**## OPIS**

Popis skupova podataka za predmet PUI

**## Sadržaj**

**## Skupovi podataka**

* bikesharing\_daily - bikesharing\_daily.csv: Skup u kojem se predviđa dnevni broj iznajmljenih bicikala (regresijski problem)
  + OPIS: Podaci predstavljaju dnevne podatke o iznajmljivanju bicikala u određenom vremenskom periodu. Cilj analize je izrada modela predikcije broja bicikala koji će biti iznajmljeni na temelju različitih značajki poput temperature, vlažnosti, brzine vjetra, dana u tjednu i vremenskih uvjeta.
  + Problematika za studente informatike: Primjena strojnog učenja (regresijski modeli, neuronske mreže) kako bi se predvidio broj bicikala koji će biti iznajmljeni na temelju vremenskih uvjeta i radnih dana. Identifikacija relevantnih značajki.
  + Problematika za studente ekonomije: Kako vremenski uvjeti i sezona utječu na potražnju za biciklima. Kako se mijenja potražnja za biciklima tijekom radnih dana u usporedbi s vikendima i praznicima. Kako prilagoditi marketinšku strategiju da se poboljša iznajmljivanje tijekom dana slabe potražnje. Može li se optimizirati prihod na temelju dinamičkog određivanja cijena (viša cijena kad je visoka potražnja).
  + Izazovi:
* credit\_data - credit\_data.csv: Skup podataka u kojem se klasificira kreditni rizik korisnika bankarskih usluga
* housing - housing.csv: Skup podataka za predviđanje cijene stana
* traffic - traffic.csv Skup podataka za predviđanje prometa
* online\_news\_pop - online\_news\_pop.csv: Skup podataka za predviđanja dijeljenja vijesti
* flight\_satisfaction - flight\_satisfaction.csv: Skup za predvižanje zadovoljstva putnika letom
* heat\_load - heat\_load.csv: Skup za predviđanje potrebe grijanja -  Npr. za prodaju grijačih jedinica
* cooling\_load - heat\_load.csv: Skup za predviđanje potrebe hlađenja - Npr. za prodaju rasladnih uređaja
* cement - cement.csv. Skup podataka za predviđanje tlačne čvrstoće betona (concrete copressive strength - MPa)
* bankruptcy - bankruptcy.csv: Skup podataka za predviđanje stečaja poduzeća
* steel\_industry\_energy\_consuption - steel\_industry\_energy\_consuption.csv: Podaci prikupljeni iz pametne čeličane za predikciju potrošnje energije u industriji čelika
* water\_potability - water\_potability.csv: Skup za predviđanje je li voda za piće ili ne
* vrtic - vrtic.csv: Skup podataka za rangiranje prijave za određeni vrtić
* cc\_approval\_cleaned - cc\_approval\_cleaned.csv: Skup podataka za predviđanja odobrenja kreditnih kartica
* hitters - hitters.csv: Skup podataka za predviđanja plaći igrača bejzbola

**## Opisi značajki**

**## Značajke bikesharing\_daily**

* instant: indeks zapisa
* dteday: datum
* season: godišnje doba (1: proljeće, 2: ljeto, 3: jesen, 4: zima)
* yr: godina (0: 2011., 1: 2012.)
* mnth: mjesec (1 do 12)
* hr: sat (0 do 23)
* holiday: pokazatelj je li dan blagdan ili ne (izvor: poveznica)
* weekday: dan u tjednu
* workingday: ako dan nije vikend ni blagdan, vrijednost je 1, inače je 0.
* weathersit:
* 1: Vedro, malo oblaka, djelomično oblačno
* 2: Magla + oblačno, magla + djelomično oblačno, magla + malo oblaka, magla
* 3: Lagani snijeg, lagana kiša + grmljavina + raspršeni oblaci, lagana kiša + raspršeni oblaci
* 4: Jaka kiša + ledene kuglice + grmljavina + magla, snijeg + magla
* temp: Normalizirana temperatura u stupnjevima Celzijusa (vrijednosti su podijeljene s 41 – maksimum)
* atemp: Normalizirana osjetna temperatura u stupnjevima Celzijusa (vrijednosti su podijeljene s 50 – maksimum)
* hum: Normalizirana vlažnost zraka (vrijednosti su podijeljene s 100 – maksimum)
* windspeed: Normalizirana brzina vjetra (vrijednosti su podijeljene s 67 – maksimum)
* casual: broj povremenih korisnika
* registered: broj registriranih korisnika
* cnt: Značajka koju predviđamo -  broj bicikala
* ukupni broj iznajmljenih bicikala, uključujući i povremene i registrirane korisnike

