

# Posudek bakalářské práce

Matematicko-fyzikální fakulta Univerzity Karlovy

**Autor práce** Jan Knížek  
**Název práce** Comparison of Approaches to Text Classification  
**Rok odevzdání** 2019  
**Studijní program** Informatika      **Studijní obor** Obecná informatika

**Autor posudku** Mgr. Barbora Vidová Hladká, Ph.D.      **Role** Oponent  
**Pracoviště** ÚFAL MFF UK

Prosím vyplňte hodnocení křížkem u každého kritéria. Hodnocení *OK* označuje práci, která kritérium vhodným způsobem splňuje. Hodnocení *lepší* a *horší* označují splnění nad a pod rámec obvyklý pro bakalářskou práci, hodnocení *nevyhovuje* označuje práci, která by neměla být obhájena. Hodnocení v případě potřeby doplňte komentářem. Komentář prosím doplňte všude, kde je hodnocení jiné než *OK*.

## K celé práci

	lepší	OK	horší	nevyhovuje
Obtížnost zadání	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Splnění zadání	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Rozsah práce ... <i>textová i implementační část, zohlednění náročnosti</i>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Komentář Cílem bakalářské práce je klasifikace slovních hodnocení restaurací pomocí metod strojového učení. Autor úlohu řeší jako binární klasifikační úlohu, ve které výstupní třídy charakterizují vstupní instanci, hodnocení restaurace hosty, jako užitečné, nebo neúžitečné. Hodnocení pochází ze serveru Yelp a užitečnost hodnocení je nastavena dle počtu uživatelů serveru, kteří shledali hodnocení užitečným. Pečlivě je uplatněna metodologie strojového učení, hlavní důraz je kladen na výběr příznaků a evaluaci. Rešerše relevantní literatury je adekvátní. Sérií experimentů se nepodařilo zlepšit úspěšnost naivního klasifikátoru. Hlavní příčinu vidím ve vágních aspektech úlohy a jejich kvantifikaci, zejména aspekty subjektivity a užitečnosti hodnocení.				

**Textová část práce**

	lepší	OK	horší	nevyhovuje
Formální úprava ... <i>jazyková úroveň, typografická úroveň, citace</i>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Struktura textu ... <i>kontext, cíle, analýza, návrh, vyhodnocení, úroveň detailu</i>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Analýza	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Vývojová dokumentace	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Uživatelská dokumentace	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<p>Komentář Struktura práce i samotný text jsou srozumitelné. Práce je psána anglicky a jistě by její úrovni prospělo ještě jedno čtení.</p> <p>V části 5.4 jsou uvedeny statistiky příznaků. Popisu dvou skupin příznaků rozumím, ovšem je pro mě obtížné zjistit celkový počet příznaků a které z nich jsou diskrétní, binární a spojitě. V části 5.5 jsou prezentovány grafy 5.8 a 5.9, ve kterých doporučuji přeformulovat popisky osy x.</p> <p>Hodnocení jsou před automatickou klasifikací předzpracována systémy pro detekci jmenných entitit a analýzu sentimentu. Myslím, že by bylo užitečné na vybraném vzorku dat vyhodnotit jejich úspěšnost.</p>				

**Implementační část práce**

	lepší	OK	horší	nevyhovuje
Kvalita návrhu ... <i>architektura, struktury a algoritmy, použité technologie</i>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Kvalita zpracování ... <i>jmenné konvence, formátování, komentáře, testování</i>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Stabilita implementace	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<p>Komentář Systém je navržen a implementován kvalitně. Autor používá k předzpracování vstupních instancí moduly společnosti Geneea.</p>				

**Celkové hodnocení** Výborně  
**Práci navrhuji na zvláštní ocenění** Ne

**Datum** 25. srpna 2019

**Podpis**