



POSUDEK OPONENTA DIPLOMOVÉ PRÁCE

Jméno studenta: Bc. Zdenko Brandejs

Název práce: Predikce ceny automobilu

Autor posudku: Ing. Karel Mls, Ph.D.

Cíl práce: Cílem práce je popsat principy tzv. hlubokého učení, vysvětlit možnosti implementace v Pythonu a realizovat ukázkovou aplikaci zaměřenou na predikování vývoje ceny automobilu od okamžiku jeho uvedení na trh.

Povinná kritéria hodnocení práce	Stupeň hodnocení (známka)					
	A	B	C	D	E	F
Práce svým zaměřením odpovídá studovanému oboru	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Vymezení cíle a jeho naplnění	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Zpracování teoretických aspektů tématu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Zpracování praktických aspektů tématu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Adekvátnost použitých metod, způsob jejich použití	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Hloubka a správnost provedené analýzy	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Práce s literaturou	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Logická stavba a členění práce	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Jazyková a terminologická úroveň	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Formální úprava a náležitosti práce	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Vlastní přínos studenta	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Využitelnost výsledků práce v teorii (v praxi)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>

Vyjádření k výsledku anti-plagiátorské kontroly

Podobnost práce s jinými texty je dle výsledku antiplagiátorské kontroly 0% - práce je tedy originální.

Dílicí připomínky a náměty:

Používání zkratk atd, atp, etc, např. v odborném textu nepovažuji za vhodné. Také používání první osoby není doporučované. Práce obsahuje poměrně značné množství pravopisných nedostatků (diakritika, shoda), autor používá originální verze oficiálních termínů (šachové postavičky, lékařský report,). Některé myšlenky jsou nešikovně vyjádřeny, což ztěžuje chápání textu („Algoritmy, které jsou závislé na kategorizování vzorků jsou náchylné i kvůli samotné kategorizaci.“). Bohužel text obsahuje i tvrzení,

s nimiž nelze souhlasit („Autoencoder je běžně stavěný tak, aby během transformací s sebou nesl všechny informace a o nic nepřišel.“), případně věcné chyby z nepozornosti („Např. mám-li sadu s 90 000 prvky podle poměru 2:1, trénovací podmnožině by připadlo 60 000 prvků a trénovací 30 000.“).

Není jasné, proč autor použil v Diagramu 2, 4 anglické popisky. Těž logika číslování/nečíslování printscreenů kódu jako obrázky mi unikla. Podobně volba představovaných konkrétních typů ANN v kapitole 2.5 nemá s následující praktickou implementací mnoho společného.

Kapitola 2.6 Matematický základ oproti očekávání neprezentuje jedinou rovnici, vše je popsáno slovy podle úvodního konstatování „Hlubší vysvětlení matematických operací by bylo nad rámec cíle této práce a k dosažení výsledku to není potřeba znát dopodrobna.“ Formátování citací je nejednotné, webové citace obsahují zavádějící informace (MIND, Data ani TEAM, Keras nejsou jména autorů).

Celkové posouzení práce a zdůvodnění výsledné známky:

Jak už jsem zmínil v Dílčích připomínkách, práce není ani po formální ani po obsahové stránce zdařilá. Autor se nejprve v teoretické části pouští do řady témat, která více či méně vzdáleně souvisí s tématem a cílem práce. To má za následek, že na poměrně skromné ploše (25 stran) jsou představeny buď jen obecné, nebo naopak velmi specifické informace bez odpovídajícího kontextu (využití logistické regrese pro doporučení císařského řezu). Potenciálně zajímavé poznatky tak vlivem autorovy nešikovnosti zapadnou (odstavec o kategorizaci pletacích stehů – zřejmě se mělo jednat o vzory). Zcela chybí teoretický úvod do softwarových prostředků a frameworků pro implementaci umělých neuronových sítí, potažmo prediktivního modelu z praktické části. Místo toho se v textu dozvíme o metodách oceňování movitého majetku a o časových řadách v ekonomické predikci.

V praktické části je implementována v prostředí Keras poměrně jednoduchá dopředná síť o jedenácti neuronech (prototyp), resp. jednadvaceti u výsledného modelu po transformaci staženého datasetu. Poměrně velká část prostoru je věnována podrobnému líčení kroků při předzpracování dat. Naopak se v textu nedočteme, proč autor použil jakou konkrétní funkci nebo proč změnil metodu adam na rmstop. Rovněž popisovaný model neumožňuje dosáhnout stanoveného cíle, což je „predikování vývoje ceny automobilu od okamžiku jeho uvedení na trh“.

Přes uvedené kritické připomínky doporučuji práci k obhajobě.

Otázky k obhajobě:

- jaký je významový rozdíl mezi Diagramem a Obrázkem (viz Diagram 1)?
- můžete blíže vysvětlit výsledek validace prototypu – obr. 13 (údajná přesnost modelu by byla 0)?
- proč byly nastaveny jako optimální parametry učení na 5 epoch a 16 vzorů v dávce?
- na kolika různých automobilech byl výsledný model testován a s jakými výsledky?
- jakým způsobem by bylo možné predikovat vývoj ceny v čase?

Práci doporučuji k obhajobě.

Navržená výsledná známka: E

V Hradci Králové, dne 3. září 2020

podpis