlostí učení 0,05. U dat Rating podniků bylo nejlepších správných výsledků dosaženo 57,63% s momentem 0,2 a rychlostí učení 0,05 a u dat Bonita zákazníků 72,50% s momentem 0,3 a rychlostí učení 0,2.

Pomocí fuzzy logických neuronových sítí bylo nejlepších výsledků dosaženo u dat Bonita obce, kde bylo zjištěno 86,92% při počtu funkce příslušnosti 4 a rychlostí učení 0,1 a s počtem podmíněných pravidel 110. U dat Rating podniků bylo zjištěno 76,94% při počtu funkce příslušnosti 5 a rychlostí učení 0,05 a s počtem podmíněných pravidel 333 a u dat Bonita zákazníků 86,20% při počtu funkce příslušnosti 5 a rychlostí učení 0,05 a počtem podmíněných pravidel 989.

Výsledky dosažené v bakalářské práci potvrzují předpoklad, že dopředné neuronové sítě i fuzzy logické neuronové sítě jsou vhodné jako metody pro klasifikaci ekonomických subjektů. Dopředné neuronové sítě mají schopnost učit se znalosti ze vstupních dat. Fuzzy logické neuronové sítě jsou schopné pracovat s nepřesností a neurčitostí.

Cíl práce, návrh modelu na klasifikaci ekonomických subjektů do tříd podle jejich bonity, byl splněn v kapitole 4. Navržený model byl verifikován na třech datových souborech. Porovnány byly výsledky klasifikace dopředné neuronové sítě a fuzzy logické neuronové sítě z hlediska správnosti klasifikace a efektivity. Experimenty byly realizovány pro různé hodnoty parametrů učení.

Použitá literatura

- [1] HÁJEK, P. *Modelování bonity obcí pomocí metod výpočetní inteligence*. Disertační práce. Univerzita Pardubice: Pardubice, 2006.
- [2] HEBÁK, P. *Vícerozměrné statistické metody*. Praha: Informatorium, 2004. ISBN 80-7333-025-3.
- [3] KVASNIČKA, V. a kol. *Úvod do teorie neuronových sítí*. Bratislava: IRIS, 1997. 285 s.
- [4] MAŘÍK, V., ŠTĚPÁNKOVÁ, O., LAŽANSKÝ, J. a kol. *Umělá inteligence (4)*. Praha: ACADEMIA, 2003. 475 s. ISBN 80-200-1044-0.
- [5] MELOUN, M., MILITKÝ, J. *Statistická analýza experimentálních dat*. Praha: Academia, 2004. 449 s. ISBN 80-200-1254-0.
- [6] NOVÁK, V. *Fuzzy množiny a jejich aplikace*. Praha: SNTL, 1986. 278 s.
- [7] NOVÁK, V. *Základy fuzzy modelování*. Praha : BEN, 2000. 175 s. ISBN 80-7300-009-1.
- [8] OLE J, V. *Modelovanie ekonomických procesov na báze výpočtovej inteligencie*. Hradec Králové: Miloš Vognar-MandV, 2003. 160 s., ISBN 80-903024-9-1.
- [9] PETR, Pavel. *Data Mining*. [s. l.] : [s.n.], 2006. 144 s. ISBN 80-7194-886-1.
- [10] TUČKOVÁ, Jana. *Úvod do teorie a aplikací umělých neuronových sítí*. ČVUT : [s.n.], 2003. 103 s. ISBN 80-01-02800-3.
- [11] RUMPLÍK, Tomáš. *Modelování bonity obcí pomocí neuro-fuzzy inferenčních systémů*. [s.l.], 2007. 86 s. UPCE - FES. Vedoucí diplomové práce Olej Vladimír.

