Posudek oponenta závěrečné práce

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

Student: Bc. Martin Beránek
Oponent práce: Ing. Tomáš Šabata
Název práce: Včasné varování před problémy infrastruktury informačního systému
Obor: Počítačové systémy a sítě

Datum vytvoření: 28. 1. 2018

Hodnotící kritérium:	*Způsob hodnocení - následující škálou 1 až 5:*
1. Náročnost a další komentář k zadání	*1=mimořádně náročné zadání,* ***2=náročnější zadání,*** *3=průměrně náročné zadání,* *4=lehčí, ale ještě dostatečně náročné zadání,* *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:
Zadání práce hodnotím jako náročnější vzhledem k propojení metod pro predikci časových řad s komplexní problematikou predictive maintanance v informačních systémech.
Jako náročnější hodnotím zadání vzhledem k oboru studenta, kde se s algoritmy, které v práci využívá, nesetkává.

Hodnotící kritérium:	*Způsob hodnocení - následující škálou 1 až 4:*
2. Splnění zadání	***1=zadání splněno,*** *2=zadání splněno s menšími výhradami,* *3=zadání splněno s většími výhradami,* *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:
Všechny body zadání byly splněny.

Hodnotící kritérium:	*Způsob hodnocení - následující škálou 1 až 4:*
3. Rozsah písemné zprávy	***1=splňuje požadavky,*** *2=splňuje požadavky s menšími výhradami,* *3=splňuje požadavky s většími výhradami,* *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:
Rozsah práce je v pořádku.

Hodnotící kritérium:	*Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):*
4. Věcná a logická úroveň práce	*55 (E)*

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:
V práci chybí jakýkoliv úvod do časových řad a jejich vlastností, na které se autor odvolává (např. stacionarita, autokorelace).

Ve třetí kapitole autor provádí Augmented Dickey–Fuller test pro ověření vlastnosti stacionarity. V tomto testu však chybně rozhodne nezamítnout nulovou hypotézu ve prospěch alternativní. Pokud jsou uvedené hodnoty testu správné, pak je signál stacionární a ne naopak, jak uvádí autor. Později autor o datech prohlásí, že jsou podobná bílému šumu víc, než aby byla stacionární. Tvrzení nedává smysl, jelikož bílý šum je stacionární proces.

Pořadí podsekcí třetí kapitoly je nevhodně zvoleno. Nejprve se v kapitole diskutuje metodika sbírání a ukládání dat z monitorovacích systémů. V této podsekci navíc autor zbytečně zachází do detailů a uvádí zde hodnoty experimentů. Následně se teoreticky uvedou modely ARIMA a LSTM, které s předchozí podsekcí nemají téměř nic společného a teprve poté přichází podsekce předzpracování dat právě pro tyto modely.

Do části teoretické analýzy jsou často zařazeny výsledky experimentů, které by pro přehlednost měly být až v kapitole s experimenty. Výsledky experimentů jsou navíc často bez smysluplné interpretace výsledků.

Některé formulace jsou nepřesné a zavádějící.

Hodnotící kritérium:	*Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):*
5. Formální úroveň práce	*68 (D)*

Popis kritéria:
Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku ZP, viz Směrnice děkana č. 26/2017, článek 3.

Komentář:
Formální úroveň práce je na horší úrovni.

Po jazykové stránce obsahuje práce větší množství vět, které jsou jazykově špatně postavené a nedávají tak smysl. Práce obsahuje velké množství vět, které jsou kostrbaté a velmi špatně čitelné. Množství překlepů v práci je nepatrné.

Některé obrázky nejsou v textu referencovány (kapitola 1) a tak se ztrácí jejich význam. Některé tabulky jsou referencovány špatným způsobem (kapitola 1). Elementy výčtů jsou věty začínající velkým písmenem bez interpunkčního znaménka na konci. V textu se objevuje špatně užitá kurzíva. Text, obrázky i tabulky místy přetékají přes okraje.

Nalezené chyby považuji v diplomové práci za nepřijatelné a velmi snižují celkovou úroveň práce.

Hodnotící kritérium:	*Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):*
6. Práce se zdroji	*80 (B)*

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:
Citacím z online zdrojů chybí datum citace.
Autor se v textu u rešerže funkcionality softwaru odvolává na jeho verzi. Verzi softwaru však nelze zjistit ani z textu, ani z citací.

Hodnotící kritérium:	*Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):*
7. Hodnocení výsledků, publikační výstupy a ocenění	*90 (A)*

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:
Vzhledem k faktu, že programové řešení vzniklo pouze v rámci diplomové práce a bylo vytvořené jednočlenným týmem, autorem práce, hodnotím vzniklý software pozitivně.
Kvalitu podtrhuje také fakt, že byl software nasazen v realném prostředí firmy, která byla předmětem účelové studie.

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:
Autor vytvořil dobrou rešerži již existujících řešení. Jak autor poukázal, využití machine learning metod v této oblasti je teprve v počátcích, a tak se autorova práce od existujících řešení odlišuje. Své navrhnuté řešení otestoval v prostředí realného informačního systému.

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:

* Predikce v sekcích 3.3.2.1 - 3.3.2.3 je velmi podezřelá, protože se jeví až na počáteční hodnoty jako konstatní. Jaká je standardní odchylka testovacích dat a predikovaných dat bez prvních deseti hodnot? Jak takové chování LSTM vysvětlujete?

* Zvažoval jste při shlukování událostí využít také nějaké jejich kategorické zařazení (např. databázové chyby, síťové chyby)?

* V práci píšete o prahu, který je využíván pro odhalování událostí. Jak jste tento práh hledal?

* Byl systém testovaný v realném provozu nebo Vám bylo vytvořeno umělé prostředí, kde jste mohli vyvolávat události a systém testovat?

Hodnotící kritérium:	*Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):*
10. Celkové hodnocení	*60 (D)*

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:

Zadaní práce v sobě skrývá velký potenciál, který se autor snažil uchopit, což se mu bohužel nepodařilo dostatečně. Hlavní část práce lze rozdělit na dvě oblasti, predikce časových řad a shlukování událostí. Pokud by se autor zaměřil více jen na jednu oblast, mohla by být práce více systematická a propracovaná. V tomto případě jsou obě oblasti pokryty pouze povrchově.

Za velký nedostatek považuji špatné rozhodnutí při testování hypotézy o stacionaritě dat. Autor na této vlastnosti v práci staví svůj další postup a špatně tak interpretuje výsledky.
Z textu práce jsou zřejmé i některé nedostatky v oblasti statistiky a oblasti machine learningu, což se dá omluvit oborem studenta.

Podpis oponenta práce: