REPORT

Assignment 2 – MLG382

MAY 15, 2023

BELGIUM CAMPUS

Table of Contents

Group Members:

- Kagiso Kgobane
- Gregory Avvakoumides
- Acsevire Mboto

Report

This assignment is directed at predicting bankruptcy within Estate market. This prediction is done using Machine leaning the coding language Python. The model used for calculating and predicting the defaults and chances of filing for bankruptcy was the Gradient Boosting Classifier.

However, the data was highly imbalanced with only a small number of bankrupt companies, which we addressed by using the RandomUnderSampler algorithm to resample the target vector. This algorithm randomly removes samples from the majority class until both classes are balanced, thus preventing the model from being biased towards the majority class.

The Following are results of the assignment showing all the Calculations and Values derived from the Machine Learning code:

- The best hyperparameters:
  - n_estimators=50,
  - max_depth=3,
  - learning_rate=0.1.
- Cross-validation score: 0.9987012987012986
- Metrics:
  - Accuracy: 1.00
  - Precision: 0.99,
  - Recall: 1.00,
  - F1-score: 0.99.
- Top 10 Most Important Features:

```
Top 10 most important features:
        Feature    Importance
0    company_id  1.000000e+00
20     feat_20  6.596300e-16
18     feat_18  6.443010e-16
5       feat_5  6.189342e-16
45     feat_45  3.432077e-16
41     feat_41  2.464738e-16
38     feat_38  2.462048e-16
8       feat_8  2.441610e-16
53     feat_53  2.389505e-16
56     feat_56  1.760143e-16
```

This analysis provides valuable insights as to which companies can be at risk of bankruptcy in Poland, with efficient accuracy. This powerful tool can be used to inform risk management strategies for businesses and investors.