Predictive Maintenance in Semiconductor Industry

#Fatima_Boujaha

MODEL BUILDING AND VALIDATION

Content

01 Problem Statement

02 Data Preprocessing

03 Feature Scaling

04 Feature Selection

05 Synthetic Data Generation

06 Parsimonious Model & Validation

1. Problem Statement

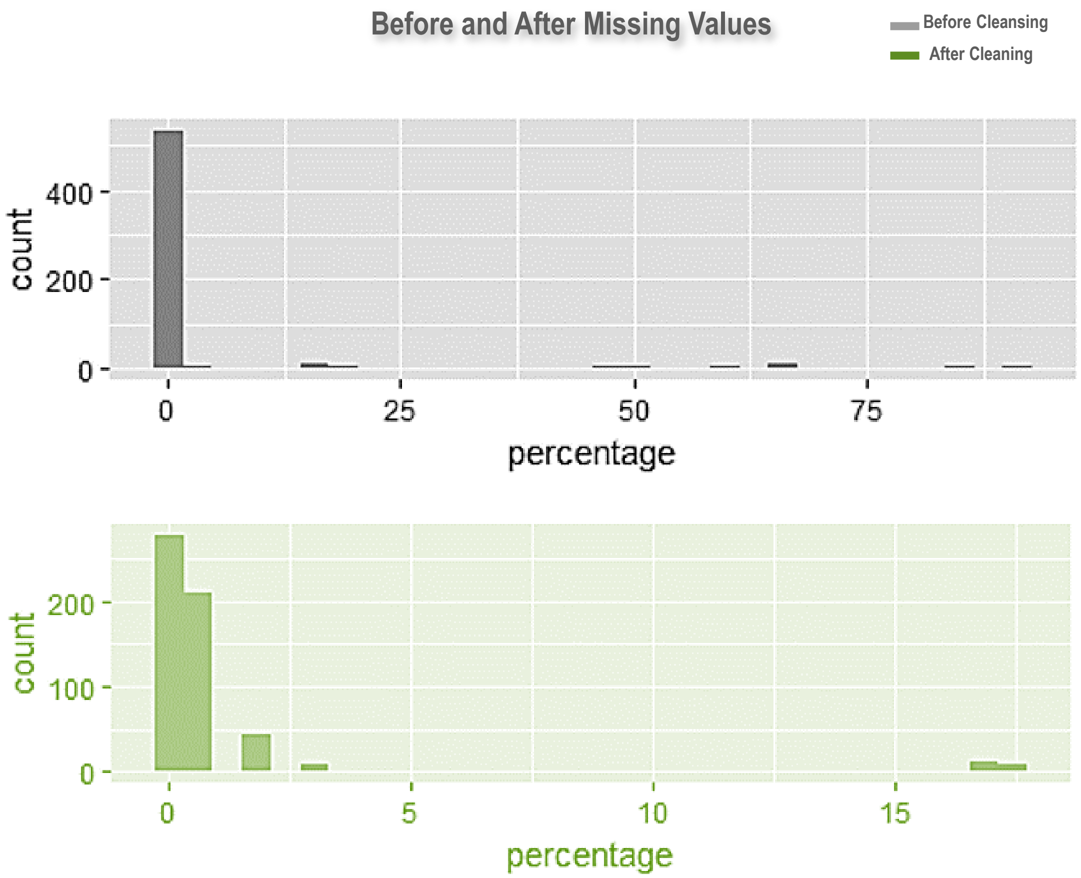
DESCRIPTION OF THE BUSINESS PROBELM

“ Predicts yield failure on a manufacturing process, This is a very important business problem for semiconductor manufacturers since their process can be complex and involves several stages from raw sand to the final integrated circuits. Given the complexity, there are several factors that can lead to yield failures downstream in the manufacturing process. ”

GOAL: To predict yield failure of a semiconductor manufacturing process in order to optimize the process.

STRATEGY: Perform data cleaning followed by feature selection techniques to preserve only most relevant signals. Compare and choose among various classification techniques to find the classifier that has the best fault detection performance.

2. Data Cleaning


Before and After Missing Values

Missing Values treatment:

kNN Imputation:

For every observation to be imputed, it identifies 'k' closest observations based on the Euclidean distance and computes the weighted average (weighted based on distance) of these 'k' obs.

41951 Total NULL values.

590 Variables excluding labels which don't contain any missing values

8823 Total NULL values after removing missing Values.

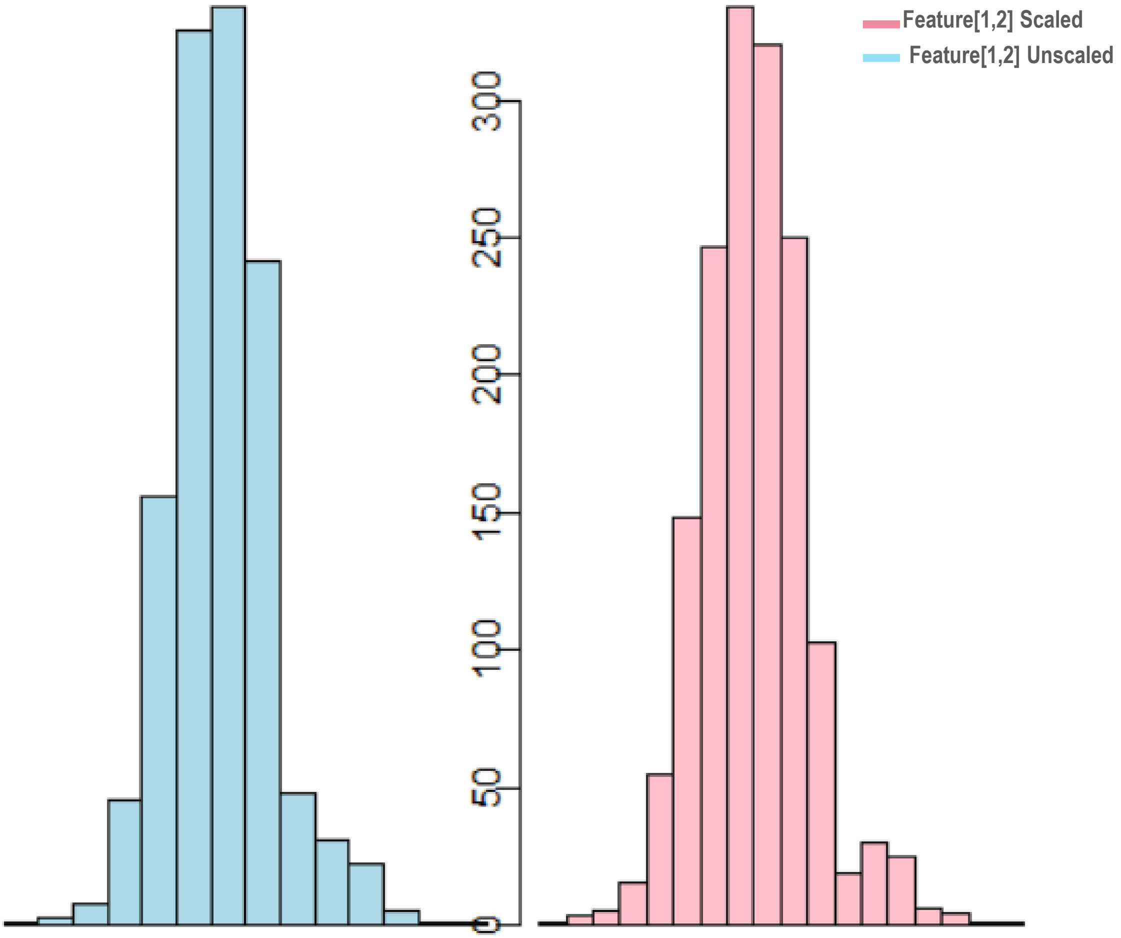
560 Variables after Missing values.

45% Threshold applied for missing Values.

3. Feature Scaling

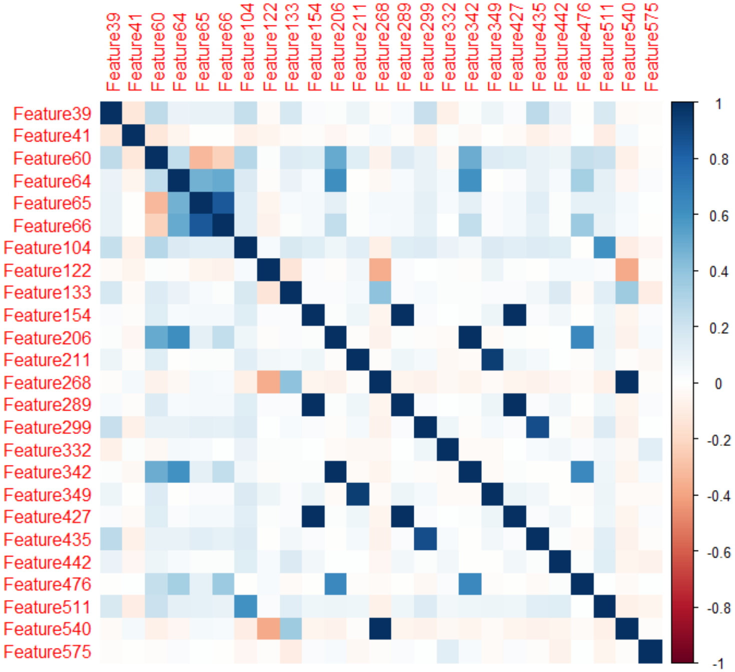
Method Applied: **MinMaxScaler**

X_std = (X - X.min(axis=0)) / (X.max(axis=0) - X.min(axis=0))

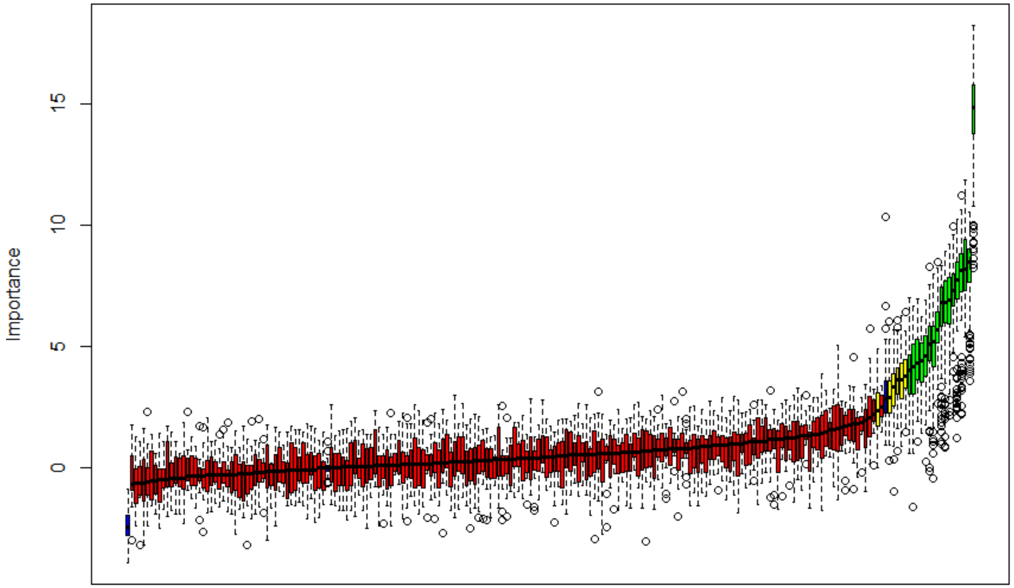


4. Feature Selection

‘ Heatmap Pearson’s Correlation ’



BORUTA



✓ 22 Variables have been selected as highly correlated at the cut-off 0.9

✓ 25 Variables have been selected with high importance

5. Synthetic Data Generation

	ORIGINAL Data	After SMOTE
Class	Training_Set	Training_Set
Pass	1024	766
Fail	73	331

SMOTE

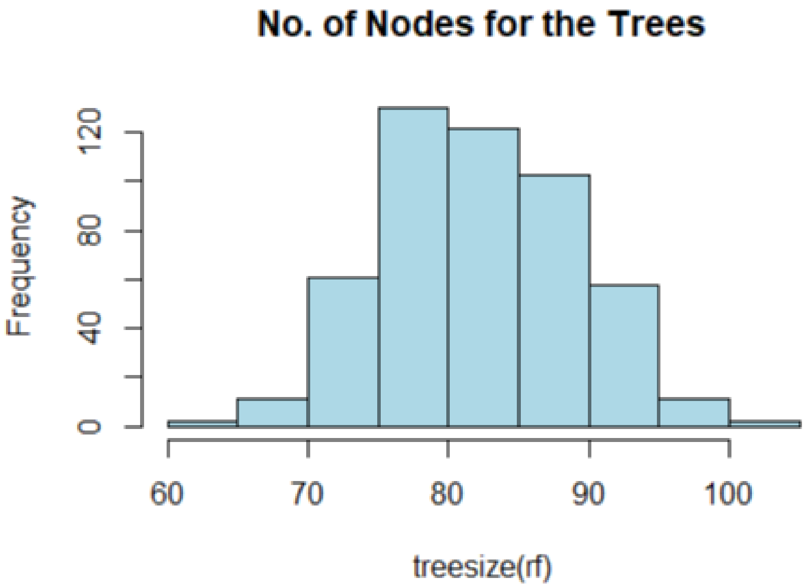
Synthetic Minority Over-sampling Technique Select random neighbors from the k minority class nearest neighbors. And forces the decision region of the minority class to become more general.

ADASYN

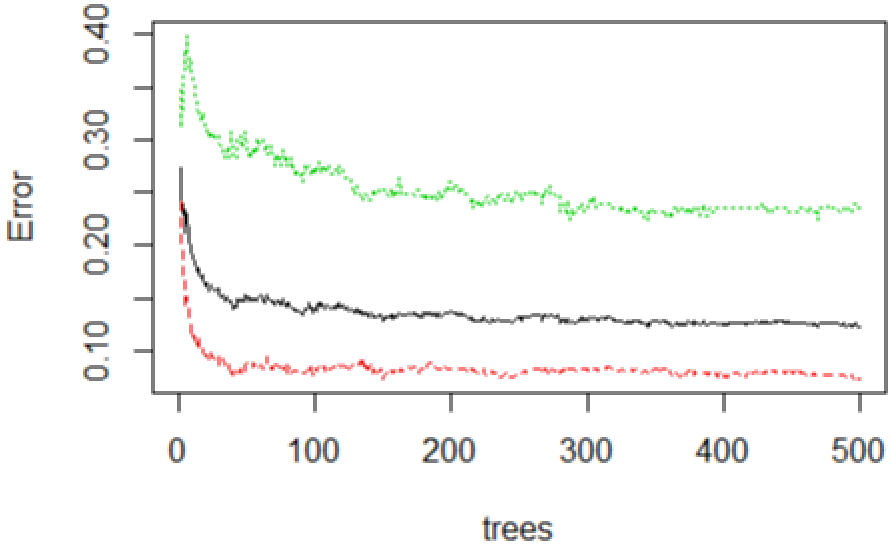
Adaptive Synthetic Sampling Approach for Imbalanced used by adaptively assigning weights to more difficult samples according to the density distribution.

6. Building PARSIMONIOUS Model

a) RANDOM Forest Model:



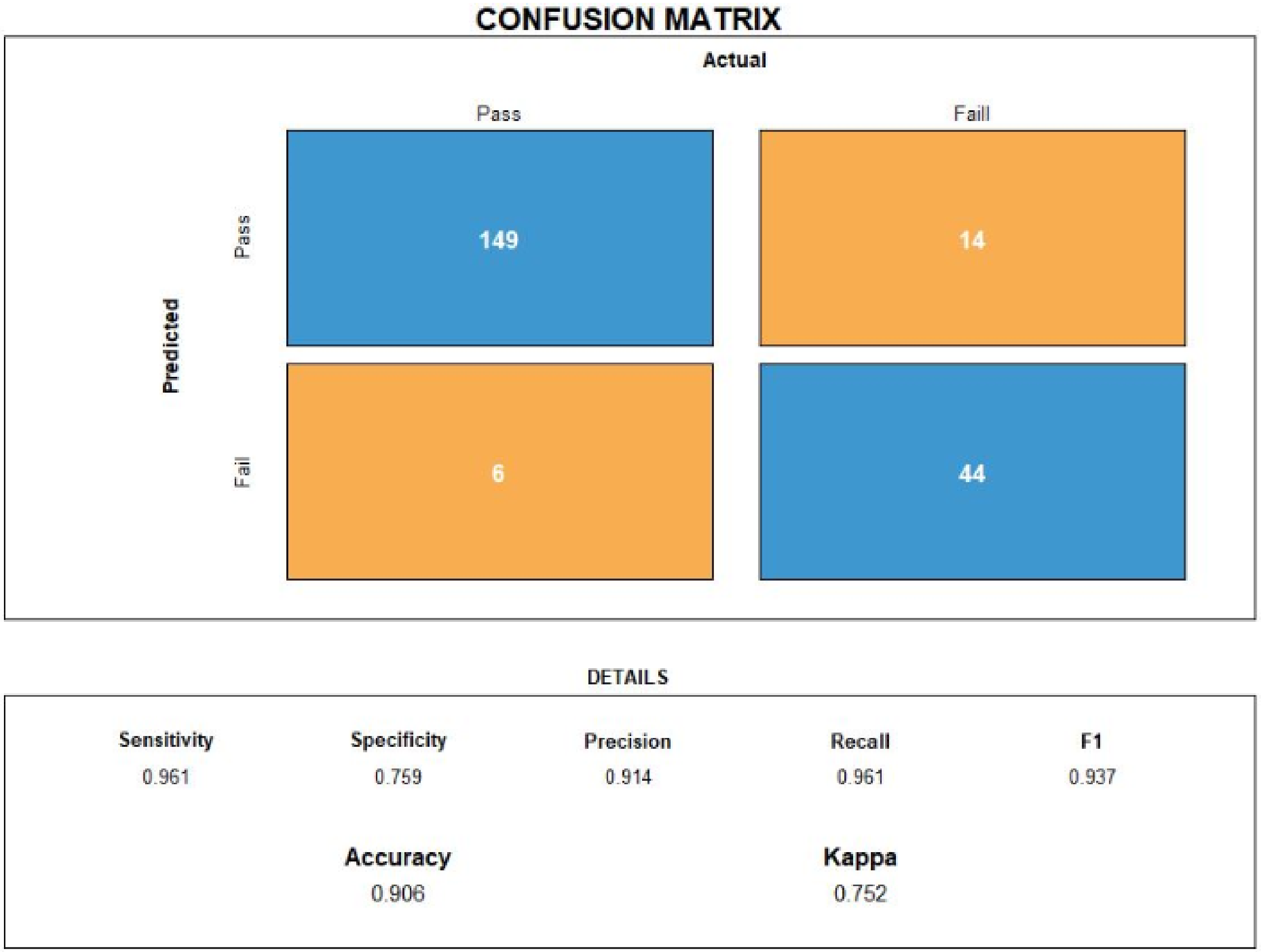
Out of bag (OOB) Error



6. Building PARSIMONIOUS Model

c) Results:

Final Model RANDOM FOREST based on BORUTA



CONFUSION MATRIX

Predicted \ Actual	Pass	Faill
Pass	149	14
Fail	6	44

DETAILS

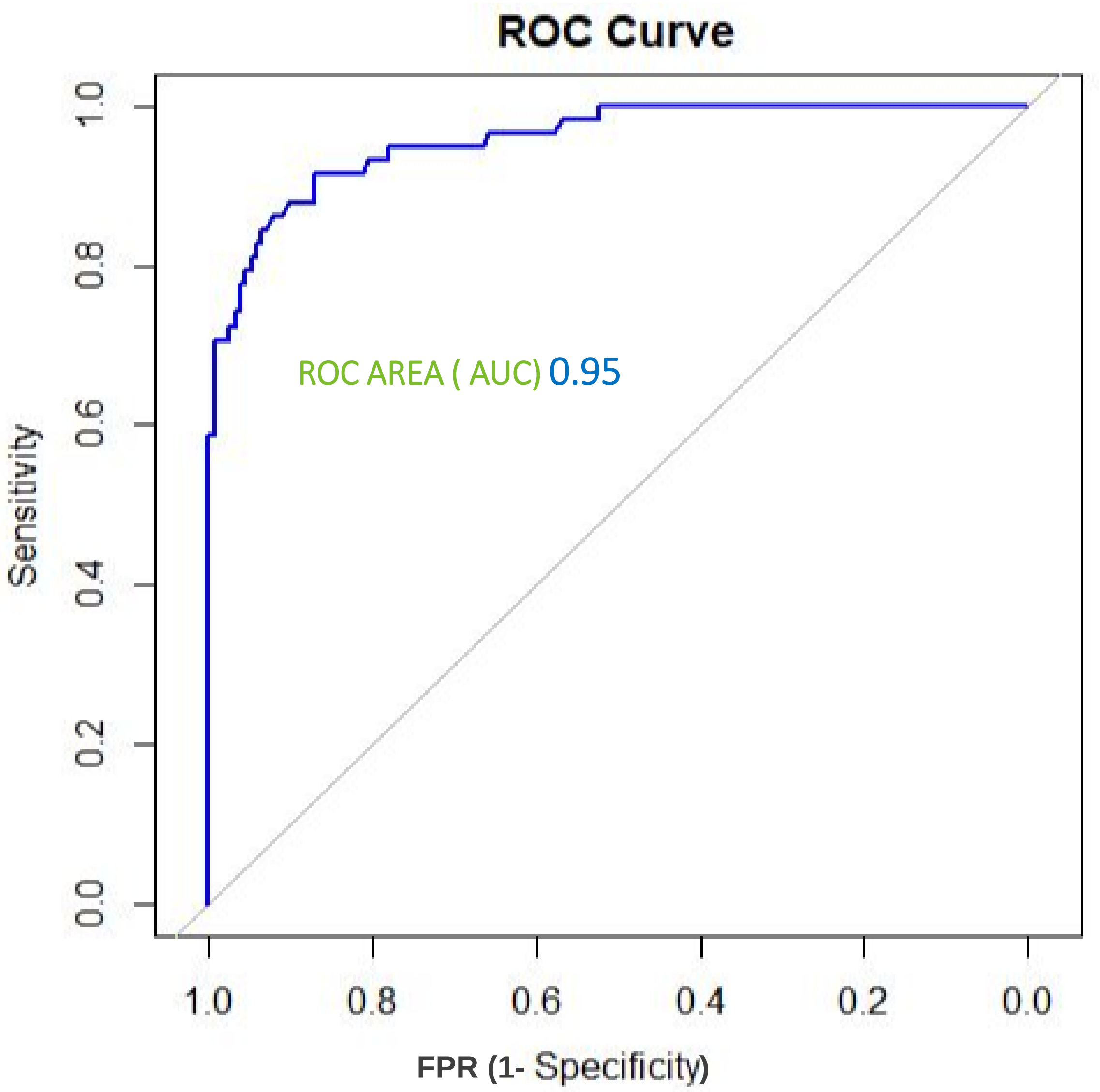
Sensitivity	Specificity	Precision	Recall	F1
0.961	0.759	0.914	0.961	0.937

Accuracy	Kappa
0.906	0.752

- K-Fold = 0.89
- TP Rate = 0.96
- FP Rate = 0.24

6. Building PARSIMONIOUS Model

b) Model Evaluation:



Thank you