Appendix

Glossary

Term	Description
AdaBoost	Adaptive Boosting
AUC	The Area Under the Curve (AUC) is the measure of the ability of a classifier to distinguish between classes and is used as a summary of the ROC curve. The higher the AUC, the better the performance of the model at distinguishing between the positive and negative classes.
Accuracy	The measure of number of correct predictions over all predictions
Binary classification	Binary classification is the task of classifying the elements of a set into two groups on the basis of a classification rule
Categorical data	Categorical data is a type of data that can be stored into groups or categories with the aid of names or labels.
Classification	classification refers to a predictive modeling problem where a class label is predicted for a given example of input data
Class Imbalance	Imbalanced data typically refers to a problem with classification problems where the classes are not represented equally
Confusion matrix	A confusion matrix is a summary of prediction results on a classification problem
Correlation matrix	A correlation matrix is simply a table which displays the correlation coefficients for different variables.
Data Preprocessing	Data preprocessing in Machine Learning refers to the technique of preparing (cleaning and organizing) the raw data to make it suitable for a building and training Machine Learning models
Dataset	a collection of data
Decision stumps	A decision stump is a machine learning model consisting of a one-level decision tree
Ensemble learning	Ensemble learning is a general meta approach to machine learning that seeks better predictive performance by combining the predictions from multiple models
F1 score	It's the harmonic mean of the precision and recall
Feature	a feature is an individual measurable property or characteristic of a phenomenon.
Hyperparameter	a hyperparameter is a parameter whose value is used to control the learning process
KernalFunction	The kernel function is what is applied on each data instance to map the original non-linear observations into a higher-dimensional space in which they become separable. It's a function that corresponds to a dot product of two feature vectors in some expanded feature space. Here it's the parameter used in fitcsvm .
Learners	Weak learners to use in the ensemble
LearnersRate	Learners rate is used to train an ensemble using shrinkage

Reference:

Charleonnan, "Credit card fraud detection using RUS and MRN algorithms," *2016 Management and Innovation Technology International Conference (MITicon)*, 2016, pp. MIT-73-MIT-76, doi: 10.1109/MITICON.2016.8025244.

MaxNumSplits	Maximal number of decision splits
Multicollinearity	Multicollinearity is the occurrence of high intercorrelations among two or more independent variables in a multiple regression model.
NumLearningCycles	Number of ensemble learning cycles where at every learning cycle, the software trains one weak learner for every template object in Learners
outliers	an outlier is a data point that differs significantly from other observations
overfitting	Overfitting refers to a model that models the training data too well
parameter	a parameter is a function argument that could have one of a range of values.
precision	Precision is a metric that quantifies the number of correct positive predictions made and is calculated as the ratio of correctly predicted positive examples divided by the total number of positive examples that were predicted
Quality dataset	Clean dataset without any outliers or noise
recall	Also known as sensitivity is the measure of our model correctly identifying True Positives.
Sensitivity	Sensitivity/recall is the measures how good a test is at detecting the positives
Shallow trees	A shallow tree is a small decision tree (most of the cases it has a small depth)
Specificity	It's the measure of how good a test is at avoiding false alarms
SVM	Abbreviation for Support Vector Machine
ROC	An ROC curve (receiver operating characteristic curve) is a graph showing the performance of a classification model at all classification thresholds. Its a probability curve
templateTree	returns a default decision tree learner template suitable for training an ensemble (boosted and bagged decision trees)
weak leaner	Weak learners are models that perform slightly better than random guessing.
Weak classifiers	A weak classifier is a model for binary classification that performs slightly better than random guessing

Intermediate results:

Initially it was planned to choose Decision tree and Naïve Bayes for the analysis. But since it didn't gave much performance improvement the analysis was done for few other models as well. The models chosen includes:
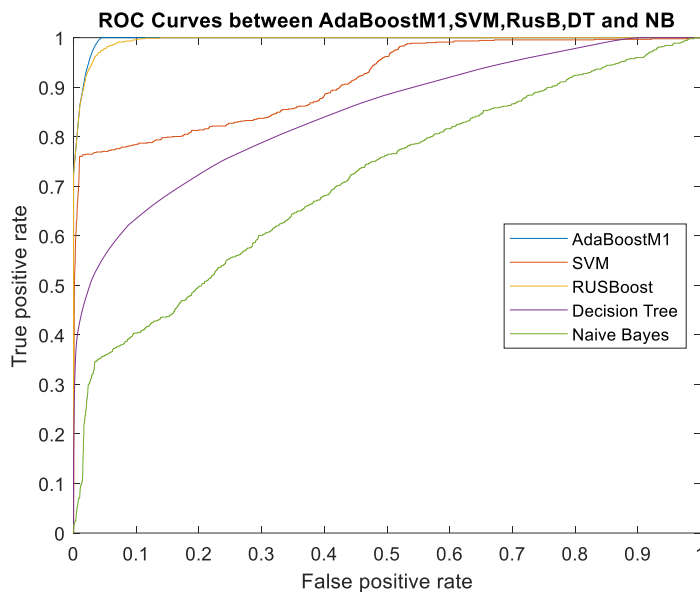
- Naïve Bayes classifier
- Decision tree
- RUSBoost
- AdaBoostM1
- Support Vector Machine

Reference:
Charleonnan, "Credit card fraud detection using RUS and MRN algorithms," *2016 Management and Innovation Technology International Conference (MITicon)*, 2016, pp. MIT-73-MIT-76, doi: 10.1109/MITICON.2016.8025244.

Observations:

- AdaBoostM1 and RUSBoost performed well when compared to rest of the models. Compared to RUSBoost, AdaBoostM1 had greater AUC.
- Naïve Bayes had to be dropped due to its poor performance and when compared to rest of the models. The only way Naïve Bayes could increase its efficiency is by balancing the dataset.
- SVM shows better performance than Decision tree
- AdaBoostM1 and SVM is chosen as models for evaluation as it was performing better when compared to rest of the models.

Best Model	AUC
AdaBoostM1	0.9963
SVM	0.9125
RUSBoost	0.9950
Decision Tree	0.8452
Naïve Bayes	0.7123

ROC Curves between AdaBoostM1,SVM,RusB,DT and NB

Main Implementation choices :

Since the dataset is highly imbalanced AUC is considered as a main factor for choosing the right model .Since AdaBoostM1 and SVM is having better AUC ,these two are chosen as implementation choices. Also AdaBoostM1 is chosen with tree as weak learner. Hence it adds as a performance improvement for Decision tree.

Reference:
Charleonnan, "Credit card fraud detection using RUS and MRN algorithms," *2016 Management and Innovation Technology International Conference (MITicon)*, 2016, pp. MIT-73-MIT-76, doi: 10.1109/MITICON.2016.8025244.