- [12] RYDVAL, Slávek. *Základy fuzzy logiky* [online]. 2005. 2000 [cit. 2009-03-23].
Dostupný z
WWW:<<http://www.rydval.cz/phprs/view.php?cisloclanku=2005061701>>.
- [13] VESELOVSKÝ , Michal. *Neuronové sítě* [online]. 1999 [cit. 2009-03-23].
Dostupný z WWW:<<http://www.avari.cz/uir/>>.
- [14] Wikipedie - otevřená encyklopedie [online]. 2008 [cit. 2008-03-31]. Dostupný z
WWW:<http://cs.wikipedia.org/wiki/Hlavn%C3%AD_strana>.

Seznam tabulek, obrázků a grafů

graf 1: Četnosti subjektů v třídách - Bonita obce	29
graf 2: Četnosti subjektů v třídách - Rating podniků	31
graf 3: Četnosti subjektů v třídách - Bonita zákazníků	33
graf 4: Porovnání četností klasifikace - Bonita obce	36
graf 5: Porovnání četností klasifikace - Rating podniků	38
graf 6: Porovnání četností klasifikace - Bonita zákazníků	40
graf 7: Porovnání četností klasifikace - Bonita obce	42
graf 8: Porovnání četností klasifikace - Rating podniků	45
graf 9: Porovnání četností klasifikace - Bonita zákazníků	48
tabulka 1: Booleova funkce. Zdroj [9]	17
tabulka 2: Analogie mezi biologickými a umělými neuronovými sítěmi. Zdroj [11]	18
tabulka 3: Třídy bonity obce. Zdroj [1]	28
tabulka 4: Vstupy proměnných pro společný rating modelování. Zdroj [1]	30
tabulka 5: Třídy ratingu. Zdroj [1]	31
tabulka 6: Popis bonity zákazníků. Zdroj [1]	32
tabulka 7: Správnost klasifikace - Bonita obce	35
tabulka 8: Matice záměn - Bonita obce	36
tabulka 9: Správnost klasifikace - Rating podniků	37
tabulka 10: Matice záměn – Rating podniků	38
tabulka 11: Správnost klasifikace - Bonita zákazníků	39
tabulka 12: Matice záměn - Bonita zákazníků	39
tabulka 13: Správnost klasifikace - Bonita obce	41
tabulka 14: Matice záměn - Bonita obce	41
tabulka 15: Správnost klasifikace - Rating podniků	44
tabulka 16: Matice záměn - Rating podniků	44
tabulka 17: Správnost klasifikace - Bonita zákazníků	47
tabulka 18: Matice záměn - Bonita zákazníků	47
obrázek 1: Funkce příslušnosti $\mu_I(x)$, $\mu_U(x)$, $\mu_{\neg A}(x)$. Zdroj [9]	13
obrázek 2: Funkce příslušnosti S a Π fuzzy množiny. Zdroj [9]	14
obrázek 3: Trojúhelníková a lichoběžníková funkce příslušnosti. Zdroj [9]	14
obrázek 4: Neuron	16
obrázek 5: Jednoduchá neuronová síť. Zdroj [9]	16
obrázek 6: Model umělého neuronu. Zdroj [15]	17
obrázek 7: Vícevrstvý perceptron. Zdroj [14]	21
obrázek 8: Model FLNS. Zdroj [13]	23
obrázek 9: Model FLNS vhodný pro klasifikaci. Zdroj [13]	25
obrázek 10: Blokové schéma modelu	33

Příloha A

Statistika dat Bonita obce

tabulka 1: Statistika Bonita obce

proměnná	průměr	směrodatná odchylka	minimum	maximum
x_1	1117.89	4667.89	27.00	88181.00
x_2	0.98	0.16	0.38	2.57
x_3	11.71	5.88	0.00	50.00
x_4	0.20	0.05	0.11	0.44
x_5	0.07	0.11	0.00	0.82
x_6	4174.69	15792.61	0.00	306480.82
x_7	0.51	0.42	0.00	1.00
x_8	1.21	0.36	0.26	2.59
x_9	0.08	0.08	0.00	0.62
x_{10}	0.22	0.18	0.00	0.83
x_{11}	0.15	0.18	0.00	0.78
x_{12}	69540.14	65203.00	5422.78	820591.81

