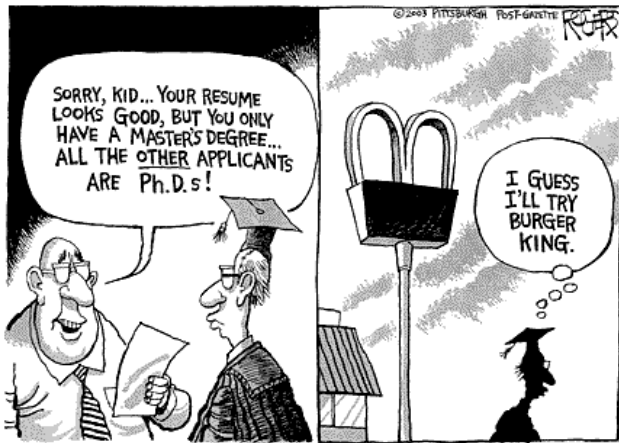


Analýza očekávané délky nezaměstnanosti na Sokolovsku

Luděk Cigler

8. května 2007

Kdo je ohrožen dlouhodobou nezaměstnaností?



1

¹<http://www.robrogers.com/cartoons/>

Nezaměstnanost na Sokolovsku – data

- Celkem více než 80 tisíc záznamů získaných z úřadů práce na Sokolovsku

Rok	RČ	Pohl.	Narozen	Nez. od	Nez. do	Vzděl.	Zdr. st.	Rod. st.	Bydl.
2000	64...	M	16/7/1964	20/3/2000	21/9/2000	M	1	2	Sokolov
2001	73...	Ž	20/3/1973	17/5/2001	5/3/2002	C	2	3	Chodov

- Někteří lidé jsou vedeni v databázi vícekrát, v případě opakované nezaměstnanosti
- Některá data jsou nekompletní (pokud je daný člověk stále nezaměstnaný)

Úprava dat

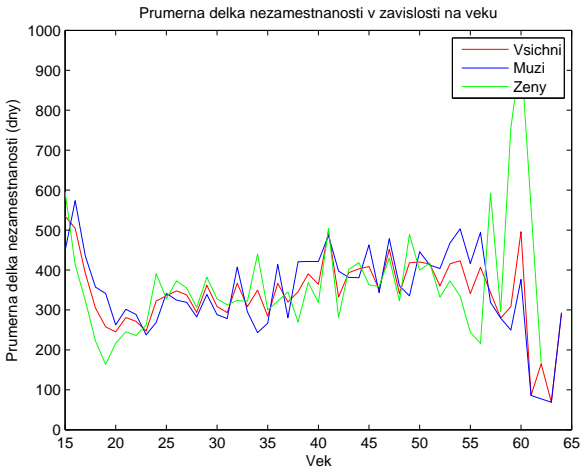
- 1 Odfiltrování neúplných dat (méně než 2% záznamů)
- 2 Vyřazení rodných čísel
- 3 Zahrnuty jsou jen záznamy ze Sokolova, Chodova a Kraslice –
≈ 27 tis. záznamů (obtížné výpočty s větším množstvím dat)
- 4 Výpočet délky nezaměstnanosti ve dnech
- 5 Převedení řetězcových hodnot na číselné
- 6 Min-max transformace na interval $< -1, 1 >$

Rok	Pohl.	Vzdelani	Zdr. st.	Rod. st.	Bydl.	Vek	Nez. dnu
0	0	17	1	1	0	43	163
1	1	2	2	3	2	34	243

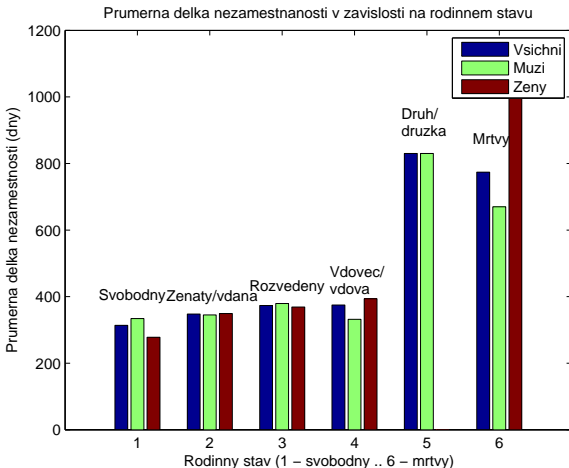
Počáteční analýza dat

- Výpočet průměrné délky nezaměstnanosti v závislosti na
 - věku
 - rodinném stavu
 - vzdělání
 - zdravotním stavu
- Samostatná analýza jednotlivých faktorů v závislosti na pohlaví

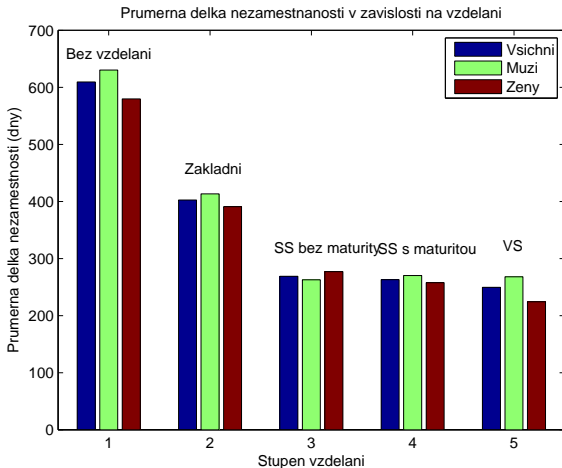
Průměrná délka nezaměstnanosti vs. Věk



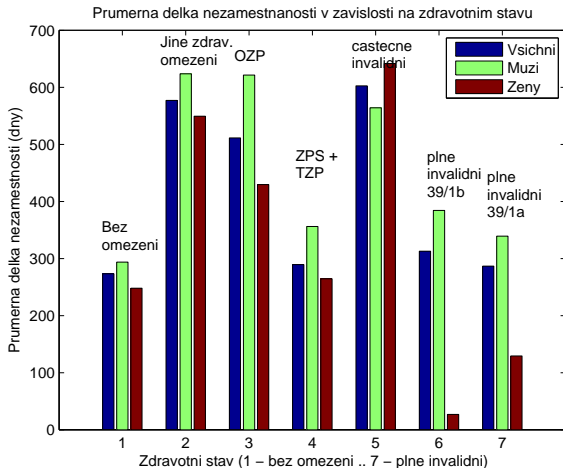
Průměrná délka nezaměstnanosti vs. Rodinný stav



Průměrná délka nezaměstnanosti vs. Vzdělání



Průměrná délka nezaměstnanosti vs. Zdravotní stav



Analýza dat – Shrnutí

- Pozvolný nárůst očekávané délky nezaměstnanosti s rostoucím věkem
- Výrazný vliv vzdělání délku nezaměstnanosti
- Částečně zdravotně omezení lidé jsou více ohroženi dlouhodobou nezaměstnaností; plně omezení se vyskytují ve statistikách zřídka
- Muži jsou v průměru nezaměstnaní déle než ženy
- Průměrná délka nezaměstnanosti je velká (přibližně 1 rok)

Odhad očekávané délky nezaměstnanosti I – BP-síť

- Dvě sítě s 10 resp. 20 skrytými neurony
- Učení algoritmem Levenberg-Marquardt
- Pro učení použito 60% dat jako trénovací, 20% jako validační a 20% jako testovací množina

Skrytá vrstva	Epochy	Přesnost
10 neuronů	82	0.0379
20 neuronů	50	0.0363

Neuronové sítě k vyzkoušení

<http://labts.troja.mff.cuni.cz/~cigl13am/dm/>

Odhad očekávané délky nezaměstnanosti II – RBF síť

- RBF = Radial Basis Function
- Třívrstvá síť
 - Skrytá vrstva se učí samoorganizací, přenosová funkce je Gaussova (= radiální funkce)
 - Výstupní vrstva používá lineární přenosovou funkci
- Implementace pomocí knihovny z Edinburgh University (<http://www.anc.ed.ac.uk/rbf/rbf.html>)
- Pro naučení použito 18% dat, 2% dat použita pro testování

Skrytá vrstva	Velikost okolí	Průměrná chyba ²
200 neuronů	2.0	0.1390
200 neuronů	1.0	0.1561
400 neuronů	2.0	0.1365

²Hodnoty jsou transformovány na interval $\langle -1, 1 \rangle$

Shrnutí

- Vzdělání a zdravotní stav mají významný vliv na očekávanou délku nezaměstnanosti
- Předpovědi neuronových sítí jsou značně nespolehlivé – různé sítě se diametrálně liší ve svých předpovědích

Možná rozšíření

- Chybí informace o nezaměstnanosti v různých odvětvích
- Porovnání dat z různých okresů
- Použití jiné techniky pro naučení očekávané délky nezaměstnanosti (např. rozhodovací stromy)