

Name: Albraa Baghdadi

Abstract

Missing appointment happens a lot and sometimes affect businesses that depend on appointments to complete the service such as Hospitals. In this project, we will build a model based on our dataset to predict the probability of the patient missing the appointment.

The benefit of this model is that the hospital could predict the probability of the patient missing the appointment and then increase the number of appointments for that day.

Design

The project contains data about patients appointments provided by Kaggle in Brazil. This project will try to clean and analyse the data and help the Hospital if we could predict if the patients are coming to their appointments or not by looking at the relation between the features.

Data

The dataset contains 100,000 records with 14 features for each, 9 of which are categorical. A few feature highlights include Gender, Appointment Day, Age, Neighbourhood, Scholarship, Hypertension, Alcoholism, Handicap, SMS_received.

Algorithms

- Removed wrong data (some records having less than zero years)
- Change typo columns name
- change the data type to date for appointment day and scheduled day columns to be able to classify the days
- Classify Age group

The Models used logistic regression binary classification and decision tree to solve our categorical problem.

Logistic regression, the categorical columns are extracted to binary by the `get_dummies` function. and the entire training dataset of 100,000 records was split into 80/20 train vs. Test. And the score is **0.80**

Decision tree, we include neighborhood columns to enhance the score by using the `LabelEncoder()` function, and the score increased by 1%. And the score is **0.81**

Tools

NumPy, Pandas, Scikit-learn, Matplotlib and Seaborn.

Communication

In conclusion, And after testing the models It seems that our models are always predicting that the patient will attend the appointment. Furthermore, the data was gathered in a short time span. The model could be improved if added more features such as :

- Forecast factors like weather and temperature
- Social factors such as marital status and employment status
- Hospitals location
- The clinic name

