**Predlog projekta**

**Predmet:** Numerički algoritmi i numerički sotver

**Student:** Mrđa Miloš, sv34-2022

**Naziv teme:**

Predikcija cene automobila i pravljenje najboljeg modela koristeći različite algoritme.

**Problem:**

Generalno govoreći, ljudima je teško izabrati koji auto kupiti. Ovaj projekat bi pomogao pri odabiru. Napravio bih 4 model koristeći sledeće algoritme: linearnu regresiju, KNN(K-nearest neighbors), Random forest i Decision trees. Prvo bih učitao i obradio podatke, napravio pomoću svakog algoritma model i na kraju uporedio i obrazložio koji je model najbolji.

**Rešavanje problema:**

1. **Zavisna promenljiva**

Cena automobila.

1. **Nezavisne promenljive**

Kompanija, model auta, godina proizvodnje, pređena kilometraža, tip goriva, konjska snaga, broj brzina, boja enterijera, boja eksterijera, broj cilindara, podržava airbag sistem, dešavanje incidenata sa automobilom na osnovu izveštaja. (Nisam još našao tačan dataset, ali bi ovo bili atributi sa sitnom izmenom).

1. **Skup podataka**

Skup podataka bi bio preuzet sa interneta. Sa nekih od poznatijih sajtova kao sto su kagle.com, dataworld/datasets/regression.com, polovniAutomobili.com. Imao bih dva skupa podatak train.csv i test.csv. Kako bih mogao da treniram model na podacima iz prvog skupa i da ga generalizujem što bolje nad nepoznatim podacima.

1. **Treniranje modela**

U ovom koraku bih vršio pregled i obradu podataka, što uključuje ispitivanje osnovnih karakteristika koje moraju biti zadovoljene(LINE + nepostojanje kolinearnosti, jer radimo za višestrukom linearnom regresijom). Kao mera za model koristiće se prilagodjeni r^2. Takođe korišćenje interpolacije i izbacivanje kolona za koje se pretpostavi da nisu potrebne za ovaj problem, popunjavanje nedostajućih vrednosti.

1. **Implementacija više modela**

Kada sam uspešno istrenirao model, uz pomoć biblioeteke scikit-learn napravio bih četiri različita modela, uz gore četiri navedena algoritma.

1. **Evaluacija modela**

Kada smo dobili odgovarajuće modele, možemo ih jos unaprediti korišćenjem RFE-a(Recursive feature elimination). To je tehnika koja se koristi sa selekciju karakteristika istreniranog modela, kako bi ga još optimizovao.

Za svaki model bih evaluirao performansu koristeći navedenu metriku prilagodjeni r^2. Ovo bi bilo upotrebljeno nad podacima iz test.csv. Na kraju programa bile bi ispisane sve informacije o modelima, koji je model najbolji i zasto(Opciono da korisnik vidi). Korisnik bi imao mogucnost da pretrazi koje automobile moze da si priušti na osnovu njegovog budžeta i želje marke i modela automobila.

**Tehnologije:**

Programski jezik: python

Biblioteke: statsmodel, pandas, matplotlib, seaborn, sklearn

**Materijali koji bi bili korisni za izradu:**

<https://www.polovniautomobili.com>

<https://archive.ics.uci.edu/datasets>

<https://www.kaggle.com>

<https://archive.ics.uci.edu>