

I. IDENTIFIKAČNÍ ÚDAJE

Název práce:	Collection and predictive analysis of student data
Jméno autora:	Jonáš Václavek
Typ práce:	diplomová
Fakulta/ústav:	Fakulta elektrotechnická (FEL)
Katedra/ústav:	Katedra počítačů
Oponent práce:	Olga Štěpánková
Pracoviště oponenta práce:	Katedra kybernetiky

II. HODNOCENÍ JEDNOTLIVÝCH KRITÉRIÍ

Zadání	náročnější
<i>Hodnocení náročnosti zadání závěrečné práce.</i>	
<p>Většina technicky zaměřených vysokých škol má v současné době velký problém v tom, že značnému procentu studentů, kteří nastoupí do prvního ročníku, se nepodaří dokončit ani bakalářský stupeň. Školy i stát mají zájem udržet kvalitu poskytovaného vzdělání a současně zvýšit průchodnost studia, což jsou dva zdánlivě antagonistické cíle. Tato práce představuje jeden z nadějných směrů, jak se oběma cílům přiblížit. Téma je vysoce aktuální a dává prostor pro návrh nových řešení.</p> <p>Zadání je poměrně komplexní a vyžaduje od studenta nejen dobrou teoretickou průpravu, ale i praktické zvládnutí řady programátorských technik a schopnost domluvit se s potenciálními uživateli cílového systému.</p>	

Splnění zadání	splněno
<i>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.</i>	
<p>Práce splňuje všechny body zadání. Součástí zadání je testování výsledné implementace na reálných datech. Velmi oceňuji to, že se podařilo získat a analyzovat úplná data o průběhu několika ročníků studia studentů bakalářského studia FS ČVUT včetně některých demografických údajů. Na základě těchto dat byly vytvořeny zajímavé modely pro predikci studijních výsledků. Analýzy provedené v diplomové práci poukazují na klíčovou roli, kterou pro zdárné ukončení studia hraje údaj o počtu získaných kreditů na začátku zkouškového období po prvním semestru studia.</p>	

Zvolený postup řešení	vynikající
<i>Posuďte, zda student zvolil správný postup nebo metody řešení.</i>	
<p>Student v souladu se zadáním implementoval elegantní webové rozhraní Seminarist pro evidenci docházky studentů, které bylo řádně otestováno v průběhu reálného nasazení v akademickém roce 2017/18. Získaná data pak zpracovával současně s daty získanými z administrativního systému KOS. Pro zpracování a analýzu dat vytvořil další interaktivní webový SW nástroj Analyst, který nabízí řadu možností pro analýzu dat od shlukování a klasifikace až po dobře pochopitelné vizualizace některých vlastností vybraných skupin dat. Oba systémy jsou implementovány tak, aby je bylo možno přirozeným způsobem dobře dále rozšiřovat a upravovat.</p>	

Odborná úroveň	A - výborně
<i>Posuďte úroveň odbornosti závěrečné práce, využití znalostí získaných studiem a z odborné literatury, využití podkladů a dat získaných z praxe.</i>	
<p>Student v Kapitolách 5 a 6 shrnuje teoretická východiska z oblasti strojového učení a analýzy dat, která používá ve svých praktických produktech. Je hezké, že použité principy vysvětluje na datech z prostředí analýzy studentských dat. Důkladně se seznámil i s aktuálními pracemi z oblasti Learning Analytics a tuto zkušenost dobře využil při návrhu implementovaných původních postupů zohledňujících specifika FS ČVUT.</p>	

Formální a jazyková úroveň, rozsah práce	B - velmi dobře
<i>Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku.</i>	
<p>Práce je psaná velmi dobrou angličtinou, do které se sem tam vloudí nějaký překlep. Typografické zpracování diplomové</p>	

práce odpovídá dobrému standardu a působivé jsou nabízené vizualizace dat.

Jen v kapitole 5. **Theory and technologies** zachází autor poněkud nestandardně s matematickou notací a s použitými pojmy. Např. na str. 22 v části 5.1.1 by mělo být lépe vysvětleno, co znamená „*samples with the same dimension*“. Neobvyklé je značení se znakem šipky místo množinového značení, které by v sekci 5.1.2 bylo víc na místě. Vzorec v sekci 5.2 na str. 23 by si zasloužil lepší vysvětlení – myslím, že správný tvar sčítané hodnoty by měl být $W(f(x_i), x_i)$.

Na str. 25 při popisu Bayesovy formule by mělo být explicitně zdůrazněno, že se týká výpočtu podmíněné pravděpodobnosti. A v dalším textu by mělo být jasně vysvětleno, čemu ve zpracovávaných datech odpovídá Evidence a čemu Observations.

Výběr zdrojů, korektnost citací

B - velmi dobře

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.

Práce účelně využívá všech informací z nalezených aktuálních zdrojů, které korektně uvádí a cituje. Reference spolu s webovými zdroji jsou řazeny abecedně a nejsou číslovány (str. 76-7), což poněkud ztěžuje hledání.

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 měření na reálných datech potvrzují, že vytvořené SW prostředí splňuje požadovaný účel práce a že stojí za to dále se tématu věnovat.

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.

Oceňuji vznik spolehlivé implementace použité ke sběru velmi cenných dat i zajímavé dosažené výsledky a neobvyklé analýzy provedené na získaných datech. Zajímalo by mne, jak student odpoví na otázku formulovanou v zadání „Je možné na základě dostupných dat formulovat nějaká doporučení pro výuku?“

Předpokládám, že se studentovi při obhajobě podaří odpovědět na uvedenou otázku a také upřesnit formulace, jejichž znění kritizují v bodě Formální a jazyková úroveň. V takovém případě si předložená závěrečná práce zaslouží hodnocení klasifikačním stupněm **A - výborně**.

Datum: 7.6.2018

Podpis: