PROBLEM STATEMENT

To develop a predictive model that combines various econometric measures to foresee a financial condition (bankruptcy or not) of a firm.

PRE-PROCESSING
Data preparation for modelling

Analyzing data variables, transform skewed variables, feature engineering

STEP 1

UPSAMPLING
Address highly imbalanced data

Balance the data set by upsampling the minority class

STEP 2

MODEL COMPARISON
To maximize the AUC

Compare AUC for models - Logistic Regression, Random Forest, Neural Network, Gradient Boost, XGBoost

STEP 3

PARAMETER TUNING
Improve accuracy

Tune the hyper parameters for the best model to increase the accuracy (AUC)

STEP 4

PROBLEM STATEMENT

To develop a predictive model that combines various econometric measures to foresee a financial condition (bankruptcy or not) of a firm.

PARAMETER TUNING

Improve accuracy

Tune the hyper parameters for the best model to increase the accuracy (AUC)

MODEL COMPARISON

To maximize the AUC

Compare AUC for models - Logistic Regression, Random Forest, Neural Network, XGBoost

UPSAMPLING

Address highly imbalanced data

Balance the data set by upsampling the minority class

PRE-PROCESSING

Data preparation for modelling

Analyzing data variables, transform skewed variables, feature engineering

STEP 1

STEP 2

STEP 3

STEP 4

FEATURE ENGINEERING

Certain financial ratios that can specifically be used to provide early warning signals of possible impending bankruptcy were used to perform feature engineering.

1) Profitability ratios
2) Leverage ratios
3) Operating performance ratios

There are key ratios that can provide such warnings well in advance, giving investors plenty of time to dispose of their equity interest before the financial roof falls in.

- (inventory * 365 / cost of products sold)/(profit on operating activities / financial expenses)
- (gross profit / total assets) – (operating expenses / total liabilities)
- ((gross profit + depreciation) / sales) – ((current assets - inventory) / short-term liabilities)
- ((gross profit + extraordinary items + financial expenses) / total assets) * ((receivables * 365) / sales)
- (profit on operating activities / total assets) * ((equity - share capital) / total assets) + (profit on operating activities / financial expenses)
- ((gross profit / total assets) – (operating expenses / total liabilities)) / ((sales - cost of products sold) / sales)
- ((gross profit + extraordinary items + financial expenses) / total assets) * ((receivables * 365) / sales)
- (sales (n) / sales (n-1)) + ((short-term liabilities *365) / sales)
- (profit on operating activities / financial expenses) + (total costs / total sales)
- (EBITDA (profit on operating activities - depreciation) / total assets) + ((short-term liabilities * 365) / cost of products sold)

PROBLEM STATEMENT

To develop a predictive model that combines various econometric measures to foresee a financial condition (bankruptcy or not) of a firm.

PRE-PROCESSING

Data preparation for modelling

Analyzing data variables, transform skewed variables, feature engineering

STEP 1

UPSAMPLING

Address highly imbalanced data

Balance the data set by upsampling the minority class

STEP 2

MODEL COMPARISON

To maximize the AUC

Compare AUC for models - Logistic Regression, Random Forest, Neural Network, XGBoost

STEP 3

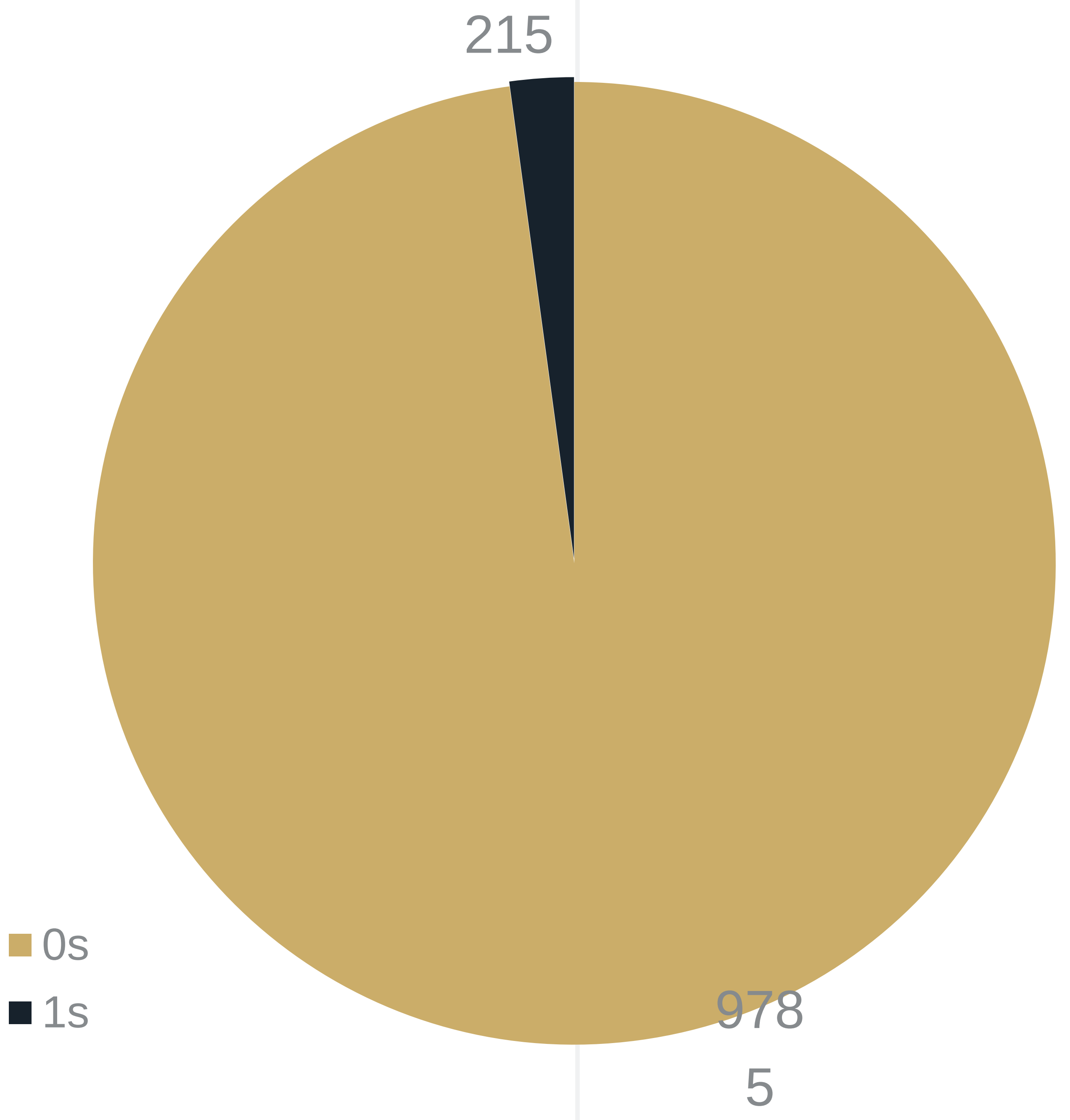
PARAMETER TUNING

Improve accuracy

Tune the hyper parameters for the best model to increase the accuracy (AUC)

STEP 4

HANDLING IMBALANCED DATA: UPSAMPLING



- Since the probability of instances belonging to the majority class is significantly high in an imbalanced dataset, algorithms are much more likely to classify new observations to the majority class
- We decided to resample the data in order to mitigate the effect caused by class imbalance
- Since under-sampling the majority class would remove observations from the original data set, it might discard useful information
- So, we use over-sampling, which increases the number of minority class members in the training set
- Synthetic Minority Over-sampling Technique: SMOTE first considers the K nearest neighbors of the minority instances. It then constructs feature space vectors between these K neighbors, generating new synthetic data points on the lines

PROBLEM STATEMENT

To develop a predictive model that combines various econometric measures to foresee a financial condition (bankruptcy or not) of a firm.

PARAMETER TUNING

Improve accuracy

Tune the hyper parameters for the best model to increase the accuracy (AUC)

MODEL COMPARISON

To maximize the AUC

Compare AUC for models - Logistic Regression, Random Forest, Neural Network, XGBoost

UPSAMPLING

Address highly imbalanced data

Balance the data set by upsampling the minority class

PRE-PROCESSING

Data preparation for modelling

Analyzing data variables, transform skewed variables, feature engineering

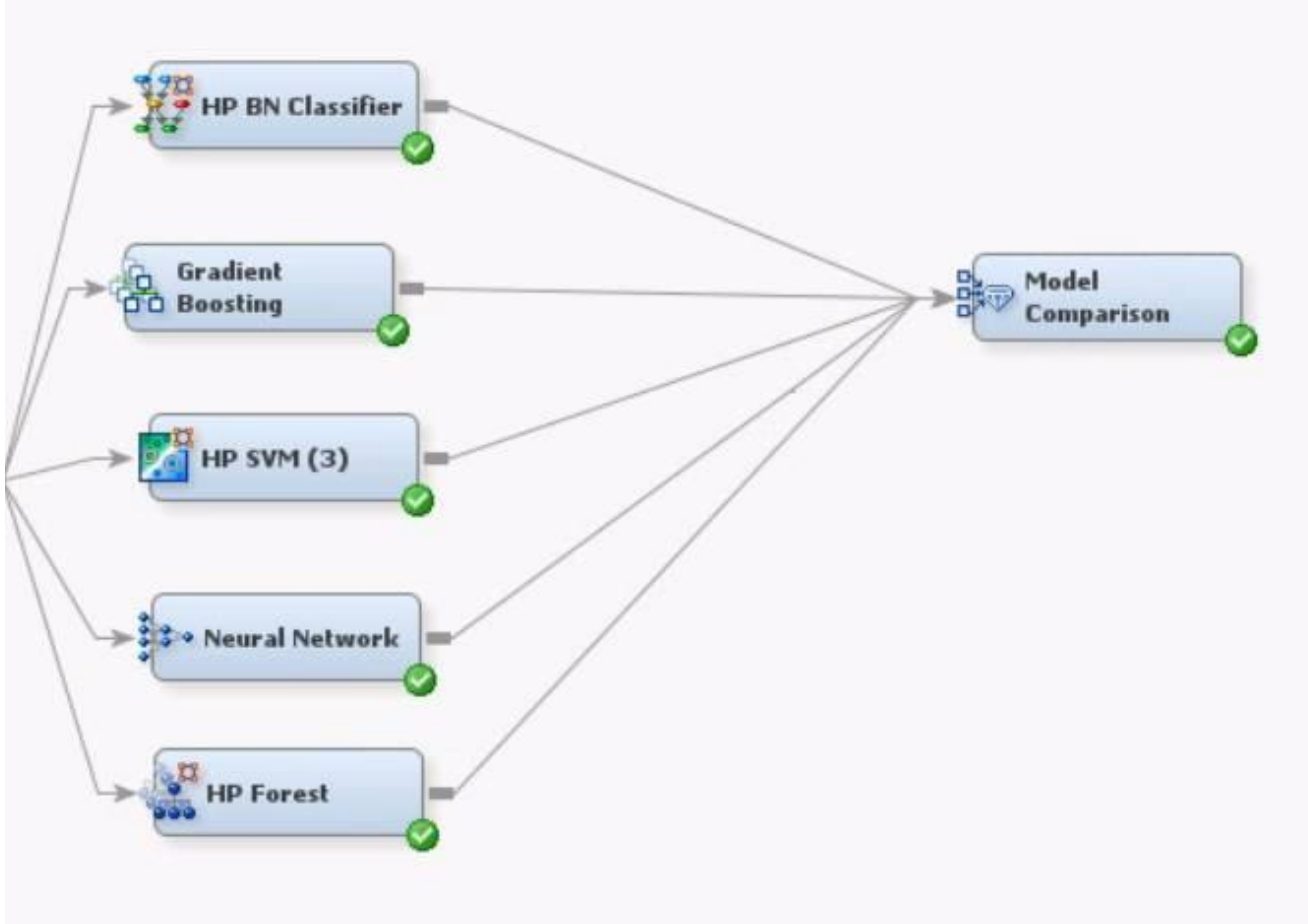
STEP 1

STEP 2

STEP 3

STEP 4

MODEL COMPARISON



We ran the following models and compared AUC results:

- Naïve Bayes
- Gradient Boosting
- SVM
- Neural Network
- Random Forest

MODEL NODE	ROC INDEX
HPDMFOREST	0.833
HPSVM3	0.895
NEURAL	0.791
HPBNC	0.808
BOOST	0.756
XGBOOST	**0.927**

PROBLEM STATEMENT

To develop a predictive model that combines various econometric measures to foresee a financial condition (bankruptcy or not) of a firm.

PARAMETER TUNING

Improve accuracy

Tune the hyper parameters for the best model to increase the accuracy (AUC)

MODEL COMPARISON

To maximize the AUC

Compare AUC for models - Logistic Regression, Random Forest, Neural Network, XGBoost

UPSAMPLING

Address highly imbalanced data

Balance the data set by upsampling the minority class

PRE-PROCESSING

Data preparation for modelling

Analyzing data variables, transform skewed variables, feature engineering

STEP 1

STEP 2

STEP 3

STEP 4

PARAMETER TUNING

Model: XGBOOST | XGB Classifier

We performed grid search CV on the hypertuning parameters in an attempt to increase AUC

```
model = XGBClassifier(
    n_estimators = 600,
    max_depth = 5,
    learning_rate = 0.05,
    subsample = 0.8,
    gamma = 0,
    min_child_weight = 11,
    colsample_bytree = .7,
    reg_alpha = 0.0000,
    scale_pos_weight = 10)
```

THANK

YOU!