

I. IDENTIFIKAČNÍ ÚDAJE

Název práce:	Product categorization using machine learning
Jméno autora:	Bc. Roman Dušek
Typ práce:	díplomová
Fakulta/ústav:	Fakulta strojní (FS)
Katedra/ústav:	Ústav přístrojové a řídicí techniky
Oponent práce:	Ing. Cyril Oswald, Ph.D.
Pracoviště oponenta práce:	Ústav přístrojové a řídicí techniky

II. HODNOCENÍ JEDNOTLIVÝCH KRITÉRIÍ

Zadání	náročnější
<i>Hodnocení náročnosti zadání závěrečné práce.</i>	
Tématem práce je autonomní kategorizace nových, či stávajících špatně kategorizovaných, položek internetového obchodu na základě porovnání metadat položky s metadaty stávajících kategorizovaných položek. Jako vstupní data jsou dostupná textová data položek (název, popis, atributy) a fotografie položek. Ačkoliv algoritmy pro tento účel jsou široce dostupné a dobře popsány, jejich implementace konkrétní úkol je náročná.	

Splnění zadání	splněno
<i>Posudte, zda předložená závěrečná práce splňuje zadání. V komentáři případně uveďte body zadání, které nebyly zcela splněny, nebo zda je práce oproti zadání rozšířena. Nebylo-li zadání zcela splněno, pokuste se posoudit závažnost, dopady a případně i příčiny jednotlivých nedostatků.</i>	
Diplomant zadání práce splnil.	

Zvolený postup řešení	vynikající
<i>Posudte, zda student zvolil správný postup nebo metody řešení.</i>	
V práci je provedena široká rešerše algoritmů předzpracování dat v podobě textu v přirozeném jazyce a algoritmů strojového učení pro kategorizaci dat v přirozeném jazyce a dat v podobě obrázků. Před samotným řešením byla provedena vhodná analýza vstupních dat. Na základě analýzy a popsaných algoritmů diplomant navrhl v praktické části větší množství různých přístupů v podobě smysluplných kombinací vhodných algoritmů které vzájemně porovnal. Zvolený postup přesně odpovídá postupu, který by měl být pro zadaný úkol použit.	

Odborná úroveň	A - výborně
<i>Posudte úroveň odbornosti závěrečné práce, využití znalostí získaných studiem a z odborné literatury, využití podkladů a dat získaných z praxe.</i>	
Jak z rešerše, tak pak z popisu praktického využití metod algoritmů lze usuzovat, že diplomant metodám porozuměl, ví jak je využít k řešení zadaného problému i jaké jsou jejich omezení. V práci je využito opravdu veliké množství zdrojů a informace z nich získané jsou po odborné stránce dobře zpracované.	

Formální a jazyková úroveň, rozsah práce	B - velmi dobře
<i>Posudte správnost používání formálních zápisů obsažených v práci. Posudte typografickou a jazykovou stránku.</i>	
Práce je psaná v anglickém jazyce. Ačkoliv se diplomant dopustil určité části překlepů a nepřesností, je text dobře čitelný a srozumitelný. Výhradu mám pouze k obrázkům 1 až 3, které by rozhodně měli být značeny jako tabulky. Také bych preferoval seřazení zdrojů dle prvního výskytu v textu.	

Výběr zdrojů, korektnost citací

B - velmi dobře

Vyjádřete se k aktivitě studenta při získávání a využívání studijních materiálů k řešení závěrečné práce. Charakterizujte výběr pramenů. Posuďte, zda student využil všechny relevantní zdroje. Ověřte, zda jsou všechny převzaté prvky řádně odlišeny od vlastních výsledků a úvah, zda nedošlo k porušení citační etiky a zda jsou bibliografické citace úplné a v souladu s citačními zvyklostmi a normami.

V textu je citováno 108 zdrojů, což je více než dostačující. Avšak, zejména u obrázků, které nejsou autorovým dílem, často chybí odcitovaný zdroj a např. u citace 106 je odkazováno na závěrečnou práci, ale citace neúplná a je psána jako pro článek či kapitolu.

Další komentáře a hodnocení

Vyjádřete se k úrovni dosažených hlavních výsledků závěrečné práce, např. k úrovni teoretických výsledků, nebo k úrovni a funkčnosti technického nebo programového vytvořeného řešení, publikačním výstupům, experimentální zručnosti apod.

Vložte komentář (nepovinné hodnocení).

III. CELKOVÉ HODNOCENÍ, OTÁZKY K OBHAJOBĚ, NÁVRH KLASIFIKACE

Shrňte aspekty závěrečné práce, které nejvíce ovlivnily Vaše celkové hodnocení. Uveďte případné otázky, které by měl student zodpovědět při obhajobě závěrečné práce před komisí.

Práce obsahuje širokou a po technické stránce kvalitně zpracovanou řešerši vhodných algoritmů a metod z dané oblasti problematiky, která nejen obsahuje široké spektrum aktuálních metod. Dále je v práci popsána vhodná a dostačující analýza vstupních dat. V implementační části je popsáno větší množství zvolených přístupů založených na v řešerši popsáných metodách. Diplomant vyzkoušel provést kategorizaci produktů založenou jak na textových informacích v přiloženém jazyce, tak na informacích získaných rozpoznáváním objektů na k produktu dodaných fotografiích. V závěru bylo použito i metod kombinace obou klasifikačních přístupů. Práce je jak po teoretické stránce tak po praktické stránce kvalitní a je z ní patrné, že diplomant problematice porozuměl a metody které popsal uměl i následně korektně aplikovat k řešení praktické úlohy.

Otázky:

V práci se zmiňujete, že k běhu algoritmů jste využil CPU, a že mnoho použitých metod je účelnější provozovat na GPU, kde je díky paralelnímu běhu možné dosáhnout násobných rychlostí. Uvažoval jste o využití cloudových výpočtů, kde je v současnosti možné na studentských účtech možné využít i výpočtů na serverech s využitím GPU?

Zamýšlel jste se nad přístupem kategorizovat produkty do podkategorií postupně, tedy nejdříve využít jeden klasifikátor pro hlavní kategorie a další klasifikátory pro následující podkategorie?

Předloženou závěrečnou práci hodnotím klasifikačním stupněm **A - výborně**.

Datum: 27.1.2022

Podpis: