Justin Elkinson

jelkinson@bryant.edu

10/15/19

AA 490 – Project Proposal

Vehicle Loan Prediction: Who will default?

Abstract:

Everyday banks struggle when deciding whether to hand out loans aimed towards the purchase of a vehicle. This decision is risky since some individuals have the potential to miss payments and ultimately default on their loans, causing extra work and stress for banks. The purpose of this project will be to identify the profile of individuals who are likely to default on a bank loan. To achieve this, I will analyze a vehicle loan dataset and use clustering to address the issue of potential loan defaults. By the end of the project I hope to identify a profile that banks can use while grading vehicle loan applications in order to avoid situations that require collection.

I am using a dataset that I found from Kaggle which includes predetermined train and test data, as well as a data dictionary. This dataset is titled L&T Vehicle Loan Default Prediction, was posted six months ago by Mamta Dhaker, and scored a usability rating of 8.2. The dataset that I am using for this project can be found here: <https://www.kaggle.com/mamtadhaker/lt-vehicle-loan-default-prediction/downloads/lt-vehicle-loan-default-prediction.zip/2>

I will be using a decision tree and random forest analysis during this project to help me reach my goal. Metrics that I will be looking for to assist me throughout this project will be the prediction accuracy score, for both my decision tree and random forest, as well as the actual accuracy of my model compared to the test data provided in the dataset. The pipeline methods that I intend to implement are a MinMaxScaler for data normalization and then passing it through a decision tree and random forest.

Gantt Chart:

