**Name: Alex Nanez**

**Project Name: Credit Card Application Approval Prediction**

1. **Introduction**

One important part of the financial industry is managing credit card applications and deciding whether the bank will issue a credit card to the applicant. This white paper is intended to provide an overview of a dataset containing information about credit card applications, which will include data preparation, exploratory data analysis, modeling, and evaluation.

1. **Dataset:**

The Credit Approval Data Set is found in the UCI Machine Learning Repository as a .data file, but it can be converted to a CSV or Excel file if needed.

1. **Data Preparation:**

This is the first step in the data mining process. I will perform methods to clean and transform the data so that it will be prepared for the next steps. These methods include transforming the features that return a Boolean variable (True or False) to 1s and 0sa nd checking for missing values, which this dataset was said to have a few.

1. **Exploratory Data Analysis:**

After preparing the data, I will perform EDA to further analyze my dataset and summarizing the main characteristics of the data. For this, I will include the summary statistics for all features of the dataset as well as incorporate visualizations, such as a pair plot, to find patterns or relationships.

1. **Modeling:**

After preparing the data and performing EDA, I will move on to the modeling portion of the data mining process. For this dataset, I chose to use three separate algorithms, which include boosting, random forest, and SVM. The goal is to predict where a credit card application was approved based on the given features. Once I reach this part, I will take the necessary steps to run these three algorithms on the dataset which includes separating the data into a training set and a testing set.

1. **Modeling:**

This is the final step of the data mining process. Once I reach this part, I will decide on the performance metrics I will use based on the modeling technique and evaluate how well the model performs. A higher accuracy is ideal, but I will also make sure to take into account overfitting of the dataset.