

Telco Customer Churn Classification – Projekat

Opis problema

Pretpostavlja se rešavanje klasifikacionog zadatka u kojem će biti **predviđeno da li će korisnik telekomunikacione usluge raskinuti ugovor** (*churn*) ili **ostati** (*no churn*), na osnovu demografskih, uslužnih i fakturačkih karakteristika. Dataset sadrži podatke o klijentima telekom usluga, a model bi trebalo da omogući identifikaciju najuticajnijih faktora koji doprinose odluci o prekidu usluge. Naglasak je stavljen na analizu, vizualizaciju i interpretaciju, koristeći isključivo *telco_data.csv*.

Pregled kolona i njihovo značenje

Dataset (*telco_data.csv*) obuhvata sledeće kolone:

- **customerID** — jedinstveni identifikator svakog korisnika.
- **gender** — pol korisnika (Male / Female).
- **SeniorCitizen** — status starog korisnika (1 = da, 0 = ne).
- **Partner** — status da li korisnik ima partnera (Yes / No).
- **Dependents** — da li korisnik ima izdržavane članove domaćinstva (Yes / No).
- **tenure** — trajanje ugovora u mesecima (koliko dugo je korisnik ostao sa kompanijom).
- **PhoneService** — dostupnost telefonske usluge (Yes / No).
- **MultipleLines** — da li korisnik koristi više linija (Yes / No / No phone service).
- **InternetService** — tip internet usluge (DSL / Fiber optic / No).
- **OnlineSecurity** — dostupnost usluge online zaštite (Yes / No / No internet service).
- **OnlineBackup** — dostupnost usluge rezervne kopije (Yes / No / No internet service).
- **DeviceProtection** — zaštita uređaja (Yes / No / No internet service).

- **TechSupport** — dostupnost tehničke podrške (Yes / No / No internet service).
- **StreamingTV** — dostupnost usluge za strimovanje TV sadržaja (Yes / No / No internet service).
- **StreamingMovies** — dostupnost strimovanja filmova (Yes / No / No internet service).
- **Contract** — vrsta ugovora (Month-to-month / One year / Two year).
- **PaperlessBilling** — bezpapirno fakturisanje (Yes / No).
- **PaymentMethod** — način plaćanja (Electronic check / Mailed check / Bank transfer (automatic) / Credit card (automatic)).
- **MonthlyCharges** — mesečni trošak usluge (u dolarima).
- **TotalCharges** — ukupni trošak do sada (u dolarima).
- **Churn** — **ciljna promenljiva** (Yes = churn, No = no churn) .

Očekivani izlazi (output):

- **Evaluacija klasifikacionog modela** predstavljena kroz ključne metrike: *tačnost, preciznost, odziv i F1-skor*.
- **Grafička analiza** najuticajnijih atributa (feature importance) na odluku modela.
- **Vizualni prikazi** distribucija i odnosa između ključnih atributa i ciljne promenljive—na primer, *churn* po tipu ugovora, mesečnoj naplati, trajanju korisničkog odnosa.
- **Interpretacija rezultata** sa naglaskom na faktore koji najviše doprinose churnu (npr. tip internet usluge, trajanje, metoda plaćanja), uz preporuke za potencijalna unapređenja modela.

Zahtevi za procesiranje podataka:

- **Nedostajući ili nekonsistentni podaci** (npr. u TotalCharges) moraju biti tretirani tako da se minimizira njihov negativni uticaj na model.
- **Kategorijske promenljive** (npr. gender, SeniorCitizen, Partner, Dependents, PhoneService, InternetService, Contract, PaymentMethod) treba da budu prikladno

pripremljene za model, uz obrazloženje odabranog pristupa.

- **Brojčani atributi** (tenure, MonthlyCharges, TotalCharges) treba da budu jasno obrađeni, uz opravdanje njihovog uključivanja i eventualnih transformacija.

Dokumentovanje rezultata:

- **Rezultati modela i vizuelni elementi** (grafici, tabele) treba da budu jasno prikazani u formatu pogodnom za Word dokument.
- **Interpretacije i zaključci** moraju biti logično i analitički utemeljeni, uz preporuke za moguća poboljšanja i dodatne pokušaje.