FAQ

**Q: Molio bih Vas da točno specificirate stupce koje će biti potrebno predvidjeti u test setu podataka jer mi to nije baš jasno iz danog opisa (https://www.estudent.hr/wp-content/uploads/2016/03/AA\_RBA\_Brief\_final\_HR.pdf), a nisam bio u prilici ići na radionicu.**

A: Training dataset koji ste dobili je sastavljen iz stupca predstavljenih u Tablica 1. Test dataseti (evaluacijski i validacijski) će biti sastavljeni iz stupaca predstavljenih u Tablica 2, od kojih je PRIJEVREMENI\_RASKID vaš target kojeg je potrebno predvidjeti.

**Q: Koja je razlika između KLIJENT\_ID i OZNAKA\_PARTIJE?**

A: Klijent ID je jedinstvena oznaka klijenta, a oznaka\_partije je jedinstvena oznaka nekog ugovora između Banke i Klijenta. Npr Marko Perić ima KLIJENT\_ID = 123, a njegov poslodavac Tvrtka ABC ima KLIJENT\_ID = 5678. Marko Perić može imati više različitih proizvoda/partija u Banci – npr ima Stambeni kredit sa oznakom partije 1212, ima depozit sa oznakom partije 2323 i Nenamjenski kredit sa oznakom partije 4545. Isto tako i tvrtka u kojoj radi, sa drugim klijent ID ima neki kredit u Banci sa koji opet ima jedinstvenu oznaku partije

**Q: Možete li pojasniti značenja kategoričkih varijabli? (npr. VALUTA je označena brojevima 1-5, VRSTA\_KLIJENTA četveroznamenkastim brojevima, VRSTA\_PROIZVODA...)**

A: Brojevi u koloni valuta su zamijenili oznake valuta npr 1= HRK. Vrste klijenata koje se kriju iza kodova su npr: Large Corporate, Mid Market, Individuals …Vrsta proizvoda je podjela na kredite I depozite.

**Q: Jesu li svi dani iznosi u kunama?**

A: Da

**Q: Kakve podatke, uz kodove, smijemo imati na validaciji? (pretpostavljam sve vanjske podatke o kojima model ovisi?)**

A: Možete koristiti sve vanjske podatke o kojima model ovisi.

**Q: Možemo li na validaciji očekivati potpune i smislene podatke (npr. u testnim se pojavljuju retci u kojima je PLANIRANI\_DATUM\_ZATVARANJA prije DATUM\_OTVARANJA)?**

A: Evaluacijski set i validacijski set imaju istu distribuciju prijevremenih raskida kao train dataset. To je cca. 30% prijevremenih raskida i cca 70% pravovremenih raskida.

**Q: Najraniji datum izvještavanja je 2011., a datum otvaranja 2000-te. Postoje klijenti za koje je datum izvještavanja nakon datuma zatvaranja, te su ova stanja na kvartalu 0, zanima me zašto imaju takve unose kada se nista nije događalo s krediom jer je bio zatvoren?**

A: Podaci koje ste dobili su od 2011 na dalje. Partije koje ste dobili na neki izvještajni datum su partije koje su bile aktivne u toj godini. Zato se može neka partija pojavit čak i nakon zatvaranja sa stanjima 0 – jer njezin rezultat (prihod/rashod) ulazi u godišnji rezultat Banke. Podatke o Prihodu/Rashodu niste dobili – osim posredno preko visine kamatne stope.

**Q: Je li ugovoreni iznos dan u kunama ili u ugovorenoj valuti (u tom slučaju nam treba značenja oznaka valuta)?**

A: Svi iznosi su u kunama

**Q: Isto pitanje kao i prethodno ali za stanja na kvartalima**

A: Svi iznosi su u kunama

**Q: Po čemu se razlikuju vrste kamatne stope?**

A: https://hr.wikipedia.org/wiki/Kamatna\_stopa

**Q: Zanima nas što točno označava stupac "Unnamed: 0.1" Pretpostavljamo da je tu kao nekakav ostatak id stupca neke originalne tablice iz koje su izbačeni neki redovi, ali nismo sigurni.**

A: Pretpostavka je pravilna, Unnamed kolone možete brisati.

**Q: U stupcu STAROST nailazimo na nekakve čudne vrijednosti poput -7994 i 908. Ovih 908 je možda čak nekako i objašnjivo, ali negativne vrijednosti su stvarno čudne. Negativnih vrijednosti ima 3367 što nije malen broj te ne vjerujemo da je to nekakva greška. Možete li nam to malo bolje objasniti?**

A: Za sve klijente koji imaju negativnu starost ili starost >= 900 godina - nemamo pravi podatak o rođenju ili otvaranju firme (u polju datum rođenja je default vrijednost)

**Q: Zanima nas na koji način ste mislili da integriramo vanjski dataset za zadanim, a da dobijemo funkcionalni model za evaluaciju?**

A: Na evaluaciji niste ograniceni sa vanjskim podatcima. Mozete jih integrirati ili upotrebiti posebno. Budite kreativni. Ako bude ovo problem, malo cemo vam pomoci (ali ne jos sada).

**Q: Možemo li dobiti točna značenja iza kodova za VRSTA\_KLIJENTA, PROIZVOD i VRSTA\_PROIZVODA?**

A: Dane su smjernice, mislim da je nepotrebno da dajemo direktan mapping koda sa značenjem.

**Q: Zanima nas i obuhvaća li naš dataset sve moguće kategorije iz stupaca "PROIZVOD", "VRSTA\_PROIZVODA" i "TIP\_KAMATE". Može li se dogoditi situacija da u testnom skupu dobijemo kategoriju koja se nije prije pojavila u ta 3 stupca?**

A: Ne možete dobiti kategoriju koja se već nije pojavila u ta tri stupca.