

OPONENTSKÝ POSUDEK

Název diplomové práce: Modelování pravděpodobnosti úvěrového selhání v rámci zobecněných lineárních modelů
Jméno diplomanta: Tomáš Magyar

Shrnutí:

Předložená diplomová práce se zabývá modelováním pravděpodobnosti úvěrového selhání dlužníka v závislosti na množství finančních charakteristik. Zkoumaný problém má velký význam pro odhadování kreditního rizika bank a finančních institucí. Diplomová práce volí přístup založený na logistické regresi s propracovanou strategií volby prediktorů a následné validaci navržených modelů na kontrolní množině dat.

Diplomant navrhuje posoudit nejprve jednotlivé prediktory samostatně s ohledem na jejich vzájemné závislosti, možné transformace a schopnosti predikce kreditního rizika. V dalším kroku přistupuje k budování vícefaktorových logistických modelů s použitím ztransformovaných prediktorů vybraných v předchozí fázi. Závěrečná část popisuje několik způsobů porovnání prediktivní síly vybraných modelů (ROC křivky, CAP křivky, Kolmogorovova-Smirnovova statistika).

Diplomová práce je napsána v angličtině, má promyšlenou logickou strukturu a je velmi pěkně graficky upravená. Obsahuje jen minimum technických chyb a překlepů. Občas obsahuje poněkud krkolomně zvolená slova a slovosled neodpovídající angličtině.

Silnou stránkou diplomové práce je, že řeší důležitý praktický problém s použitím v zásadě vhodně zvolených a rozumně uplatněných metod. Hlavní slabinu vidím v tom, že v práci jsou prezentovány jen teoretické základy jednotlivých komponent postupu bez detailů jejich konkrétního použití. I po prostudování celé práce není jasné, jak byly tyto postupy aplikovány a s jakým úspěchem. Částečně to může být způsobeno tím, že skutečná data a výsledky jsou předmětem obchodního tajemství, ale i tak by bylo možné prezentovat ilustrativní výsledky spočítané na pozměněných či neúplných datech a zařadit daleko víc detailů o klíčových bodech celého postupu. Teoretické výsledky zařazené v práci jsou do značné míry převzaty ze standardních přednášek magisterského studia (χ^2 -statistika, Spearmanův korelační koeficient, logistická regrese, Kolmogorovova-Smirnovova statistika). Ty není třeba tak podrobně popisovat. Snad jen pasáž o ROC křivkách a AUC jde nad rámec metod přednášených během studia.

Práce Tomáše Magyara odpovídá nárokům kladeným na diplomovou práci a proto ji doporučuji přijmout k obhajobě.

Konkrétní připomínky:

- Str. 8, 1.1.2 Někdy je rozumné rozlišovat chybějící hodnoty, které měly být měřeny,

ale z nějakého důvodu nebyly („not available“) od těch, které nebyly vyplněny, protože v dané situaci nemají smysl („non-applicable“).

- **Str. 8–9, 1.1.3** Podle jakých konkrétních kritérií byly faktory zařazovány do analýz na základě své diskriminační schopnosti a závislostí? Bylo by dobré sdělit, jak přesně byly k těmto účelům použity popisované charakteristiky (AUC, míry asociace).
- **Str. 11–14** K důkazu nerovnosti $0 \leq \chi_T^2 \leq 1$ nestačí ukázat, že $\chi_T^2 = 1$ v nějaké situaci (i když se nám jeví jako situace extrémní), ale je třeba doložit, že $\chi_T^2 > 1$ nemůže v žádné situaci nastat.
- **Str. 18, 1.2.2** Jak bylo nakládáno se „suspicious obligors“? Muselo jich být hodně, když mezi nimi bylo 10% všech pozorování pro každý spojitý prediktor.
- **Str. 19, 1.2.3** (i) Chybějící hodnoty, jež jsou „non-applicable“, není třeba nahrazovat; tato pozorování tvoří svébytnou skupinu a je lépe nechat model sám, aby odhadl její parametr. (ii) Nenašel jsem žádnou následnou zmínku o „*cutoff points described below*“.
- **Str. 20–21** Proč k určení transformací nepoužít GAM (zobecněný aditivní model) zmiňovaný na str. 6? Ještě k transformacím: ve vícefaktorovém modelu mohou mít regresní závislosti úplně jiný tvar než v jednofaktorovém. Například původně nemonotónní vztah se může změnit v monotónní a naopak. Je toto bráno v úvahu?
- **Str. 23–30** Byly uvažovány interakce?
- **Str. 31** Metody popsané v Kap. 3 byly aplikovány na validační data nebo na data použitá k vývoji modelů? Pokud na validační data, jaké manipulace byly s validačními daty provedeny (nahrazování chybějících hodnot, „*cutoff points*“, „*suspicious obligors*“, atd.)?
- **Str. 31–52** Která z popisovaných metod byla skutečně použita pro výběr konečného modelu a jak přesně? Z kolika modelů se vybíralo?

Mgr. Michal Kulich, PhD.
KPMS MFF UK
3. února 2006

