MET CS 777
Eljan Ibrahimli
Date: 10-11-2022
Project Proposal

In the past 2 years I have worked mainly in credit card dataset and I am going to analyze Credit approval data set which is available publicly. The link for the dataset: https://archive.ics.uci.edu/ml/datasets/Credit+Approval
Goal: my goal in this dataset is to show visualization related to data distribution, apply some machine learning models to predict if any person submits the application and that application can be approved or denied. Also, I would like to analyze which model will provide higher accuracy particularly in this dataset.

This dataset has 690 instances and 15 attributes. Attribute characteristics are categorical, integer and real.

I will take the following steps in order to reach the above goals.

1. Once I import the data, I will do data preprocessing. In the data description, it shows that there are some missing values. I need to deal with them before analyzing the data.
2. Once I clean the data, I would like to check if the data is balanced or imbalanced. If it is too much imbalanced, I will balance the data.
3. The dataset has 15 attributes, I will do feature selection and choose the most important attributes. I will also analyze in which model which attributes are more important
4. Before moving to different classifiers, I will divide the data into training and testing set
5. I will implement different classifiers such as K-NN, Logistic Regression and others.
6. After receiving the results of the classifiers, I plan to create confusion matrices table and compare the results to highlight my findings
7. Conclusion will show us which classifier has higher accuracy