**Analiza podataka i primena algoritama mašinskog učenja na podacima o proizvodnji solarne energije**

**Opis problema**

Problem koji se rešava jeste analiza podataka o proizvodnji solarne energije na području Čačka u Srbiji. Cilj je identifikovati ključne faktore koji utiču na proizvodnju solarne energije i kreirati model koji će moći da predviđa proizvodnju za određene vremenske intervale, počevši od satnog intervala. Ovo rešenje omogućava optimizaciju proizvodnje i planiranje potrošnje energije, čime se doprinosi efikasnijem korišćenju solarne energije u pametnim gradovima.

Razlog za odabir ovog problema leži u značaju solarne energije kao obnovljivog izvora i njenoj ulozi u održivom razvoju i smanjenju emisije štetnih gasova.

**Skup podataka**

Podaci koji će se koristiti su preuzeti sa platforme Kaggle, a odnose se na solarne elektrane na području Čačka. Skup podataka sadrži sledeće karakteristike:

* AirTemperature (°C): Temperatura vazduha, ključni faktor koji utiče na efikasnost solarnih panela.
* CloudOpacity (%): Količina oblačnosti izražena u procentima.
* DHI (kW/m²): Difuzno horizontalno zračenje koje dolazi od celokupnog neba.
* DNI (kW/m²): Direktno zračenje koje dolazi direktno od sunca.
* EBH (kW/m²): Ekstraterestrijalno horizontalno zračenje (bez atmosfere).
* GHI (kW/m²): Globalno horizontalno zračenje, uključuje direktne i difuzne komponente.
* Proizvodnja – Lokacija 1 (kWh): Proizvodnja električne energije na lokaciji 1.
* Proizvodnja – Lokacija 2 (kWh): Proizvodnja električne energije na lokaciji 2.
* Proizvodnja – Lokacija 3 (kWh): Proizvodnja električne energije na lokaciji 3.

Podaci su preuzeti sa sledeće adrese: [Kaggle - Serbia Solar Energy Production](https://www.kaggle.com/datasets/mexwell/serbia-solar-energy-production).

**Algoritmi**

Za rešavanje problema koristiće se sledeći algoritmi:

* Random Forest Regressor: Algoritam zasnovan na drveću odlučivanja, koristi se za predikciju kontinuiranih vrednosti.
* Linear Regression: Algoritam za regresiju koji modelira linearne odnose između nezavisnih i zavisnih varijabli.

Kroz testiranje, analiziraju se performanse ovih algoritama kako bi se odabrao najpogodniji za predviđanje proizvodnje solarne energije.

**Tehnologije**

Projekat će koristiti sledeće tehnologije:

* Python: Programski jezik za analizu i obradu podataka.
* Pandas: Biblioteka za manipulaciju i analizu podataka.
* Scikit-learn: Biblioteka za implementaciju algoritama mašinskog učenja.
* Matplotlib: Biblioteka za vizualizaciju podataka.
* Flask: Okvir za kreiranje API-ja i backend-a.
* React: JavaScript biblioteka za frontend razvoj i vizualizaciju podataka.

Razvojni alat biće Visual Studio Code.

**Cilj**

Cilj projekta je da se analiziraju podaci i identifikuju ključni faktori koji utiču na proizvodnju solarne energije. Nakon analize, potrebno je razviti model koji predviđa proizvodnju energije za satni interval na osnovu ulaznih podataka. Implementacija uključuje:

1. Prikaz grafičke analize podataka.
2. Kreiranje modela mašinskog učenja za predikciju proizvodnje.
3. Razvoj aplikacije sa korisničkim interfejsom koji omogućava unos parametara i prikaz predikcija.

Projekat započinje sa satnim predikcijama, a nakon uspešne implementacije razmotriće se proširenje na dnevne, mesečne i godišnje intervale.