The impact of women's driving on “WOSUL” program

Abstract

The purpose of this project is to build a classification model that helps to looking for” The **impact of women's driving on “WOSUL” program”**. I worked with a human open resource dataset provided by https://data.gov.sa/ "Saudi government".

First, I started exploring this project goal, using Matplotlib and Seaborn tools to visualize the data .then , I used **a Random Forest model** with one feature/target “*status*” to describe impact of women's driving on “WOSUL” program as a function of the status of “WOSUL” program .

Design

This project helps to figure out witch feature has a direct and clear impact on **impact of women's driving on “WOSUL” program based on status.** and predict whether **the number of women's reduce as the women's start driving. Finally,** Our Target users is Women's in Saudi Arabia.

Data

The dataset contains **123871 data point** and **8 fields.** As the link

The data is from **https://data.gov.sa/** open data source. And Our data sample is women in Saudi Arabia that can get “WOSUL” program after women driving.

Algorithm

Our project is **category classification,** so we choose a *Decian Tree* and *Random Forest algorithm.*

Model Evaluation and Selection:
the entire dataset of **123871 record** was split into **70/30 Train vs. Test**.

Below the evaluation of each model:

Decian Tree

- Accuracy: `0.788 (0.036)`

Random Forest

- Accuracy: `0.903 (0.023)`

Tolls

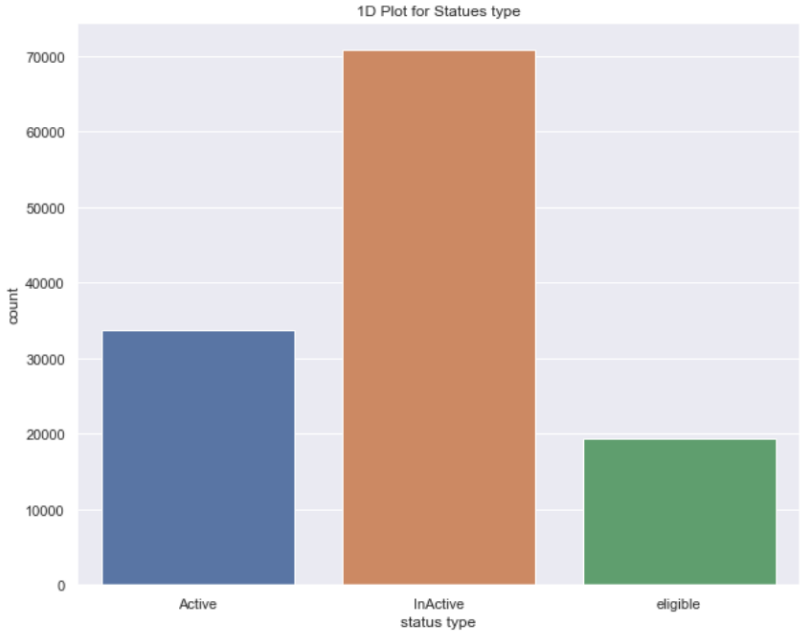
I used:

- **jupyter environment .**
- **python programming language** with : (*Numpy, Pandas, MATPLOTLIB, Seaborn, Datetime, RandomForestClassification, DecisionTreeClassifier*) **libraries**

Communication

Presentation that includes visuals for communicating the objectives and findings

According to the plot below , we can see that number of “**InActive**” women’s statues is **70000 out of 123871**. So we can it’s the higher number.



According to the plot , we can see that number of “*Abha’s*” women’s region is the **lowest value**, And “*Makkah’s*” women’s region is the **highest value** based on *Region*.

- According to the plot, we can see that number “InActive” women’s statues the higher number, based on the *Last flight date*.