

TERMÍN: 05.11.2020

xx38670xx
Recenzia C
Andrej Iring
andrej.iring@cognexa.com

*Prosím nezasahujte do tejto tabuľky*RECENZENT/KA (meno a priezvisko, pozícia, inštitúcia): **Andrej Iring**NÁZOV MATERIÁLU: **Profilácia UoZ pomocou strojového učenia**TYP VÝSTUPU\*[1]: **Analýza**

(pri spoločných výstupoch uviesť aj typy individuálnych vkladov):

ANALYTICKÝ ÚTVAR, REZORT: **Ministerstvo práce, sociálnych vecí a rodiny SR - Inštitút sociálnej politiky**AUTORI/KY: **Ján Komadel, Tadeáš Chujac, Martin Demeter;**

SPOLUAUTORI/KY: - - ; - - ; - - ; - -

RECENZNÝ FORMÁT\*[2]: **2****PRIPOMIENKY:**

P.č.	Pripomienka sa vzťahuje k (strana, odsek):	Text pripomienky*[3]	Odôvodnenie pripomienky	Vysporiadanie sa s pripomienkou*[4]
1	Strana 2/ odsek 1	LightGBM je implementácia Gradient Boosting trees, samotne LightGBM nie je model ako taky. Odkazoval by som na LightGBM ako na implementáciu nie ako na model.	LightGBM ponúka v podstate viacero možností voľby modelov, konkrétne gradient boosting classification/regression tree ale ponúka (hoci menej používanú) možnosť random forests model, ktorý je značne odlišný od gradient boosting tree	Pripomienka akceptovaná. Upresnená formulácia v odseku 18.

			<p>modelov. Preto si nemyslím že je najšťastnejšie označovať váš model ako „model LightGBM“, preto si myslím že by bolo dobré to na začiatku upresniť. Kľudne by stačilo na začiatku spomenúť že používate model Gradient Boosting (classification) trees ktorý je implementovaný v knižnici LightGBM. Ďalej sa na náš model budeme odkazovať ako na LightGBM model.</p>	
2	Strana 12/ odsek 1	Vyslovene spomenúť že skóre bolo počítané na validačnej vzorke. Nie len ako poznámku pod čiarou.	Táto informácia bola poskytnutá v poznámke pod čiarou avšak myslím si že by bolo dobré to spomenúť aj priamo v časti s vypočítanými hodnotami v časti kde sa vyhodnocuje model.	Pripomienka akceptovaná.
3	Strana 12/ odsek 1	Vzhľadom na použitie F_beta skóre pre rôzne hodnoty, napísať krátke zdôvodnenie pre voľbu F_1 skóre	Zdôvodnenie môže byť len veľmi strohé, ale vzhľadom na to, že ste uverejnili hodnoty pre rôzne F_beta skóre tak by som aj očakával zdôvodnenie prečo ste sa rozhodli pre F_1 ktoré váhuje presnosť a citlivosť rovnako (stačí z pohľadu end usera rovnako zlé ak je ).	Pripomienka neakceptovaná. Maximalizáciu F1 štatistiky neprezentujeme ako najlepšiu možnosť, len ako možnú kompromisnú voľbu, na ktorej ilustrujeme úspešnosť modelu. Konkrétnu voľbu hraničnej rizikovosti odporúčame nastaviť podľa potreby a konkrétneho využitia modelu.

4	Strana 17/ odsek 3	Návrh. Doplniť odporúčania aj o opakovanie tejto analýzy po uskutočnení cieľných projektov na zníženie rizikovosti ako spôsob vyhodnocovania úspešnosti týchto projektov.	Jedná sa skôr o návrh na doplnenie odporúčaní, keďže vďaka tejto analýze by bolo teoreticky možné odmerať vplyv rôznych programov na zníženie dlhodobej nezamestnanosti. Napríklad vplyv na zníženie rizikovosti absolventov v prípade projektu na ich ľahšie začlenenie do pracovného trhu.	Pripomienka akceptovaná. Doplnený odsek 47.
5	Strana 6/ odsek 3	Spomínate že dátový súbor je nevyrovnaný avšak nespomínate či to bol problém pre model a bolo treba ďalšie kroky, alebo nevyrovnanosť nepredstavovala väčší problém.	Nespomínate, či ste sa venovali aj nevyrovnanosti datasetu, či už pri tvorbe datasetov ako takých (v prípade náhodného výberu validačnej vzorky či ste zachovali pomer jednotlivých tried v dátových výberoch , v prípade časového rozdelenia dát je to bezpredmetné) alebo v konštrukcii modelu (nastavenie väčšieho váhovania chyby pri minoritnej triede) alebo či nevyrovnanosť síce bola prítomná avšak nebola až tak veľká aby ovplyvnila model ako taký.	Pripomienka akceptovaná. Pridané upresnenie v časti 4 o zachovaní zastúpenia hodnôt $y$ v jednotlivých datasetoch a o váhovaní chyby.
6	Strana 6/ odsek 3	Nejednotné uvádzanie hodnôt pri dátovom súbore.	Pre hodnoty $y=0$ uvádzate percentá pre $y=1$ uvádzate absolútne hodnoty.	Pripomienka akceptovaná.
7				
8				

9				
10				

**CELKOVÉ HODNOTENIE (recenzent/ka vyplní túto časť po vysporiadaní sa s pripomienkami analytickou jednotkou):**

Celkovo veľmi dobre spracovaná analýza, využívacia najnovšie metódy a poznatky zo strojového učenia. Analýza priniesla zaujímavé výsledky ktoré boli dobre spracované a prezentované. Ako najväčší prínos analýzy považujem, že poskytla aj možnosti na ďalšie rozšírenie a využitie vytvorených modelov v realnej praxi.

[1] Výber medzi: 1. analýza (komplexný analytický materiál s návrhmi konkrétnych systémových opatrení); 2. komentár (rozsahovo menší analytický materiál venujúci sa konkrétnemu čiastkovému problému); 3. manuál (metodické usmernenie vyplývajúce z potreby zjednotenia procesov a postupov v konkrétnej oblasti).

[2] Formát 1 pre komentár/manuál (2 recenzenti bez povinného odborného workshopu); Formát 2 pre analýzu (3 recenzenti a povinný odborný workshop).

[3] Do tabuľky značiť pripomienky zásadného metodologického a obsahového charakteru (nie štylistické či gramatické opravy).

[4] Vyplní analytická jednotka: pripomienka bola akceptovaná / pripomienka nebola akceptovaná a zdôvodnenie / pripomienka bola čiastočne akceptovaná a zdôvodnenie.