

POSUDEK OPONENTA BAKALÁŘSKÉ PRÁCE

Název: Míry diskriminace v kreditním riziku

Autor: Michal Polak

SHRNUTÍ OBSAHU PRÁCE

Práce pojednává o mírách diskriminace v kreditním riziku, tj. mírách snažících se ohodnotit, jak se modelům kreditního rizika daří oddělit tzv. dobré zákazníky od tzv. špatných. Kvalitní modely kreditního rizika jsou jednou ze základních a zásadních podmínek, aby banky mohly dlouhodobě dosahovat zisku. Práce nejprve shrnuje základní poznatky o modelování kreditního rizika, dále se věnuje definici několika diskriminačních měr a jejich vlastnostem, aby v poslední kapitole soustředěně dovednosti užila na datech a pojednala o jejich užitečnosti a úspěšnosti.

CELKOVÉ HODNOCENÍ PRÁCE

Téma práce. Téma práce je jedním ze standardních úvodních kroků k tématu kreditního rizika. Je v oblasti kreditního rizika dobře ukotvené, neboť řešitele nutí porozumět základním kamenům problematiky a zároveň poskytuje možnost ihned aplikovat teoretické poznatky na data. Jedná se o téma praxí motivované, ale z hlediska matematické teorie spíše jednodušší.

Práce zevrubně plní zadání - obecný úvod, míry diskriminace, i aplikace na data se v práci objevují.

Vlastní příspěvek. Vlastní příspěvek autora je potřeba vyjasnit. Vzhledem k velmi nekvalitní práci se zdroji není jasné, které pasáže jsou převzaté, ale zdá se, že vlastním příspěvkem autora je hlavně práce s daty. Teoretické poznatky se zdají být převzaté z relevantních pasáží textů, které jsou sice uvedeny v seznamu literatury, ale již nejsou citovány v kritických místech. Domněnce o převzatých pasážích nahrává skutečnost, že ačkoliv místy je text poměrně srozumitelný, jednotlivé srozumitelné pasáže působí velmi odděleně. Odstavce v podstatě jen slohové, které by je měly přirozeně spojovat, již často vyznívají značně rozpačitě a pro neznalého a nedomýšlivého čtenáře někdy až nepochopitelně.

Matematická úroveň. Matematická úroveň práce je průměrná. Korektně zformulovaného matematického textu je relativně málo, což je ale samozřejmě částečně dáno volbou tématu a vedoucím práce patrně očekávaným přínosem v podobě umné aplikace na data. Bohužel i v relativně krátkých pasážích matematického textu lze najít problematické pasáže, ve kterých si člověk musí domýšlet a opravovat. Označení odhadu se bere jako samozřejmé, jako samozřejmé se bere asi i to, že Y_i je všude v práci binární a označuje konkrétní náhodnou veličinu? Připouštím, že to může být moje nechápavost, ale nepřijde mi zřejmé ani např. co konkrétně, či zda je to obecný pojem, se myslí pod "skóre klienta" na straně 12. Značení indexů je také na důvtipu čtenáře, vizme např. str. 12 dole či 13 nahoře, kde se sice v textu říká, že indexujeme počínaje indexem i , ale matematický text to reflektuje pouze na náhodných místech.

Práce se zdroji. Práce se zdroji je dle mého názoru katastrofální. Není respektován jednotný styl, autoři jsou citováni jednou příjmením, jednou křestním jménem. Když je autorů více, vzniká změt jmen či slov nejasného původu. Příklady můžeme najít např. na str. 3 (Teplý Petr a Magda) nebo straně 17 (Jr. and Stanley). Je zde zřejmé naprosté nezvládnutí funkcionality, kterou LaTeX poskytuje prakticky zadarmo automaticky. Taktéž je dle mého citováno nedostatečně, neboť mám dle náhledů do nabízené literatury pocit, že kapitoly 3 a 4 jsou převzaté (z prací

“Applied logistic regression” a “Postupná výstavba modelů kreditního rizika”, které jsou uvedené v seznamu literatury na str. 24 práce.) Pakliže je tomu tak, očekával bych jasné vypíchnutí tohoto faktu společně s pečlivou citací dle standardů, které se obvykle vyžadují (tj. u knih i uvedení konkrétních stran, ze kterých se čerpá).

Formální úprava. Formální úroveň práce je velmi špatná. Typografických problémů, překlepů, drobných matematických chyb i určitých formulačních nesrovnalostí se v práci najde mnoho desítek. V podstatě na první pohled je vidět, že autor před tiskem práce nevěnoval žádné úsilí finální korektuře a odhalení alespoň do očí bijících problémů. Uveďme pár příkladů:

1. Překlepy: z dlouhé řady např. tikvidní, skoringové vs. skóringové, LC vs. Lc, chybný zápis SSE na str. 9, atp.
2. Není konzistence v počest'ování vícejmenných názvů: najdeme jak Kolmogorov–Smirnovova, tak Kolmogorovova–Smirnovova.
3. Čárky a tečky mezi vzorci chybí v podstatě úplně, v libovolném jazyce (sám CSLaTeX nepoužívám, proto to nechci vyžadovat, i když v bakalářské práci by to asi žádoucí bylo..) špatné uvozovky (“takto používat uvozovky v LaTeXu nelze”).
4. O více než 3 znaky přetečené řádky se vyskytují na stranách 2,6,7,18. Menší přetečení se objevuje v podstatě na každé druhé straně.

PŘIPOMÍNKY A OTÁZKY

1. Na str. 6 se ve třetím odstavci hovoří o lineární regresi. Proč? Co má obecně $E(Y|x)$ společného s linearitou? Stojí za to druhý i třetí odstavec v sekci 3.1 reformulovat tak, aby byly určité správně a bylo i jasné, co např. znamená výrok obsahující formulaci “naš model má rozdělení”.
2. Prvnímu odstavci v 3.4 také úplně nerozumím a rád bych, aby autor vysvětlil, co má na mysli. Vypočtené zde znamená odhadnuté?
3. Na str. 9 dále najdeme formulaci “získali výraz, pro nějž známe distribuční funkci...” Za jakých předpokladů to platí?
4. Kdo a jak vybral míry diskriminace uvedené v práci ve čtvrté kapitole?
5. V aplikaci na data v kapitole 5 je několik nejasností - ve zdrojovém kódu chybí vybírání regresorů. Užívá se zde jádrový odhad. Víme, co to přesně je? Jak se závěry zde liší od necitovené práce “Postupná výstavba modelů kreditního rizika”? V čem ji případně doplňují či na ni navazují? Kromě využití jiného softwaru samozřejmě.

ZÁVĚR

Práci považuji za podprůměrnou, jak moc podprůměrnou závisí na vyjasnění, nakolik jsou podstatné pasáže skutečně studentovou samostatnou prací s porozuměním. Vzhledem k tomu, že práce z vědeckého hlediska nepřináší nic nového, je viditelnější nezvládnutý cíl alespoň napsat rozumně čtivý text, který není přeplněn chybami a dodržuje jakási základní formální pravidla. Přestože však k práci mám mnoho výhrad, lze ji snad ještě doporučit a uznat jako bakalářskou.

Návrh klasifikace oponent sdělí předsedovi zkušební (sub)komise.

Petr Zahradník
KPMS, MFF UK a oddělení ekonometrie, ÚTIA AVČR
20. srpna 2013