# Posudek oponenta závěrečné práce

**České vysoké učení technické v Praze** **Fakulta informačních technologií**

**Student:** Bc. Matúš Tóth
**Oponent práce:** Ing. Martin Kopp
**Název práce:** Analýza leteckých dat a hledání anomálních pasažérů
**Obor:** Znalostní inženýrství

**Datum vytvoření:** 6. 6. 2017

| *Hodnotící kritérium:* | *Způsob hodnocení - následující škálou 1 až 5:* |
|---|---|
| **1. Náročnost a další komentář k zadání** | *1=mimořádně náročné zadání,*<br>*2=náročnější zadání,*<br>***3=průměrně náročné zadání,***<br>*4=lehčí, ale ještě dostatečně náročné zadání,*<br>*5=nedostatečně náročné zadání* |

*Popis kritéria:*
Podrobněji charakterizujte diplomovou (bakalářskou) práci a její případné návaznosti na předchozí nebo běžící projekty. Dále posuďte, čím je zadání této ZP náročné. (U obtížnější ZP lze dále tolerovat některé nedostatky, které by u ZP standardní obtížnosti tolerovány nebyly; a naopak u jednoduché ZP mohou být zjištěné nedostatky hodnoceny přísněji.)

*Komentář:*
Zadani je velice zajimave a vzhledem k tematu i velmi dulezite. Rozsahem odpovida prumerne diplomove praci.

| *Hodnotící kritérium:* | *Způsob hodnocení - následující škálou 1 až 4:* |
|---|---|
| **2. Splnění zadání** | *1=zadání splněno,*<br>*2=zadání splněno s menšími výhradami,*<br>***3=zadání splněno s většími výhradami,***<br>*4=zadání nesplněno* |

*Popis kritéria:*
Posuďte, zda předložená ZP splňuje zadání. V komentáři uveďte body zadání, které nebyly zcela splněny, případně rozšíření ZP oproti původnímu zadání. Nebylo-li zadání zcela splněno, pokuste se posoudit závažnost, dopady a případně i příčiny jednotlivých nedostatků.

*Komentář:*
Zadani bylo splneno jen castecne. Prvni bod zadani byl nastudovat metody detekce anomalit z grafovych dat. Jedina zminka o grafech je v sekci 1.16.2, kde student pise ze tato oblast je neprozkoumana. Toto tvrzeni rozporuje napriklad: "Akoglu (2015). Graph based anomaly detection and description: a survey", coz je aktualni prehled metod detekce anomalii z grafovych dat, ktery ma 68 stran a obsahuje citace na desitky relevantnich clanku.

Hlavnim vystupem prace mela byt odpoved na analyticke otazky zformulovane spolecne s policii CR. V zaveru misto jasnych odpovedi student popisuje pouze kroky potrebne k jejich zodpovezeni. Nektere otazky jsou sice zodpovezeny v experimentalni casti, ale bohuzel ne vsechny a analyza nutna pro jejich zodpovezeni je prinejlepsim nedostatecna a neprukazna, zejmena vezmeme-li v potaz zavaznost tematu.

| *Hodnotící kritérium:* | *Způsob hodnocení - následující škálou 1 až 4:* |
|---|---|
| **3. Rozsah písemné zprávy** | *1=splňuje požadavky,*<br>*2=splňuje požadavky s menšími výhradami,*<br>***3=splňuje požadavky s většími výhradami,***<br>*4=nesplňuje požadavky* |

*Popis kritéria:*
Zhodnoťte přiměřenost rozsahu předložené ZP vzhledem k obsahu, tj. zda všechny části ZP jsou informačně bohaté a ZP neobsahuje zbytečné části.

*Komentář:*
Prace ma sice 75 cislovanych stran, bohuzel relevantniho obsahu je zde minimum. Kapitoly 2,3,4 a polovina kapitoly 5 popisuje stale dokola totez a i dalsi casti se opakuji. Hlavni teziste prace, kterym by mela byt vlastni analyza dat je naopak nedostatecna.

| *Hodnotící kritérium:* | *Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):* |
|---|---|
| **4. Věcná a logická úroveň práce** | *20 (F)* |

*Popis kritéria:*
Posuďte, zda předložená ZP je po věcné stránce v pořádku, případně vyskytují-li se v práci věcné chyby nebo nepřesnosti. Zhodnoťte dále logickou strukturu ZP, návaznosti jednotlivých kapitol a pochopitelnost textu pro čtenáře.

*Komentář:*
Prace je naprosto nelogicky strukturovana, viz cela prvni kapitola. Student zde naprosto jasne ukazuje, nejen ze nerozumi dane problematice, ale ani vztahum mezi jednotlivymi obory do kterych zasahuje.

Prace se velmi casto opakuje cimz komplikuje citelnost.

Dale prace obsahuje velke mnozstvi nepresnosti viz. napriklad kapitola o neparametrickych metodach, u ktere nasleduje popis parametru ktere je potreba optimalizovat. Prace casto rozporuje sama sebe a to i v ramci jedne kapitoly, viz 1.8.3 vs 1.8.3.2 kde nejprve student tvrdi, ze pro detekci anomalii je tezsi sehnat labelovana normalni data, nacez tvrdi ze normalni data lze sehnat snadneji nez data anomalni.

Dalsi nepresnosti napr: str.21 student tvrdi, ze regresni model je statisticke rozdeleni; jako nevyhody neuronovych siti a nahodnych lesu uvadi, ze vraci primo tridy (anomalni/normalni) namisto miry anomalie, nicmene oba modely se daji pouzit i tak aby vraceli miru anomalie; u metody zalozene na hledani asociacnich pravidel navrhuje pouzit min support threshold, coz je standardni technika pro klasifikaci nebo doporucovaci systemy ktera odstrani malo caste vzory z dat, coz je naopak naprosto nezadouci pro detekci anomalii.

| *Hodnotící kritérium:* | *Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):* |
|---|---|
| **5. Formální úroveň práce** | *60 (D)* |

*Popis kritéria:*
Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku ZP, viz Směrnice děkana č. 14/2015, článek 3.

*Komentář:*
I po formalni strance ma prace nekolik zasadnich nedostatku jako je nekonzistentni znaceni vzorcu, ktere student jednoduse odnekud opsal aniz by se nad nimi prilis zamyslel. Kapitola 1.12.1.1 je jasnym prikladem. Student zde zavedl dva znaky pro prumernou hodnotu, ci smerodatnou odchylku, pro hodnoty pouziva male x i velke X a podobne. Naopak nektere znaky jako velke S nezavedl vubec. Notace je nekonzistentni v ramci cele prace, velka pismena nekde znaci mnoziny, jinde hodnoty atp.

Dalsi nekozistence, tektokrat typograficka je napr. v sekci 1.15.3. Obvykle student popisuje vyhody a nevyhody jako cislovane seznamy, zde jsou v textu odstavce.

Jazykovou uroven bohuzel nemohu prilis posoudit, jelikoz je prace psana slovensky. Nicmene mi prijde, ze je prace psana ponekud hovorove. Dale prace obsahuje preklepy a opakujici se slovesa.

| *Hodnotící kritérium:* | *Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):* |
|---|---|
| **6. Práce se zdroji** | *40 (F)* |

*Popis kritéria:*
Vyjádřete se k aktivitě studenta při získávání a využívání studijních materiálů k řešení ZP. Charakterizujte výběr studijních pramenů. Posuďte, zda student využil všechny relevantní zdroje nebo zda se pokoušel řešit již vyřešené problémy. Ověřte, zda jsou všechny převzaté prvky řádně odlišeny od vlastních výsledků a úvah, zda nedošlo k porušení citační etiky a zda jsou bibliografické citace úplné a v souladu s citačními zvyklostmi a normami.

*Komentář:*
Cela sekce 1.7.2 je opsana z clanku: "Chandola, Banerjee and Kumar 2009 Anomaly detection: A survey" bez jedineho odkazu na puvodni praci. Clanek je sice v seznamu literatury, nicmene v teto kapitole citace uvedena neni.

Obrazky a tabulky nejsou obvykle v textu referovany vubec.

Citovana literatura je k obsahu prace relevantni, bohuzel fakticke chyby v textu naznacuji, ze ji student ve skutecnosti necetl.

