

Posudek oponenta na diplomovou práci

Bc. Adéla Mrázková

Comparison of statistical methods for the scoring models development

Předložená práce se zabývá srovnáním tří metod pro vytvoření skóringového modelu pro modelování kreditního rizika úvěrů. Postupně je představen model logistické regrese, neuronových sítí a náhodných lesů (rozhodovací stromy).

Vlastním příspěvkem je třetí kapitola práce, ve které autorka aplikuje zmíněné metody na reálná data české společnosti poskytující úvěry.

Práce je psaná v anglickém jazyce na velmi dobré jazykové úrovni. Práci ale chybí větší matematická preciznost, zejména při představování jednotlivých metod, viz následující dotazy a připomínky:

- 1, Kozmík (2006) je nesprávná citace.
- 2, V Definicí 1 jsou uvažovány nějaké kategorie a proměnné, ale není jasné, co se tím míní, ani proměnné ani kategorie nejsou předtím definovány.
- 3, Str. 12: Nad definicí 5 je vektor β uvažován jako $n+1$ rozměrný, ale pak v definici 5 už jenom jako n rozměrný. Je tam schován nějaký další předpoklad, např. $\beta_0 = 0$?
- 4, na str. 12 je hezká interpretace parametru β_1 . Jaká je interpretace parametru β_0 ?
- 5, Str. 13: jak se volí parametre p_E a p_R v stepwise algoritmu?
- 6, Kapitola 1.2.3 není dobře napsaná. Obsahuje mnoho pojmů, které nejsou vysvětleny. Např. Def.12 je nejasná. Není vysvětleno co to je charakteristika, co to znamená, že má úroveň i a jak to souvisí s tím, že klient má vlastnost i . Podobně Def. 15 není srozumitelná.
- 7, Vysvětlení specifity a sensitivity pod Def. 21 nekoresponduje s Def. 21.
- 8, Druhý odstavec kapitoly 2.1. není jasný. Jak se tedy uvažují klienti, kteří přestanou splácet již v prvním roce?
- 9, Nad obrázkem 3.1. se píše, že klienti, kteří nežijí ve sdílené domácnosti budou spíše špatní, ale z obrázku 3.1. a 3.2 plyne opak. Jak je to tedy správně?
- 10, Nejsou ty data trochu divné ve smyslu, že podíl špatných klientů je skoro tak velký jako podíl dobrých klientů (viz. Tabulka 3.1)?
- 11, Jak vypadá finální model logistické regrese? Které všechny proměnné jsou signifikantní? Jak vychází Hosmer-Lemeshow test a McFadden R^2 pro tento model?
- 12, Jak si vysvětlujete, že model logistické regrese dal tak špatné výsledky, když je to nástroj který se používá nejčastěji?

Celkově považuji předloženou práci za poměrně kvalitní a doporučuji ji uznat jako diplomovou.

V Praze 3.6.2018

Doc. RNDr. Ing. Miloš Kopa, Ph.D.