Příloha B

Statistika dat Rating podniků.

tabulka 2: Statistika Rating podniků

proměnná	průměr	směrodatná odchylka	minimum	maximum
x ₁	8.85	1.14	5.00	10.00
x ₂	44.74	44.10	0.00	792.00
x ₃	18722.97	44080.42	130.30	511719.20
x ₄	21070.89	54375.93	341.00	732751.90
x ₅	19.91	18.95	4.63	277.59
x ₆	3.07	3.68	-0.34	63.28
x ₇	4.08	7.24	0.33	129.26
x ₈	1.71	1.65	0.05	11.97
x ₉	8.33	4.14	1.77	31.08
x ₁₀	0.15	0.17	-0.25	0.89
x ₁₁	1.17	0.64	-0.67	4.99
x ₁₂	0.23	0.10	0.06	0.84
x ₁₃	0.26	0.09	0.10	0.64
x ₁₄	0.67	0.13	0.31	0.80
x ₁₅	0.21	0.23	0.00	1.77
x ₁₆	0.20	0.14	0.00	1.60
x ₁₇	0.09	0.07	0.00	0.48
x ₁₈	5407.33	13106.38	65.80	115540.00
x ₁₉	1155.44	15861.99	-11455.00	331379.00
x ₂₀	1333.05	4888.61	-636.32	87864.24
x ₂₁	0.32	0.09	0.00	0.60
x ₂₂	486.57	1346.47	5.80	14545.00
x ₂₃	13790.17	31977.42	233.40	367416.00
x ₂₄	1112.53	2882.32	2.30	38790.00
x ₂₅	308.52	1066.50	0.00	12743.98
x ₂₆	0.20	0.16	0.00	0.87
x ₂₇	0.37	0.18	0.00	0.93
x ₂₈	0.01	0.01	0.00	0.08
x ₂₉	0.68	0.31	0.00	1.00
x ₃₀	0.38	0.29	0.00	2.03
x ₃₁	1.55	0.95	0.20	7.69
x ₃₂	29.46	161.74	0.14	2088.93
x ₃₃	1.15	0.72	0.17	4.85
x ₃₄	69.60	230.17	1.03	3173.29

Příloha C

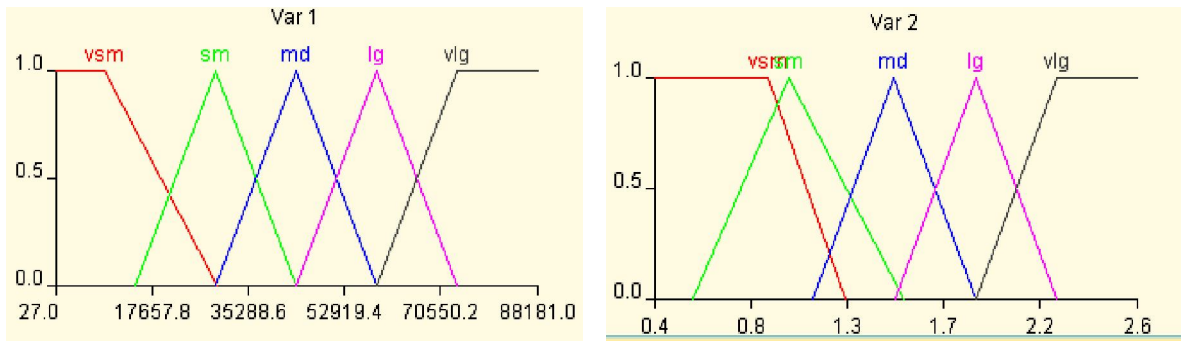
Statistika Bonita zákazníků.

