**Capstone Project Submission**

**Instructions:**

i) Please fill in all the required information.

ii) Avoid grammatical errors.

|  |
| --- |
| **Team Member’s Name, Email and Contribution:** |
| |  |  |  | | --- | --- | --- | | **NAMES** | **E-MAIL** | **CONTRIBUTION** | | Charan | charanbzact9865@gmail.com | All project worked | |  |  | |  |  |  | |
| **Please paste the GitHub Repo link.** |
| **GitHub link:-**  **https://github.com/Charanpoojary/CUSTOMER-SEGMENTATION-UNSUPERVISED-ML-** |
| **Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)** |
| This Project summary is entitled as “**Online Retail Customer Segmentation”.** The objective of this project is to train various Clustering algorithms to predict the client’s behavior  Customer segmentation is the practice of dividing a company’s customers into groups that reflect similarity among customers in each group. The goal of segmenting customers is to decide how to relate to customers in each segment in order to maximize the value of each customer to the business.  This project is an attempt to identify credit card customers who are more likely to default in the coming month. A lot of credit card issuing companies are working on predictive models which would help them predict the payment status of the customer ahead of time using the customer’s credit score, credit history, payment history and other factors.  This project is aimed at using customer’s personal and financial information like credit line, age, repayment and delinquency history for the past 6 months to predict the probability of the particular customer to become default next month. Many statistical and data mining techniques will be used to build a binary predictive model. If the credit card issuing companies can effectively predict the imminent default of customers beforehand, it will help them to pursue targeted customers and take calculated efforts to avoid the default, to overcome future losses efficiently.  The data, in any sense, does not directly reveal the identity of any individual or provide information that could be decrypted to connect to an individual. In this project, the plan is to predict the probability of credit-card holders to go default in the next month by using payment data from October 2015 to March 2016. Among the total 30,000 observations,To determine the binary variable – default payment in April 2016 (Yes = 1, No = 0), as the response variable.  In the end, looking at the accuracy scores of each model was not enough information to choose which model performed best when trying to predict whether or not a person would default on their credit card. Above one can clearly see how close all these accuracy scores fall. In fact, by just seeing the accuracy scores the Voting Classifier seems to perform the best. But if looked at performance via ROC curves the Logistic Regression model seems to perform the best and the Voting Classifier does significantly worse.For future work, we think it would be interesting to develop more complex models, such as implementing a neural network and seeing if there could be a better performance of an 82% accuracy score since all our previous techniques did not seem to affect the accuracy. |