



ADACTA



Client Behavior Prediction: A Machine Learning Challenge

OPIS NAGRADNOG ZADATKA

Temeljem 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. Također je moguće vizualizirati utjecaj vanjskih faktora na ponašanje klijenata. Temeljem dobivenih podataka potrebno je pripremiti nove personalizirane ponude klijentima.

Banka je također primijetila ovisnost ponašanja prema različitim faktorima, internim i eksternim - kao npr. demografija (starost klijenata, broj djece itd), makroekonomski podaci (snaga ekonomije, stope zaposlenosti), a i naravno visina kamatnih stopa na tržištu.

Banka nastoji aktivno upravljati rizicima proizašlim iz takvog ponašanja te prediktivno kreirati moguće scenarije kroz određeni vremenski horizont. Cilj je prilagoditi ponudu predviđenim novonastalim uvjetima te na taj način umanjiti negativne efekte na svoj financijski rezultat.

ŠTO JE TOČNO NAGRADNI ZADATAK

Temeljem 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. Temeljem dobivenih podataka potrebno je pripremiti nove personalizirane ponude klijentima.

EDUKATIVNA KOMPONENTA

Natjecatelji imaju priliku naučiti kako se u bankarskom sektoru koriste nove metode prediktivne analitike kako bi se poboljšala usluga klijentima, ali i smanjila varijabilnost samog poslovanja te smanjio sveukupni rizik poslovanja.

Natjecatelji imaju priliku naučiti i dobiti iskustva u strojnom učenju na realnim podacima.



ADACTA



POTENCIJALNE PRIMJENE

Oba opisana primjera u uvodnom dijelu mogu imati snažan utjecaj na financijski rezultat banke i poslovnog segmenta u kojem se nalaze definirani klijenti.

POTENCIJALNI PROBLEMI KOJI SE MOGU POJAVITI PRILIKOM RJEŠAVANJA NAGRADNOG ZADATKA

Adacta će ažurirati dokument s FAQ odgovorima uz pomoć kojih će se rješavati tehnički problemi ako oni nastanu tijekom rješavanja zadatka. Navedene podatke RBA već koristi u simulacijama za predviđanje zatvaranja kredita tako da će učenje iz tih podataka biti izvedivo. Trenutno predviđamo, da bi mogli problemi da se pojave na samom početku zbog specifičnosti podataka i zbog potrebe za pred procesiranje.

Specifičnost podataka: Banka prati svaki ugovor kroz kalendarsku godinu, pa i nakon što je zatvoren, zato što kumulirani rezultat i po tom ugovoru ulazi u ukupan godišnji rezultat. Iz tog razloga ista partija se više puta pojavljuje sa oznakom da je prijevremeno zatvorena - u principu do kraja godine.

Mogući problemi u pretprocesiranju: Problemi se mogu pojaviti u fazi pretprocesiranja zbog velike količine informacija i zbog "missing values". To ćemo rješavati analizom strukture podataka i metodama za pretprocesiranje. Ako se ukaže da je količina podataka ipak prevelika za analizu, uklonit ćemo dio.

PREPORUČENI PROGRAMSKI JEZICI I/ILI ALATI ZA RJEŠAVANJE NAGRADNOG ZADATKA

Preporučujemo Qlik i Python, ali konačan odabir je na samim natjecateljima.

KORISNI ALATI:

- Jupyter Notebook (python)
- Scikit-learn (python)
- Bokeh (vizualizacija)
- Orange (python)
- Weka (java)

PRIMJER SKUPA PODATAKA KOJE ĆE NATJECATELJI DOBITI S NAZNAKOM KOLIČINE PODATAKA

Lista partija transakcijskih računa pravnih osoba i/ili stambenih kredita fizičkih osoba u nekoliko vremenskih serija. Podaci će biti pohranjeni u csv datotekama.

- Količina podataka neće biti veća od 1GB
- Broj stupaca je 17 (16 parametra i 1 kategorija)

**ADACTA****Raiffeisen
BANK****OPIS ULAZNIH VARIJABLI KOJE DOBIVAJU NATJECATELJI**

Podaci o partijama su sljedeći:

Ime stupca	Tip podatka	Opis
DATUM_IZVJESTAVANJA	datetime	Datum vremenske serije
KLIJENT_ID	numeric	ID
OZNAKA_PARTIJE	numeric	Partija
DATUM_OTVARANJA	datetime	Datum otvaranja
PLANIRANI_DATUM_ZATVARANJA	datetime	Planiran datum zatvaranja
DATUM_ZATVARANJA	datetime	Stvarni datum zatvaranja
UGOVORENI_IZNOS	numeric	Originalni iznos
STANJE_NA_KRAJU_PRETH_KVARTALA	numeric	Preostali iznos kredita na kraju prethodnog kvartala
STANJE_NA_KRAJU_KVARTALA	numeric	Preostali iznos kredita 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)

MAKROEKONOMSKI POKAZATELJI ZA PRAVNE OSOBE:

BDP, prosječni uvoz, Foreign Direct Investment – utjecaj na tekući račun poduzeća



ADACTA



MAKROEKONOMSKI POKAZATELJI ZA FIZIČKE OSOBE:

Stopa nezaposlenosti, Prosječna plaća, inflacija, BDP – utjecaj na retail & stambene kredite

IMA LI NARUČITELJ ISKUSTVA U RJEŠAVANJU ISTOG ILI SLIČNOG ZADATKA?