tabulka 3: Statistika Bonita zákazníků

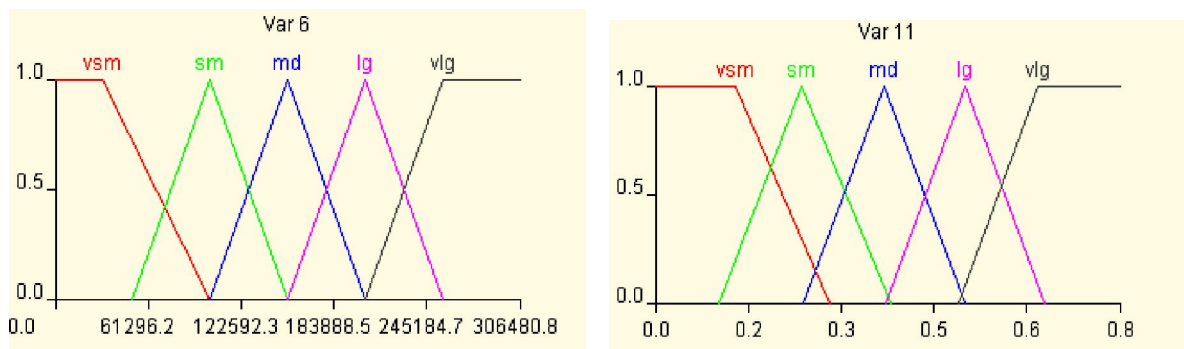
proměnná	průměr	Směrodatná odchylka	minimum	maximum
x₁	2.58	1.26	1.00	4.00
x₂	20.90	12.05	4.00	72.00
x₃	2.54	1.08	0.00	4.00
x₄	32.71	28.24	2.00	184.00
x₅	2.10	1.58	1.00	5.00
x₆	3.38	1.21	1.00	5.00
x₇	2.68	0.71	1.00	4.00
x₈	2.84	1.10	1.00	4.00
x₉	2.36	1.05	1.00	4.00
x₁₀	35.55	11.37	19.00	75.00
x₁₁	2.68	0.71	1.00	3.00
x₁₂	1.41	0.58	1.00	4.00
x₁₃	1.16	0.36	1.00	2.00
x₁₄	1.40	0.49	1.00	2.00
x₁₅	1.04	0.19	1.00	2.00
x₁₆	0.23	0.42	0.00	1.00
x₁₇	0.10	0.30	0.00	1.00
x₁₈	0.91	0.29	0.00	1.00
x₁₉	0.04	0.20	0.00	1.00
x₂₀	0.18	0.38	0.00	1.00
x₂₁	0.71	0.45	0.00	1.00
x₂₂	0.02	0.15	0.00	1.00
x₂₃	0.20	0.40	0.00	1.00
x₂₄	0.63	0.48	0.00	1.00

Příloha D

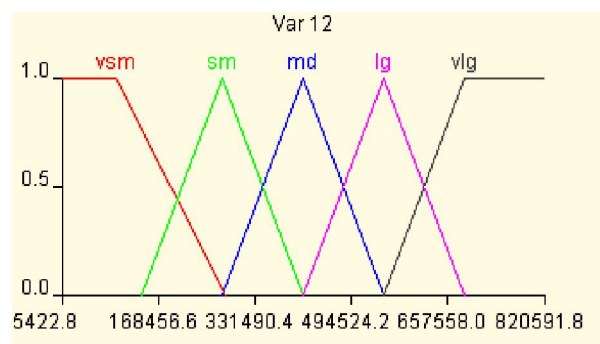
Funkce příslušnosti pro vstupní proměnné Bonita obce. Např. x_1 daňové příjmy obce (VSM = velmi nízké, SM = nízké, MD = střední, LG = vysoké, VLG = velmi vysoké).