**## Značajke credit\_data**

* Age: Starost
* Sex: Spol
* Job: broj poslova
* Housing: Ima svoju nekretninu
* Saving accounts: Visina štednje (kategorična)
* Checking account: Visina iznosa na tekućem računu (kategorična)
* Credit amount: Visina iznosa kredita
* Duration: Rok otplate
* Purpose: Za što se podiže kredit
* Risk: Značajka koju predviđamo - rizični korisnik

**## Značajke housing**

* price: Značajka koju predviđamo -  cijena nekretnine
* area: veličina nekretnine
* bedrooms: broj spavaćih soba
* bathrooms: broj kupaonica
* stories: broj trgovina u blizini
* mainroad: blizina glavne ceste
* guestroom: broj gostinjskih soba
* basement: podrum
* hotwaterheating: grijač tople vode?
* airconditioning: klima?
* parking: parking?
* prefarea: preferirano područje kupca
* furnishingstatus: stanje namještenosti

**## Značajke traffic**

* date\_time
* is\_holiday
* air\_pollution\_index
* humidity
* wind\_speed
* wind\_direction
* visibility\_in\_miles
* dew\_point
* temperature
* rain\_p\_h
* snow\_p\_h
* clouds\_all
* weather\_type
* weather\_description
* traffic\_volume: Značajka koju predviđamo - broj vozila

**## Značajke online\_news\_pop**

* url : URL članka (ne-prediktivno)
* timedelta : Broj dana između objave članka i preuzimanja skupa podataka (ne-prediktivno)
* n\_tokens\_title : Broj riječi u naslovu
* n\_tokens\_content : Broj riječi u sadržaju
* n\_unique\_tokens : Postotak jedinstvenih riječi u sadržaju
* n\_non\_stop\_words : Postotak riječi koje nisu zaustavne (stop words) u sadržaju
* n\_non\_stop\_unique\_tokens : Postotak jedinstvenih riječi koje nisu zaustavne u sadržaju
* num\_hrefs : Broj poveznica (linkova)
* num\_self\_hrefs : Broj poveznica na druge članke objavljene od strane Mashable-a
* num\_imgs : Broj slika
* num\_videos : Broj videozapisa
* average\_token\_length : Prosječna duljina riječi u sadržaju
* num\_keywords : Broj ključnih riječi u metapodacima
* data\_channel\_is\_lifestyle : Je li članak u kategoriji 'Lifestyle'?
* data\_channel\_is\_entertainment : Je li članak u kategoriji 'Zabava' (Entertainment)?
* data\_channel\_is\_bus : Je li članak u kategoriji 'Poslovanje' (Business)?
* data\_channel\_is\_socmed : Je li članak u kategoriji 'Društveni mediji' (Social Media)?
* data\_channel\_is\_tech : Je li članak u kategoriji 'Tehnologija' (Tech)?
* data\_channel\_is\_world : Je li članak u kategoriji 'Svijet' (World)?
* kw\_min\_min : Najlošija ključna riječ (min)
* kw\_max\_min : Najlošija ključna riječ (max)
* kw\_avg\_min : Najlošija ključna riječ (prosjek)
* kw\_min\_max : Najbolja ključna riječ (min)
* kw\_max\_max : Najbolja ključna riječ (max)
* kw\_avg\_max : Najbolja ključna riječ (prosjek)
* kw\_min\_avg : Prosjek
* kw\_max\_avg : Prosjek
* kw\_avg\_avg : Prosjek
* self\_reference\_min\_shares : Min
* self\_reference\_max\_shares : Max
* self\_reference\_avg\_sharess : Prosjek
* weekday\_is\_monday : Je li članak objavljen u ponedjeljak?
* weekday\_is\_tuesday : Je li članak objavljen u utorak?
* weekday\_is\_wednesday : Je li članak objavljen u srijedu?
* weekday\_is\_thursday : Je li članak objavljen u četvrtak?
* weekday\_is\_friday : Je li članak objavljen u petak?
* weekday\_is\_saturday : Je li članak objavljen u subotu?
* weekday\_is\_sunday : Je li članak objavljen u nedjelju?
* is\_weekend : Je li članak objavljen vikendom?
* LDA\_00 : Blizina LDA (Latent Dirichlet Allocation) temi 0
* LDA\_01 : Blizina LDA temi 1
* LDA\_02 : Blizina LDA temi 2
* LDA\_03 : Blizina LDA temi 3
* LDA\_04 : Blizina LDA temi 4
* global\_subjectivity : Subjektivnost teksta
* global\_sentiment\_polarity : Polaritet sentimenta teksta
* global\_rate\_positive\_words : Udio pozitivnih riječi u sadržaju
* global\_rate\_negative\_words : Udio negativnih riječi u sadržaju
* rate\_positive\_words : Udio pozitivnih riječi među ne-neutralnim tokenima
* rate\_negative\_words : Udio negativnih riječi među ne-neutralnim tokenima
* avg\_positive\_polarity : Prosjek
* min\_positive\_polarity : Min
* max\_positive\_polarity : Max
* avg\_negative\_polarity : Prosjek
* min\_negative\_polarity : Min
* max\_negative\_polarity : Max
* title\_subjectivity : Subjektivnost naslova
* title\_sentiment\_polarity : Polaritet naslova
* abs\_title\_subjectivity : Apsolutna razina subjektivnosti
* abs\_title\_sentiment\_polarity : Apsolutna razina polariteta
* shares : Značajka koju predviđamo - Broj dijeljenja članka

**## Značajke flight\_satisfaction**

* Gender: Spol putnika (Žensko, Muško)
* Customer Type: Tip kupca (Vjerni kupac, Nevjerni kupac)
* Age: Stvarna dob putnika
* Type of Travel: Svrha leta putnika (Osobno putovanje, Poslovno putovanje)
* Class: Klasa putovanja u avionu (Poslovna, Ekonomična, Ekonomična Plus)
* Flight distance: Udaljenost putovanja leta
* Inflight wifi service: Razina zadovoljstva WiFi uslugom u letu
* Departure/Arrival time convenient: Razina zadovoljstva vremenom polaska/dolaska
* Ease of Online booking: Razina zadovoljstva online rezervacijom
* Gate location: Razina zadovoljstva lokacijom izlazne kapije
* Food and drink: Razina zadovoljstva hranom i pićem
* Online boarding: Razina zadovoljstva online prijavom na let
* Seat comfort: Razina zadovoljstva udobnošću sjedala
* Inflight entertainment: Razina zadovoljstva zabavom u letu
* On-board service: Razina zadovoljstva uslugom na brodu
* Leg room service: Razina zadovoljstva prostorom za noge
* Baggage handling: Razina zadovoljstva rukovanjem prtljagom
* Check-in service: Razina zadovoljstva uslugom prijave na let
* Inflight service: Razina zadovoljstva uslugom u letu
* Cleanliness: Razina zadovoljstva čistoćom
* Departure Delay in Minutes: Kašnjenje pri polasku u minutama
* Arrival Delay in Minutes: Kašnjenje pri dolasku u minutama
* Satisfaction: Značajka koju predviđamo - Razina zadovoljstva zrakoplovnom kompanijom (Zadovoljstvo, Neutralno, Nezadovoljstvo)

**## Značajke heat\_load**

* Relative Compactness: Relativna kompaktnost stana
* Surface Area: Površina
* Wall Area: Površina zida
* Roof Area: Površina krova
* Overall Height: Ukupna visina
* Orientation:  Orijentacija
* Glazing Area: Površina ostakljenja
* Glazing Area Distribution: Raspodjela površine ostakljenja
* Heating load: Značajka koju predviđamo - Toplinski učinak grijanja

**## Značajke cooling\_load**

* Relative Compactness: Relativna kompaktnost stana
* Surface Area: Površina
* Wall Area: Površina zida
* Roof Area: Površina krova
* Overall Height: Ukupna visina
* Orientation:  Orijentacija
* Glazing Area: Površina ostakljenja
* Glazing Area Distribution: Raspodjela površine ostakljenja
* Cooling load: Značajka koju predviđamo - Toplinski učinak hlađenja

**## Značajke cement**

* Cement: Cement
* Blast Furnace Slag: Troska visoke peći
* Fly Ash: Količina pepela
* Water: Količina vode
* Superplasticizer: Količina superplastifikatora
* Coarse Aggregate: Količina grubih agregata (šoder)
* Fine Aggregate: Količina finih agregata (pijesak)
* Age: starost cementa
* Concrete compressive strength: Značajka koju predviđamo - Tlačna čvrstoća betona

