ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE

# POSUDEK OPONENTA ZÁVĚREČNÉ PRÁCE

## I. IDENTIFIKAČNÍ ÚDAJE

| | |
|---|---|
| **Název práce:** | **Získávání informací z rozsáhlých časových dat** |
| **Jméno autora:** | **Bc. Sebastian Skoupý** |
| **Typ práce:** | diplomová |
| **Fakulta/ústav:** | Fakulta elektrotechnická (FEL) |
| **Katedra/ústav:** | Katedra počítačů |
| **Oponent práce:** | Ing. Jiří Kubalík, Ph.D. |
| **Pracoviště oponenta práce:** | CIIRC, ČVUT |

## II. HODNOCENÍ JEDNOTLIVÝCH KRITÉRIÍ

**Zadání** **náročnější**

*Hodnocení náročnosti zadání závěrečné práce.*

Cílem práce bylo navrhnout metody pro získávání informací z rozsáhlých dat s velkým počtem příznaků, pocházející z reálného systému papírny. Konkrétně šlo o odhalení závislostí v datech, které by mohly být využity při konstrukci systému včasného varování před blížícím se kolizním stavem na výrobní lince - přetržením papíru.
Jedná se o netriviální data s mnoha, pravděpodobně redundantními, proměnnými, kde jednotlivé proměnné jsou vzorkovány s ruznou frekvencí a jejichž hodnoty v záznamech často chybí.
Zpracování takovýchto dat vyžaduje netriviální přístup. Z tohoto pohledu hodnotím zadání jako náročnější.

**Splnění zadání** **splněno**

*Posuďte, zda předložená závěrečná práce splňuje zadání. V komentáři případně uveďte body zadání, které nebyly zcela splněny, nebo zda je práce oproti zadání rozšířena. Nebylo-li zadání zcela splněno, pokuste se posoudit závažnost, dopady a případně i příčiny jednotlivých nedostatků.*

Student navrhnul několik metod pro zpracování dat z oblasti selekce příznaků a extrakce příznaků. Všechny přístupy v kombinaci s učícím algoritmem Support Vector Machine experimentálně otestoval a výsledky statisticky vyhodnotil. Student splnil zadání ve všech bodech.

**Zvolený postup řešení** **vynikající**

*Posuďte, zda student zvolil správný postup nebo metody řešení.*

Student postupoval systematicky. Každý svůj krok při analýze dat předem zdůvodnil, dosažené výsledky analyzoval a pokusil se je vysvětlit.
K postupu řešení nemám výhrad.

**Odborná úroveň** **A - výborně**

*Posuďte úroveň odbornosti závěrečné práce, využití znalostí získaných studiem a z odborné literatury, využití podkladů a dat získaných z praxe.*

Odbornou úroveň práce hodnotím velice pozitivně. Z textu je zřejmé, že student proniknul do řešené problematiky, k čemuž musel uplatnit poměrně široké znalosti zejména z oblasti strojového učení a statistiky. O tom svědčí i rozsáhlý seznam použité literatury.

**Formální a jazyková úroveň, rozsah práce** **B - velmi dobře**

*Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku.*

Po formální a typografické stránce nemám výhrad. Text je dobře strukturován, obsah je vyvážený.
Práce je psána dobrou angličtinou s přiměřeným množstvím překlepů a chybiček, které nijak nesnižují srozumitelnost textu.
Výhradu mám k tomu, že student nedůsledně definuje význam symbolů použitých ve formulích. Čtenář si tyto věci sice domyslí, ale přesto by měl být každý symbol explicitně zaveden. Podobná výhrada se týká i Algoritmu 1 na straně 29, který je velice nepřehledný, se vstupními a výstupními parametry vysvětlenými jen pomocí strohých komentářů.

| **Výběr zdrojů, korektnost citací** | **A - výborně** |
|---|---|
| *Vyjádřete se k aktivitě studenta při získávání a využívání studijních materiálů k řešení závěrečné práce. Charakterizujte výběr pramenů. Posuďte, zda student využil všechny relevantní zdroje. 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.* | |
| Student uvádí 45 referencí, což považuji za zcela dostačující výběr zdrojů. | |

| **Další komentáře a hodnocení** |
|---|
| *Vyjádřete se k úrovni dosažených hlavních výsledků závěrečné práce, např. k úrovni teoretických výsledků, nebo k úrovni a funkčnosti technického nebo programového vytvořeného řešení, publikačním výstupům, experimentální zručnosti apod.* |
| Výsledky dosažené s daty pocházejícími ze senzorů výrobní linky nejsou tak dobré, aby mohly být reálně aplikovatelné pro predikci přetržení papíru v dané papírně. Student svojí prací potvrdil závěry předchozí analýzy, která dopadla podobně. Takže i tento „negativní" výsledek hodnotím pozitivně.<br>Na druhou stranu, student v kapitole 6 ukazuje, že data popisující chemické vlastnosti zpracovávané buničiny, která dostal až v průběhu řešení své diplomové práce, mohou obsahovat více informací využitelných pro predikci přetržení papíru. Což by stálo za další zkoumání. |

## III. CELKOVÉ HODNOCENÍ, OTÁZKY K OBHAJOBĚ, NÁVRH KLASIFIKACE

*Shrňte aspekty závěrečné práce, které nejvíce ovlivnily Vaše celkové hodnocení. Uveďte případné otázky, které by měl student zodpovědět při obhajobě závěrečné práce před komisí.*

Diplomová práce Bc. Sebastiana Skoupého se mi líbila. Ukázal, že dokáže systematicky řešit zadaný problém a analyzovat dosažené výsledky.

Do diskuze mám následující otázky:

- Jestli se nemýlím, tak jste úlohu klasifikace definoval tak, že se snažíte rozpoznat stav, který předchází přetržení papíru od ostatních „normálních" stavů. Přičemž stavem rozumíte jedno pozorování systému v konkrétním čase, reprezentované aktuálními hodnotami senzorů. Toto podle mého nepostihne dynamiku systému. Nezkoušel jste stav definovat např. pomocí rozdílu hodnot u každého senzoru?
- Opět na podobné téma. Podle mého nemusí být správně už původní předpoklad, že je možné odhalit blížící se přetržení papíru pouze na základě aktuálních hodnot senzorů v posledním měřeném okamžiku před přetržením. Možná za přetržení může jev, který nastal už dřívě, ovšem jeho důsledek se projeví až s větším časovým zpožděním, kdy už na odpovídajících senzorech není pozorován. Můžete se k tomu vyjádřit?

Předloženou závěrečnou práci hodnotím klasifikačním stupněm **B - velmi dobře.**

Datum: 27.1.2016

Podpis: Kubah