Dle zadani se mel student venovat detekci anomalii na grafovych datech, kde tvrdi ze oblast neni dostatecne prozkoumana coz neni pravda viz bod 2.

| *Hodnotící kritérium:* | *Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):* |
|---|---|
| **7. Hodnocení výsledků, publikační výstupy a ocenění** | *40 (F)* |

*Popis kritéria:*
Vyjádřete se k úrovni dosažených hlavních výsledků ZP, např. k úrovni teoretických výsledků, nebo k úrovni a funkčnosti technického nebo programového vytvořeného řešení, apod. Případně také zhodnoťte, zda software nebo zdrojové texty, které nevytvořil sám student, byly v ZP použity v souladu s licenčními podmínkami a autorským právem. Popište případnou publikační činnost a získaná ocenění související s řešením této ZP.

*Komentář:*
Vystupy prace jsou nedostatecne a nepresvecive. V kapitole 6.2 student obhajuje svuj binarni klasifikator s presnosti 46%, coz je horsi nez nahodny klasifikator. Nasleduje porovnani s rozhodovacim stromem, ktery na temer vsechny vstupy vraci tridu 1 a s neuronovou siti, ktera se rozhoduje na zaklade jednoho vstupniho atributu obr. 6.2 str. 58.

U modelu nejsou typicky uvedene zadne parametry ani nastaveni. Napriklad pro SVM je volba jadrove funkce naprosto zasadni.

Nasledna analyza nebezpecnych letist a letu, vice mene zachranuje alespon castecne splneni zadani.

| *Hodnotící kritérium:* | *Způsob hodnocení - nehodnotí se* |
|---|---|

### 8. Komentář o využitelnosti výsledků

Popis kritéria:
Uveďte, zda hlavní výsledky ZP rozšiřují již publikované známé výsledky a/nebo přinášející zcela nové poznatky. Uveďte možnosti využití výsledků ZP v praxi.

*Komentář:*

Vysledky ohledne nebezpecnych letist a nebezpecnych letu jsou vcelku prukazne zalozeny na dostupnych datech. Jejich analyza spocivala v prostem vypocteni frekvence nebezpecnych pasazeru.

Vysledky, ktere se tykaji oznaceni anomalnich pasazeru jsou naopak naprosto nedostatecne a jejich vyuziti v praxi je nerealne, viz predchozi bod hodnoceni.

| *Hodnotící kritérium:* | *Způsob hodnocení - nehodnotí se* |
|---|---|

### 9. Otázky k obhajobě

*Popis kritéria:*
Uveďte případné dotazy, které by měl student zodpovědět při obhajobě ZP před komisí (body oddělte odrážkami).

*Otázky:*

V kapitole 7.2 uvadite, ze by bylo mozne data dale analyzovat algoritmy detekce anomalii. Mel jste na mysli nejake konretni algoritmy? Jake a proc?
Proc jste z duvodu anonymizace vynechal i parametr typ dokladu?

| *Hodnotící kritérium:* | *Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):* |
|---|---|
| **10. Celkové hodnocení** | *40 (F)* |

*Popis kritéria:*
Shrňte stránky ZP studenta, které nejvíce ovlivnily Vaše celkové hodnocení. Celkové hodnocení **nemusí** být aritmetickým průměrem či jinou hodnotou vypočtenou z hodnocení v předchozích jednotlivých kritériích 1 až 9.

*Text hodnocení:*

Prace obsahuje velke mnozstvi logickych, formalnich i typografickych chyb.
Analyza je nedostatecna a prezentovane modely pro detekci anomalii (ve skutecnosti binarni klasifikatory) nejsou pouzitelne ani teoreticky natoz pak v realne situaci.
Navzdory nazvu i zadani, prace ve skutecnosti nepouziva metody detekce anomalii, kterym se venuje v teoreticke casti, ale spoleha se na binarni klasifikaci.
Student neuvadi zdroj pro sekci 1.7.2, ktera je opsana vcetne formatovani.

Vzhledem k tomu, ze zde uvedene pripominky jsou jen spicka ledovce a vzhledem k zavaznosti tematu hodnotim praci znamkou F a navrhuji aby student praci prepracoval.

Podpis oponenta práce: