

# Predikcija predstečaja poduzeća

David Bojanić, Domagoj Demeterfi

Zagreb, 2018.

# Motivacija

Od interesa je za:

- lokalnu zajednicu
- sudionike u industriji i investitore
- zakonodavce te globalnu ekonomiju.

# Ciljevi

- Postaviti prediktivni model koji, koristeći ekonometrijske mjere poduzeća u Hrvatskoj, predviđa ulazak tog poduzeća u postupak predstećajne nagodbe (pet zadataka binarne klasifikacije u ovisnosti o periodu predviđanja)
- Model interpretirati s ekonomskog stajališta

# Podaci

Skup financijskih izvještaja za 80-ak tisuća poduzeća u Hr. od 2007 do 2014 godine.

Sastoje se:

- općenitih podataka poduzeća
- sektora poduzeća (Nace Rev)
- izvještaji: bilance, RDG, novčani tok

Problem: asimetričnost podataka

Koncentrirali smo se na građevinski sektor.

Podaci za 5 klasifikacijskih problema su tada:

- *Nakon1Godine* - 301 u predstečaju, 29358 nije (ukupno 29659)
- *Nakon2Godine* - 153 u predstečaju, 29625 nije (ukupno 29778)
- *Nakon3Godine* - 151 u predstečaju, 30231 nije (ukupno 30382)
- *Nakon4Godine* - 153 u predstečaju, 29808 nije (ukupno 29961)
- *Nakon5Godine* - 129 u predstečaju, 28549 nije (ukupno 28678)

# Metodologija

Ansambl boosted stabala trenirana tehnikom extreme gradient boostinga s generacijom sintetičkih značajki (EXGB model)

# Sintetičke značajke

Sintetičke značajke računaju se u svakom koraku boostinga kombinirajući postojeće značajke nekom od osnovnih aritmetičkih operacija.

Svrha sintetičkih značajki je kombinirati ekonomske indikatore predložene od stručnjaka u kompleksne značajke.

# Metode evaluacije

Tehnike probira koje koristimo:

- Train & test
- Unakrsna validacija

Metoda usporedba različitih modela: Prosjek površina ispod ROC krivulje 10-fold unakrsne validacije (average auc roc 10-fold cross validation)



# Korišteni modeli

Baseline modeli:

- Logistička regresija
- Stablo odlučivanja
- Slučajna šuma

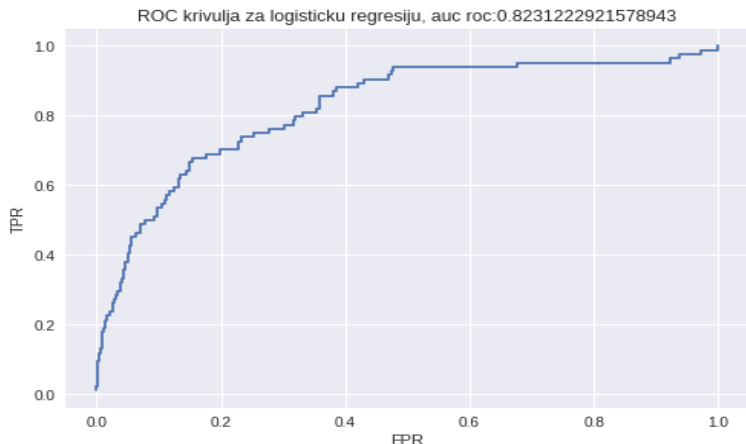
Modeli:

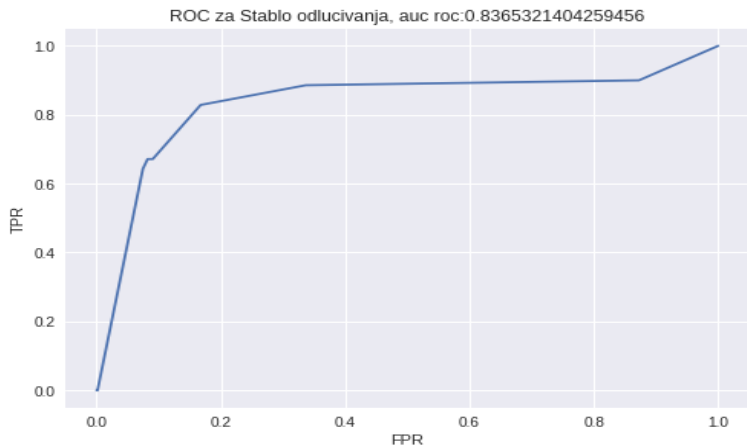
- XGboost
- XGboost sa sintetičkim značajkama

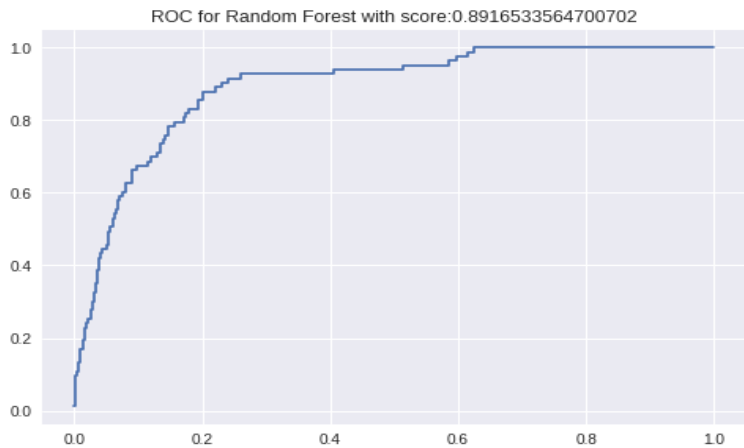
# Rezultati

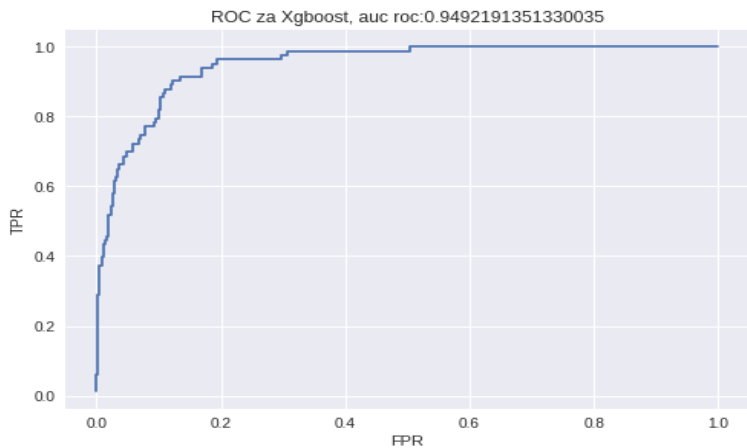
	1 god.	2 god.	3 god.	4 god.	5 god.
LR	0.700	0.628	0.562	0.630	0.674
DT	0.852	0.792	0.766	0.719	0.705
RF	0.899	0.844	0.806	0.785	0.790
XGB	0.935	0.878	0.817	0.811	0.844
XGBSZ	0.636	0.573	0.562	0.573	0.550

# ROC krivulje za problem Nakon1Godine (na cijelom train skupu)









# Sintetičke značajke

U financijama kao prediktori propadanja poduzeća često koriste omjeri stavki iz financijskog izvještaja. Zbog manjka preprocesiranja u podatkovnom skupu nalazi se puno nula što je onemogućilo da se stvaraju sintetičke značajke s operacijom djeljenja.

# Najvažnije značajke

Značajka	Važnost
Dugotrajna imovina	0.122
Obveze za poreze, doprinose i slična davanja	0.045
Kratkotrajna imovina	0.023
Promjena obveza prema zaposlenicima	0.020
Promjena obveza za poreze, doprinose i slična davanja	0.017
Novac u banci i blagajni	0.017
Promjena odgođene porezna	0.016
Postrojenja i oprema	0.014
Novac na početku razdoblja	0.013
Nematerijalna imovina	0.012



Projekt inspiriran radom:

Maciej Zieba, Sebastian K. Tomczak, Jakub M. Tomczak:

*Ensemble boosted trees with synthetic features generation in application to bankruptcy prediction,*

Expert Systems with Applications, str: 93-101

# Proširenje

- predviđati trajanje predstečajne nagodbe i/ili iznos tražbine u nagodbi
- dodavanje makroekonomskih uvjeta
- izloženost poduzeća drugim poduzećima