

# Uvodno poročilo projekta - Modeliranje in vizualizacija časa od oddaje vloge do prejema kredita

Anže Mramor

31. 3. 2022

## Opis projekta

Za projekt pri predmetu Matematika z računalnikom bom obdelal, modeliral in vizualiziral podatke o kreditih. Analiziral bom sledeče čase:

- čas od oddaje vloge do odobritve ("time to yes")
- čas od oddaje vloge do podpisa pogodbe ("time to contract")
- čas od oddaje vloge do prejema sredstev ("time to money")

**DISCLAIMER:** ker gre za občutljive podatke, bodo vsi podatki anonimizirani oziroma popolnoma izmišljeni (simulirani tako, da bodo sicer odsevali realnost), podobnosti z resničnimi podatki pa ne bo, saj so vsi resnični podatki poslovna skrivnost banke.

## Načrt dela

Za začetek bom pridobil anonimizirane podatke. Odvisno od tega kakšna bo njihova struktura (po potrebi) jih bom preoblikoval v *tidydata*. Sledilo bo grupiranje in analiziranje podatkov. V podatkih bom zgoraj že navedene čase (time to yes, time to contract in time to money) poskušal povezati z različnimi atributi posameznih kreditov - za kakšen tip kredita gre (osebni, študentski, avtomobilski, startup, izobraževalni, hipotekarni, investicijski), vrednosti kredita (znesek, merjen v izmišljeni denarni enoti "pandacoin"), v kateri regiji (izmed 2 - vzhodna in zahodna) je bila prošnja obravnavana, ter v kateri poslovalnici posamezne regije (označene z zaporednoštevilko 1-7) in v katerem časovnem obdobju (mesec leta). Poračunal bom različne deleže, povprečne vrednosti, mediane in jih primerjal glede na omenjene lastnosti. Sledila bo vizualizacija, kjer bom vse navedene in poračunane podatke poskušal čim bolj smiselno prikazati, da bom dobil in predstavil celosten vpogled v to, kako se časi spreminjajo glede na različne attribute.

V nadaljevanju se bom ukvarjal s sledečimi vprašanji:

- kako bi postavili sistem data quality (izpolnjeni-neizpolnjeni podatki)? - v praksi tabele pogosto niso tako popolno izpolnjene kot moja s podatki?
- kako bi postavili sistem za manjkajoče podatke?
- kako bi našel osamelce (ne nujno tiste ki so samo po času najkrajši in najdaljši)?
- kako bi dali povratno informacijo za izboljšanje poslovnega procesa?

- kako bi napovedali trend?

V zadnjem delu projekta bom pripravil uporabniški vmesnik (verjetno s paketom *shiny*), s pomočjo katerega bom omogočil sortiranje in prikazovanje vseh poračunanih lastnosti in grafov. Poleg tega bom z metodami strojnega učenja sestavljal napovedni model (ali več njih - odvisno od časa ter količine in konkretnega tipa podatkov), s katerim bom poskušal predvideti trend - predvsem čas od oddaje vloge do prejema sredstev (pa tudi oba vmesna časa). Končni cilj projekta bo sestaviti program, ki bo uporabniku omogočal vnos svojih podatkov v vlogo za odobritev kredita, vrnil pa mu bo pričakovani čas (datum), ko bo vloga odobrena, pričakovan datum podpisa pogodbe in pričakovan datum prejema sredstev na svoj transakcijski račun.