

Bank Marketing

Podaci se odnose na direktne marketinške kampanje (telefonski pozivi) Portugalske banke. Cilj klasifikacije je da predvidi da li će se klijent pretplatiti za oročeni depozit.

Informacije o atributima

- age (numerički) – starost.
- job (kategorijalni) – tip posla.
- marital (kategorijalni) – bračno stanje,
- education (kategorijalni) – obrazovanje.
- default (binarni) – neisplaćeno kreditno zaduženje
- balance (numerički) – srednji godišnji balans u €.
- housing (binarni) – stanbeni kredit.
- loan (binarni) – lični zajam.
- contact (kategorijalni) – tip komunikacije sa klijentom (fiksni/mobilni telefon)
- month (kategorijalni) – poslednji mesec kontakta sa klijentom.
- day_of_week (kategorijalni) – poslednji dan kontakta sa klijentom.
- duration (numerički) – trajanje poslednjeg poziva u sekundama. **Važna napomena:** ovaj atribut jako utiče na izlaz jer ako je duration = 0 onda je y = 'no'. Trajanje poziva je nepoznato dok se poziv ne završi, a kada se poziv završi y je poznato, tako da ovaj atribut se koristi kao benchmark.
- campaign (numerički) – broj kontaktiranja klijenta tokom kampanje.
- pdays (numerički) – broj dana koji je prošao od poslednjeg kontakta klijenta iz prošle kampanje (-1 znači da klijent nije prethodno kontaktiran).
- previous (numerički) – broj izvršenih kontakata pre ove kampanje.
- poutcome (kategorijalni) – ishod prethodne marketinške kampanje.
- y (binarni) – da li se klijent prijavio za oročeni depozit.

- Prosečna starost ljudi u datasetu je 41 sa standardnom devijacijom od 10.61.
- Minimum je 18 godina, a maksimum 95 godina.
- 75% klijenata ima 48 ili manje godina.

- Prosečni balans u datasetu je 1.362€ sa standardnom devijacijom od 3.045€.
- Minimum je -8.019€, a maksimum 102.127€.
- Kod 75% klijenata srednji godišnji balans je 1.428€ ili manje.

- Prosečan broj kontaktiranja klijenta pre kampanje je 0.6 puta sa standardnom devijacijom od 2.3.
- Minimalni broj kontaktiranja je 0, a maksimalan je 275.
- 75% klijenata nije uopšte kontaktirano pre kampanje.

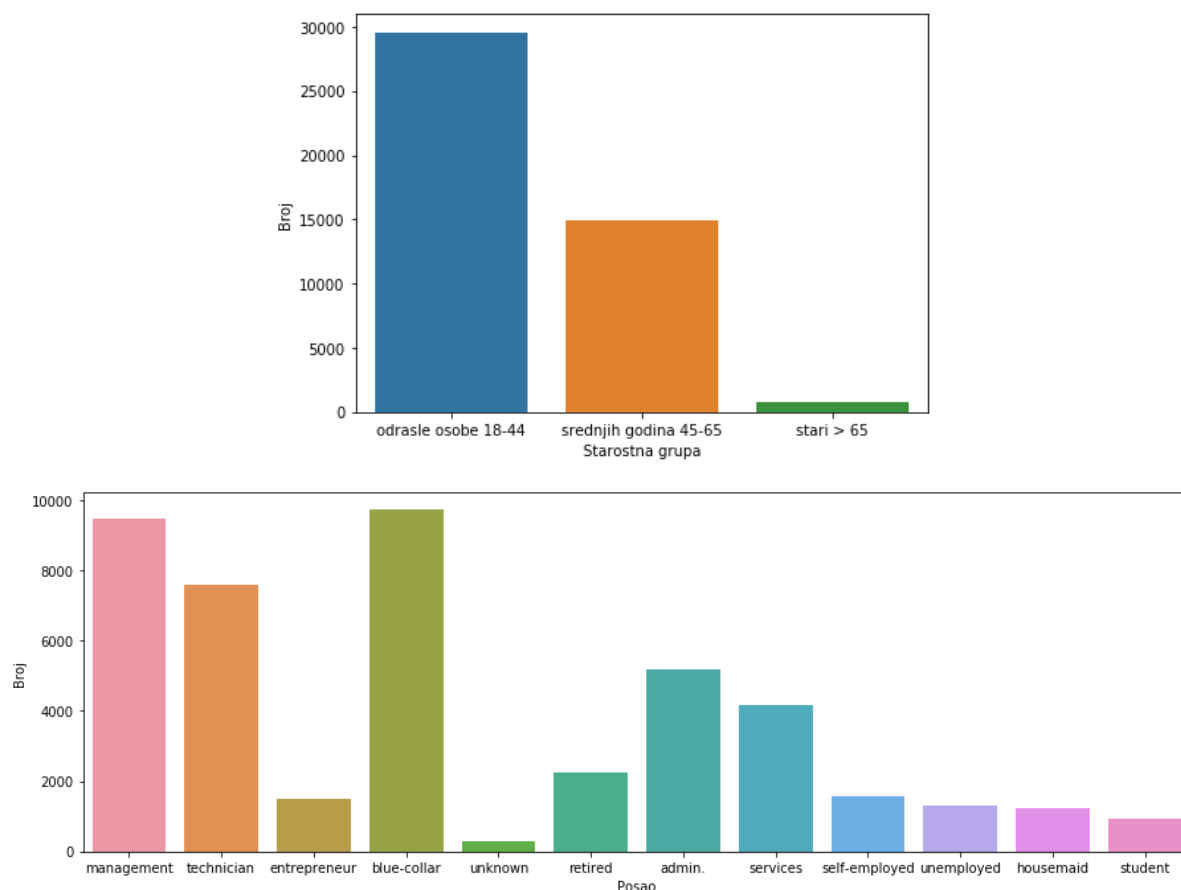
- Prosečan broj dana koji je prošao od poslednjeg kontakta klijenta iz prošle kampanje je 40 sa standardnom devijacijom od 100.
- Minimum je -1 dana (nije uopšte kontaktiran u prošloj kampanji), a maksimum 871 dan.
- 75% klijenata nije uopšte kontaktirano tokom prošle kampanje.

- Prosečna dužina trajanja poziva je 258 sekundi (4,3 minuta) sa standardnom devijacijom od 257 sekundi (4,28 minuta).
- Minimum je 0 sekundi, a maksimum 4918 sekundi (1h 22 minuta).
- 75% klijenata je pričalo 319 sekundi (5,32 minuta) ili manje .

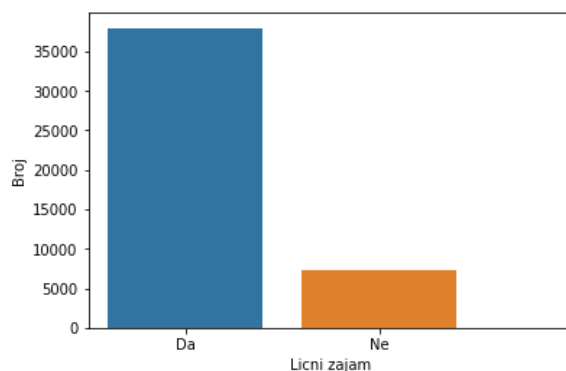
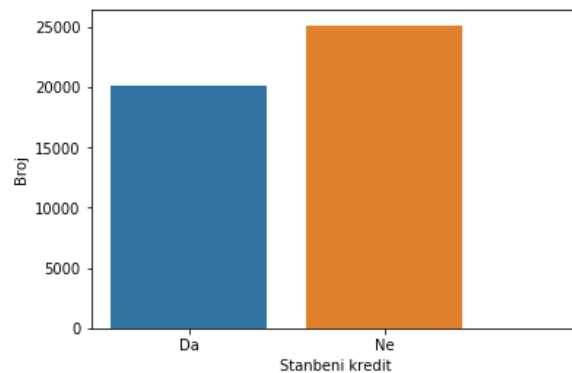
- Prosečan broj kontaktiranja klijenta tokom kampanje je 2.8 puta sa standardnom devijacijom od 3.1.
- Minimalni broj kontaktiranja je 1, a maksimalan je 63.
- 75% klijenata je kontaktirano 3 puta.

Kod atributa kod kojih je standardna devijacija visoka znači da je taj atribut široko rasprostranjen u datasetu.

Analiza klijenata



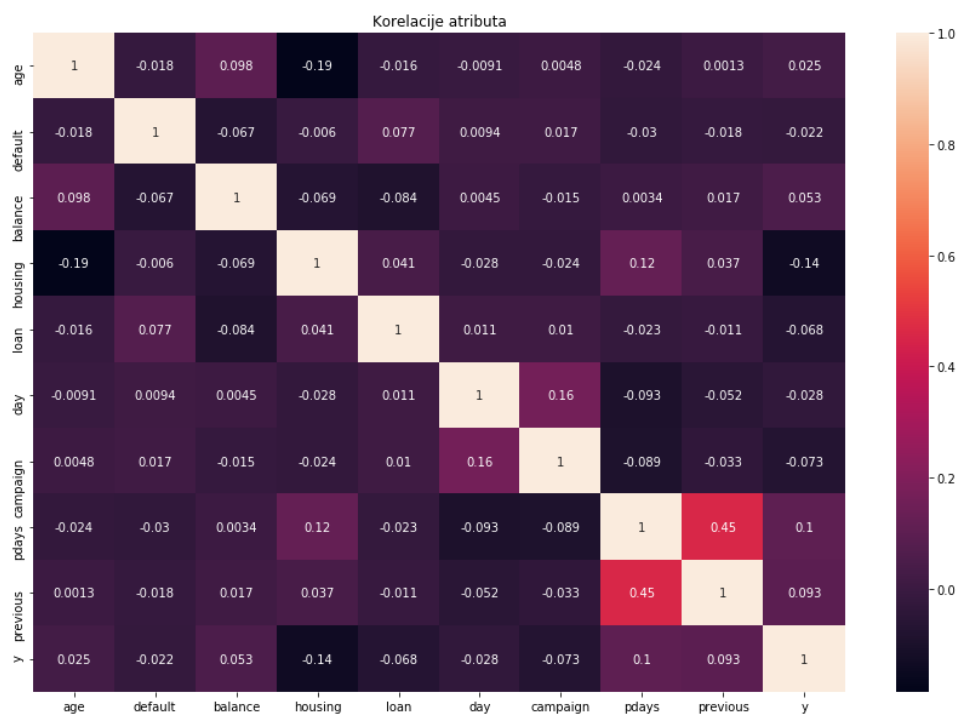
Iz dijagrama sa starosnim grupama možemo zaključiti da imamo najviše klijenata između 18 i 44 godina, a najmanje sa preko 65 godina. Iz dijagrama poslova imamo najviše klijenata iz radničke klase, menadžmenta, i tehničara, a najmanji broj su klijenti koji se bave nama nepoznatim poslom i studenti.



Kreditni dijagrami nam pokazuju da većina klijenata ima neko kreditno zaduženje, da oko 45% klijenata je uzelo stambeni kredit i više od 80% klijenata je uzelo lični zajam.

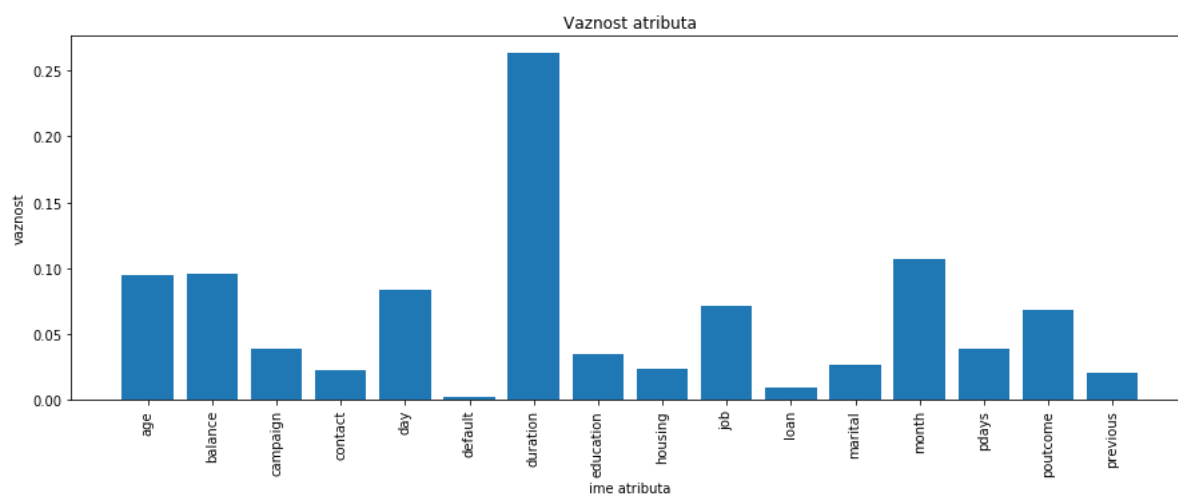
Korelacije atributa

Atributi nisu u korelaciji ni sa jednim drugim atributom u setu.

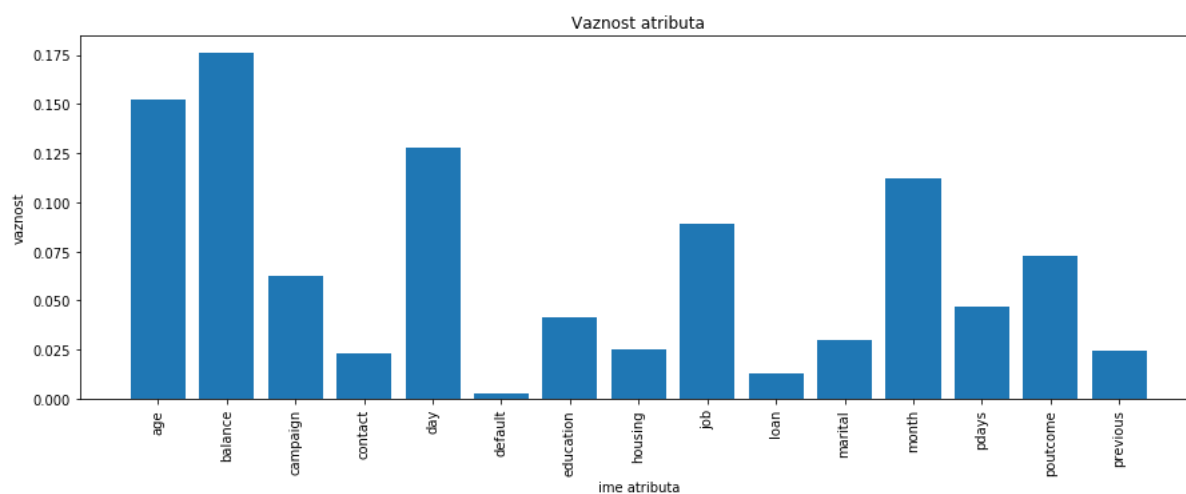


Važnost atributa za predikciju sa Random Forest Klasifikatorom

Da bi model bio realan moramo izbaciti atribut duration jer on najviše utiče na izlaz. U realnom slučaju nikad nećemo znati dužinu poziva dok ga ne obavimo, a kada ga obavimo znamo traženi rezultat y tj. da li se pretplatilo. Tako da duration koristimo kao benchmark za model.

