QF624

Machine Learning and Financial Application

Group Project Guideline

Project Objective:
Able to research and dig into a real-world application of machine learning in finance, including risk management, portfolio optimization, or asset pricing.

Project Deliverables:
1. Project Report in Word, PDF, Markdown, or web application format.
2. Datasets, preferably in CSV format, or provide link to datasets if size is too big to upload. Provide reference to the source if the data is directly downloaded via scripts.
3. Execution scripts for data processing, modeling, and analysis (R/Python scripts or other open-source languages)
4. Presentation slides in PowerPoint format (write the name of the presenter on each slide)
5. Only one member in a group needs to submit the deliverables. The file needs to be properly named

Project Report General Requirements and Guidelines:
1. Format: Max 20 pages (excl. appendix and executive summary).
2. One-page executive summary that summarizes the most important findings.
3. Clearly define the Business problem (or opportunity) statement(s).
4. Explain the business outcome measures and desired targets. How would we know if the business outcome(s) had been achieved?
5. Explain the ML solution and how it solves the business problem or opportunity.
6. Define and explain the model performance measures and evaluation metrics.
7. Find suitable dataset(s) that can be used to develop at least two different ML models.
8. Demonstrate and explain at least two ML models using software (R/Python or other languages). The script will need to be submitted too. The submitted dataset and analysis files/scripts must be sufficient for someone to reproduce and verify your work.
10. The entire ML solution (if multiple components) must be explained and cover at least one model covered in class. The dataset and demonstration can choose to focus on just one component of the solution due to limitations of time and access to actual data.
11. Additional bonus for the following:
    - Tackle real-world challenges (ongoing competitions or industry projects)
    - Explore additional ML techniques behind class coverage

Presentation Slides General Requirements and Guidelines:
1. A deck of PowerPoint presentation slides in PPTX format that summarizes your important findings and recommendations. Guideline: **max 20 minutes presentation and 5 minutes Q&A.**
2. All students must speak and present their slides.
3. Write the slide speaker's name in the corner of each slide so that marks can be attributed to that speaker.
4. Screenshots of the most important visualization/dashboard in the slides appendix section.

Submission Deadlines:
1. Project proposal: **end of week 4**
2. Project slides: **one day before presentation**
3. Project report, slides, and scripts: **end of week 10**

Final presentation schedule:
The presentation is scheduled for week 8, 9, and 10.
Please indicate your presentation preference to TA by the end of week 5. Random assignment will be conducted if more than five groups choose to present in the same week.

Reference data and codes:
Data set 1: Credit agency - company bankruptcy prediction
Data file:
https://www.kaggle.com/datasets/fedesoriano/company-bankruptcy-prediction
Sample code by SANJOY MONDAL:
https://www.kaggle.com/code/sanjoymondal0/company-bankruptcy-prediction-acc-97

Data set 2: Retail bank - predicting churn for bank customers
Data file:
https://www.kaggle.com/datasets/adammaus/predicting-churn-for-bank-customers
Sample code by AHMET CAN KARAOĞLAN:
https://www.kaggle.com/code/ahmetcankaraolan/churn-prediction-using-machine-learning

Data set 3: Equity trading - predicting stock closing price
Data file: no data file. Price data of any Stock Ticker can be read from Yahoo directly.
Sample code by AKSHAY SHARMA:
https://www.kaggle.com/code/akshaysharma001/predicting-stock-closing-price-99-accuracy/notebook

Data set 4: Diamond trading - diamond price prediction
Data file: https://www.kaggle.com/datasets/shivam2503/diamonds
Sample code by SURAJ JHA:
https://www.kaggle.com/code/surajjha101/regression-models-diamond-price-prediction/notebook