

Predlog Projekta

Naziv Teme:

Analiza i optimizacija sistema iznajmljivanja bicikala: Višekontekstualni pristup sa fokusom na korelaciju između gradova.

Definicija Problema:

Sistemi iznajmljivanja bicikala u urbanim sredinama suočavaju se sa izazovima kao što su sezonske fluktuacije, promene u potražnji tokom dana i različiti vremenski uslovi. Ovaj projekat će istražiti da li se modeli predikcije iz jednog grada, kao što je Seul, mogu primeniti na druge gradove sa sličnim servisima, na primer London, kao i mogućnost razvoja sveobuhvatnog modela koji integriše podatke iz više gradova radi boljeg predviđanja.

Cilj Projekta:

Razviti prediktivne modele koji odgovaraju na sledeća pitanja:

- Koji faktori najviše utiču na potražnju za biciklima u različitim gradovima?
- Da li je moguće koristiti modele jednog grada za predikciju potražnje u drugim gradovima?
- Kako se potražnja za biciklima menja u zavisnosti od vremenskih uslova, sezone i dana u nedelji?
- Može li integrisani model, baziran na podacima iz više gradova, dati bolje rezultate od pojedinačnih modela?

Metodologija:

1. Prikupljanje podataka:

Prvi korak u realizaciji projekta je prikupljanje podataka o sistemima iznajmljivanja bicikala. Ovi podaci uključuju informacije o broju iznajmljenih bicikala, vremenskim uslovima (temperatura, vlažnost, brzina vetra), sezoni, radnim i neradnim danima, kao i praznicima. Podaci će biti prikupljeni iz više gradova, kao što su Seul i London, korišćenjem javno dostupnih setova podataka, čiji linkovi se nalaze na dnu predloga projekta.

2. Pretprocesiranje podataka:

U ovoj fazi, podaci će biti pripremljeni za analizu i modeliranje.

To uključuje:

- Standardizaciju numeričkih podataka, kako bi svi podaci bili u istom formatu.
- Enkodiranje kategorijalnih podataka, kao što su dan u nedelji ili tip vremenskih uslova.
- Obrada nedostajućih podataka, uz zamenu ili eliminaciju nepotpunih vrednosti kako bi se očuvala tačnost analize.

3. **Analiza podataka (EDA):**

Istraživačka analiza podataka (EDA) biće ključna faza u razumevanju podataka i identifikovanju obrazaca. U ovoj fazi ćemo proučiti osnovne karakteristike podataka, kao što su distribucija broja iznajmljenih bicikala, korelacije između vremenskih uslova i potražnje, kao i sezonske varijacije. Takođe, biće izvršena komparativna analiza između različitih gradova, kako bi se utvrdile sličnosti i razlike u obrascima potražnje za biciklima.

4. **Modeliranje i evaluacija:**

Na osnovu dobijenih uvida iz EDA, razvijaćemo prediktivne modele za analizu potražnje za biciklima. Biće korišćeni različiti modeli regresije, kao što su Linearna regresija, XGBoost i Random Forest, kako bi se predvidela potražnja u budućnosti. Takođe, razvijaćemo integrisani model koji koristi podatke iz više gradova kako bi se testiralo da li se modeli iz jednog grada mogu primeniti na druge. U fazi evaluacije, modeli će biti ocenjeni korišćenjem metrika kao što su RMSE (Root Mean Squared Error) i adj. R^2 (prilagođeni koeficijent determinacije).

Način Evaluacije:

Rezultati modela će se evaluirati pomoću sledećih metrika:

- RMSE (Root Mean Squared Error);
- adj. R^2 (prilagodjeni koeficijent determinacije).
- Komparacija performansi između pojedinačnih i integrisanih modela.

Tehnologije:

- Programski jezici: Python.
- Biblioteke: Pandas, NumPy, Scikit-learn, XGBoost, Matplotlib, Seaborn.

Primeri i Literatura:

Primeri modela koji se bave istom tematikom:

- Kaggle bike-sharing datasets: <https://www.kaggle.com/competitions/bike-sharing-demand/code>
- Projekat iste tematike: <https://github.com/sinhabishal77/Kaggle-Bike-Sharing-Demand>

Literatura:

- Seoul bike sharing demand:

<https://archive.ics.uci.edu/dataset/560/seoul+bike+sharing+demand>

- London Cycle Hire Data: <https://www.kaggle.com/datasets/hmavrodiy/london-bike-sharing-dataset>