obrázek 1: Funkce příslušnosti x_1 a x_2 - Bonita obce



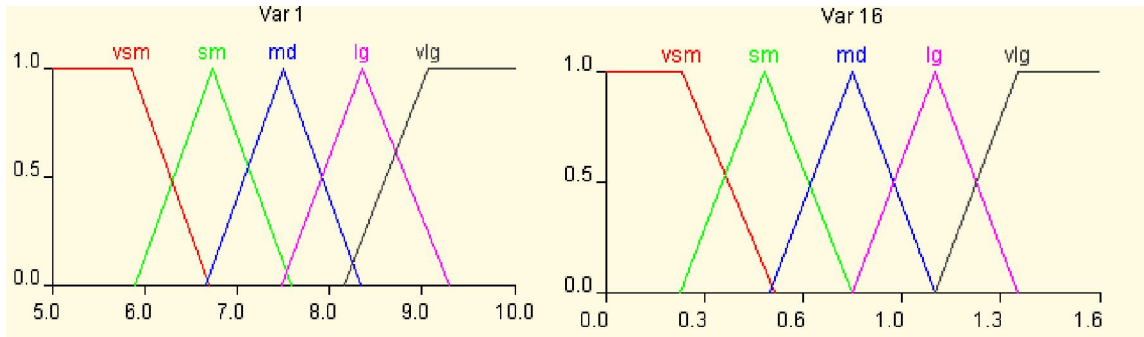
obrázek 2: Funkce příslušnosti x_6 a x_{11} - Bonita obce



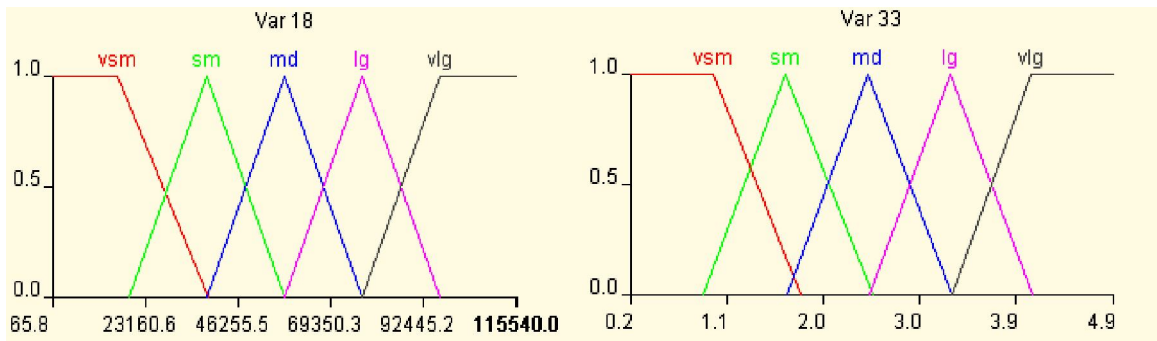
obrázek 3: Funkce příslušnosti x_{12} - Bonita obce

Příloha E

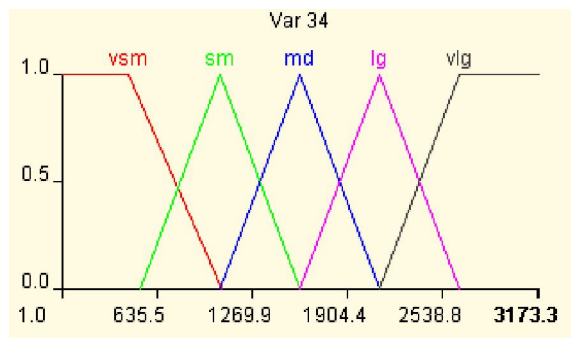
Funkce příslušnosti pro vstupní proměnné Ratingu podniků.



