**FoMLSS/ DEPARTMENT**

**MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE IN FINANCE, AFE 7518-B**

Assessment: Individual Coursework (80% of Final Mark)

**Submission Date: 3.00 PM, April 30th, 2024.**

**INSTRUCTIONS TO CANDIDATES**

1. The maximum word count of the work should be about 2000 words, excluding the title page, contents page, tables, figures, and appendices.
2. The work should be written in easy-to-read English using an academic style and without journalistic hyperbole.
3. Pagination: each page is to be numbered consecutively from 1, except the first.
4. The first page is the title page, and the second page is the contents page.
5. The title page must contain the title, UB number and the word count.
6. The body of contents is to be sectionalized and numerically ordered.
7. The minimum font size is Times Roman 12 pt. or equivalent, with double or 1.5 line spacing.
8. Text is to be justified both left and right.
9. Tables and figures must have a title and be numerically labelled. If convenient, they are to be placed near the relevant text, or if large, exhibited following the conclusion.
10. All external sources of information are to be cited using Harvard style. Complete details of all citations are to be collected under the “List of References”.
11. Equations such as formulas should be entered using an equation editor and numbered. Complex derivations should be assigned to an appendix.
12. The complete work is to be organized in the following way:

* Title page
* Contents page
* Introduction – Body Text – Conclusion
* Any large tables and figures
* Any Appendices
* List of References

1. The following items must be submitted:

* **One word document including** a written report on main results, analysis, and discussion.
* **Python codes:** containing your data and analyses.

1. You **MUST** share the link to codes in Google Colab.

**Developing distress prediction models**

**Aim**:

The object of this task is to develop three models to predict the probability of distress of UK firms and classify firms into two groups of distress vs healthy.

* 1. **Data cleaning and creating features.**

You need to follow Mousavi, Ouenniche and Tone (2023) and take the following steps to collect the data set for your empirical analysis.

* 1. The raw dataset is provided on Canvas. Data belongs to UK firms which are listed on LSE for the period of 2015 to 2020.
  2. Generate financial ratios listed in table 1. Corporate governance items are collected directly from Eikon Refinitive.
  3. Since developing some models require minimum historical data, exclude companies that have been listed for less than two years.
  4. To minimise any bias related to excluding companies with missing data, you need to discard those companies with missing values for the main accounting variables (e.g., total assets, sales) and market-based variables (e.g., price) which are necessary for computing a variety of accounting and market-based ratios.
  5. To deal with the extreme values of any variable, the outlier values should be winsorised between the 1st and 99th percentile of each variable.
  6. To mark distress firms, follow Pindado et al. (2008) in defining a variable, say D, that equals 1 for financially distressed firms and 0 otherwise. A firm will be classified as financially distress if: (1) for two succeeding years, the company’s interest expense is more than its earnings before interest, taxes, depreciation and amortization (EBITDA), and (2) for two succeeding years, the company suffers a negative growth in market value.
  7. Divide sample into two sub-samples: Training sample (2015 – 2018) and Test sample (2019-2020).
  8. **Feature Selection & Models Development**
  9. Develop **three** optional supervised or unsupervised learning techniques. Each model should be developed twice (once without feature selection by choosing your own features, and once with feature selection). Therefore, you will end up with **six** models. You can try a different combination of features (explanatory variables) to select the best ones for developing your models.
  10. **Models Evaluation**
  11. Show if your model is over-fit or under-fit using the **test sample**.
  12. Test the accuracy of models using both the in-sample (training sample) and out of sample (test sample) approach.
  13. Compare the performance of models using Type 1 error (T1), Type 2 error (T2), Total accuracy, and **AUC** as measures. (Cut-off point is 50%)

Table 1: Financial Ratios and Information

|  |  |  |  |
| --- | --- | --- | --- |
| **Category** | **Ratio or item** | **Category** | **Ratio or item** |
| Liquidity | Current assets / Total assets  Cash and equivalent / current liabilities.  Current assets / current liabilities  Sales / inventory  Current assets / sales  Current assets / total liabilities | Solvency | Total Liabilities / sales  Total liabilities / assets  The book value of equity / Total liabilities  Interest/EBITDA |
| Market information | Lag (excess return)  Market capital  Market value / Total Equity | Cash flow | Operating cash flow / Total liabilities  Operating cash flow / Total assets |
| Asset utilisation | Working capital / sales  Quick assets / sales | Profitability | Net income / Total assets  Net income / Total Equity |

Marks will be awarded based on the following criteria:

Table 2: Assessment evaluation criteria

|  |  |  |
| --- | --- | --- |
| Criteria | Explanation | Weight |
| Date cleaning and creating features | Did the student create appropriate financial ratios (listed in table 1) correctly? Was the data cleaned and pre-processed correctly? | 20% |
| Feature Engineering and Selection | Did the student perform appropriate feature engineering and selection techniques? Did the student justify the choice of features used in the model? | 20% |
| Model Building and Evaluation | Did the student build and evaluate multiple supervised learning models? Did the student use appropriate evaluation metrics? Did the student use time series analysis to forecast stock prices? | 40% |
| Presentation and Report | Did the student’s report include appropriate background information and a clear explanation of the methodology used? Did the student’s report include a thorough evaluation of the model's performance and future recommendations? | 20% |

**References:**

Mousavi, M. M., Ouenniche, J., & Tone, K. (2023). A dynamic performance evaluation of distress prediction models. Journal of Forecasting, 42(4), 756-784.

Pindado, J., Rodrigues, L., & De la Torre, C. (2008). Estimating financial distress likelihood. Journal of Business Research, 61(9), 995-1003.

**-----End----**