

Posudek oponenta bakalářské práce

Student: Kocinec Patrik
Téma: Zpracování logovacích informací na platformě TeskaLabs (id 23106)
Oponent: Jeřábek Kamil, Ing., UIFS FIT VUT

- 1. Náročnost zadání** **obtížnější zadání**
Na bakalářskou práci se jednalo o složitější zadání vzhledem k nutnosti provést analýzu dat, navrhnout výslednou pipeline pro jejich zpracování a porozumět a vhodně použít vybrané metody strojového učení
- 2. Splnění požadavků zadání** **zadání splněno s drobnými výhradami**
Zadání bylo splněno.
- 3. Rozsah technické zprávy** **splňuje pouze minimální požadavky**
Práce včetně obrázků dosahuje minimálního počtu normostran.
- 4. Prezentací úroveň předložené práce** **60 b. (D)**
Prezentací úroveň technické zprávy je na dobré úrovni. Jednotlivé kapitoly na sebe vhodně navazují a práce je pochopitelná. Praktická část, tedy návrh a implementace je vhodně, avšak ne nijak rozsáhle v práci zdokumentována. Část věnující se datům, jejich zpracování, použití algoritmů strojového učení a vyhodnocení dosažených výsledků je zmatečná. Jsou zde letmo popsány použité algoritmy, kterým není věnována ani zmínka v kapitole zabývající se teorií strojového učení (kde je například popsán algoritmus K-means, který v práci vůbec nefiguruje). Podčást kapitoly věnující se testování a vyhodnocení přesnosti zvolených algoritmů je velice omezená a krátká, nicméně obsahuje naměřené výsledky a jejich slovní zhodnocení.
- 5. Formální úprava technické zprávy** **70 b. (C)**
Práce je po typografické a jazykové stránce v pořádku. Je však možné si všimnout některých nevhodných tvarů slov či překlepů. V druhé části práce se dokonce objevují hovorová vyjádření.
- 6. Práce s literaturou** **55 b. (E)**
Použité literární zdroje jsou pro práci vhodně zvoleny. Místy však nebylo úplně přesně patrné, odkud byly informace čerpány. Pro samotné citace není použita jednotná forma.
- 7. Realizační výstup** **80 b. (B)**
Nástroj pro zpracování logů využívá vhodně připravené pipeline LogMan.io v návaznosti na další technologie. Kombinace technologií a jednotlivé části jsou vhodně navrženy a implementovány. Navržen byl taktéž postup, přeučování modelu, který však již není součástí implementace. Práce byla studentem demonstrována jako funkční.
- 8. Využitelnost výsledků**
Dle tvrzení studenta bude práce využita externím zadavatelem v praxi.
- 9. Otázky k obhajobě**
-V práci uvádíte, že k jednotlivým zprávám přidáváte typ, na základě kterého s nimi pak pracujete, nicméně je pak ukládáte v systému Apache Kafka pod jeden topic. Můžete zdůvodnit toto rozhodnutí? Případně byla by zde alternativa jak toto vylepšit za pomoci Apache Kafka?
- Byly při práci použity nějaké metody pro "cross validation"? Tedy testování více modelů s různým nastavením parametrů? Případně jak bylo dosaženo výsledného modelu prezentovaného v této práci?
- Jak jste určil, že zrovna Vámi vybrané vlastnosti, použité pro natrénování modelu jsou ty nejvhodnější?
- 10. Souhrnné hodnocení** **66 b. uspokojivě (D)**
Jedná se o zajímavé a netriviální zadání, které vyžaduje nastudování určité a poměrně rozmanité skupiny dat a způsobu jejího zpracování v oblasti bezpečnosti za pomoci strojového učení. Vytvořený výsledek je v praxi použitelný a vhodně navržený. Avšak text práce, především druhá polovina věnující se praktické části kazí celkový dojem.

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

V Brně dne: 25. června 2020

Jeřábek Kamil, Ing.
oponent