

Client behavior prediction

Uvodni opis problema:

Banke su oduvijek htjele predvidjeti ponašanje svojih klijenata. Naime, klijenti banaka se vrlo često ponašaju drugačije od ugovorenih uvjeta. Na primjer, znaju povući svoje depozite prije dogovorenog vremena ili prijevremeno otplatiti svoj kredit.

Bankama je u cilju imati kontrolu nad takvim situacijama te pravovremeno reagirati na njih. To zapravo znači da žele predvidjeti hoće li klijent prekršiti ugovorne uvjete ili ne.

U sklopu studentskog natjecanja Mozgalo rješavamo ovaj problem.

Skup podataka:

Ime stupca	Tip podataka	Opis
DATUM_IZVJESTAVANJA	datetime	Datum vremenske serije
KLIJENT_ID	numeric	Id
OZNAKA_PARTIJE	numeric	Partija
DATUM_OTVARANJA	datetime	Datum otvaranja
PLANIRANI_DATUM_ZATVARANJA	datetime	Planirani datum zatvaranja
DATUM_ZATVARANJA	datetime	Stvarni datum zatvaranja
UGOVORENI_IZNOS	numeric	Originalni iznos
STANJE_NA_KRAJU_PRETH_KVARTALA	numeric	Preostali iznos na kraju prethodnog kvartala
STANJE_NA_KRAJU_KVARTALA	numeric	Preostali iznos na kraju kvartala
VALUTA	numeric	Valuta
VRSTA_KLIJENTA	numeric	Klijentski segment
PROIZVOD	categorical	Produkt
VRSTA_PROIZVODA	categorical	Vrsta produkta
VISINA_KAMATE	numeric	Iznos visine kamate (postotak)
TIP_KAMATE	categorical	Vrsta kamatne stope
AGE	numeric	Starost klijenta
PRIJEVREMENI_RASKID	categorical	Da/Ne (Y/N)

Ovo su značajke koje imamo u trening skupu. U evaluacijskom skupu ćemo imati sve ove značajke osim STANJE_NA_KRAJU_PRETH_KVARTALA, STANJE_NA_KRAJU_KVARTALA i DATUM_ZATVARANJA.

Naš je cilj predvidjeti klasifikaciju kroz značajku PRIJEVREMENI_RASKID

Potiče se korištenje nekih “vanjskih” značajki koje bi mogle pomoći u klasifikaciji. Neke od tih značajki su naprimjer prosječna plaća u tom razdoblju, inflacija, BDP...

Cilj i hipoteze istraživanja problema:

Temeljem danih, povijesnih i vanjskih podataka (opći ekonomski pokazatelji) potrebno je pokušati predvidjeti klijente koji će potencijalno promijeniti ugovoreni odnos s bankom u smislu prijevremene otplate kredita ili produženja oročenog depozita.

Dodatno je kao opcija moguće vizualizirati utjecaj vanjskih faktora na ponašanje klijenata.

Materijali, metodologija i plan istraživanja:

Trening skup smo dobili od kreatora ovog problema. Dobit ćemo i evaluacijski skup nad kojim ćemo evaluirati naše rješenje.

Cilj nam je napraviti što bolji *feature engineering* te naposljetku isprobavati razne modele te izabrati onaj koji nam daje najbolji rezultat. Uz to sve, posebno ćemo paziti na interpretabilnost cijelog modela.

Očekivani rezultati predloženog projekta:

Očekujemo dobiti vrlo precizan model (nadamo se preciznosti od barem 80%) te očekujemo da naš model bude interpretabilan i stavljamo naglasak na to.

Želimo da naš model (ili barem neka naša ideja) bude upotrebljiv u bankarskom sektoru.

Literatura:

- Kolegij Strojno učenje vježbe i predavanja – PMF, Tomislav Lipić i Matija Piškorec

<https://web.math.pmf.unizg.hr/nastava/su/materijali/>

- Trevor Hastie, Robert Tibshirani, Jerome Friedman: “The elements of statistical learning”, kolovoz 2008.

- Coursera, Machine learning – Stanford University