

Predviđanje uspjeha bankarskog marketinga

Projekt iz kolegija Strojno učenje

Davorin Gradečak, Petra Jambriško, Jakov Krunić

24. lipnja 2020.

Opis problema

Predviđanje
uspjeha
bankarskog
marketinga

Davorin Gradečak,
Petra Jambriško,
Jakov Krunić

Predviđanje uspjeha bankarskog marketinga na temelju popisnih podataka:

- ▶ *train* podaci - 41188 instanci/ispitanih korisnika
- ▶ *test* podaci - 4119 instanci/ispitanih korisnika

Svaki korisnik je opisan s 20 atributa i atributom *y* koji predviđamo (target).

Atribut *y* nam govori je li korisnik ugovorio oročeni depozit ili ne - *y* je kategorijska varijabla.

Opis i cilj problema

Analiza podataka

Metodologija

Algoritmi

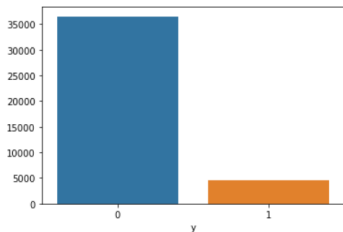
Funkcija troška

Rezultati

Osvrt na druge
pristupe

Mogući nastavak
istraživanja

Analiza podataka



U *train* podacima, od svih korisnika, njih 36548 nije ugovorilo oročeni depozit, a njih 4640 jest. Negativnih slučajeva je skoro 8 puta više od pozitivnih.

Predviđanje
uspjeha
bankarskog
marketinga

Davorin Gradečak,
Petra Jambriško,
Jakov Krunić

Opis i cilj problema

Analiza podataka

Metodologija

Algoritmi

Funkcija troška

Rezultati

Osvrt na druge
pristupe

Mogući nastavak
istraživanja

Analiza podataka

Predviđanje
uspjeha
bankarskog
marketinga

Davorin Gradečak,
Petra Jambriško,
Jakov Krunić

Opis i cilj problema

Analiza podataka

Metodologija

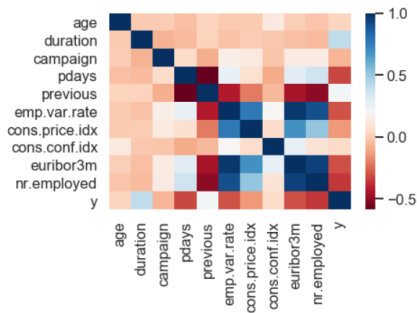
Algoritmi

Funkcija troška

Rezultati

Osvrt na druge
pristupe

Mogući nastavak
istraživanja



Korelacijska
matrica atributa
pokazuje da je
target varijabla y
najviše korelirana s
trajanjem poziva.

Algoritmi koje smo primijenili:

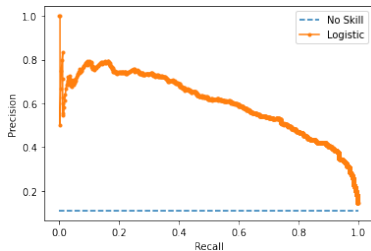
- ▶ logistička regresija
- ▶ naivni Bayes
- ▶ SVM
- ▶ stabla odluke, slučajne šume i XGBoost

Evaluacija:

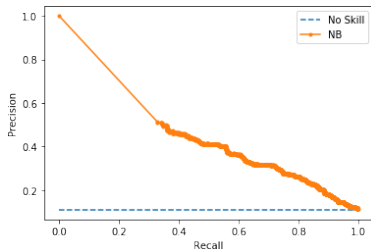
- ▶ AUPRC (area under precision-recall curve)

Koristili smo One Hot Encoding.

Logistička regresija i Naivni Bayes



Precision-recall
krivulja za
logističku regresiju,
 $AUPRC=0.60$



Precision-recall
krivulja za naivni
Bayes,
 $AUPRC=0.46$

Predviđanje
uspjeha
bankarskog
marketinga

Davorin Gradečak,
Petra Jambriško,
Jakov Krunić

Opis i cilj problema

Analiza podataka

Metodologija

Algoritmi

Funkcija troška

Rezultati

Osvrt na druge
pristupe

Mogući nastavak
istraživanja

- ▶ dugo vrijeme potrebno za učenje na danom skupu
- ▶ nema poboljšanja u odnosu na logističku regresiju

	Metric	Score Linear Without PCA	Score RBF Without PCA	Score RBF With PCA (31)
0	Accuracy	0.881282	0.900947	0.914300
1	Precision	0.455814	0.630303	0.658065
2	Recall	0.434590	0.230599	0.452328
3	F1 Score	0.444949	0.337662	0.536137
4	AUPRC	0.448848	0.582164	0.586098

Stabla odluke, slučajne šume i XGBoost

Predviđanje
uspjeha
bankarskog
marketinga

Davorin Gradečak,
Petra Jambriško,
Jakov Krunić

Opis i cilj problema

Analiza podataka

Metodologija

Algoritmi

Funkcija troška

Rezultati

Osvrt na druge
pristupe

Mogući nastavak
istraživanja

- ▶ Train & Test — 20% od primjera za učenje odabrano za validacijski skup
- ▶ Stablo odluke — $AUPRC = 0.56$
- ▶ Slučajne šume — $AUPRC = 0.60$
- ▶ XGBoost — $AUPRC = 0.67$

Promjena funkcije troška

Predviđanje
uspjeha
bankarskog
marketinga

Davorin Gradečak,
Petra Jambriško,
Jakov Krunić

Opis i cilj problema

Analiza podataka

Metodologija

Algoritmi

Funkcija troška

Rezultati

Osvrt na druge
pristupe

Mogući nastavak
istraživanja

- ▶ Potrebno uzeti u obzir ekonomsku realnost - troškovi za pokušaj stjecanja odnosno propust klijenata
- ▶ Propust dobrog klijenta znatno skuplji ($C_{FN} \gg C_{FP}$)

Stvarna klasa

		Stvarna klasa	
		Pozitivni	Negativni
Predviđeno	Pozitivni	0	C_{FP_i}
	Negativni	C_{FN}	0

► Koristili smo COSTCLA paket

	Naivni	Cost-sensitive
Logistička regresija	265 492.99925	19 366.3308
Decision Tree	466 392.99005	35 452.0257
Slučajne šume	584 525.402	36 512.73465

Usporedba rezultata

	Accuracy	AUPRC
Logistička regresija	0.913086	0.601339
Naivni Bayes	0.854819	0.463814
SVM Linear	0.881282	0.448848
SVM RBF	0.900947	0.582164
SVM RBF (PCA)	0.914300	0.586098
Slučajne šume	0.903594	0.601897
Bootstrap slučajne šume	0.903473	0.611274
XGBoost	0.911789	0.667031

- ▶ XGBoost smatramo najuspješnijim algoritmom
- ▶ Naivni Bayes smatramo najlošijim algoritmom

Predviđanje
uspjeha
bankarskog
marketinga

Davorin Gradečak,
Petra Jambriško,
Jakov Krunić

Opis i cilj problema

Analiza podataka

Metodologija

Algoritmi

Funkcija troška

Rezultati

Osvrt na druge
pristupe

Mogući nastavak
istraživanja

Osvrt na druge pristupe

- ▶ korištena je metrika AUC
- ▶ naglasak je na usporedbi algoritama
- ▶ često korištena metoda je XGBoost
- ▶ praktična realizacija je nedovoljno istražena

Predviđanje
uspjeha
bankarskog
marketinga

Davorin Gradečak,
Petra Jambriško,
Jakov Krunić

Opis i cilj problema

Analiza podataka

Metodologija

Algoritmi

Funkcija troška

Rezultati

Osvrt na druge
pristupe

Mogući nastavak
istraživanja

Mogući nastavak istraživanja

- ▶ potrebno je više atributa
- ▶ preciznija analiza profita i isplativosti
- ▶ proučavanje karakteristika agenata
- ▶ veza između agenata i njihovih uspješno kontaktiranih korisnika

Predviđanje
uspjeha
bankarskog
marketinga

Davorin Gradečak,
Petra Jambriško,
Jakov Krunić

Opis i cilj problema

Analiza podataka

Metodologija

Algoritmi

Funkcija troška

Rezultati

Osvrt na druge
pristupe

Mogući nastavak
istraživanja