RBA kontinuirano provodi slične simulacije, a Adacta ima dugogodišnje iskustvo u analizi podataka. Na dnevnoj razini oni se suočavaju sa sličnim problemima koji zahtijevaju analizu i profiliranje klijenata, analizu interakcija između klijenata i tvrtke, treniranje machine-learning modela kako bi se predvidjele buduće interakcije ili odgovori klijenata i izgradnja integriranih rješenja u formi *recommendation systems* ili *decision support systems*.

DODATNE INFORMACIJE I LITERATURA

- <https://www.hnb.hr/statistika>
- <https://www.dzs.hr/>
- <https://www.hnb.hr/statistika/sdds>
- <https://www.hnb.hr/statistika/statisticki-podaci>
- <https://www.hnb.hr/statistika/glavni-makroekonomski-indikatori>
- <https://www.hnb.hr/statistika/informacije-za-korisnike-statistickih-podataka>
- Elements of Statistical Learning - Jerome H. Friedman, Robert Tibshirani, and Trevor Hastie
- Python: <https://www.dataquest.io/blog/sci-kit-learn-tutorial/>
- Java: <https://www.cs.waikato.ac.nz/ml/weka/book.html>
- Coursera - Machine learning by Andrew Ng
- Machine Learning - Mitchell
- Pattern Recognition, 2nd ed - Duda, Hart & Stork
- Macroeconomics, 3rd ed - O. Blanchard
- Macroeconomics-Study Guide - D.W. Findlay
- Modern econometrics - R. L. Thomas
- Time Series Analysis - J. D. Hamilton
- Reducing Customer Attrition with Predictive Analytics for Financial Institutions, Nate Derby, Mark Keintz, MWSUG 2016, Paper BI04
- Maximizing Cross-Sell Opportunities with Predictive Analytics for Financial Institutions, Nate Derby
- Using SAS® GTL to Visualize Your Data When There is Too Much of It to Visualize, Perry Watts, Nate Derby
- <http://stakana.com/>



mozgalo

ADACTA



Raiffeisen
BANK

OBAVEZNE STAVKE I FORMATI RJEŠENJA

Korisnička i tehnička dokumentacija (pripremljen template), prezentacija, izvještaj o postignutim rezultatima na dnevnoj razini (barem jednom), izvještaj o postignutim validacijskim rezultatima (finalna ocjena točnosti), kodovi.

KRITERIJI BODOVANJA

PREZENTACIJA

Prezentacija bi trebala sadržavati slijedeću strukturu koja će biti prezentirana natjecateljima na radionici:

- Motivacija
- Definicija problema i analiza dataseta
- Izabrana metoda
- Rezultati
- Zaključak

DOKUMENTACIJA – PDF DOKUMENT

Dokumentacija bi trebala slijediti dani predložak.

SOFTVER - SW

Softver treba biti strukturiran i pregledan te dokumentiran na način da ga i nezavisni promatrač koji posjeduje tehnička znanja može pratiti i čitati.

OCJENA TOČNOSTI TRENIRANOG MODELA (EVALUACIJA)

Dnevno rangiranje:

Adacta će osigurati web servis putem kojeg će natjecatelji moći uploadati rezultate svojih modela, te će jednom dnevno vidjeti trenutačni raspored timova uzimajući u obzir preciznost rezultata treniranog modela. Natjecatelji će morati barem jednom evaluirati svoj model da bi pridobili ocjenu za evaluaciju.



ADACTA



OCJENA TOČNOSTI TRENIRANOG MODELA (VALIDACIJA)

Rangiranje na licu mjesta:

Natjecatelji će također na ad-hoc upit organizatora prilikom prezentacije rješenja testirati preciznost svog modela nad validacijskim setom podataka koji će Adacta pripremiti. Rezultat ad-hoc validacije također će ući u strukturu ukupne ocjene.

KVALITETA RJEŠENJA

Ocijenit će se kvaliteta i izvedivost rješenja.

INOVATIVNOST RJEŠENJA

Inovativnost rješenja ocjenjivat će se kroz dvije komponente:

- U slučaju da rezultati natjecatelja daju novi uvid u podatke koji se može upotrijebiti za unapređenje trenutnog procesa koji banka koristi
- Upotreba inovativnih metoda

Kriterij	Ocjena	Doprinos ukupnoj ocjeni
Prezentacija	0-10	10%
Dokumentacija	0-10	10%
SW – kvaliteta rješenja	0-35	35%
Inovativnost rješenja	0-10	10%
Ocjena točnosti treniranog modela (evaluacija)	0-15	15%
Ocjena točnosti treniranog modela (validacija)	0-20	20%

DODATNE POGODNOSTI

Adacta i RBA pružaju mogućnost odrađivanja prakse ili studentskog rada.