

## POSUDEK OPONENTA DIPLOMOVÉ PRÁCE

Student: **Mikhail Larkin**

Oponent: **prof. RNDr. PaedDr. Eva Volná, PhD.**

Studijní program: **Inženýrská informatika**

Studijní obor/Specializace: **Softwarové inženýrství**

Akademický rok: **2022/2023**

Téma diplomové práce: **Analýza vlivu reprezentace textu na kvalitu klasifikace pomocí metod umělé inteligence**

### Hodnocení práce:

#### Úplnost vypracování, aktuálnost a obtížnost řešeného úkolu

Téma diplomové práce je aktuální. Práce se zabývá analýzou vlivu metod reprezentace textu na kvalitu klasifikace pomocí metod umělé inteligence. Byl analyzován vliv následujících metod reprezentace textu: Bag of Words, TF-IDF, Word2Vec, GloVe, FastText a Doc2Vec. Jako klasifikátory byly zvoleny: Random Forest, Support Vector Machine, Multilayer Perceptron a Convolutional Neural Network.

#### Způsob a úroveň pojetí řešeného úkolu

V průběhu experimentální studie byly využity volně dostupné nástroje a technologie z oblasti umělé inteligence a strojového učení. Aplikace byla vyvinuta v jazyce Python. Bylo provedeno a vyhodnoceno velké množství experimentů. Výstupy z experimentální studie byly v závěrečné části diskutovány, vzájemně porovnány a vyhodnoceny. Tímto byl tedy hlavní cíl práce ve splnění.

#### Úroveň zpracování tématu, přínos diplomanta

V teoretické části diplomové práce jsou podrobně popsány zvolené metody reprezentace textu a zvolené klasifikátory. Praktická část práce prezentuje návrh a implementaci použité metodiky pro analýzu vlivu metod reprezentace textu na kvalitu jeho klasifikace.

#### Formální náležitosti práce, chyby a omyly v technické zprávě

Výběr literatury je dostatečný a odpovídá tématu diplomové práce. Většina obrázků nemá odpovídající kvalitu. Práce má formální nedostatky obsahuje překlepy a nesmyslné formulace – převážně v teoretické části.

#### Překlepy:

Abstrakt: Tato práce se zabývá analýzou vlivu metod reprezentace textu na kvalitu **kvalifikace** pomocí metod umělé inteligence.

Str. 9: První část tohoto **článku** je věnována ...

Str. 9: ... pomocí klasifikátorů uvažovaných v tomto **článku**.

Str. 13: ... **velký text** ....

Str. 15: Textová data **se** obvykle **skládají z dokumentů** ...

Str. 22: GloVe je **učební** algoritmus ...

Str. 35: ...**sáčkové** rozhodovací stromy (Bagging) ...

Str 38: ... je podobná tradiční vícevrstvé **perseptronové** neuronové síti.

Další nepřesnosti a formální nedostatky:

Číslování vzorců není průběžné, např. vzorec (1) se vyskytuje na str. 16, 18, 23, 31, 40, 46, 50

Str. 16: Oba vzorce 1 a 2 ukazují „výpočet termínovou frekvenci (TF)“, ale jsou různé.

Str. 18: není vysvětleno, co reprezentují některé symboly použité ve vzorcích

Str. 37: Listový uzel je tedy takový, který nemá žádnou entropii nebo jehož entropie je nulová.

Str. 43: Skryté vrstvy neuronové sítě jsou konfigurovány odlišně. **V některých případech jsou vážené vstupy přiřazeny náhodně. V jiných případech jsou vyladěny a kalibrovány pomocí procesu zvaného zpětné šíření chyby.**

Str. 43: Ačkoli se vytvářejí stejným způsobem jako ostatní umělé neurony v neuronové síti, **neurony výstupní vrstvy mohou být konstruovány nebo pozorovány odlišně**, vzhledem k tomu, že jsou posledními uzly sítě.

Str. 43: Typická tradiční neuronová síť má tři typy vrstev: **jednu nebo více vstupních vrstev**, jednu nebo více skrytých vrstev **a jednu nebo více výstupních vrstev.**

Str. 43: Složitější, inovativní neuronové sítě **mohou mít více než jednu vrstvu libovolného typu** - a jak bylo zmíněno ...

Str. 43: Stojí za zmínku, že neurony ve vstupní vrstvě musí mít velikost trénovacích instancí a **výstupní vrstva musí mít velikost výstupních štitků.**

Str. 45: ... naše skrytá vrstva neobsahuje **zkreslující člen.**

**Dotazy k obhajobě**

1. Jaké je praktické využití vaší práce?
2. Jak by bylo potřeba vytvořenou aplikaci upravit, aby pracovala s cizojazyčným textem.

**Shrnutí**

Student vykonal spoustu práce, ale s přihlédnutím k formální úpravě navrhuji hodnocení B - velmi dobře.

**Celkové hodnocení práce:**

Známku uvede oponent dle svého uvážení dle klasifikační stupnice ECTS:

A – výborně, B – velmi dobře, C – dobře, D – uspokojivě, E – dostatečně, F – nedostatečně.

Stupeň F znamená též „nedoporučuji práci k obhajobě“.

**Předloženou diplomovou práci doporučuji k obhajobě a navrhuji hodnocení**

**B - velmi dobře.**

**V případě hodnocení stupněm „F – nedostatečně“ uveďte do připomínek a slovního vyjádření hlavní nedostatky práce a důvody tohoto hodnocení.**