



Posudek oponenta závěrečné práce

Oponent práce:	Ing. Ivo Petr, Ph.D.
Student:	Kostiantyn Romanov
Název práce:	Využití analýzy medializovaných událostí pro vylepšení predikce vývoje cen akcií
Obor / specializace:	Znalostní inženýrství
Vytvořeno dne:	6. února 2023

Hodnotící kritéria

1. Splnění zadání

[1] zadání splněno

► [2] zadání splněno s menšími výhradami

[3] zadání splněno s většími výhradami

[4] zadání nesplněno

Student ve své práci propojuje několik oblastí strojového učení, když se zabývá spojením metod zpracování přirozeného jazyka a analýzy časových řad. Analýzou sentimentu ve zprávách objevujících se v médiích se student pokusil vylepšit predikční schopnosti modelů stojících na rekurzivních neuronových sítích. Menší nesoulad se zadáním sledávám v tom, že nejsou dostatečně diskutovány možnosti získávání dat z jiných trhů než NYSE a jako zdroj informací je použit výlučně web finance.yahoo.com. V rámci snahy o potvrzení či vyvácení domněnky o vlivu médií považuji za rozumnou volbu několika predikčních horizontů. Doporučil bych ale také zahrnutí dalšího delšího horizontu, aby zprávy mohly být do cen dostatečně promítnuty.

2. Písemná část práce

95 /100 (A)

Text práce je psán srozumitelně a čtivě. Práce je velmi dobře logicky strukturována a čtenáře až překvapí vysoká jazyková uroveň většiny práce. Ta je bohužel trochu nižší v závěru, kde se objeví několik překlepů (např. "analysis shoes"). Student podrobně popisuje použité nástroje a doplňuje je ilustracemi. U nich bych doporučil podrobnější popisky. Na práci bych zejména ocenil, že se student nesnaží za každou cenu ukázat přínos jimi použitých metod, ale otevřeně přiznává, že nemusí vést ke zlepšení predikce.

3. Nepísemná část, přílohy

100 /100 (A)

Softwarové řešení využívá jazyka Python a jeho knihoven. Zdrojový kód je kvalitně připraven. Celkové řešení má podobu Jupyter notebooku a je zpracováno přehledně a srozumitelně.

4. Hodnocení výsledků, jejich využitelnost

90 /100 (A)

Práce hezky popisuje několik konceptů oblasti zpracování přirozeného jazyka, architekturu transformerů a "self-attention". Jako hlavní přínos práce vidím testování hypotézy o vlivu médií na vývoj trhu. Pokud by se tuto snahu podařilo dále zpracovat, má práce i publikační potenciál. Výsledky prezentované v závěru lze interpretovat tak, že v krátkém horizontu predikce převažují metody založené čistě na ceně aktiva. Taková predikce se ale s prodlužujícím se oknem pro predikci bude přirozeně zhoršovat. Není tedy zas tak překvapivé, že vliv zpráv a sentimentu může vylepšit právě předpověď v delším horizontu.

Celkové hodnocení

95 /100 (A)

Práce je velmi dobře strukturována, použití zvolených metod dobře zdůvodněno a SW řešení kvalitně provedeno. Práce je na vysoké úrovni a i přes drobné výhrady navrhuji známku A.

Otázky k obhajobě

Hodnocení výsledků predikce probíhalo čistě porovnáním MSE? Je možno stanovit a testovat hypotézu o vlivu mediálních zpráv na přesost predikce?

Jak se vypořádáte s vlivem zásadních kurzotvorných zpráv k jednotlivým aktivům nebo trhům. Pozoroval jste v datasetu nějaké?

Jak se váš model může vypořádat se situací známou jako "good news is bad news" (např. dobré zprávy z trhu práce -> zvýšení úrokových sazeb FEDem -> pokles cen aktiv)?

Instrukce

Splnění zadání

Posudte, zda předložená ZP dostatečně a v souladu se zadáním obsahově vymezuje cíle, správně je formuluje a v dostatečné kvalitě naplňuje. V komentáři uveďte body zadání, které nebyly splněny, posudte závažnost, dopady a případně i příčiny jednotlivých nedostatků. Pokud zadání svou náročností vybočuje ze standardů pro daný typ práce nebo student případně vypracoval ZP nad rámec zadání, popište, jak se to projevilo na požadované kvalitě splnění zadání a jakým způsobem toto ovlivnilo výsledné hodnocení.

Písemná část práce

Zhodnoťte přiměřenost rozsahu předložené ZP vzhledem k obsahu, tj. zda všechny části ZP jsou informačně bohaté a ZP neobsahuje zbytečné části. Dále posudte, zda předložená ZP je po věcné stránce v pořádku, případně vyskytují-li se v práci věcné chyby nebo nepřesnosti.

Zhodnoťte dále logickou strukturu ZP, návaznosti jednotlivých kapitol a pochopitelnost textu pro čtenáře. Posudte správnost používání formálních zápisů obsažených v práci. Posudte typografickou a jazykovou stránku ZP, viz Směrnice děkana č. 52/2021, článek 3.

Posudte, zda student využil a správně citoval relevantní zdroje. Ověřte, zda jsou všechny převzaté prvky řádně odlišeny od vlastních výsledků, zda nedošlo k porušení citační etiky a zda jsou bibliografické citace úplné a v souladu s citačními zvyklostmi a normami. Zhodnoťte, zda převzatý software a jiná autorská díla, byly v ZP použity v souladu s licenčními podmínkami.

Nepísemná část, přílohy

Dle charakteru práce se případně vyjádřete k nepísemné části ZP. Například: SW dílo – kvalita vytvořeného programu a vhodnost a přiměřenost technologií, které byly využité od vývoje až po nasazení. HW – funkční vzorek – použité technologie a nástroje, Výzkumná a experimentální práce – opakovatelnost experimentů.

Hodnocení výsledků, jejich využitelnost

Dle charakteru práce zhodnoťte možnosti nasazení výsledků práce v praxi nebo uveďte, zda výsledky ZP rozšiřují již publikované známé výsledky nebo přinášející zcela nové poznatky.

Celkové hodnocení

Shrňte stránky ZP, které nejvíce ovlivnily Vaše celkové hodnocení. Celkové hodnocení nemusí být aritmetickým průměrem či jinou hodnotou vypočtenou z hodnocení v předchozích jednotlivých kritériích. Obecně platí, že bezvadně splněné zadání je hodnoceno klasifikačním stupněm A.