

Razumjevanje biznis problema

U svijetu finansija, svaka priča koja počinje sa "beskamatni krediti" odmah privlači pažnju. Predstavljam vam ne samo jedan takav neobičan scenario, već priliku da ga detaljno istražite i analizirate. Moje ime je Nekamatar Nekamatić i vodim banku koja se odlučila na hrabar eksperiment – nudimo beskamatne kredite!

Ova inovacija nije proizvod mašte, već temeljito promišljanje o tome kako promijeniti finansijske usluge i istinski pomoći našim klijentima. Kako bismo osigurali uspjeh ovog projekta, kreirali smo dataset koji sadrži podatke o klijentima koji su se prijavili za naše beskamatne kredite. Vaš zadatak? Koristiti ovaj dataset da razvijete model koji predviđa ko bi trebao biti odobren za kredit.

Dataset obuhvata različite demografske i finansijske karakteristike aplikanta, kao što su starost, pol, zanimanje, obrazovanje, bračni status, godišnji prihod, i kreditni skor. Vaša misija je da primijenite logističku regresiju i identifikujete ključne faktore koji utječu na odluku o odobravanju kredita.

Ovaj pristup nam omogućava da bolje razumijemo naše klijente i kreiramo finansijske proizvode koji odgovaraju njihovim potrebama. Vi, kao vrhunski inženjeri mašinskog učenja, dobićete priliku da radite sa stvarnim podacima, rješavate stvarne probleme i naučite kako vaše analitičke vještine mogu donijeti promjene u svijetu finansija.

Pomoću vaših analiza, možemo definisati budućnost bankarstva, gdje su krediti pristupačni i pravedni, bez dodatnog tereta kamata. Vaša analiza i preporuke biće ključne u oblikovanju naših politika i ponuda. Radujem se vašim inovativnim rješenjima i energiji koju ćete unijeti u ovaj projekt.

Sretno!

Zadatak 1: Istraživanje dataset-a

Analizirati osnovne informacije o dataset-u, uključujući vrste podataka, broj redova i kolona, te prikazati prvih nekoliko redova za bolje razumijevanje sadržaja dataset-a. Potrebno je napisati kratak komentar o strukturi i potencijalnim karakteristikama ili izazovima u dataset-u koji bi mogli utjecati na daljnju analizu.

Zadatak 2: Vizualizacija podataka

Za numeričke varijable kreirati histogram dijagrama kako bi se vizualizirala distribucija svake numeričke varijable. Za kategoričke varijable, potrebno je kreirati countplot dijagrame. Ovaj zadatak zahtijeva analizu i komentarisanje rezultata vizualizacija, uključujući uočene raspodele, izražene tendencije ili anomalije u podacima.

Zadatak 3: Statistička analiza numeričkih kolona

Provjeriti statističke karakteristike numeričkih kolona, uključujući srednje vrijednosti, standardnu devijaciju, minimalne i maksimalne vrijednosti, i kvartilne vrijednosti. Nakon toga, analizirati i komentirati rezultate te identificirati potencijalne izvore varijabilnosti ili stabilnosti unutar podataka.

Zadatak 4: Priprema podataka za model

Izvršiti pripremu podataka za analitički model. Ovo uključuje primjenu One-hot encodinga i ILbel encodinga. Također, provesti normalizaciju ili standardizaciju numeričkih podataka kako bi se osigurala ista skala vrijednosti za numeričke podatke.

Zadatak 5: Primjena logističke regresije

Primijeniti logističku regresiju na pripremljene podatke i evaluirati rezultate modela. Ovo uključuje ispitivanje preciznosti, matrice konfuzije, te drugih relevantnih metrika performansi. Analizirati rezultate i identificirati ključne pokazatelje koji utječu na ishod modela.