obrázek 4: Funkce příslušnosti x_1 a x_{16} - Rating podniků



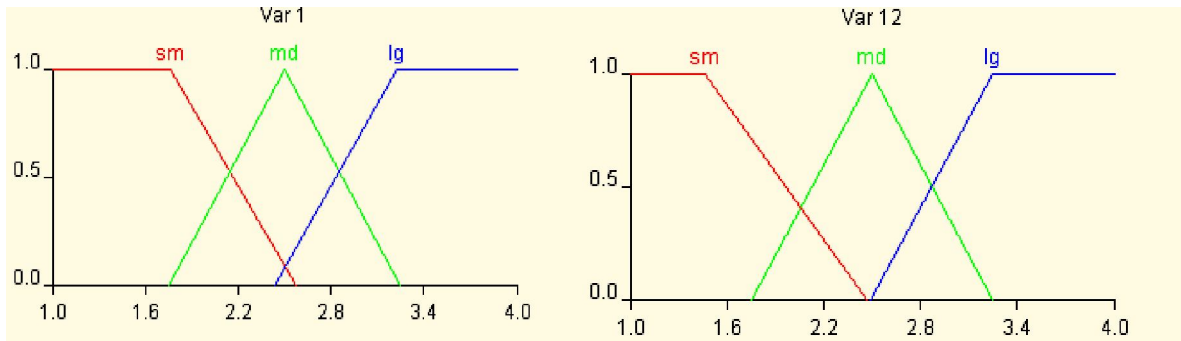
obrázek 5: Funkce příslušnosti x_{18} a x_{33} - Rating podniků



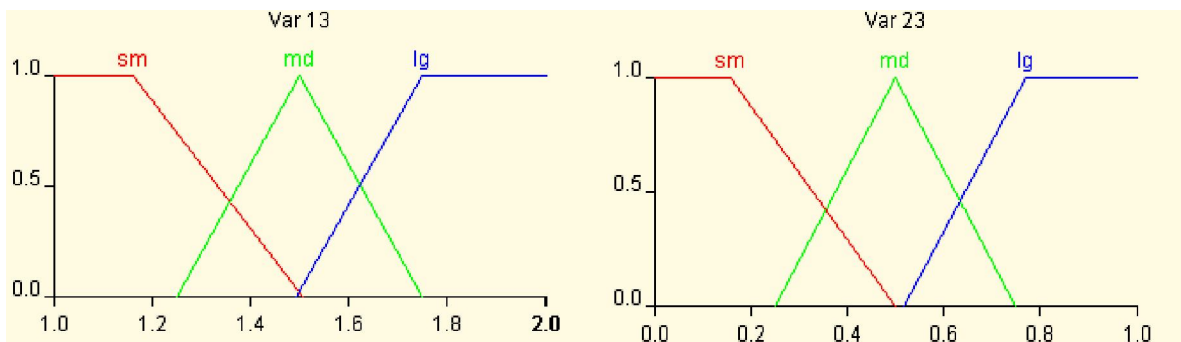
obrázek 6: Funkce příslušnosti x_{34} - Rating podniků

Příloha F

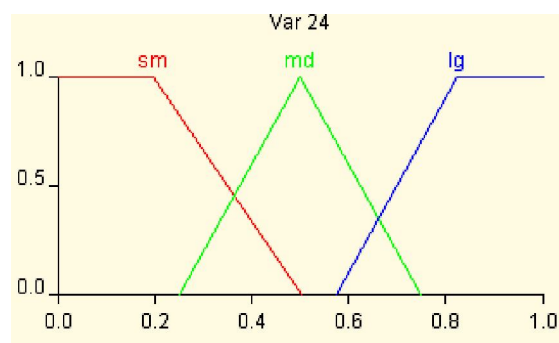
Funkce příslušnosti pro vstupní proměnné Bonita zákazníků.



obrázek 7: Funkce příslušnosti x_1 a x_{12} - Bonita zákazníků



obrázek 8: Funkce příslušnosti x_{13} a x_{23} - Bonita zákazníků



obrázek 9: Funkce příslušnosti x_{24} - Bonita zákazníků