

Klasifikacija lokacija u klimatske zone

Projekat iz predmeta *Računarska inteligencija*

Nikola Savić SV21/2020

Opis projekta

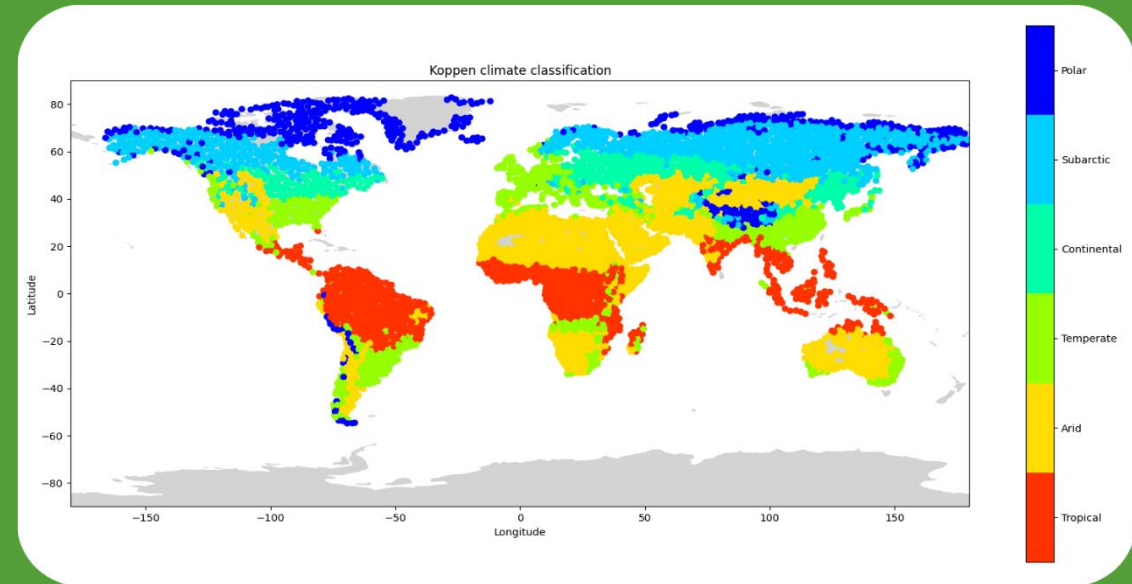
Cilj projekta je klasifikacija lokacija na planeti Zemlji u jednu od šest postojećih klimatskih zona: polarna, subpolarna, kontinentalna, umerena, pustinjska i tropska.

Podaci korišćeni za projekat su dobijeni korišćenjem dva API-ja:

- Za datu lokaciju dobijaju se parametri poput temperature, vlažnosti, pritiska itd. za svaki sat za poslednjih sedam dana – korišćeno za sve lokacije
- Za datu lokaciju dobija se kojoj klimatskoj zoni pripada – korišćeno za trening skup lokacija (veličine oko 90%)

Algoritmi

- **K-means algoritam** – lokacije koje su ulaz za API su dobijene tako što je generisan određen broj slučajnih lokacija unutar nekog mnogougla koji predstavlja granice kontinenta. Kako bi skup slučajnih lokacija bio ravnomernije raspoređen na njega je primenjen K-means algoritam klasterovanja tako što centri klastera predstavljaju izlazne lokacije.
- **Random forest algoritam** – na osnovu trening skupa lokacija, njihovih vremenskih prilika i klimatske zone, za test skup lokacija sa vremenskim prilikama dobija se predikcija klimatske zone kojoj lokacija pripada. Izvršena je standardizacija uzorka. Zbog same strukture *Random forest* algoritma validacioni skup nije korišćen. Hiperparametri algoritma koji su korišćeni radi postizanja boljih rezultata su navedeni u delu *Rezultati*.



Rezultati

Metrike korišćene za analizu rezultata:

- **Tačnost** – postignuto ≈85%
- **Preciznost** – postignuto ≈85% u proseku
- **Odziv (recall)** – postignuto ≈80% u proseku

Korišćeni hiperparametri:

- **n_estimators** = 400, broj stabala
- **min_samples_split** = 2 i **min_samples_leaf** = 1 koji definišu stepen razgranatosti stabala
- **class_weight** – proporcionalan uzorku