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

**Opis problema**

Projekat se fokusira na analizu podataka o proizvodnji solarne energije na području Čačka u Srbiji, sa ciljem identifikacije ključnih faktora koji utiču na proizvodnju solarne energije. Glavni cilj je kreirati modele koji mogu precizno predviđati proizvodnju energije na različitim vremenskim intervalima: satnom, dnevnom i mesečnom.

Ovo rešenje omogućava:

* Optimizaciju proizvodnje energije.
* Bolje planiranje potrošnje energije.
* Efikasnije korišćenje solarne energije u pametnim gradovima.

Razlog za odabir problema proizilazi iz značaja solarne energije kao obnovljivog izvora, njene uloge u smanjenju emisije štetnih gasova i postizanju održivog razvoja.

**Skup podataka**

Opis podataka:

Podaci korišćeni u projektu preuzeti su sa platforme [AI4EU](https://www.ai4europe.eu/research/ai-catalog/solar-energy-production-dataset). Skup podataka sadrži sledeće karakteristike:

* AirTemperature (°C): Temperatura vazduha koja utiče na efikasnost solarnih panela.
* CloudOpacity (%): Količina oblačnosti na nebu izražena u procentima.
* DHI (kW/m²): Difuzno horizontalno zračenje.
* DNI (kW/m²): Direktno normalno zračenje koje dolazi direktno od sunca.
* EBH (kW/m²): Ekstraterestrijalno horizontalno zračenje bez atmosferskog uticaja.
* GHI (kW/m²): Globalno horizontalno zračenje koje 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.

Izvori podataka: [AI4EU - Solar Energy Production Dataset](https://www.ai4europe.eu/research/ai-catalog/solar-energy-production-dataset).

**Algoritmi i modeli**

Za rešavanje problema korišćeni su sledeći algoritmi mašinskog učenja:

1. Random Forest Regressor:
   * Ensemble metoda sa više stabala odluka za predviđanje kontinuiranih vrednosti.
   * Robustna je i može modelovati nelinearne odnose.
2. Pipeline sa StandardScaler:
   * Normalizacija podataka za efikasno treniranje modela.

Evaluacija modela uključuje:

* MAE (Mean Absolute Error): Prosečna apsolutna greška.
* MSE (Mean Squared Error): Prosečna kvadratna greška.
* R² Score: Koeficijent determinacije.

**Tehnologije i alati**

Projekat koristi sledeće tehnologije:

* Python: Glavni programski jezik.
* Pandas: Manipulacija i analiza podataka.
* NumPy: Numeričke operacije.
* Scikit-learn: Implementacija algoritama mašinskog učenja.
* Plotly: Vizualizacija podataka.
* Flask: Backend framework.
* HTML/CSS/Bootstrap: Frontend za aplikaciju.
* Jinja2: Templating engine.
* Visual Studio Code: Razvojno okruženje.

**Implementacija**

1. Preprocesiranje podataka
   * Uklanjanje nedostajućih vrednosti.
   * Grupisanje podataka po vremenskim intervalima: satnom, dnevnom i mesečnom.
2. Razvoj modela
   * Trenirani su zasebni modeli za svaku lokaciju i ukupnu proizvodnju.
   * Upoređena je ukupna proizvodnja dobijena sabiranjem predikcija i iz posebnog modela.
3. Evaluacija i vizualizacija
   * Prikaz grafova za predikcije i uticaj ulaznih parametara na proizvodnju.
4. Razvoj korisničkog interfejsa
   * Dinamički frontend sa grafičkim prikazima rezultata.
   * Responzivni dizajn za prikaz podataka na svim uređajima.

**Ciljevi projekta**

Primarni ciljevi:

* Analiza podataka i identifikacija ključnih faktora.
* Razvoj modela za predikciju proizvodnje energije.
* Implementacija aplikacije sa korisničkim interfejsom.

Sekundarni ciljevi:

* Vizualizacija uticaja ulaznih parametara kroz interaktivne grafove.
* Unapređenje korisničkog iskustva kroz moderni dizajn aplikacije.

**Rezultati**

Na osnovu evaluacije različitih modela i intervala:

1. Satna predikcija:
   * Sabiranje predikcija po lokacijama pruža tačnije rezultate.
2. Dnevna predikcija:
   * Poseban model za ukupnu proizvodnju daje bolje performanse.
3. Mesečna predikcija:
   * Sabiranje predikcija pokazuje bolje rezultate.

Vizualizacija uticaja parametara je omogućila uvid u značaj pojedinačnih karakteristika poput GHI, DNI i CloudOpacity.

**Zaključak**

Projekat pruža alat za analizu i predikciju solarne energije sa sledećim benefitima:

* Bolje razumevanje faktora koji utiču na proizvodnju.
* Optimizacija energetskog sistema.
* Vizualizacija rezultata i intuitivno korisničko iskustvo.

Budući pravci razvoja:

* Proširenje na druge geografske lokacije.
* Integracija sa podacima u realnom vremenu.
* Primena naprednijih algoritama kao što su duboke neuronske mreže.

**Reference**

1. [AI4EU - Solar Energy Production Dataset](https://www.ai4europe.eu/research/ai-catalog/solar-energy-production-dataset)
2. Scikit-learn Documentation
3. Flask Documentation
4. Plotly Documentation