Analiza bankovnih marketinških podataka  
Tehnička dokumentacija

Verzija 1.0

Studentski tim: Tina Bakić

Mateo Elez

Josip Prpić

Nastavnik: Damir Pintar  
Mihaela Vranić

Sadržaj

1. Opis razvijenog proizvoda 4

2. Tehničke značajke 5

3. Upute za korištenje 6

4. Literatura 7

Tehnička dokumentacija

***Na koji način koristiti predložak?***

Dokument se po potrebi može prilagoditi potrebama pojedinog projekta promjenom predloženih naslova predloženih poglavlja, kao i eventualnim dodavanjem novih poglavlja i potpoglavlja.

Cilj dokumenta je opisati rezultat rada studentskog tima, problem koji je riješen u okviru projekta, korištenu tehnologiju, mogućnosti i značajke dobivenog proizvoda i sl. Razinu detalja opisanu u ovom dokumentu studentski tim treba dogovoriti s nastavnikom.

***Literatura:***

U tekstu rada treba biti navedena literatura svugdje gdje je tekst, slika ili grafički prikaz preuzet ili se temelji na nekom pisanom predlošku. Literatura se navodi iza zaključka. U tekstu se literatura navodi unutar zagrada s navođenjem prvog autora i godine izdanja, npr. (Martinis, 1998).

***Primjer citiranja knjige:***

Prezime, inicijal(i) imena autora. Naslov: podnaslov. Podatak o izdanju. Mjesto izdavanja: Nakladnik, godina izdavanja.

***Primjer citiranja članka u časopisu:***

Prezime, inicijal(i) imena autora. Naslov članka: podnaslov. Naziv časopisa. Oznaka sveska/godišta, broj(godina), str. početna-završna.

***Primjer citiranja rada sa konferencije:***

Prezime, inicijal(i) imena autora. Naslov rada: podnaslov. Naslov zbornika, mjesto održavanja konferencije, (godina), str. početna-završna.

***Primjer citiranja doktorskog, magistarskog ili diplomskog rada:***

Prezime, inicijal(i) imena autora. Naslov. Vrsta rada. Ustanova na kojoj je rad obranjen, godina.

***Primjer citiranja www izvora:***

Ime(na) autora (ako je/su poznata), naslov dokumenta, datum nastanka (ako se razlikuje od datuma pristupa izvoru), naslov potpunog djela (italic), potpuna http adresa, datum pristupa dokumentu.

***Ostale upute***

U svim dokumentima obvezno primjenjivati SI jedinice. Slike, formule i tablice potrebno je numerirati. Opis tablice stavlja se iznad, a opis slike ispod nje. U opisu slike ili tablice pišu se samo podaci neophodni za njeno razumijevanje (npr. Slika 6. Pojačalo s promjenljivim pojačanjem). Dodatna objašnjenja daju se u tekstu uz povezivanje sa slikom ili tablicom. Osi i parametri na slikama i grafičkim prikazima trebaju biti obilježeni. Daljnji opis tog grafičkog prikaza treba se nalaziti u tekstu rada. Formule se obilježavaju brojevima u zagradi, uz desni rub stranice, a u tekstu se poziva na broj formule.

# Uvod

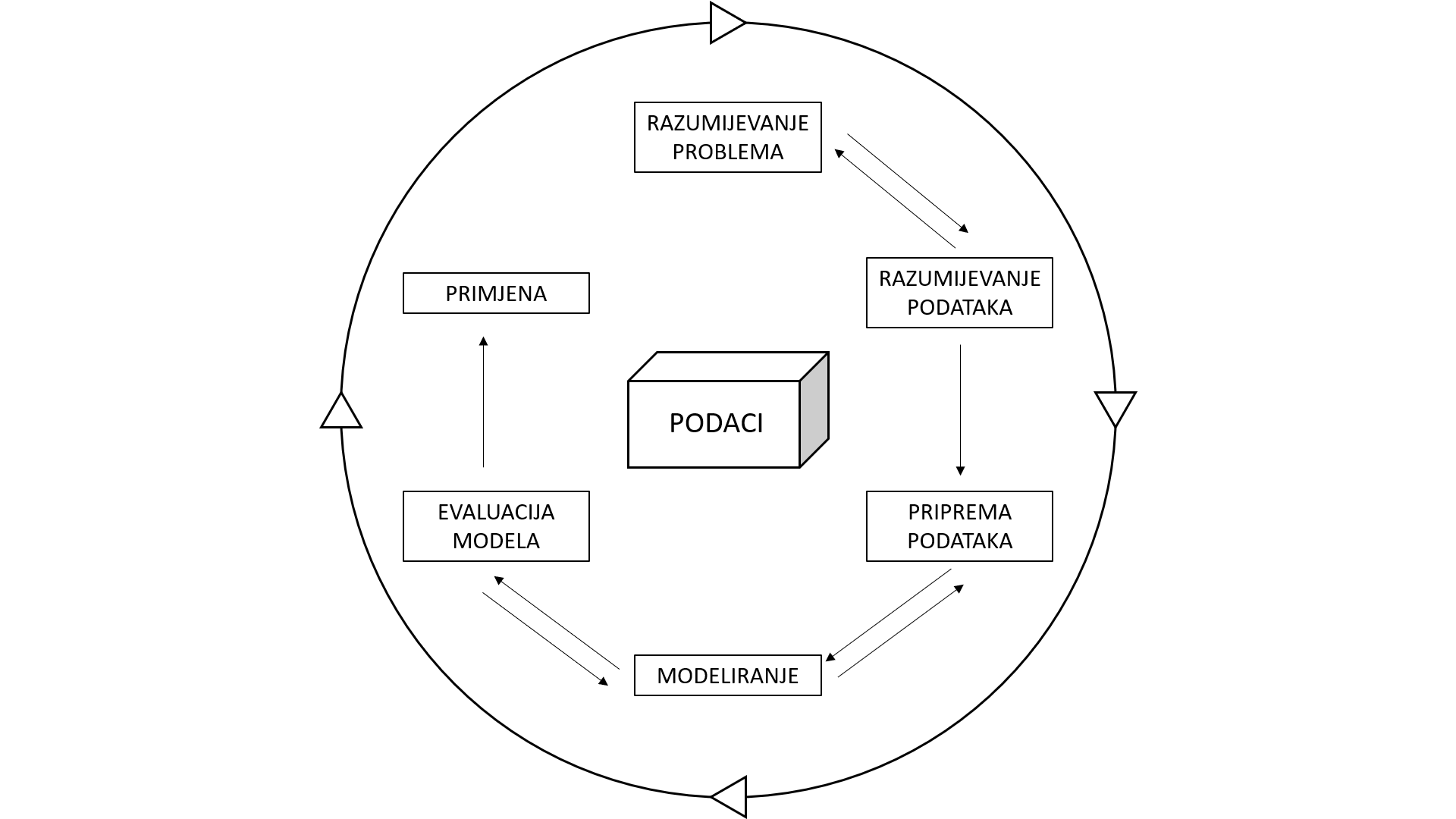
Danas, u svakoj sferi života postoji velika količina podataka. Područje za koje smo se odlučili u sklopu Projekt R je analiza poslovnih podataka. Cilj je bio razviti proizvod pomoću kojega bi ovisno o karakteristikama osobe kao što je dob, obrazovanje, bračni status i sl., bilo moguće odrediti kreditne usluge koje bi pojedina osoba mogla koristiti te ujedno provjeriti što nam još može biti dobar pokazatelj uz naše početne pretpostavke.

Skup podataka korišten u sklopu projekta preuzet je s platforme [www.kaggle.com](http://www.kaggle.com). Radi se o podacima prikupljenim telefonskim anketiranjem za potrebe Portugalske banke. Anketiranje je provedeno u periodu od svibnja 2008. godine do studenog 2011. godine. Cijeli skup podataka te njegove značajke moguće je proučiti na sljedećoj poveznici [Bank Marketing Data Set 🏦 | Kaggle](https://www.kaggle.com/berkayalan/bank-marketing-data-set).

Uz važnost prikupljanja, važnija je kvalitetna obrada i tumačenje podataka jer tek time imamo mogućnost razumijevanja i shvaćanja što nam prikupljeni podaci ustvari označavaju te je upravo to ono što smo se potrudili što bolje razviti u sklopu našeg projekta.

Najveću ulogu u tome imala je dubinska analiza podataka. To je proces otkrivanja globalnih pravilnosti i odnosa u velikim podatkovnim skupovima, u kojima su zbog prevelikog broja i nepreglednosti upravo te značajke skrivene. Za analizu podataka, dubinska analiza podataka služi se metodama iz računarske znanosti, statistike i strojnog učenja.

Slika 1.: prikaz presjeka na kojem se nalazi dubinska analiza podataka



Slika 2.: osnovne faze dubinske analize podataka

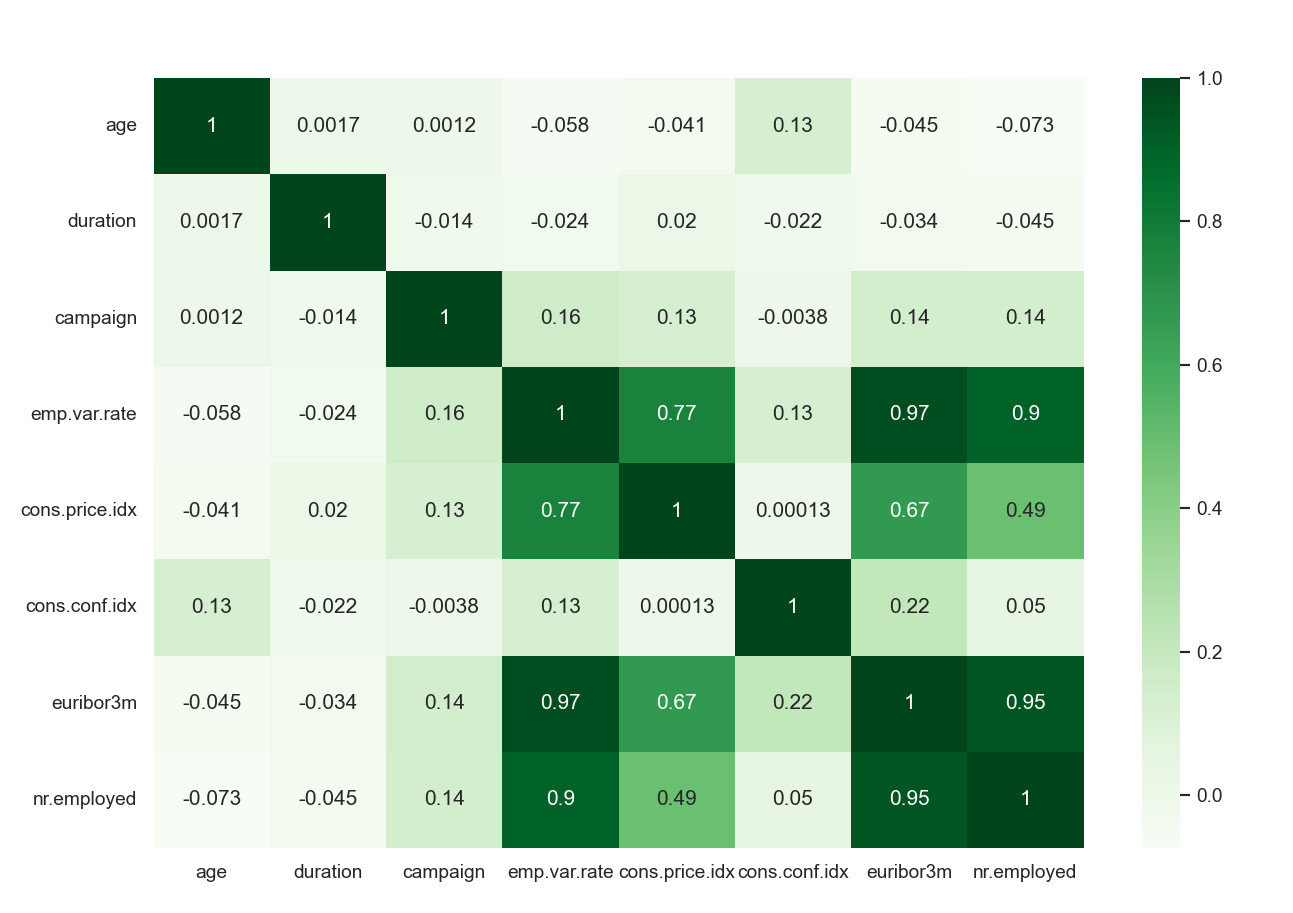
# Eksploratorna analiza i pretprocesiranje

