STAFF TURN OVER

A SUPERVISED MACHINE LEARNING APPROACH

July 6th 2022
Mohamed MZAOUALI

OUTLINE

- Introduction
- Exploratory Data Analysis
- Feature Engineering
- Model Selection
- Features Importances
- Conclusion
- References

INTRODUCTION

Objective:

- Classification:
  - Employee will leave
  - Employee will not leave
- Classification Algrithms Variations

Data Source:

- HR Data Set

EDA

DATA SET:

Shape:

- 15781 Lines by 16 Columns.

Null Values:

- None.

Missing Values:

- None.

NaN Values:

- None.

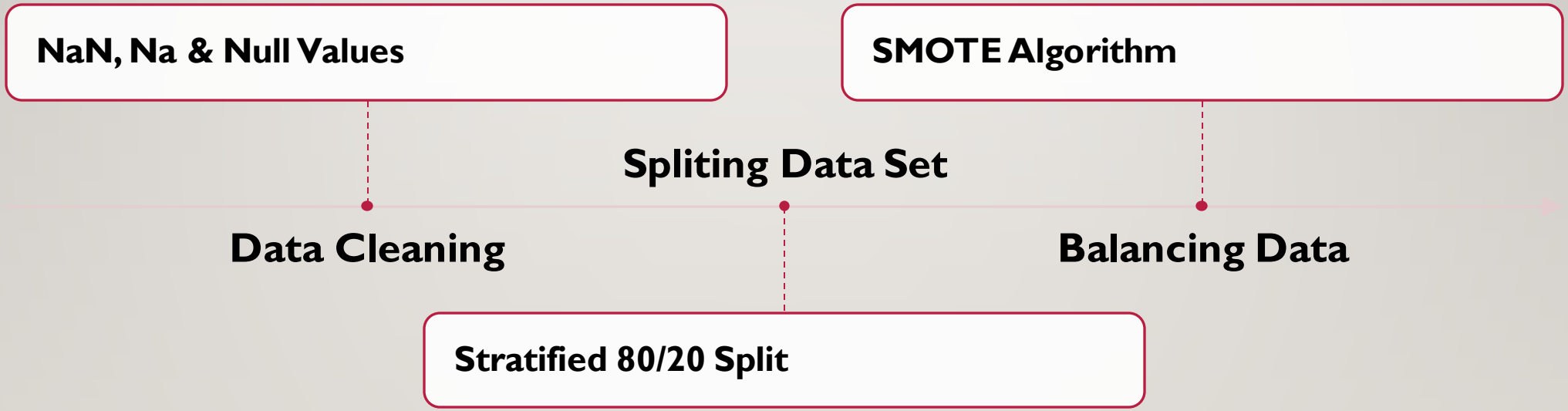
Independent Variables:

15 numerical features.

Dependent Variable:

'Class' Column with 1 or 0 values.

FEATURE ENGINEERING



MODEL SELECTION

Classification Models

Logistic Regression

Decision Tree Classifier

XGBoost Classifier

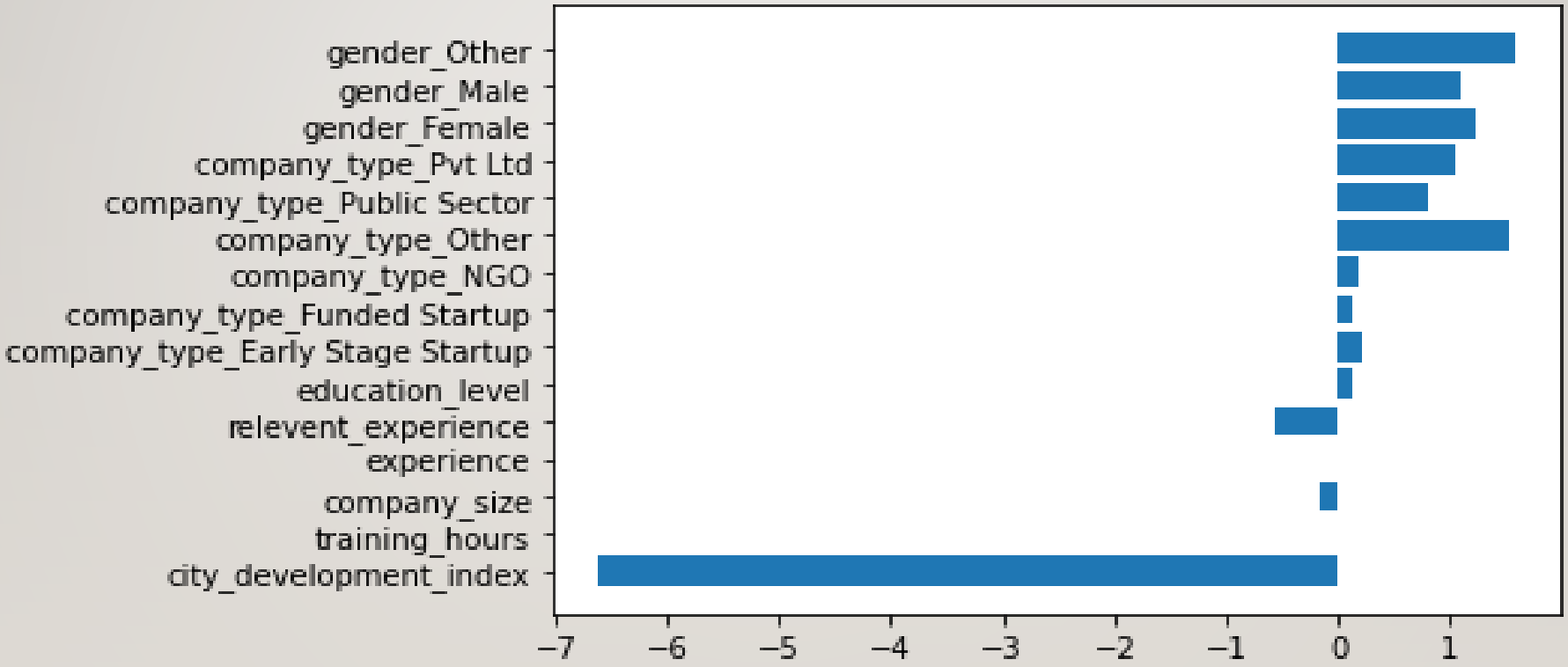
Best results

Logistic Regression

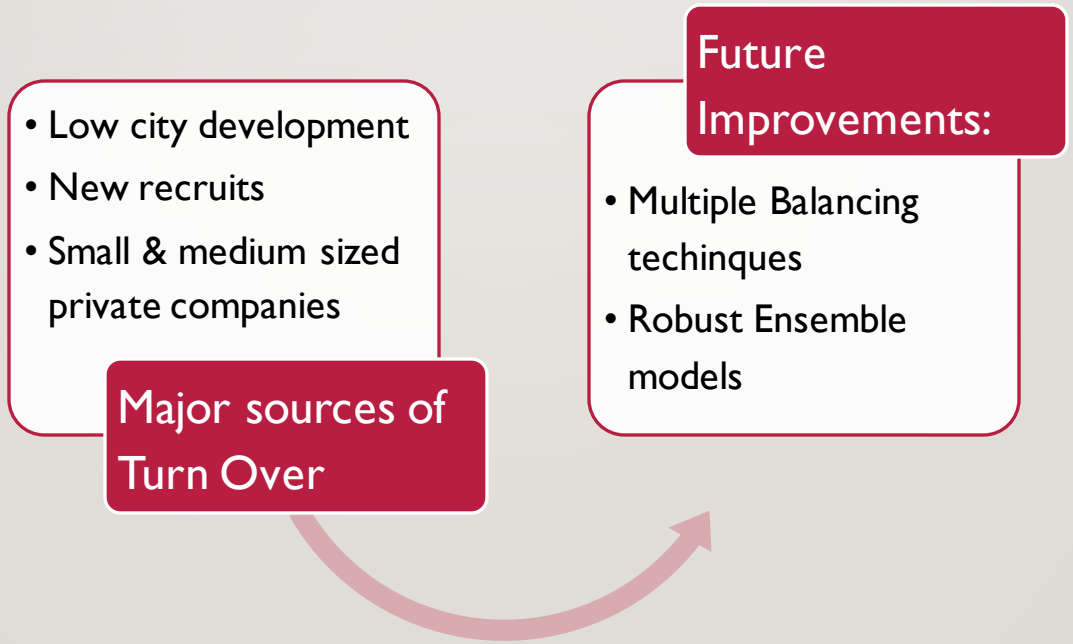
- Accuracy: 71%
- Recall:52.5%

GridSearchCV

FEATURES IMPORTANCES



CONCLUSION


Major sources of Turn Over
• Low city development
• New recruits
• Small & medium sized private companies
Future Improvements:
• Multiple Balancing techinques
• Robust Ensemble models

REFERENCE

- The full approach is presented in the following ipynb:
- https://github.com/mzaoualim/Coursera_IBM_Machine_Learning_Professional_Certificate/blob/main/supervised-machine-learning-classification/Project_Classification.ipynb