**## Značajke bankruptcy**

* Bankrupt :  Značajka koju predviđamo - bankrot ili ne
* ROA(C) before interest and depreciation before interest: Return On Total Assets(C) : ROA(C) prije kamata i amortizacije prije kamata- Povrat na ukupnu imovinu (C)
* ROA(A) before interest and % after tax: Return On Total Assets(A) : ROA(A) prije kamata i % nakon oporezivanja- Povrat na ukupnu imovinu (A)
* ROA(B) before interest and depreciation after tax: Return On Total Assets(B) : ROA(B) prije kamata i amortizacije nakon oporezivanja- Povrat na ukupnu imovinu (B)
* Operating Gross Margin: Gross Profit/Net Sales : Bruto profitna marža poslovanja- Bruto dobit/Neto prodaja
* Realized Sales Gross Margin: Realized Gross Profit/Net Sales : Ostvarena bruto profitna marža prodaje- Ostvarena bruto dobit/Neto prodaja
* Operating Profit Rate: Operating Income/Net Sales : Stopa operativnog profita- Operativni prihod/Neto prodaja
* Pre-tax net Interest Rate: Pre-Tax Income/Net Sales : Stopa neto interesa prije oporezivanja- Neto prihod prije oporezivanja/Neto prodaja
* After-tax net Interest Rate: Net Income/Net Sales : Stopa neto interesa nakon oporezivanja- Neto prihod/Neto prodaja
* Non-industry income and expenditure/revenue: Net Non-operating Income Ratio : Prihodi i rashodi izvan osnovne djelatnosti- Neto omjer neoperativnog prihoda
* Continuous interest rate (after tax): Net Income-Exclude Disposal Gain or Loss/Net Sales : Kontinuirana kamatna stopa (nakon oporezivanja)- Neto prihod bez dobiti/gubitka od otuđenja/Neto prodaja
* Operating Expense Rate: Operating Expenses/Net Sales : Stopa operativnih troškova- Operativni troškovi/Neto prodaja
* Research and development expense rate: (Research and Development Expenses)/Net Sales : Stopa troškova istraživanja i razvoja- (Troškovi istraživanja i razvoja)/Neto prodaja
* Cash flow rate: Cash Flow from Operating/Current Liabilities : Stopa novčanog toka- Novčani tok iz poslovanja/Tekuće obveze
* Interest-bearing debt interest rate: Interest-bearing Debt/Equity : Stopa kamata na dug s kamatom- Dug s kamatom/Kapital
* Tax rate (A): Effective Tax Rate : Stopa poreza (A)- Efektivna porezna stopa
* Net Value Per Share (B): Book Value Per Share(B) : Knjigovodstvena vrijednost po dionici (B)
* Net Value Per Share (A): Book Value Per Share(A) : Knjigovodstvena vrijednost po dionici (A)
* Net Value Per Share (C): Book Value Per Share(C) : Knjigovodstvena vrijednost po dionici (C)
* Persistent EPS in the Last Four Seasons: EPS-Net Income : Postojani EPS u posljednje četiri sezone- EPS-neto dobit
* Cash Flow Per Share : Novčani tok po dionici
* Revenue Per Share (Yuan ¥): Sales Per Share : Prihod po dionici (¥ juan)- Prodaja po dionici
* Operating Profit Per Share (Yuan ¥): Operating Income Per Share : Operativni profit po dionici (¥ juan)- Operativni prihod po dionici
* Per Share Net profit before tax (Yuan ¥): Pretax Income Per Share : Neto dobit prije poreza po dionici (¥ juan)- Neto prihod prije poreza po dionici
* Realized Sales Gross Profit Growth Rate : Stopa rasta ostvarene bruto dobiti prodaje
* Operating Profit Growth Rate: Operating Income Growth : Stopa rasta operativne dobiti- Rast operativnog prihoda
* After-tax Net Profit Growth Rate: Net Income Growth : Stopa rasta neto dobiti nakon oporezivanja- Rast neto prihoda
* Regular Net Profit Growth Rate: Continuing Operating Income after Tax Growth : Stopa rasta redovite neto dobiti- Rast kontinuiranog operativnog prihoda nakon oporezivanja
* Continuous Net Profit Growth Rate: Net Income-Excluding Disposal Gain or Loss Growth : Stopa rasta kontinuirane neto dobiti- Rast neto prihoda bez dobiti/gubitka od otuđenja
* Total Asset Growth Rate: Total Asset Growth : Stopa rasta ukupne imovine- Rast ukupne imovine
* Net Value Growth Rate: Total Equity Growth : Stopa rasta neto vrijednosti- Rast ukupnog kapitala
* Total Asset Return Growth Rate Ratio: Return on Total Asset Growth : Stopa rasta povrata na ukupnu imovinu
* Cash Reinvestment %: Cash Reinvestment Ratio : Postotak reinvesticije gotovine- Omjer reinvesticije gotovine
* Current Ratio : Trenutni omjer likvidnosti
* Quick Ratio: Acid Test : Brzi omjer likvidnosti- Kiselinski test
* Interest Expense Ratio: Interest Expenses/Total Revenue : Omjer troškova kamata- Troškovi kamata/Ukupni prihod
* Total debt/Total net worth: Total Liability/Equity Ratio : Ukupni dug/Ukupna neto vrijednost- Omjer ukupnih obveza i kapitala
* Debt ratio %: Liability/Total Assets : Omjer duga %- Obveze/Ukupna imovina
