

Posudek oponenta bakalářské práce

Autor/autorka práce: **Irina Osovschi**

Název práce: **Azure Machine Learning**

Obsah práce

Práce obsahuje přehled principů spjatých s učením obecně a specificky strojovým učením, základní koncepty (metody) strojového učení, přehled možností využití nástrojů Microsoft Azure Machine Learning a následně konkrétní vlastní návrh a implementaci modelu v Microsoft Azure vhodného pro modelování kapacitního vytížení pracovníků call centra firmy CCA Group a.s.

Kvalita řešení a dosažených výsledků

Koncept řešení, dosažené výsledky

Autorka se poměrně významně (na 20 stranách) věnuje prvnímu bodu zadání (seznámení se se strojovým učením a nástrojem MS Azure Machine Learning). Ve druhé kapitole se věnuje učení jako takovému, což je náročné téma, které nelze v rozsahu bakalářské práce příliš kvalitně zpracovat. Při vysvětlování konceptů (metod, stylů) strojového učení autorka tyto relativně detailně (na úrovni rovnic) popisuje, odkaz na jakýkoli spolehlivý zdroj by byl v určité míře detailu zřejmě dostačující. Délka tohoto přípravného popisu může čtenáře odrazovat, neb ten je dlouho napínán tím, které z popisovaných konceptů nástrojů autorka posléze použije a jak zní konkrétní praktické zadání/úkol. Popis konkrétního úkolu (samotný úkol není příliš rozsáhlý/složitý) a jeho zpracování začíná na straně 27; v úvodním odstavci je dobře popsán. Porozumění zadání je podpořeno popisem životního cyklu a zejména finální distribuce softwaru zákazníkům, neboť modelování vytížení pracovníků call centra je časem distribuce softwaru významně ovlivněno. Autorka získala od firmy CCA relativně kvalitní data týkající se časů distribuce sw a počtu hovorů v call centru v jednotlivých dnech od roku 2013. Autorka data rozumně analyzovala a předzpracovala, následně pak vybrala i vhodné metody strojového učení a vytvořila příslušné (dva) modely pro predikci předpokládaného množství hovorů během jednotlivých dnů a optimálního množství pracovníků v call centru. Jestliže předzpracování a anotování dat poměrně detailně vysvětluje, výběr metod komentuje, evaluaci modelu věnuje nejprve pouze odstavec a pak se k ní více vrací v části 5.6. Ve zbývajícím textu se autorka věnuje popisu způsobu aplikace zvolených metod a postupu zpracování dat v nástroji MS Azure Machine Learning studio.

Zdrojové kódy v jazyce R jsou relativně krátké (desítky řádek), a tudíž stále čitelné, nicméně jejich komentování by bylo vhodné. Překvapilo mne, že manuál (instructions v terminologii autorky) není k dispozici jako textová příloha výtisku práce.

Formální úroveň

Text bakalářské práce má v rozumné úpravě celkem 37 stran (číslování 7 až 43), ve skutečnosti je to méně, neboť např. některé obrázky (např. 5.11 a 5.12) jsou na samostatných stranách. Prvních dvacet stránek textu je obsahově naddimenzovaných vůči samotnému praktickému úkolu, nicméně toto je nakonec vzhledem k jeho rozsahu pochopitelné.

Práce je psána slušnou angličtinou, vyskytují se v ní překlepy a nevhodně formulované věty, nicméně jejich počet je akceptovatelný. U převzatého obrázku 4.1 a možná i 4.2 není uveden zdroj.

Práce s literaturou

Autorka v práci cituje 18 zdrojů (většinou internetových), mezi nimi jsou odkazy na některé základní práce týkající se machine learningu, devět odkazů na Wikipedii a odkazy na MS Azure Machine Learning nástroje. V českém akademickém světě obecně nejsou odkazy na Wikipedii pozitivně vnímány, nicméně práce tímto nebyla negativně ovlivněna. Formát citací není zcela konzistentní.

Splnění zadání

Zadání bylo splněno.

Doplňující informace k práci

Ač bakalářské práci chybí větší provázanost mezi přípravnou částí a samotným zpracováním konkrétního úkolu, studentka se úspěšně vypořádala s pro ni neznámou a obtížnou doménou. To jí jistě stálo nemalé množství úsilí a času.

Dotazy k práci

1. Z textu práce usuzuji, že popisované modely byly nasazeny jako webové služby, nicméně v závěru práce píšete, že vylepšením by bylo právě veřejné nasazení těchto služeb. Znamená to, že modely jako webové služby byly otestovány, nicméně nebyly reálně nasazeny?
2. Jak jste simulovala délku hovorů při výpočtu optimálního množství pracovníků call centra? Jak jste při této simulaci využívala data dodaná firmou CCA?

Navrhuji hodnocení známkou **výborně** a práci doporučuji k obhajobě.

V Plzni 10. 7. 2020

Ing. Roman Mouček, Ph.D.