

Predikcia prosperity poľnohospodárskych fariem aplikáciou metód data miningu

(abstrakt)

Mgr. Eva Uhrinčat'ová

*Výskumný ústav ekonomiky poľnohospodárstva a potravinárstva v Bratislave
Katedra štatistiky, Fakulta hospodárskej informatiky, Ekonomická univerzita v Bratislave
– externý doktorand*

Abstrakt a ciele práce:

Predikcie finančného vývoja podnikov sú bežné v trhovom rozvinutom prostredí. Ich mechanická aplikácia na slovenské podmienky môže viesť k nesprávnym záverom z dôvodov existencie javov, ktoré sú špecifické pre transformačný proces (v slovenských podmienkach iste ešte existujúci). Vývoj nových metódik na Slovensku v tomto smere pozorujeme u niektorých autorov (metódy sú prevažne založené na vývoji jednotného integrovaného ukazovateľa na základe konštrukcie rôznych finančných ukazovateľov, zväčša metódou viacnásobnej diskriminačnej analýzy). Cieľom prezentujúcej štúdie bol pokus o riešenie uvedenej problematiky modelovými nástrojmi data miningu. Pomocou štatistických postupov a klasifikačných metód (rozhodovacích stromov, logistickej regresie, neurónových sietí) sme vytvorili finálny model s atribútmi, ktoré sú aj podľa expertného názoru adekvátne. Parametre finálneho modelu sme optimalizovali.

Cieľ uvedenej štúdie je aplikačno – metodický

- Vyvinúť model, na základe ktorého možno predikovať prosperitu poľnohospodárskych fariem
- Samozrejme vystupuje tu do popredia otázka validnosti predikcií pri nestacionárnosti a nehomogénosti ekonomických javov, znásobenými aj inými špecifikami v prípade poľnohospodárskych fariem – metodický aspekt štúdie môže byť ale potvrdený bez ohľadu na aktuálnosť dát

Podrobnejší popis:

Výber znakov pre predikčnú analýzu economickej prosperity fariem

- Model 1 - závisle premenná alternatívneho typu je znak kód hospodárskeho výsledku fariem po zdanení (farmy s kladným hospodárskym výsledkom a farmy so záporným hospodárskym výsledkom)
- Model 2 - závisle premenná trichotomického typu je znak skupina prosperity fariem; pri tvorbe hodnôt znaku skupina prosperity sú farmy podľa kombinácie dvoch znakov finančnej situácie – celkovej likvidity a rentability tržieb – rozdelené na skupiny prosperujúcich, priemerných a neprosperujúcich podnikov. Zohľadňujeme úroveň ukazovateľov v poľnohospodárskych podnikoch na Slovensku a upravujeme podľa nej všeobecne odporúčaný štandard

Nezávislé premenné

- Relatívne ukazovatele
- Významnejšie absolútne ukazovatele prepočítané na 1 ha
- Premenné, ktoré sme získali inšpiráciou z parciálnych finančno – ekonomických indikátorov z indexov (vykazovali významné rozdiely v súbore prosperujúcich podnikov a v súbore neprosperujúcich podnikov), ktoré sa používajú na analýzu „ex ante“ finančnej situácie podnikov (napr. Altmanov index, index bonity, Ch-index, G-index atď.)
- Spolu sme prvotne selektovali 101 znakov
- Formálna adekvátnosť ukazovateľov bola prevedená štandardizáciou dát. Vzhľadom k tomu, že nemáme skúsenosti s faktom, či existujú rozličné dôležitosti znakov pre účely tejto analýzy,

nezaraďujeme použitie štatistických váh. Pri predikcii uvažujeme o celkovej dôležitosti celku znakov, takže je obtiažne určiť dôležitosť jedného znaku

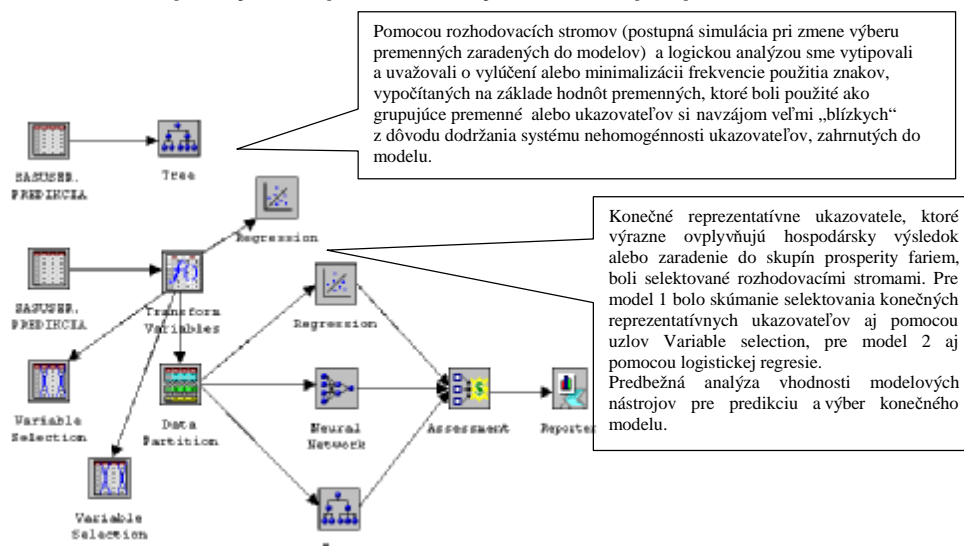
Posúdenie významnosti jednotlivých ukazovateľov pre predikciu prosperity fariem v modeli 1 na základe testu zhody dvoch stredných hodnôt

