

# Posudek oponenta závěrečné práce

České vysoké učení technické v Praze

Fakulta informačních technologií

**Student:** Bc. Magda Friedjungová

**Oponent práce:** Ing. Pavel Kordík, Ph.D.

**Název práce:** Predikce studijních výsledků studentů bakalářského programu Informatika FIT ČVUT

**Obor:** Znalostní inženýrství

**Datum vytvoření:** 14. 6. 2016

<b>Hodnotící kritérium:</b>	<b>Způsob hodnocení - následující škálou 1 až 5:</b>
<b>1. Náročnost a další komentář k zadání</b>	<b>1=mimořádně náročné zadání, 2=náročnější zadání, 3=průměrně náročné zadání, 4=lehčí, ale ještě dostatečně náročné zadání, 5=nedostatečně náročné zadání</b>
<b>Popis kritéria:</b> 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.)	
<b>Komentář:</b> Úspěšné splnění zadání vyžaduje nejen technické dovednosti, ale i práci z lidmi a schopnost interpretace problémů v doméně.	
<b>Hodnotící kritérium:</b>	<b>Způsob hodnocení - následující škálou 1 až 4:</b>
<b>2. Splnění zadání</b>	<b>1=zadání splněno, 2=zadání splněno s menšími výhradami, 3=zadání splněno s většími výhradami, 4=zadání nesplněno</b>
<b>Popis kritéria:</b> 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.	
<b>Komentář:</b> V celém rozsahu.	
<b>Hodnotící kritérium:</b>	<b>Způsob hodnocení - následující škálou 1 až 4:</b>
<b>3. Rozsah písemné zprávy</b>	<b>1=splňuje požadavky, 2=splňuje požadavky s menšími výhradami, 3=splňuje požadavky s většími výhradami, 4=nesplňuje požadavky</b>
<b>Popis kritéria:</b> Zhodnoťte přiměřenost rozsahu předložené ZP vzhledem k obsahu, tj. zda všechny části ZP jsou informačně bohaté a ZP neobsahuje zbytečné části.	
<b>Komentář:</b> Nadstandardní.	
<b>Hodnotící kritérium:</b>	<b>Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):</b>
<b>4. Věcná a logická úroveň práce</b>	<b>92 (A)</b>
<b>Popis kritéria:</b> 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:**

Práce je rozsáhlá, dotýká se několika oblastí, a tak lze najít i drobné nepřesnosti. Například v části, která se věnuje modelování:

Formulace kolem obrázku 5.3. by mohla nepozorného čtenáře svést k tomu, že vidí reálná data a skutečnou cílovou proměnnou. Ve skutečnosti je na vizualizaci predikce modelu.

"Metoda Random Forest spočívá ve vytvoření několika rozhodovacích stromů a následně v jejich složení (en. ensemble), díky čemuž se metoda lépe vyrovnává se změnami v distribuci dat." Toto tvrzení je odvážné a já bych to asi takto nenapsal. Samozřejmě většinou díky ensemblování vznikne robustnější model, ale jsou i případy, kdy tomu tak být nemusí. Nicméně v rámci popularizace technik je toto tvrzení asi akceptovatelné.

"kombinování modelů (en. ensemble methods) [30], která má za úkol ověřit prediktivní sílu modelů a utvrdit nás v přesnosti predikovaných výsledků" tato formulace vysloveně podivná.

Hodnotící kritérium:

Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):

**5. Formální úroveň práce**

100 (A)

Popis kritéria:

Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku ZP, viz Směrnice děkana č. 12/2014, článek 3.

Komentář:

Velmi pečlivě zpracováno.

Hodnotící kritérium:

Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):

**6. Práce se zdroji**

100 (A)

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 etikety a zda jsou bibliografické citace úplné a v souladu s citačními zvyklostmi a normami.

Komentář:

Citace bez výhrad.

Hodnotící kritérium:

Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):

**7. Hodnocení výsledků, publikační výstupy a ocenění**

100 (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:

Velmi rozsáhlá práce - cenná je tím, že hledí na skutečnou hodnotu a význam dat. Proto i výsledky jsou korektní a mají velkou hodnotu. Prediktivní modely byly nejen vyzkoušeny, ale výsledky byly skutečně dotaženy do praxe a navíc pečlivě zhodnoceny. To se běžně nevidí.

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:

Práce je nesmírně zajímavá a dá se na ní stavět. A to nejen technicky - příští semestr modely dále na fakultě využijeme, ale i teoreticky - oblast "student churn rate reduction" zajímá snad každou školu a není zatím z pohledu datové vědy uspokojivě prozkoumána.

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:

Dají se aktuální modely použít pro skórování studentů v reálném čase - tedy kdykoli v průběhu semestru, nebo je třeba postavit modely jiné. Jak?

Hodnotící kritérium:

Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):

**10. Celkové hodnocení**

99 (A)

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:*

Vynikající práce - rozsahem i kvalitou zpracování a dotažením výsledků do reálného života. I kdyby prediktivní model pomohl jen jedinému studentovi, už se to vyplatilo.

Podpis oponenta práce: