

Data Analytics

CS40003

Project Assignment -3

2017-2018

14MA20003

Aakash Singh

Customer Churn Prediction

CCP data with 5000 observation)

Tools used : Python-Programming, Spyder, Ubuntu, Python Libraries (pandas, matplotlib, sklearn, etc.)

Methodology :

- Dropped the useless columns and changed the Boolean groups into numeric groups (True,False -> 1,0)
- Used Stratified Shuffle Split for splitting data into testing and training sets(25% for testing and 75% for training for 3 iterations)
- Gaussian Naïve Bayes, Decision Tree and Support Vector Machine is used for validation on each iteration.
- Accuracy and Precision is calculated used each model.
- ROC curves are plotted for each model and at every iteration.
- Finally selected Decision tree as best suited model for Customer Churn Prediction.

Reasonable assumptions : Assumed proper shuffling during Stratified Shuffle Split.