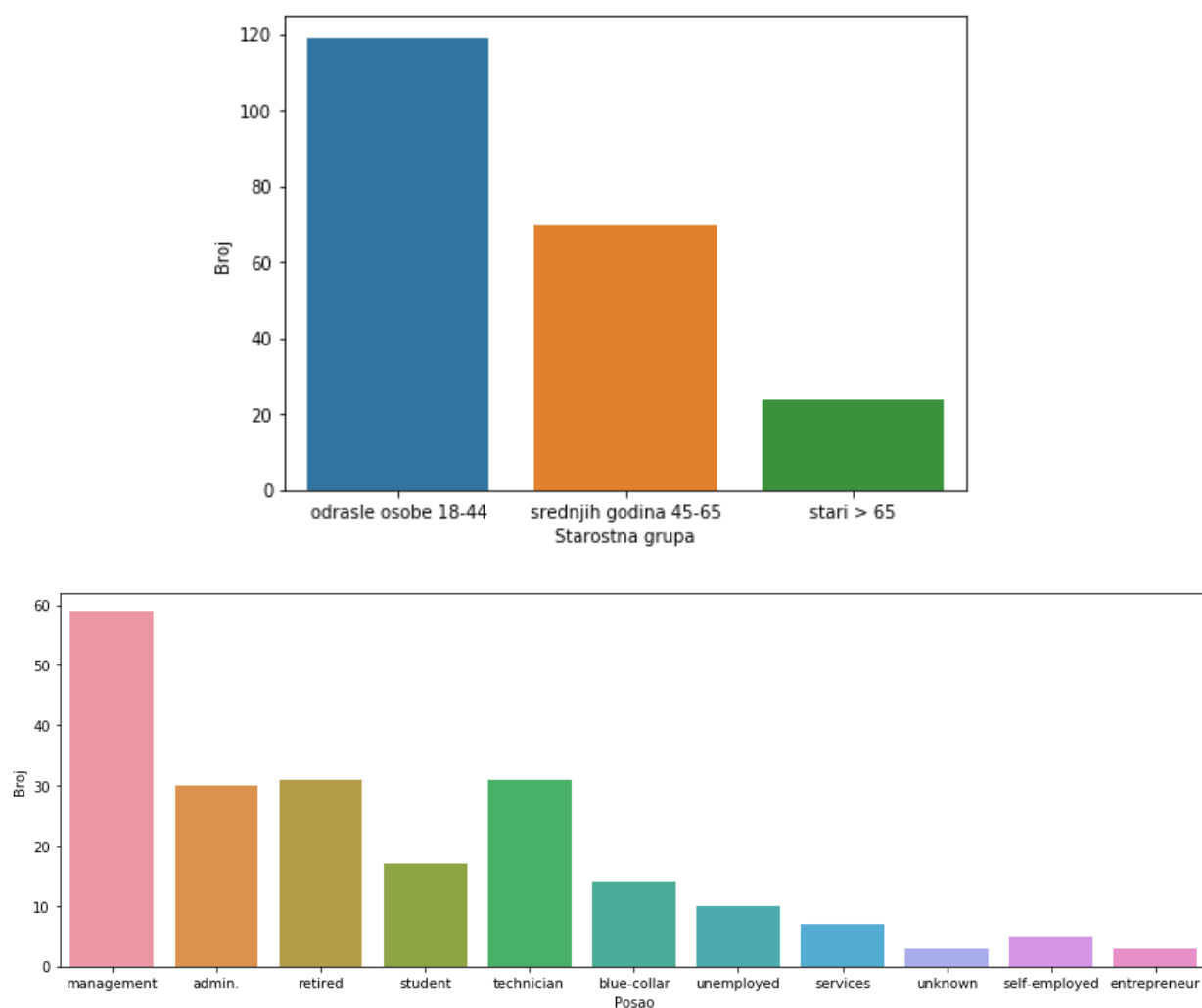


sa duration vs. bez duration atributa

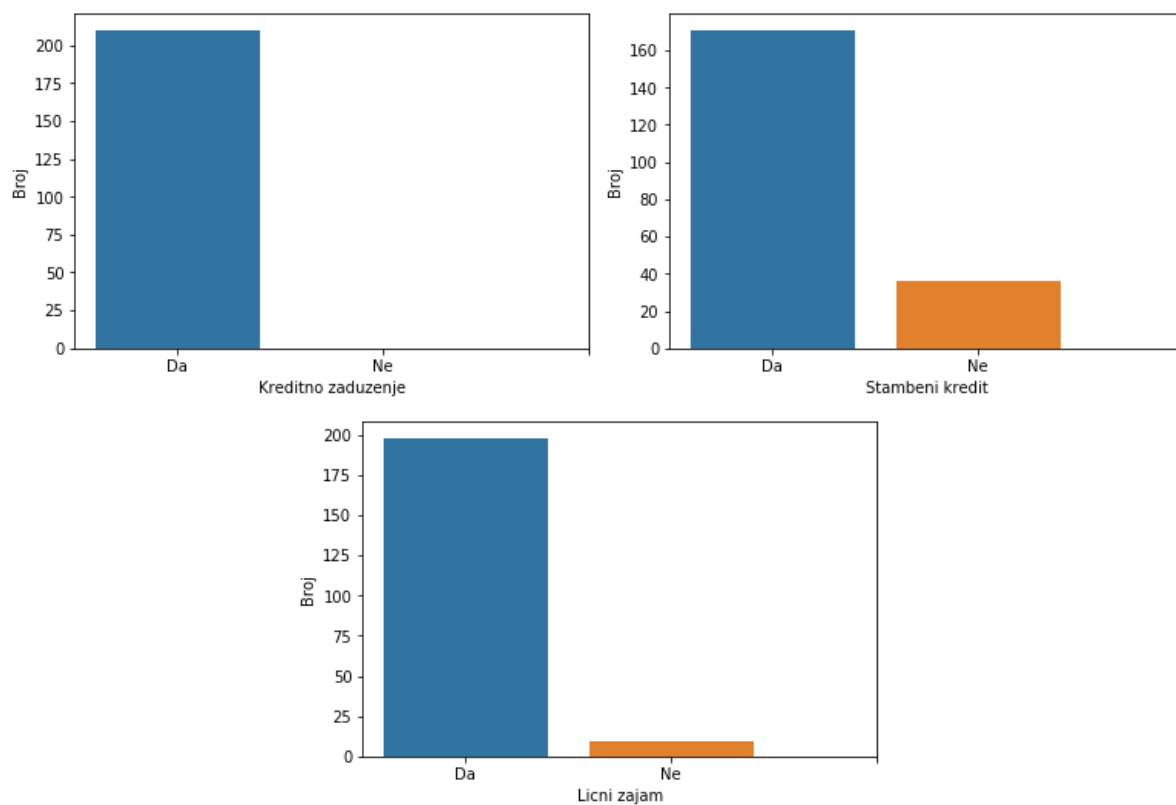


Gledajući drugi dijagram, vidimo da na izlaz najviše utiču godine i godišnji balans na računu, a najmanje neisplaćeno kreditno zaduženje.

Analiza klijenata koji su predviđeni da će se pretplatiti



Dijagram starosnih grupa je sličan dijagramu svih klijenata, gde imamo najviše klijenata između 18 i 44 godina, dok se dijagram poslova pretplaćenih klijenata drastično razlikuje od dijagrama svih klijenata, gde sad imamo najviše klijenata koji se bave menadžmentom, tehničari, penzioneri, i administratori, dok su klijenti iz radničke klase pali na šesto mesto.



Kreditni dijagrami predviđenih pretplaćenih klijenata nam sada govore da svi imaju kreditno zaduženje, 85% ima stambeni kredit na sebi, dok je 90% klijenata je uzelo lični zajam.