



# Posudek oponenta závěrečné práce

**Student:** Lukáš Langr  
**Oponent práce:** Mgr. Petr Novák, Ph.D.  
**Název práce:** Product review sentiment analysis in the Czech language  
**Obor:** Znalostní inženýrství

**Datum vytvoření:** 10. 6. 2019

<i>Hodnotící kritérium:</i>	<i>Způsob hodnocení – následující škálou 1 až 4:</i>
<b>1. Splnění zadání</b>	<b><u>1=zadání splněno,</u> 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> Zadání práce bylo splněno. Více prostoru mohlo být věnováno možnostem dalšího vylepšení zkoumaných metod.	
<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>80 (B)</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> V práci jsou přehledně představeny metody strojového učení pro analýzu sentimentu a reprezentaci textu. Přesnost metod je zkoumána na reálných datech z uživatelských recenzí v češtině. Text práce je dobře členěn a zkoumané metody jsou náležitě citovány.  Experimentální část práce by mohla být doplněna o podrobnější analýzu výsledků a zejména o diskusi možných vylepšení metod.  V práci se občas vyskytují neformální výrazy a přechází se mezi singulární a plurální vypravěčskou formou. Na některých místech je použito značení, které není předem zavedeno (např. v popisu TF a IDF, str.13).	
<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>90 (A)</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	
<i>Komentář:</i> Práce využívá existující implementace metod pro reprezentaci textu a zkoumá možnosti jejich využití a případné úpravy pro analýzu sentimentu.	
<i>Hodnotící kritérium:</i>	<i>Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):</i>
<b>4. Hodnocení výsledků, jejich využitelnost</b>	<b>85 (B)</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 aplikuje a srovnává nové metody strojového učení pro analýzu sentimentu uživatelských recenzí v češtině. Může sloužit jako základ pro další výzkum zlepšování těchto metod.

*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 odrážkami).

**Otázky:**

- Jaký může být důvod, že klasifikátory využívající reprezentaci Word2vec klasifikují většinu recenzí jako pozitivní (tabulky 3.5 a 3.6)?

- Jaký může být důvod, že klasifikátory využívající reprezentaci TF-IDF vykazují nejvyšší přesnost?

- Jaké jsou další možnosti pro dosažení vyšší přesnosti metod? Některé postupy byly nastíněny (nastavování parametrů u Word2vec nebo fine-tuning klasifikátoru BERT), ale nejsou šířeji diskutovány.

*Hodnotící kritérium:*

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

## 6. Celkové hodnocení

85 (B)

**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:**

V předložené práci jsou čtivým způsobem představeny, prozkoumány a otestovány metody strojového učení pro analýzu sentimentu. Práce by mohla být vylepšena jak po formální stránce, tak zejména co se týče diskuse výsledků experimentů a prozkoumání dalších možností pro zpřesnění metod, nicméně splňuje požadavky kladené na bakalářskou práci.

Podpis oponenta práce: