

# Detekcia poisťovacích podvodov pomocou nástrojov SAS

(Abstrakt)

**ING. ĽUBOSLAV GABAL' (vedúci diplomovej práce: doc.RNDr.JURAJ PANČÍK,PhD. )**

**Školský rok 2006 -2007**

**Katedra kvantitatívnych metód a informatiky, Ekonomická fakulta, Univerzita Mateja Bela v Banskej Bystrici**

## Abstrakt a ciele práce:

Diplomová práca sa zaoberá data miningom a jeho praktickou aplikáciou v oblasti detekcie poisťovacích podvodov. Teoreticky popisuje kroky dolovania dát ako je príprava dát, transformácia dát, metódy dolovania dát a vyhodnocovanie úspešnosti jednotlivých metód. Analyzuje výberový súbor dát teoreticky popísané metódy sú aplikované na tento súbor. Prezентuje výsledky tejto analýzy a vzájomne ich porovnáva. Výsledkom skúmania je zostavenie logicky štruktúrovanej práce predstavením spomínaných poznatkov a ich aplikácie na reálne dáta. Zistili sme vhodnosť použitia data mining-ových metód v tejto oblasti s vysokou vypovedacou schopnosťou a vysokou úsporou nákladov.Úplný text diplomovej práce je dostupný na URL adrese :<http://diplomovka.sme.sk/praca/3166/detekcia-poisťovacich-podvodov-pomocou-nastrojov-sas.php>

**Táto diplomová práca bola ocenená 1. miestom v rámci národnej súťaže SIBAF® award 2007 ako „Najlepšia diplomová práca v oblasti poisťovníctva“**

## Podrobnejší popis:

Hlavným cieľom predkladanej diplomovej práce bolo zosumarizovať postupy a proces data mining-u ako takého a priblížiť konkrétne metódy používané na detekciu poisťovacích podvodov. Tento cieľ sa nám podarilo naplniť popisom tohto nástroja a výberom poznatkov, obsiahnutých v dostupnej literatúre, časopisoch a iných materiáloch týkajúcich sa danej problematiky. Cieľom druhej a tretej kapitoly tejto práce bola aplikácia postupov a metód na reálne poisťovacie dáta a skúmanie, či je tieto metódy možné úspešne využiť a napomôcť tak finančnému hospodáreniu poisťovacej spoločnosti.

V súlade so zadaním diplomovej práce bola detekcia podvodov postavená na modeloch rozhodovacích stromov, neurónových sietí a logistickej regresie. Jej informačným a dátovým základom boli informácie týkajúce sa poisťných udalostí havarijného a povinného zmluvného poistenia austrálskej poisťovne. K dispozícii sme mali dáta o rozsahu 15 500 záznamov, ktoré mali 33 atribútov – charakteristík poisťných udalostí, ktoré plne postačovali na úspešnú detekciu. Hlbšia a podrobnejšia analýza by si vyžadovala rozšírený atribútový súbor, avšak zámerom diplomovej práce bolo dokázať využiteľnosť týchto metód, čo sa podarilo aj pri nižšom počte atribútov.

Ako parciálny cieľ, ktorý mal napomôcť dosiahnutiu hlavného cieľa, bol popis procesu samotného data mining-u. V tejto časti sme sa venovali príprave dát, dátovým skladom ako esenciálnemu zdroju dát pre data mining. Väčší priestor bol venovaný samotným metódam data miningu a ich aplikácii. Nakoniec sme sa venovali samotnému hodnoteniu výkonnosti jednotlivých metód pri ich aplikácii. Tieto metódy a postupy boli v plnej miere uplatnené aj v druhej a tretej kapitole tejto práce. V rámci nej sa nám podarilo dokázať, že využitie data mining-u v tejto oblasti je viac než vhodné, pričom bola demonštrovaná relatívne významná úspora pri výplate náhrad poisťných udalostí.

### **Záver:**

Z prezentovaných výsledkov skúmania a analýzy vyplynula vhodnosť aplikácie všetkých použitých metód – rozhodovacích stromov, neurónových sietí ako aj regresie. Všetky spomínané metódy vykazovali relatívne podobné a stabilné výsledky ako na tréningových, tak aj testovacích dátach. Následne boli natrénované modely použité na skórovanie – predikciu podvodu na základnom výberovom súbore, na základe čoho bolo možné odhadnúť spomínanú nákladovú úsporu.

Pre spoľahlivejšiu detekciu podvodov odporúčame využiť na skórovanie jednotlivých záznamov všetky, prípadne najvýkonnejšie metódy a zvýšiť tak šance na úspešnú detekciu. Na samotné tréningovanie by bol vhodnejší širší dátový súbor, ktorý však nebol k dispozícii. V praxi by však prístup k týmto dátam nemal predstavovať problém. V opačnom prípade je vhodné vytvoriť testovací súbor dát, ktoré vôbec nebudú využívané pri tréningu metód a tento použiť na skórovanie a následný odhad skutočných výsledkov.

Napriek tomu, že analýzu, ktorá bola uskutočnená na použitých dátach nemôžeme považovať za vyčerpávajúcu, rovnako ako nemôžeme považovať za vyčerpávajúci počet použitých metód, myslíme si, že hlavný cieľ diplomovej práce bol v rozhodujúcej miere dosiahnutý a práca tak splnila svoj zámer

### **Použité SAS moduly a nástroje:**

Pri písaní diplomovej práce boli použité produkty SAS Enterprise Guide 4 a SAS Enterprise Miner 4.3.

### **Zdroj údajov:**

Údaje boli získané ako voľne prístupný textový súbor, ktorý bol súčasťou výskumu spracovania zošikmených dát pomocou data miningu vid. aj :

- Phua, C. – Alahakoon, D. – Lee, V.: Minority Report in Fraud Detection: Classification of Skewed Data, School of Business Systems, Faculty of Information Technology, Monash University, Clayton, Australia

### **Literatúra a referencie:**

Úplný text diplomovej práce je dostupný na URL adrese :<http://diplomovka.sme.sk/praca/3166/detekcia-poitovacich-podvodov-pomocou-nastrojov-sas.php>

- [1] Witten Ian H. - Frank E.: Data Mining – Practical Machine Learning Tools and Techniques, Second Edition, Morgan Kaufmann 2005
- [2] Han J. – Kamber M.: Data Mining – Concepts and Techniques, Second Edition, Morgan Kaufmann 2006
- [3] Berry M.J.A. – Linoff G. S.: Data Mining Techniques for Marketing, Sales, and Customer Relationship Management, Second Edition, Wiley Publishing, Inc. 2004
- [4] Wu T. – Li X.: The Handbook of Data Mining, Lawrence Erlbaum Associates, Inc. 2003
- [5] Inmon W.H.: Building the Data Warehouse, John Wiley & Sons, 1996
- [6] Tang Z. – MacLennan J. – Data Mining with SQL Server 2005, Wiley Publishing, Inc. 2005

### **Kontakt autori:**

Ing. Luboslav Gabal, data mining & DWH consultant, ADASTRA, s.r.o., Francisciho 4, 811 08 Bratislava, [Luboslav.Gabal@adastracorp.com](mailto:Luboslav.Gabal@adastracorp.com)

doc. RNDr. JURAJ PANČÍK, PhD., Katedra kvantitatívnych metód a informatiky, Ekonomická fakulta, Univerzita Mateja Bela v Banskej Bystrici, Tajovského 10, 976 50 Banská Bystrica, [juraj.pancik@umb.sk](mailto:juraj.pancik@umb.sk)