BFF5555 Project

Semester 2 2024

Project Overview

This capstone project assesses your ability to apply machine learning concepts and frameworks covered in this unit to build a predictive model for a financial application.

- **Assessment Type**: Individual project
- **Total Marks**: 100 (Contributes 40% to the final grade)
- **Due Date**: 11:55 pm, Wednesday, October 9th
- **Late Penalty**: 5 marks per day for late submissions without prior approval

Submission Requirements:

1. A PDF report documenting your model building process, including analysis and conclusions.
2. A Jupyter Notebook containing all Python code for your implementation. The notebook should be self contained, meaning the marker can run all the codes therein without making any adjustment. Clear all outputs before submitting the notebook.
3. A data file.
4. Use of AI statement (if applicable). See below.

Each of these files should be named as FirstnameLastName. For example, XiaoWang_REPORT.pdf, XiaoWang_CODE.ipynb, XiaoWang_data.csv, XiaoWang_AI.pdf.

There is no set number of pages and styling. Some students prefer to present in bullet points, others choose a more narrative style.

Project Objective

You are required to develop a machine learning model to predict positive market movements (uptrend). This prediction task will be treated as a binary classification problem, where the target variable is binary [0, 1].

Key Tasks:

- Select one ticker symbol (stock or ETF) of interest.
- Focus on predicting short-term returns (e.g., daily or weekly).
- Follow the **six-step model building process** discussed in class.

- Perform all computation, plotting and model implementation in a Jupyter Notebook. Document the model building process in a PDF file. The PDF file should contain tables and plots generated from the Notebook.

Additional guidance:

Data and preprocessing

- For daily predictions, a dataset of 5 years should be sufficient. For weekly predictions, select a suitably longer timeframe.
- The features must be derived from OHLCV data (Open, High, Low, Close, Volume) available from Yahoo Finance. You are expected to:
- Construct features such as intra-period price range, sign and magnitude of past returns:

Feature	Formula	Description
O-C, H-L	Open - Close, High - Low	Intraperiod price range
Sign	sign $[r_t = \ln\left(\frac{P_t}{P_{t-1}}\right)]$	Sign of return or momentum
Past Returns	$r_{t-1}, r_{t-2}, \ldots$	Lagged returns

- Construct additional technical indicators using `Pandas-TA`.
- The total number of initial features is your design choice.
- Apply feature selection techniques such as feature importance ranking and regularization, to refine the feature set.

Algorithm and model training and selection

- Include all suitable machine learning algorithms covered in this unit. Perform hyperparameter tuning and model selection on the training set. Select the final model based on cross validation.
- The train-test split ratio is your design choice.
- Define the response variable based on your selected ticker and prediction frequency. For example, you may choose to label small positive returns (below 0.25%) as negative for weekly returns.

Model evaluation

- The evaluation must include relevant metrics plus a backtest, the latter should report the annualised return and Sharpe ratio for your strategy compared to a buy-and-hold benchmark.

Support

- Post any project-related queries on the Assessments Forum.

Use of Generative AI

You may use Generative AI tools to assist with the Python coding aspects of this assessment. If you choose to do so, you must submit a separate document that includes:

- Acknowledgment of AI usage, with a clear explanation of how and where it was used.
- Documentation of the AI tool employed, including screenshots of the prompts and any interactions with the AI.

For guidance on how to complete this document, please refer to this link.

While the use of Generative AI is permitted, it is not mandatory.

Marking rubric

Criteria	Marks	Description
1. Ability to follow the six-step process	**10**	- Clearly structured workflow following the six-step process discussed in class. - Logical progression and adherence to all key steps (e.g., data collection, preprocessing, etc.).
2. Competent execution of technical aspects	**10**	- Data collection: Sufficient, relevant data collected and explained. - Preprocessing: Correct handling of missing data, normalization, and transformations. - Model evaluation: Robust validation methods (train-test split, cross-validation, backtest) applied and evaluated clearly. - Coding: efficient Python programming with appropriate use of relevant packages.
3. Creative application in feature engineering and model selection	**10**	- Thoughtful feature engineering (going beyond basic features demonstrated). - Exploration of advanced techniques or creative use of domain knowledge. - Novel algorithm/model selection and tuning (attempt to innovate or tailor to dataset).
4. Quality of documentation	**10**	- Well-commented code in Jupyter notebook. - Clear, concise, and thorough report, explaining choices, findings, and reflections. - Screenshots of AI tools used and interactions documented (if applicable).