* Net worth/Assets: Equity/Total Assets : Kapital/Imovina- Kapital/Ukupna imovina
* Long-term fund suitability ratio (A): (Long-term Liability+Equity)/Fixed Assets : Omjer dugoročnih izvora financiranja (A)- (Dugoročne obveze + kapital)/Stalna imovina
* Borrowing dependency: Cost of Interest-bearing Debt : Ovisnost o posudbama- Trošak duga s kamatom
* Contingent liabilities/Net worth: Contingent Liability/Equity : Uvjetne obveze/Neto vrijednost- Uvjetne obveze/Kapital
* Operating profit/Paid-in capital: Operating Income/Capital : Operativni profit/Uplaćeni kapital- Operativni prihod/Kapital
* Net profit before tax/Paid-in capital: Pretax Income/Capital : Neto profit prije oporezivanja/Uplaćeni kapital- Neto prihod prije poreza/Kapital
* Inventory and accounts receivable/Net value: (Inventory+Accounts Receivables)/Equity : Zalihe i potraživanja/Neto vrijednost- (Zalihe + Potraživanja)/Kapital
* Total Asset Turnover : Obrt imovine
* Accounts Receivable Turnover : Obrt potraživanja
* Average Collection Days: Days Receivable Outstanding : Prosječni dani naplate- Dani nenaplaćenih potraživanja
* Inventory Turnover Rate (times) : Stopa obrta zaliha (puta)
* Fixed Assets Turnover Frequency : Frekvencija obrta stalne imovine
* Net Worth Turnover Rate (times): Equity Turnover : Stopa obrta kapitala (puta)- Obrt kapitala
* Revenue per person: Sales Per Employee : Prihod po zaposleniku- Prodaja po zaposleniku
* Operating profit per person: Operation Income Per Employee : Operativni profit po zaposleniku- Operativni prihod po zaposleniku
* Allocation rate per person: Fixed Assets Per Employee : Omjer alokacije po zaposleniku- Stalna imovina po zaposleniku
* Working Capital to Total Assets : Radni kapital/Ukupna imovina
* Quick Assets/Total Assets : Brza imovina/Ukupna imovina
* Current Assets/Total Assets : Tekuća imovina/Ukupna imovina
* Cash/Total Assets : Novac/Ukupna imovina
* Quick Assets/Current Liability : Brza imovina/Tekuće obveze
* Cash/Current Liability : Novac/Tekuće obveze
* Current Liability to Assets : Tekuće obveze/Imovina
* Operating Funds to Liability : Operativna sredstva/Obveze
* Inventory/Working Capital : Zalihe/Radni kapital
* Inventory/Current Liability : Zalihe/Tekuće obveze
* Current Liabilities/Liability : Tekuće obveze/Obveze
* Working Capital/Equity : Radni kapital/Kapital
* Current Liabilities/Equity : Tekuće obveze/Kapital
* Long-term Liability to Current Assets : Dugoročne obveze/Tekuća imovina
* Retained Earnings to Total Assets : Zadržana dobit/Ukupna imovina
* Total income/Total expense : Ukupni prihodi/Ukupni troškovi
* Total expense/Assets : Ukupni troškovi/Imovina
* Current Asset Turnover Rate: Current Assets to Sales : Stopa obrta tekuće imovine- Tekuća imovina/Prodaja
* Quick Asset Turnover Rate: Quick Assets to Sales : Stopa obrta brze imovine- Brza imovina/Prodaja
* Working capitcal Turnover Rate: Working Capital to Sales : Stopa obrta radnog kapitala- Radni kapital/Prodaja
* Cash Turnover Rate: Cash to Sales : Stopa obrta novca- Novac/Prodaja
* Cash Flow to Sales : Novčani tok/Prodaja
* Fixed Assets to Assets : Stalna imovina/Imovina
* Current Liability to Liability : Tekuće obveze/Obveze
* Current Liability to Equity : Tekuće obveze/Kapital
* Equity to Long-term Liability : Kapital/Dugoročne obveze
* Cash Flow to Total Assets : Novčani tok/Ukupna imovina
* Cash Flow to Liability : Novčani tok/Obveze
* CFO to Assets : Novčani tok od operacija/Imovina
* Cash Flow to Equity : Novčani tok/Kapital
* Current Liability to Current Assets : Tekuće obveze/Tekuća imovina
* Liability-Assets Flag: 1 if Total Liability exceeds Total Assets, 0 otherwise : Zastavica obveze-imovina- 1 ako su ukupne obveze veće od ukupne imovine, 0 inače
* Net Income to Total Assets : Neto prihod/Ukupna imovina
* Total assets to GNP price : Ukupna imovina/GNP cijena
* No-credit Interval : Interval bez kredita
* Gross Profit to Sales : Bruto dobit/Prodaja
* Net Income to Stockholder's Equity : Neto prihod/Vlasnički kapital dioničara
* Liability to Equity : Omjer obveza i kapitala
* Degree of Financial Leverage (DFL) : Stupanj financijske poluge (DFL)
* Interest Coverage Ratio (Interest expense to EBIT) : Pokazatelj pokrića kamata (Trošak kamata prema EBIT-u)
* Net Income Flag: 1 if Net Income is Negative for the last two years, 0 otherwise : Zastavica neto prihoda- 1 ako je neto prihod negativan u posljednje dvije godine, 0 inače
* Equity to Liability : Kapital/Obveze

**## Značajke steel\_industry\_energy\_consuption**

* Date: datum
* Usage\_kWh:  Industrijska potrošnja energije (kontinuirano, kWh)
* Lagging\_Current\_Reactive.Power\_kVarh: Zaostala jalova snaga (kontinuirano, kVarh)
* Leading\_Current\_Reactive\_Power\_kVarh: Prednjačeća jalova snaga (kontinuirano, kVarh)
* CO2(tCO2): Koncentracija CO2 (kontinuirano, ppm)
* Lagging\_Current\_Power\_Factor:
* Leading\_Current\_Power\_Factor
* NSM: Broj sekundi od ponoći (kontinuirano, s)
* WeekStatus:  Kategorična varijabla (Vikend (0) ili Radni dan (1))
* Day\_of\_week: Kategorična varijabla - dan u tjednu
* Load\_Type:  Značajka koju predviđamo - Kategorična varijabla (Malo opterećenje, Srednje opterećenje, Maksimalno opterećenje)

**## Značajke water\_potability**

* ph: kiselost vode
* Hardness: tvrdoća vode
* Solids: količina krutih stvari u vodi
* Chloramines: količina kloramina u vodi
* Sulfate: količina sulfata u vodi
* Conductivity: električna vodljivost vode
* Organic\_carbon: količina organskog ugljik u vodi
* Trihalomethanes: Trihalometani  u vodi
* Turbidity: turbiditet označava zamućenost vode uzrokovanu prisutnošću suspendiranih čestica
* Potability: Značajka koju predviđamo - Pitka ili ne

**## Vrtic**

* parents: kategorije zanimanja roditelja
* has\_nurs: status vrtića
* form: status roditelja
* children: broj dijece roditelja
* housing: uvjeti stanovanja
* finance: materijalno stanje roditelja
* social: socijalno stanje familije
* health: zdrastveno stanje djeteta
* final evaluation: Značajka koju predviđamo - preporuka za upis u vrtic

**## cc\_approval\_cleaned**

* Gender: spol
* Age: starost
* Debt: dug
* Married: bračni status
* BankCustomer: je li korisnik te banke
* EducationLevel: razina obrazovanja
* Ethnicity: etnička pripadnost (kategorična)
* YearsEmployed: Godina zaposlenja
* PriorDefault: je li ikad imao osobni bankrot
* Employed: Zaposlenje
* CreditScore: Kreditna ocjena
* Citizen: je li građanin države ili stranac
* Income: prihodi
* ApprovalStatus: Značajka koju predviđamo - preporuka za kreditnu karticu

**## hitters**

* AtBat: Broj udaraca bejzbolskom palicom tijekom sezone
* hITS: Broj pogodaka u sezoni
* HmRun: Najvrjedniji pogoci u sezoni
* rUNS: Bodovi koje je zaradio za svoju momčad u sezoni
* RBI: Broj igrača koje je udarač trčao kada je pogodio u sezoni
* Walks: Broj pogrešaka koje je napravio protivnički igrač
* Years: Vrijeme igranja igrača u višoj ligi (u godini)
* CAtBat: Broj udaraca bejzbolskom palicom u karijeri
* CHITOVI: Broj pogodaka napravljenih u karijeri
* CHmRun: Najvrjedniji pogoci u karijeri
* CRuns: bodovi koje je zaradio za svoj tim u svojoj karijeri
* CRBI: Broj igrača koje je udarač trčao kada je pogodio u karijeri
* CWwalks: Broj pogrešaka koje je napravio protivnički igrač u karijeri
* League: faktor s razinama A i N koji pokazuje ligu u kojoj je igrač igrao do kraja sezone
* Division: faktor s razinama E i W koji označava poziciju koju je igrač igrao na kraju sezone
* PutOuts: Pomaganje suigraču u igri
* Assists: Broj asistencija koje je igrač ostvario
* Errors: Pogreške igrača u sezoni
* Salary:  Značajka koju predviđamo -Plaća igrača u sezoni
* NewLeague: faktor s razinama A i N koji prikazuje ligu igrača na početku sezone