

## Posudek oponenta diplomové práce

**Student:** Drápela Karel, Bc.  
**Téma:** Predikce povahy spamových krátkých textů textovým klasifikátorem (id 20199)  
**Oponent:** Křena Bohuslav, Ing., Ph.D., UITs FIT VUT

1. **Náročnost zadání** průměrně obtížné zadání
2. **Splnění požadavků zadání** zadání splněno
3. **Rozsah technické zprávy** splňuje pouze minimální požadavky  
Odhaduji celkový rozsah práce na 65 ns (55 ns text a 10 ns obrázky), což s dostatečnou rezervou překračuje požadavky kladené na diplomovou práci.
4. **Prezentační úroveň předložené práce** 85 b. (B)  
Technická zpráva je členěna logicky a jednotlivé kapitoly na sebe dobře navazují. Text je čitelný a pochopitelný. Menší výhradu mám ke vzorcům uvedeným v kapitole 4, které by měly být buď lépe vysvětleny, aby je běžný čtenář mohl pochopit bez studia literatury, nebo vynechány (pro pochopení zbytku práce nejsou nezbytné).
5. **Formální úprava technické zprávy** 85 b. (B)  
Práce je zpracována pečlivě. Několik drobností (např. nekonzistentní označení chí-kvadrát testu ve vzorcích 3.12, 3.13 a 3.14; mísení strukturálního a behaviorálního pohledu v obrázku 6.1; nesoulad komentáře k obrázku 7.2 s jeho obsahem) nenarušuje celkově velmi dobrý dojem z práce.
6. **Práce s literaturou** 98 b. (A)  
Student využívá řadu relevantních zdrojů a řádně je cituje.
7. **Realizační výstup** 70 b. (C)  
Výstupem je konzolová aplikace pro klasifikaci zpráv, jejíž funkčnost mi student předvedl. Důležitější pro její použitelnost je přesnost klasifikace než způsob ovládní, nicméně propracovanější ošetření neočekávaných stavů by uživatel určitě ocenil. Při připojování aplikace k databázi by bylo vhodnější nepředávat heslo v otevřené podobě.
8. **Využitelnost výsledků**  
Aplikace dosahuje reálné přesnosti klasifikace 60-90 %. To je v řadě oblastí nedostačující, nicméně s ohledem na charakter problému je to akceptovatelný výsledek. Domnívám se, že aplikaci s touto přesností lze s výhodou používat pro předzpracování krátkých zpráv, když finální rozhodnutí zůstane na člověku.
9. **Otázky k obhajobě**  
Při vyhodnocování škálovatelnosti aplikace (kapitola 7.9) uvádíte, že program zvládne natrénovat klasifikátor až podle 100.000 vzorků. V testovacích sadách však máte nejvýše 1.000 anotovaných vzorků. Kde se těch dalších 99.000 vzorků vzalo? A pokud jste je získal duplikací, neovlivní to věrohodnost výsledků?
10. **Souhrnné hodnocení** 78 b. dobře (C)  
Student touto prací bezesporu prokázal inženýrské kvality - vytvořil funkční a užitečnou aplikaci podle zadání i související technickou zprávu velmi dobré úrovně. K celkovému pouze dobrému hodnocení se kloním proto, že vyhodnocování jednotlivých experimentů probíhalo pouze po jednotlivých dimenzích (odpovídajících jednotlivým technikám), což nezaručuje nalezení globálního optima. Výsledek pak navíc závisí na pořadí, v jakém byly techniky vyhodnocovány.

Prohlášení: Uděluji VUT v Brně souhlas ke zveřejnění tohoto posudku v listinné i elektronické formě.

V Brně dne: 4. června 2018

.....  
podpis