

I. IDENTIFIKAČNÍ ÚDAJE

Název práce:	Využití strojového učení při obchodování na energetických trzích
Jméno autora:	Nazariy Shukatka
Typ práce:	bakalářská
Fakulta/ústav:	Fakulta elektrotechnická (FEL)
Katedra/ústav:	Katedra počítačů
Oponent práce:	Petr Ryšavý
Pracoviště oponenta práce:	IDA, Katedra počítačů, FEL ČVUT

II. HODNOCENÍ JEDNOTLIVÝCH KRITÉRIÍ

Zadání	průměrně náročné
<i>Hodnocení náročnosti zadání závěrečné práce.</i>	
Zadání práce spočívalo v návrhu systému, který by umožňoval predikci vývoje cen plynu na energetickém trhu a tím následné strojové obchodování s touto komoditou. Určitá náročnost byla v nutnosti nastudovat si zákonitosti obchodování a základy strojového učení, které nejsou součástí programu SIT.	

Splnění zadání	splněno s většími výhradami
<i>Posuďte, 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>	
Student zadání splnil, nicméně některé části jsou nedostatečné. Podle zadání se měl student při studiu strojového učení zaměřit na algoritmy pro klasifikaci dat, kam spadají např. neuronové sítě, SVM klasifikátory, rozhodovací stromy a jiné. Místo toho student prezentuje pouze obecný popis, co je strojové učení, a jeho rozdělení na učení s učitelem, bez učitele a posilované učení. Podobně nedostatečná je evaluace dosažených výsledků. Ta je provedena v práci na jedné stránce, data jsou prezentovaná pouze ve formě otisku obrazovky z terminálu bez vysvětlení použitých metrik (zda jde o trénovací či testovací chybu, co znamená metrika PNL).	

Zvolený postup řešení	částečně vhodný
<i>Posuďte, zda student zvolil správný postup nebo metody řešení.</i>	
Zvolený postup vedl k cíli, nicméně obsahuje absenci mnoha důležitých rozhodnutí. Chybí například vysvětlení, jaké okno student pro učení použil, proč zvolil jako klasifikátor <code>RandomForestClassifier</code> či způsob rozdělení dat na trénovací a testovací data. Po nahlédnutí do kódu na <code>machine_learning_strategy.py:12</code> je vidět, že student použil klasifikátor z knihovny <code>scikit</code> jakožto hotové řešení bez jakéhokoliv nastavování metaparametrů či kontroly, zda nedošlo k přeučení.	

Odborná úroveň	D - uspokojivě
<i>Posuďte úroveň odbornosti závěrečné práce, využití znalostí získaných studiem a z odborné literatury, využití podkladů a dat získaných z praxe.</i>	
Text není příliš srozumitelný. Jako příklad mohu uvést kapitolu 4.2.3 vysvětlující posilované učení. Některé části ukazují nepochopení použitých termínů. Například po přečtení úvodu kapitoly 4.2.2 se může zdát, že jediný způsob učení bez učitele je pomocí neuronových sítí. Myšlenka použitého algoritmu <code>RandomForestClassifier</code> není vysvětlena.	

Formální a jazyková úroveň, rozsah práce	E - dostatečně
<i>Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku.</i>	
V práci jsem sice neobjevil mnoho překlepů (akorát chybějící zalomení řádku na straně 25), nicméně práce není vždy dobře srozumitelná. Rozsahem je kratší, pouze 25 stran, část z nich obsahuje ale nic neříkající kapitoly jako 1.1, 1.3 či 2.2. Další přibližně dvě strany obsahují přejaté obrázky, práce je tedy na hranici minimálního rozsahu.	

Čitelnosti nepřidává ani míchání českých a anglických termínů (UI vs. AI, Bollingerova pásma vs. Bollinger bands).

Výběr zdrojů, korektnost citací

D - uspokojivě

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.

Neobjevil jsem porušení citační etiky, nicméně citování je provedeno nevhodně, především v rešeršní části. Místo tradičního odkazu při vysvětlování důležitých termínů je pouze na konci práce seznam použitých referencí. Použité reference převážně odkazují na webové stránky a tutoriály, nikoliv na vědecké publikace. Například zmínka Bollingerových pásem by mohla být doplněna citací

- Bollinger, John. "Using bollinger bands." *Stocks & Commodities* 10.2 (1992): 47-51.

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.

Dodané kódy se zcela vyhýbají použití technik OOP. Obsahují minimum komentářů, jsou krátkého rozsahu a na některých místech obsahují kód, který by ve finálním programu neměl být. Například `machine_learning_strategy.py` obsahuje na řádce 4 `import IRIS` datasetu, který bývá v tutoriálech používán pro prezentaci algoritmů strojového učení, nicméně s bakalářskou prací nemá nic společného. Kódy navíc neobsahují žádné unit testy.

Jako velký problém poskytnutých kódů vidím časté tzv. *copy-paste* kódu. V souboru `strategies_testing.py` jsou metody `get_dictionary_with_metrics` a `test_statistical_strategy`, které sdílejí 44 řádků a liší se pouze tím, že jedna vypočtené hodnoty vytiskne a druhá vrátí jako slovník. Další příklad je v souboru `statistical_strategy.py`, kde se opakují řádky 119-134 a 172-187.

Podobně v souboru `machine_learning_strategy.py` lze najít opakované operace nad podobně pojmenovanými proměnnými na řádcích 88, 92, 96 a 100. Každé z volání představuje jedno období obchodování, vzhledem k rozšiřitelnosti by tedy bylo vhodnějším řešením mít data v poli a využít smyčku.

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.

Text práce neobsahuje vysvětlení některých použitých pojmů, testování není dostatečné, ani neobsahuje analýzu vhodnosti použití zvoleného klasifikátoru. Nicméně i přesto bych hodnotil text práce známkou D vzhledem k tomu, že jde o studenta program SIT. Na základě poskytnutých kódů bych navrhol celkové hodnocení na pomezí D a E, student se vyhýbá použití OOP, v kódu je časté *copy-paste* a použitý klasifikátor je použit pouze jako předpřipravená funkce bez přizpůsobení na problém studovaný v práci.

Na studenta mám následující dotazy:

1. Popište, jak byste mohl upravit Vaše kódy, aby využívaly OOP. Zaměřte se především na soubory `machine_learning_strategy.py` a `statistical_strategy.py`.
2. Co vás vedlo k použití klasifikátoru pomocí náhodného lesa? Proč jste nezkusil jiné klasifikátory?
3. Na stranách 20 a 31 říkáte, že by bylo lepší mít více příznaků pro učení. Je možné, že více příznaků bude znamenat horší přesnost klasifikátoru?

Předloženou závěrečnou práci hodnotím klasifikačním stupněm **D - uspokojivě**.

Datum: 26.8.2021

Podpis: