Glossary

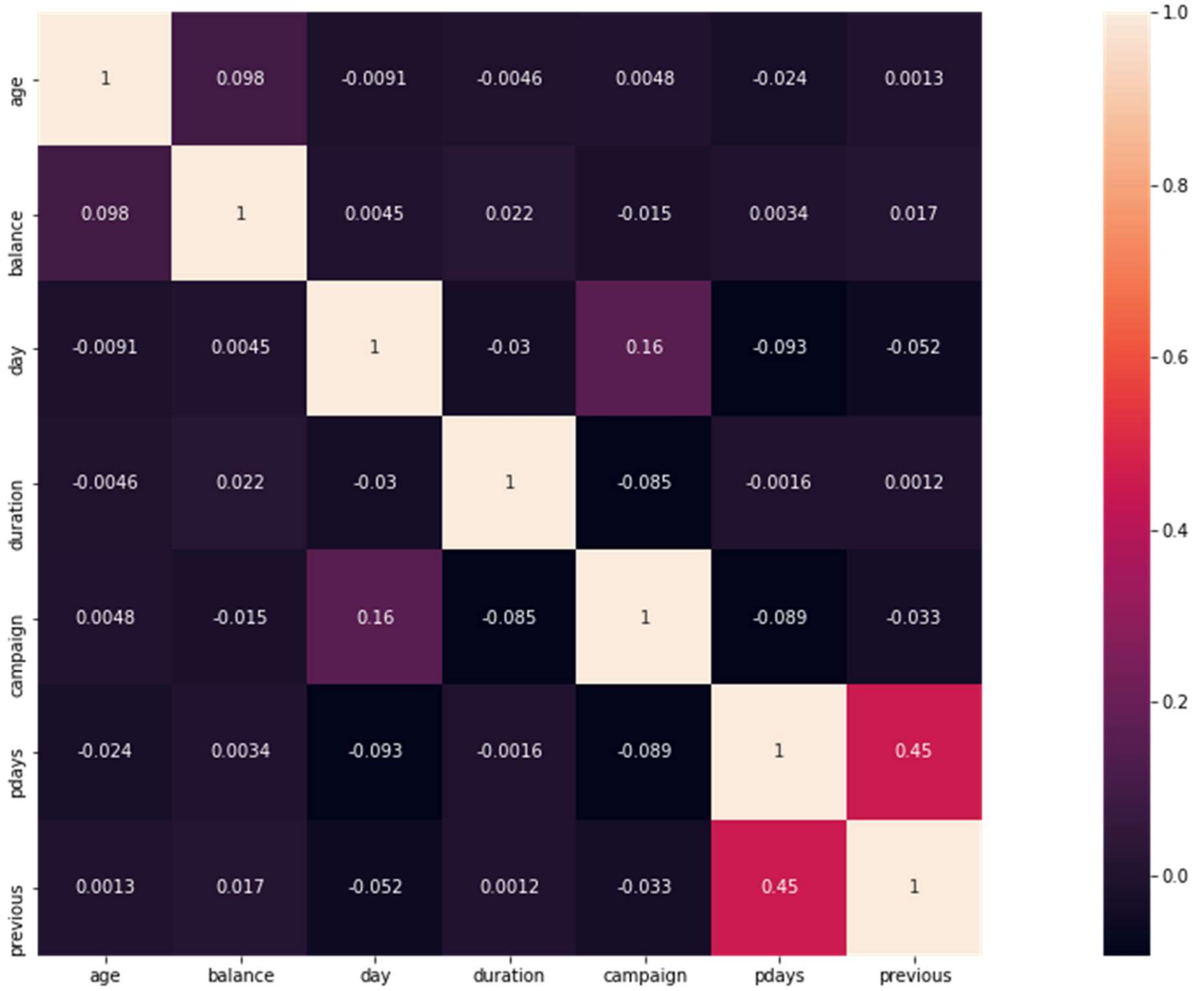
Area under curve (AUC) The measure of the ability of a classifier to distinguish between classes.
Accuracy Ratio of correctly predicted observation to the total observations.
Bayes Theorem The probability of an event, based on prior knowledge of conditions that might be related to the event.
Binning The transformation of continuous variables into discrete. Binning groups related values together in bins to reduce the number of distinct values.
Cross Validation Procedure used to evaluate machine learning models on a limited data sample.
Confusion Matrix is a specific table layout that allows visualization of the performance of an algorithm.
ECOC or Error Correcting Output Code is a technique that allows a multi-class classification to be given as multiple binary classification problem.
F Score(F1) The weighted average of Precision and Recall.
Holdout Cross Validation Technique where the dataset is split into training and testing datasets.
Hyperparameters Optimisation Optimal Parameters where values are set to observe best performance models. In any machine learning algorithm, the parameters would be initialised before training a model.
Kernel Distribution Non parametric representation of the probability density function of a random variable. Kernels are defined by smoothing functions and bandwidth value, which controls the smoothness of the outputted density curve.
K-Fold Cross Validation Technique to evaluate models on a limited dataset. K refers to the number of groups the data is split into.
Leaf Size Leaf are essentially end nodes of a decision tree.
OnevsAll K number of binary learners where each learner has one positive class whereas the rest are negative.
OnevsOne Predicts one class label and the one with the most predictions is voted for.
Principal Component Analysis The process of computing the principal components
Prior A probability of an event based on pre established knowledge before data is collected.
Precision Ratio of correctly predicted positive observations to the total predicted positive observations.
Recall Ratio of correctly predicted positive observations to all observations in actual class.
ROC Curve ROC curve is a performance measurement for the classification problems at various threshold settings.
Specificity (FPR) Defines how many incorrect positive results occur among all negative samples available during the test.
Sensitivity (TPR) TPR defines how many correct positive results occur among all positive samples available during the test
Z-Score Numerical value that shows a value relationship to the mean of a group of values. Z-Score is measured as standard deviations away from the mean. If the score is 0 then the data point's score is identical to the mean score.
Overfitting Occurs when the neural network has so much information processing capacity but limited information contained in the training set, not enough to train all the neurons in the hidden layers.
Underfitting When model performs poor on training data and well on evaluation data.

Implementation Details

The steps taken to process the data are:

- Data Pre-processing, Computations, Visualisation, Model Building, Tuning and Optimising was done on Python and MATLAB (Version R2021B)
- Imported the data and organised and pre-processed
- We carried out exploratory data analysis and feature engineering
- Built the Model and evaluated further
- Tuning Model Parameters before evaluating further
- Built most optimised model
- The classification that we identify is whether the customer will subscribe to term deposit or not (‘Y’) and is a categorical variable.
- Yes = Client subscribed to Term Deposit
- No = Client did not subscribe to Term deposit

Numerical Column Correlation Matrix

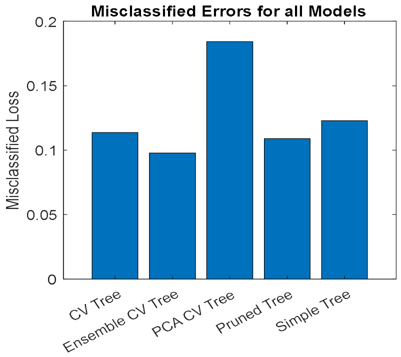


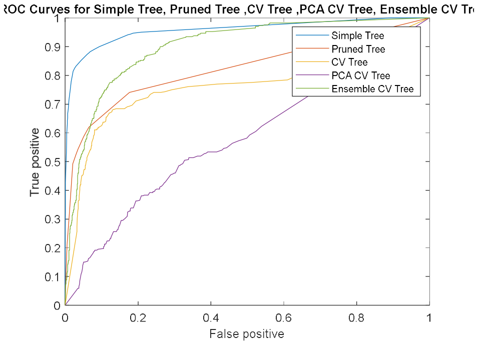
	age	balance	day	duration	campaign	pdays	previous
age	1	0.098	-0.0091	-0.0046	0.0048	-0.024	0.0013
balance	0.098	1	0.0045	0.022	-0.015	0.0034	0.017
day	-0.0091	0.0045	1	-0.03	0.16	-0.093	-0.052
duration	-0.0046	0.022	-0.03	1	-0.085	-0.0016	0.0012
campaign	0.0048	-0.015	0.16	-0.085	1	-0.089	-0.033
pdays	-0.024	0.0034	-0.093	-0.0016	-0.089	1	0.45
previous	0.0013	0.017	-0.052	0.0012	-0.033	0.45	1

• Models were built using Decision Tree (DT) and Naïve Bayes (NB) classifiers. Data was given a holdout and split into testing and training datasets. We experimented by building various models with different parameters and techniques in order to get the optimal final model. Our goal was to aim for low error during the model building process with high AUC and accuracy values. We showed Improvement with each technique in order to find the ideal model.

Decision Tree Model

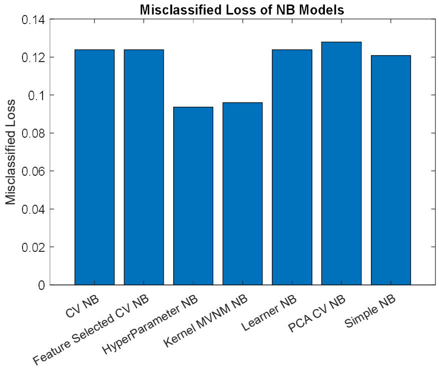
- We built a basic tree model with split criteria as 'GDI' and maximum number of splits as 100 was used.
- We carried out level 12 Pruning for the basic model. Whilst we found the error to have reduced, the accuracy and precision performance metrics were not the best.
- Principal Component Analysis (PCA) was applied to transform the features which contributed to over 95% of the variance. The labels were then predicted by fitting this into the model. The classes here however resulted in larger misclassifications.
- Ensemble tree was used to combine weak learners into a more efficient stronger learner. The AUC value improved via the 'Bagging' Technique. This was very complex and the model took hours to run.

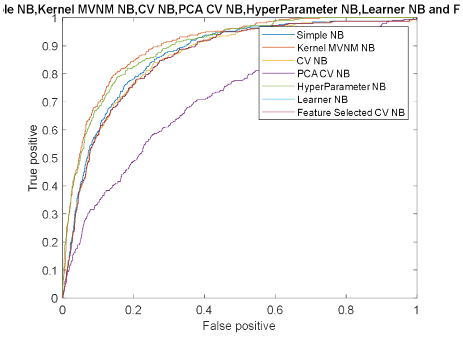




Naïve Bayes Model

- Gaussian Distribution was used in order to create our basic Naïve Bayes Model.
- Default width size classifier was used.
- Tuning the model by changing the priors to the model. This technique was somewhat successful.
- ECOC model was used with a Naïve Bayes Learner however the misclassified error rose and accuracy/precision metrics dropped.
- K-Fold was applied for cross validating the Naïve Bayes Model. We selected the K-Fold with the least losses. The K-Fold loss was more when compared to other models which maybe due to the nature of our large dataset.


Misclassified Loss of NB Models
Misclassified Loss
CV NB
Feature Selected CV NB
HyperParameter NB
Kernel MVNM NB
Learner NB
PCA CV NB
Simple NB


le NB,Kernel MVNM NB,CV NB,PCA CV NB,HyperParameter NB,Learner NB and F
True positive
False positive
Simple NB
Kernel MVNM NB
CV NB
PCA CV NB
HyperParameter NB
Learner NB
Feature Selected CV NB