- Na základe výsledku F testu na overenie predpokladu o zhodnosti rozptylov sme použili na overenie predpokladu o zhodnosti stredných hodnôt t-test, previedli sme aj overenie predpokladu normality pre znaky, ktoré sú potenciálnymi kandidátmi pre predikciu prosperity fariem za podmienok oboch hodnôt znaku kód hospodárskeho výsledku fariem po zdanení

Posúdenie významnosti jednotlivých ukazovateľov pre predikciu prosperity fariem v modeli 2 analýzou rozptylu

- Na základe výsledku testu na identifikáciu variančnej homogenity (v tomto prípade sme použili Levene test) sme v prípade homogenity rozptylov použili na testovanie homogenity strednej hodnoty pre jednotlivé triedy grupujúcej premennej ANOVA F test, ak podmienka homogenity rozptylov nebola splnená, použili sme Welchov test

Procesný tok modelovania pre výber reprezentatívnych znakov pre predikciu



Vzhľadom k tomu, že výsledky simulácií v modeli 2 neboli uplatnené vo finálnom modeli na predikciu prosperity fariem z dôvodu horších štatistík kvality jednotlivých simulácií v porovnaní so štatistikami simulácií v modeli 1, v ďalšom uvádzame postupy a výsledky pre model 1. Činitele, ktoré výrazne ovplyvňujú hospodársky výsledok a spĺňajú predpoklady, uvedené v diagrame, sú pomer cash flow k celkovým pasívam, pomer tržieb ku kapitálu, hospodárnosť nákladov, konkrétne nákladovosť variabilných nákladov vzhľadom k výnosom, investície na rozvoj na hektár a priame platby na hektár.

Analýza vhodnosti modelových nástrojov pre predikciu prosperity fariem

- Model logistickej regresie (zaujíma nás iba účinnosť modelu ako prediktívneho nástroja, nie ekonomická interpretovateľnosť hodnôt estimátorov)
- Model rozhodovacieho stromu (klasifikačný strom CHAID)
- Model neurónovej siete (architektúra viacvrstvového perceptrónu: jedna skrytá vrstva, s jedným druhom vstupných premenných a binárnym výstupom; počet neurónov v skrytej vrstve sme určili pomocou používanej heuristiky, ktorá odporúča, aby počet neurónov v skrytej vrstve bol dvojnásobný

ako počet neurónov vo vstupnej vrstve; spomedzi variantov učenia siete sme vyskúšali štandardnú verziu algoritmu spätného šírenia chyby, v tomto prípade bez použitia momentu (konštanta udávajúca veľkosť zotrvačnosti zmeny), learning rate (konštanta udávajúca rýchlosť zmeny váh) mal hodnotu 0,1)

Resumé

- Najkvalitnejšia je simulácia pomocou neurónovej siete v skupine modelov 1 (presnosť predpovede obidvoch hodnôt modelovanej premennej bola na validačnej množine dát 95 %, presnosť predpovede kladného hospodárskeho výsledku farmy bola 98 %, presnosť predpovede záporného hospodárskeho výsledku farmy bola 73 %; takže ukazovateľ podielu chybných predpovedí na validačnej množine dát bol 5 %, na tréningovej množine dát bol 7 %)
- Neúspešnosť simulácii v skupine modelov 2 pravdepodobne súvisí s prvotným začlenením podnikov do skupín prosperity fariem, v ktorej bolo najviac priemerných podnikov, pričom do tejto skupiny priemerných podnikov bolo začlenených aj 85 % podnikov, ktoré mali kladný hospodársky výsledok, z fariem so záporným hospodárskym výsledkom bolo začlenených do skupiny priemerných fariem 53 %. Rozdelenie podnikov do skupín prosperity fariem prebehlo správne podľa zvolených atribútov, ktoré boli konzultované aj s expertom. Evidentne ale vhodnejšie delenie (viac jednoznačnejšie) je v prípade delenia podľa dosiahnutého hospodárskeho výsledku. Rok 2004 bol pre poľnohospodárov úspešný – neboli nepriaznivé klimatické výkyvy, bol to rok vstupu do Európskej únie s úpravou aj dotačnej politiky. Každopádne pre toto naše statické skúmanie môžeme prehlásiť, že prejavom ekonomickej diferenciácie fariem bolo kritérium zisku

Pri **optimalizácii finálneho modelu** sme sa snažili nájsť optimálny model viacvrstvého perceptrónu

- Optimálnu architektúru siete (počet skrytých neurónov, prípadne pridanie druhej skrytej vrstvy neurónov, vhodné aktivačné funkcie neurónov)
- Optimálny variant učenia (aj optimálnu zmenu parametrov variantu učenia pomocou spätného šírenia chyby (štandardná verzia))

Ako optimálny model vyšiel perceptrón s jednou skrytou vrstvou s počtom neurónov 9, s algoritmom spätného šírenia chyby s parametrami learning rate 0,1 a momentom 0, s aktivačnou funkciou hyperbolický tangens (charakteristiky sa zlepšili oproti pôvodnému modelu približne o 1 %, okrem presnosti predpovede záporného hospodárskeho výsledku – pri tejto predpovedi nastalo zlepšenie o 9 %).

Záver:

Príklady ďalších možností uplatnenia nástrojov data miningu v poľnohospodárskom sektore – vyhodnotenie stavov a trendov ekonomických a produkčných indikátorov v oblastiach

- Analýza nákladovosti poľnohospodárskej výroby (napr. intenzita výroby – komoditné zameranie alebo nákladová intenzita výroby komplexne)
- Analýza tvorby pridanej hodnoty v poľnohospodárstve
- Analýza tvorby výnosov v poľnohospodárstve (napr. vplyv faktorov diferenciácie na dosahované hospodárske výsledky – výnosy, výrobu, tržby za predaj vlastných výrobkov a služieb, podnikový výkon a pod.)
- Syntetické zhodnotenie vybraných ekonomických a produkčných ukazovateľov
- Analýza výsledkov hospodárenia poľnohospodárskych subjektov (napr. kvantifikácia vplyvu zmien analytických ukazovateľov na zmenu hospodárskeho výsledku)
- Analýza základných výrobných prostriedkov – pôdy, kapitálu a práce (napr. modelovanie produkčných funkcií a modelovanie produkčnej a nájomnej renty pôd)
- Analýza dotácií (napr. generovanie efektu úrovne podpôr)

- Analýza nekomoditných výstupov (z hľadiska ekonomických finančných tokov podmienené existenciou satelitných informačných blokov (ktoré napr. obsahujú informácie o environmentálnych dopadoch ekonomických aktivít) ku klasickému podnikovému účtovníctvu na úrovni mikroekonomie alebo k národnému účtovníctvu na úrovni makroekonomie)
- Atd...

Použité SAS moduly a nástroje:

SAS Enterprise Guide

- Data – Filter and Query
- Analyze – ANOVA

SAS Enterprise Miner

- Sample – Input Data Source
- Sample – Data Partition
- Explore – Multiplot
- Explore – Variable Selection
- Modify – Transform Variables
- Model – Regression
- Model – Tree
- Model – Neural Network
- Assess – Assessment
- Assess – Reporter
- Scoring – Score

Zdroj údajov:

Vstupné informácie boli čerpané najmä z časti výkazu ziskov a strát a súvah, doplnené o informácie o dotáciách a niektorých vstupoch a výstupoch v selekcii na rastlinnú a živočíšnu výrobu, všetko zo šetrenia Informačnej siete poľnohospodárskeho účtovníctva Slovenskej republiky pre rok 2004 – s použitím výberového súboru právnických osôb. V čase spracovania štúdie boli prístupné dáta pre rok 2004, tohto času sú prístupné dáta pre rok 2005, ktoré sme použili pre skórovanie modelu.

Literatúra a referencie:

Berka, P.: Dobývaní znalostí z databází. Academia Praha, 2003

Gurčík, Ľ.: G-index – metóda predikcie finančného stavu poľnohospodárskych podnikov. In Agricultural Economics 2002, volume 48, 2002, No. 8, p. 373 - 378

Chajdiak, J.: Ekonomická analýza stavu a vývoja firmy. Statis. Bratislava, 2004. ISBN 80–85659–32–8, 349 s.

Chrastinová, Z.: Metodika hodnotenia bonity a návrh "systému včasného varovania", t.j. predikovania vývoja finančnej situácie poľnohospodárskych podnikov. Štúdia VÚEPP. Bratislava, 1997

Klímeš, J. a kol.: Výberové zisťovanie výsledkov hospodárenia v Informačnej sieti poľnohospodárskeho účtovníctva SR - FADN SR za rok 2004. Štúdia VÚEPP. Bratislava, 2005. ISBN 80-8058-395-1, 250 s.

Zalai, K. a kol.: Finančno – ekonomická analýza podniku. Sprint vfra. Bratislava, 2004. ISBN 80–88848–89–1, 305 s.

Kontakt autori:

eva.uhrincatova@vuepp.sk