DATA MINING PROJECT REPORT
Medical Appointment No-Show

171210005, 171210009, 171210030

April 9, 2020

Contents

1 Abstract

People book appointments in medical facilities and institutions for healthcare. Sometimes they do not show up for the appointment scheduled by them. This number is not insignificant and so it is required to analyze this. This project uses dataset built from records of a medical facility and includes fine techniques of machine learning to predict if an individual who books an appointment, will show up or not.

2 Introduction

In this work we study dataset of a medical facility and understand the behavior of different factors and their importance in determining whether a person will show up for an appointment he/she booked or not.

Dataset link: Medical Appointment No Shows

3 Related Work

Following is the list of similar datasets:
1. Medical Appointment
2. Hair Salon No-Show Dataset 3. medical-appointment-no-shown

4 Methodology

Dataset is studied and observed which helps in deciding the type of problem - it is a classification problem. In data preprocessing, inconsistent values and unnecessary columns are removed. Data is visualized to understand the distribution of values better and an optimal algorithm is selected. Decision Tree Classifier is the optimal algorithm under selection and it is used to predict the output. The accuracy is measured and the algorithm is run for different values of tree-depth to get the best results.

5 Results and Discussion

The classification problem was solved successfully by implementing Decision Tree Classifier with an accuracy of 79.80%.
Multiple acccuracies are achieved for different 'max-depth' of decision tree and the highest accuracy is achieved when max-depth value is 4.

6 Conclusion

The dataset is pre-processed successfully as well as processed and worked upon to predict a class using Decision Tree Classifier with accuracy of 79.80%.

7 Future Work

This can be further extended to study the losses of doctors and whole medical sector which occurs due to absence of patients at scheduled time. If the classifier predicts that the next person scheduled for an appointment will not show up, then the patient after him/her can be called early, saving both time and effort in overall scenario. In medical databases we can start rating patients as well as doctors on their punctuality and it can further improve the accuracy of the model.