



# Posudek oponenta závěrečné práce

**Student:** Daniel Schmidt  
**Oponent práce:** Mgr. Petr Šimánek  
**Název práce:** The decision tree on a stream of features  
**Obor:** Znalostní inženýrství

**Datum vytvoření:** 31. 8. 2020

<i>Hodnotící kritérium:</i>	<i>Způsob hodnocení – následující škálou 1 až 4:</i>
<b>1. Splnění zadání</b>	<b>1=zadání splněno, 2=zadání splněno s menšími výhradami, 3=zadání splněno s většími výhradami, 4=zadání nesplněno</b>
<i>Popis kritéria:</i> Posuďte, zda předložená ZP dostatečně a v souladu se zadáním obsahově vymezuje cíle, správně je formuluje a v dostatečné kvalitě naplňuje. V komentáři uveďte body zadání, které nebyly splněny, posuďte závažnost, dopady a případně i příčiny jednotlivých nedostatků. Pokud zadání svou náročností vybočuje ze standardů pro daný typ práce nebo student případně vypracoval ZP nad rámec zadání, popište, jak se to projevilo na požadované kvalitě splnění zadání a jakým způsobem toto ovlivnilo výsledné hodnocení.	
<i>Komentář:</i> Předložená práce naplňuje požadované cíle práce. Rešerše literatury by mohla více obsahovat i metody pracující se "stream of features". Algoritmy jsou implementovány. Experimenty student provedl korektně a výsledky popsal. Diskuze výsledků by mohla být důkladnější, kapitola "3.6 Discussion and future work" se věnuje především future work.	
<i>Hodnotící kritérium:</i>	<i>Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):</i>
<b>2. Písemná část práce</b>	<b>78 (C)</b>
<i>Popis kritéria:</i> Zhodnoťte přiměřenost rozsahu předložené ZP vzhledem k obsahu, tj. zda všechny části ZP jsou informačně bohaté a ZP neobsahuje zbytečné části. Dále 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. Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku ZP, viz Směrnice děkana č. 26/2017, článek 3. Posuďte, zda student využil a správně citoval relevantní zdroje. Ověřte, zda jsou všechny převzaté prvky řádně odlišeny od vlastních výsledků, zda nedošlo k porušení citační etiky a zda jsou bibliografické citace úplné a v souladu s citačními zvyklostmi a normami. Zhodnoťte, zda převzatý software a jiná autorská díla, byly v ZP použity v souladu s licenčními podmínkami.	
<i>Komentář:</i> Rozsah práce je dostatečný. Struktura práce je smysluplná.  Velká část práce je věnována popisu existujících metod a rešerše literatury, tato kapitola je pěkně zpracovaná. V rešerši se autor věnuje především metodám rozhodovacích stromů, ale nejsou zde popsána existující řešení problému s postupně přibývajícími příznaky (feature streams).  Kapitola popisující implementaci by mohla být více rozpracovaná, například popis jednotlivých algoritmů by mohl být proveden v pseudokódu, vysvětlení by pak bylo srozumitelnější.  Kapitola obsahující experimenty je popsána korektně, výsledky jsou z popisu patrné a jsou důveryhodné.  Práce je psána poměrně srozumitelnou angličtinou, s poměrně velkým počtem chyb (především chybějící členy, neanglická větná stavba). Některé věty nejsou srozumitelné či dokonce dokončené (jako věta na straně 21). Některé chyby jsou při čtení rušivé, jako velká písmena uprostřed věty. Práci by prospělo ještě jedno přečtení před odevzdáním a kontrola gramatiky.  Citace jsou korektní a relevantní (možná by ale bylo vhodné citovat i sci-kit).	
<i>Hodnotící kritérium:</i>	<i>Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):</i>
<b>3. Nepísemná část, přílohy</b>	<b>80 (B)</b>
<i>Popis kritéria:</i> Dle charakteru práce se případně vyjádřete k nepísemné části ZP. Například: SW dílo – kvalita vytvořeného programu a vhodnost a přiměřenost technologií, které byly využité od vývoje až po nasazení. HW – funkční vzorek – použité technologie a nástroje, Výzkumná a experimentální práce – opakovatelnost experimentů	

*Komentář:*

SW - program je kvalitní, nástroje vhodně zvolené. Jejich volbu (proč python) bylo vhodné více vysvětlit v písemné části. Experimenty jsou opakovatelné.

Nutnost manuálního přidávání příznaků limituje využitelnost kódu. Počet příznaků je omezen na 15, což je v praxi nedostatečné (i když autor tvrdí opak). Důvod tohoto omezení není v práci vysvětlen.

*Hodnotící kritérium:*

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

**4. Hodnocení výsledků, jejich využitelnost**

85 (B)

*Popis kritéria:*

Dle charakteru práce zhodnoťte možnosti nasazení výsledků práce v praxi nebo uveďte, zda výsledky ZP rozšiřují již publikované známé výsledky nebo přinášející zcela nové poznatky.

*Komentář:*

Práce se snaží řešit důležitý a praktický problém při využití variant rozhodovacích stromů v praxi, výsledky jsou zajímavé.

Výsledky jsou zajímavé spíše z teoretického pohledu, pro praktické využití by bylo nutné udělat mnoho z kroků, které jsou popsány v práci v odstavci future work. Není jisté, že budou zjištěné poznatky platit i po těchto úpravách.

*Hodnotící kritérium:*

*Způsob hodnocení – nehodnotí se*

**5. Otázky k obhajobě**

*Popis kritéria:*

Uveďte případné dotazy, které by měl student zodpovědět při obhajobě ZP před komisí (body oddělte odřádkami).

*Otázky:*

Bylo by možné postupy pro stream of features (nebo jejich varianty) použít i v případě v praxi častěji používaných metod jako je xgBoost nebo random forest?

*Hodnotící kritérium:*

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

**6. Celkové hodnocení**

78 (C)

*Popis kritéria:*

Shrňte stránky ZP, které nejvíce ovlivnily Vaše celkové hodnocení. Celkové hodnocení nemusí být aritmetickým průměrem či jinou hodnotou vypočtenou z hodnocení v předchozích jednotlivých kritériích. Obecně platí, že bezvadně splněné zadání je hodnoceno klasifikačním stupněm A.

*Text hodnocení:*

Práce popisuje praktický ale často opomíjený problém při použití strojového učení praxi. Rychlost algoritmů, které je potřeba často přetrénovávat na nových datech a příznacích, je v praxi velmi důležitá. Experimenty provedené v práci jsou kvalitní a dobře popisují problém, implementace algoritmů je dobrá. Kvalitu práce poněkud snižují chyby v angličtině, nižší srozumitelnost textu (především ve výsledcích a závěru) a další nedostatky psaného textu. Rešerše literatury by mohla být mírně rozšířena a použité algoritmy přehledněji popsány.

Podpis oponenta práce: