

PRIRODOSLOVNO MATEMATIČKI FAKULTET, ZAGREB

STROJNO UČENJE

Mozgalo: Predviđanje ponašanja klijenata

ANA PARAĆ, SARA PUŽAR, PAULA VUJASIĆ

Podsjetnik

ZADATAK

- predvidjeti klijente koji će promijeniti ugovoreni odnos s bankom (kredit ili depozit) prijevremenim raskidom ugovora

SKUP PODATAKA

- podaci – informacije o ugovorima klijenata s bankom



Obrada podataka

PRETPROCESIRANJE PODATAKA

- Uklonjene nelogičnosti iz skupa podataka
- Uklonjeni višestruki izvještaji
- Outlieri

DATUM_IZVJESTAVANJA	KLIJENT_ID	OZNAKA_PARTIJE	DATUM_OTVARANJA
---------------------	------------	----------------	-----------------

2013-12-31	412140	7146814	2012-08-27
------------	--------	---------	------------

PLANIRANI_DATUM_ZATVARANJA	DATUM_ZATVARANJA	UGOVORENI_IZNOS
----------------------------	------------------	-----------------

2018-04-30	NaT	629155.97
------------	-----	-----------

STANJE_NA_KRAJU_PRETH_KVARTALA	STANJE_NA_KRAJU_KVARTALA	VALUTA
--------------------------------	--------------------------	--------

639373.81	641562.01	2
-----------	-----------	---

VRSTA_KLIJENTA	PROIZVOD	VRSTA_PROIZVODA	VISINA_KAMATE	TIP_KAMATE	STAROST	PRIJEVREMENI_RASKID
----------------	----------	-----------------	---------------	------------	---------	---------------------

1420	TM0109	L	1.50	A	22	N
------	--------	---	------	---	----	---

Obrada podataka

DODAVANJE ZNAČAJKI

- Značajke izvedene iz poznatih atributa

mjeseci, prijevremeno i neprijevremeno otplaćeni ugovori, nadvrsta klijenta, jednodnevni, godine zatvaranja i otvaranja, kamate nula

- Makroekonomske značajke

GDP, prosječna plaća, nezaposlenost, price index, tečajevi

Metode rješavanja

POČETNE METODE - Mozgalo

- Logistička regresija
- Random Forest

Podjela na kredite i depozite, podjela na pravne i fizičke osobe...



Metode rješavanja

KONAČNO RJEŠENJE

- XGBoost

Odabarne značajke :

- vrsta proizvoda, visina kamate i kategoričke vrijednosti tip kamate, proizvod, valuta i vrsta klijenta
- gotovo sve kreirane značajke

Evaluacija modela

TEHNIKE PROBIRA

- Podjela skupa: 80% za treniranje, 20% za testiranje
- Na skupu za treniranje korištena unakrsna validacija (5-fold CV)

MJERE USPJEŠNOSTI

- Točnost
- F1 mjera

Rezultati

XGBoost	Točnost	F1 mjera
Unakrsna validacija	89.61%	88.75%
Testni skup	89.72%	88.83%

Mozgalo	Točnost	F1 mjera
Testni skup	70%	73%
Validacijski skup	69%	69%

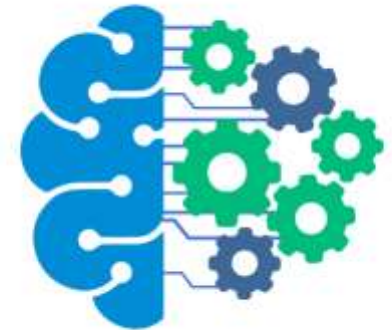
Rezultati

MATRICA KONFUZIJE

		stvarno	
		Prijevremeno	Neprijevremeno
predviđeno	Prijevremeno	38625	4301
	Neprijevremeno	5409	46203

Budući nastavak

- Ansambli
- Podjela modela na fizičke i pravne osobe
- Kreiranje dodatnih značajki



Hvala na pažnji!
