

## Posudek oponenta bakalářské práce

**Student:** Kubica Jan

**Téma:** Nástroj pro predikci atributů životního stylu na základě metagenomických dat z tlustého střeva (id 22114)

**Oponent:** Hon Jiří, Ing., UIFS FIT VUT

- 1. Náročnost zadání** **obtížnější zadání**  
Cílem práce bylo klasifikovat metagenomické vzorky odebrané z lidského střeva do různých tříd podle životního stylu. Vzhledem k tomu, že analýza metagenomických dat vyžaduje velké množství bioinformatických znalostí, považuji zadání za obtížnější a je spíše na úrovni diplomové práce.
- 2. Splnění požadavků zadání** **zadání splněno**
- 3. Rozsah technické zprávy** **je v obvyklém rozmezí**  
Rozsah teoretické a praktické části je vyvážený.
- 4. Prezentací úroveň předložené práce** **70 b. (C)**  
Kapitoly jsou uspořádány v logickém sledu a práce je z velké části napsána přehledně. Výjimku tvoří osmá kapitola věnující se analýze výsledků, ve které schází hlubší vysvětlení provedených experimentů a jejich pečlivější interpretace. Je náročné se v této kapitole dobře orientovat. Z textu např. není jasné, jaká konkrétně metrika byla použita v algoritmu kNN.
- 5. Formální úprava technické zprávy** **75 b. (C)**  
Práce je psána česky na slušné úrovni. Občas se v práci objevují překlepy a výjimečně i výraznější nepřesnost (např. chybný vzorec 3.2). Po typografické stránce je práce kvalitní. Text je doplněn názornými obrázky, schémata a grafy. Na druhou stranu, některé grafy jsou příliš malé a především jejich textové popisky jsou špatně čitelné. Občas chybí popis os grafu (grafy 8.1 a 8.2). Dále by mnohdy bylo vhodnější uvést podrobnější vysvětlení tabulek a grafů přímo do jejich popisku a neuvádět je volně v textu.
- 6. Práce s literaturou** **80 b. (B)**  
Práce s literaturou je na dobré úrovni. Čerpáno bylo z kvalitních časopiseckých a konferenčních publikací z oblasti analýzy metagenomických dat. Převzaté části textu a obrázky jsou řádně označeny a odděleny od vlastního přínosu.
- 7. Realizační výstup** **85 b. (B)**  
Hlavní realizační výstup tvoří poměrně rozsáhlé zdrojové kódy v jazyce Python zajišťující zpracování datové sady, trénování klasifikačních nebo regresních modelů a následnou predikci. Dalším výstupem je množství experimentů s různými predikčními rysy a jejich zhodnocení.
- 8. Využitelnost výsledků**  
Práce staví na publikovaných metagenomických datech a snaží se je využít k predikci životního stylu. Výsledný prediktor vykazuje spíše nízkou přesnost a pravděpodobně naráží na velkou obtížnost daného predikčního problému související nejspíše s velkým šumem v trénovacích datech.
- 9. Otázky k obhajobě**
  - V práci používáte algoritmus kNN. Zkoušel jste více metrik pro posouzení vzdálenosti mezi vzorky? Jak dobře by fungovala např. kosinova vzdálenost, resp. Pearsonova korelační vzdálenost?
  - Zkoušel jste pro výběr rysů použít nějakou z již publikovaných metod? Např. LASSO?
  - Jak si vysvětlujete, že algoritmus LDA byl schopen za daných podmínek oddělit třídy atributu diet\_type? Dosáhl by natrénovaný model stejného výsledku i na dosud nepozorovaných datech?
- 10. Souhrnné hodnocení** **75 b. dobře (C)**  
Student prokázal, že je schopen se orientovat v obtížnější problematice analýzy metagenomických dat. Navrhl a implementoval řadu variant predikčních modelů a provedl mnoho experimentů, které ve výsledku dobře demonstrují současné limity predikce životního stylu z metagenomických dat. Prezentací a formální stránka práce je však na průměrné až podprůměrné úrovni a schází pečlivá interpretace některých výsledků. S ohledem na tyto skutečnosti navrhuji hodnocení stupněm **dobře (C)**.

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

V Brně dne: 30. května 2019

Hon Jiří, Ing.  
oponent