## Pročišćavanje podataka

Početni skup podataka sastojao se od 21 (dvadeset i jednog) atributa te s preko 41000 (četrdeset i jedne tisuće) uzoraka. Nakon ukupne obrade koja će detaljnije biti opisana u nastavku sveli smo podatke na 15 (petnaest) atributa i nešto malo više od 30000 (trideset tisuća) uzoraka, s kojima smo nastavili raditi projekt.

Najprije smo izbacili atribute koji se odnose na ranije provedene ankete, čiji podaci nisu trenutni podaci koje koristimo. Zatim uzorke koji ukazuju na neuspješnost provedbe ankete kao što je trajanje poziva manje od 35 sekundi za koje vrijeme anketa ne može biti ispunjena te ostale nepoznate vrijednosti označili smo i sve uzorke koje sadrže za barem jedan atribut nepoznatu vrijednost izbacili smo iz skupa podataka. Za podatke koji su za bračni status stavili nepoznatu vrijednost postavili smo da su slobodni, za nepoznatu vrijednost zaposlenja postavili smo da je osoba nezaposlena, a za nepoznatu vrijednost kod obrazovanja postavili smo da osoba nije išla u školu.

Nakon uređivanja unesenih podataka, provjerili smo koji su atributi korelirani. Obzirom da bi nas prekorelirani prediktori mogli navesti na krivi zaključak, izračunom korelacije odlučili smo određene atribute izbaciti te ostaviti samo jedan od tih koreliranih.

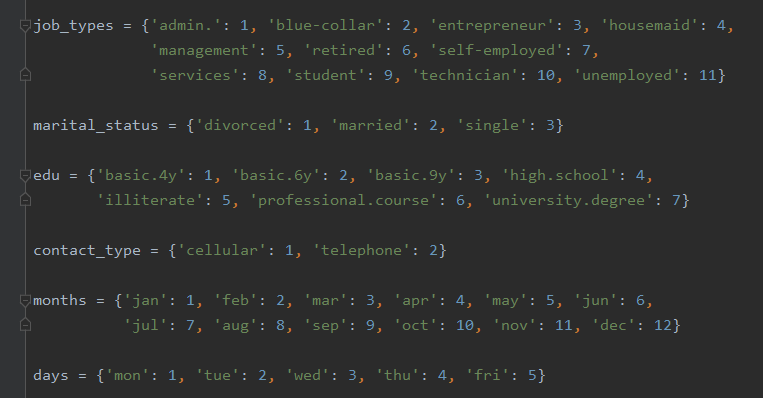


Slika 3.: Matrica koreliraonosti

Po dobivenim rezultatima prikazanim na Slika 3., izbacili smo *emp.var.rate, euribor3m* i *nr.employed* te ostavili *cons.price.idx*.

Za jednostavniji rad s podacima kategorijske podatke kao što je vrsta posla, razina obrazovanja, dan u tjednu i ostale vrijednosti unesene u obliku stringa, prebacili smo u brojčane vrijednosti što konkretno možete vidjeti dijelom koda prikazanom na Slika 4.

Slika 4.: Isječak programskog koda koji prikazuje koje numeričke vrijednosti će zamijeniti koje kategoričke vrijednosti



Istu stvar bilo je potrebno napraviti s binarnim podacima o korištenju kreditnih usluga, koji mogu poprimiti samo vrijednost „yes“ ili vrijednost „no“, stoga smo sve vrijednosti „yes“ pretvorili u numeričku vrijednost 1, a sve „no“ u numeričku vrijednost 0.

## Vizualizacija

## 

# Opis razvijenog proizvoda

# Tehničke značajke

# Upute za korištenje

# Literatura

1. [Filip Ujevic - Postupci i tehnike dubinske analize podataka (uniri.hr)](https://www.inf.uniri.hr/images/studiji/poslijediplomski/kvalifikacijski/Filip_Ujevic_-_Postupci_i_tehnike_dubinske_analize_podataka.pdf)