CLASSIFICATION PROJECT REPORT

MEDICAL APPOINTMENT NO SHOWS

PREPARED BY
ALANOUD ALOSAIMI
SHAIKHA BIN ATEEQ

ABSTRACT

One of the problems that medical clinics face is when a person books an appointment but does not show up. Every appointment that is not attended deprives another patient of the opportunity to benefit from it.

DESIGN

In this project, we will predict whether the patient will show or not, the aim is to help clinics in Brazil know if the clinic or the patient is responsible for this problem and understand the causes that led to it.

DATA

The dataset collects information on more than 100k medical appointments in Brazil and 14 features which are: PatientId, AppointmentID, Gender, ScheduledDay, AppointmentDay, Age, Neighborhood, Scholarship, Hipertension, Diabetes, Alcoholism, Handcap, SMS_received and the prediction value is No-show.

ALGORITHMS

- Data Preprocessing.
- Features Engineering.
- Handle the outliers.
- Resampling the data.
- Scaling the data.

The implemented models are:

1. Logistic Regression
2. K-Nearest Neighbors
3. Decision Tree
4. Extra Trees
5. Random Forest
6. Support Vector Machine
7. Naive Bayes (Gaussian)
8. Naive Bayes (MultinomialNB)
9. XGBoost

TOOLS

Here are the tools that used in this project:

- Software platform: Jupyter Notebook
- Programming language: Python
- Includes libraries:
  - Pandas
  - Numpy
  - Seaborn
  - Matplotlib
  - Sklearn

CONCLUSION

- The best model in the term of performance is XGBoost.
- The train accuracy was 0.99 which the test accuracy was 0.97