

COS30018 - Option C - Task 7: Extension - Sentiment-Based Stock Price Movement Prediction

In this task, you will extend your existing stock price prediction model by incorporating sentiment analysis from external textual sources (such as financial news or social media). The goal is to develop a classification model that predicts whether the stock price will rise or fall on the following trading day.

Your work should address the following components:

- **Data Collection & Preprocessing (5 marks)**
Collect relevant textual data using appropriate APIs or publicly available sources. Ensure the data is time-aligned with your historical stock price dataset. Clearly document your data sources, any filtering or cleaning steps, and how the data was prepared for analysis.
- **Sentiment Analysis (5 marks)**
Use sentiment analysis tools to generate sentiment scores from the collected text. These scores should be aggregated at a daily level to match the frequency of stock data. You may experiment with different tools or configurations and discuss their suitability for your dataset.
- **Feature Engineering & Modelling (5 marks)**
Create input features that combine both historical stock data (e.g., technical indicators, lagged values) and sentiment scores. Train a classification model using these features to predict whether the next day's closing price will be higher or lower than the current days. Justify your model choice and preprocessing steps.
- **Evaluation (5 marks)**
Evaluate your model using appropriate metrics such as accuracy, precision, recall, F1-score, and confusion matrix. Compare its performance against a baseline model that excludes sentiment features to assess the added value of sentiment data.
- **Independent Research Component (5 marks)**
Explore and integrate at least one advanced enhancement — such as alternative data sources, domain-specific sentiment techniques, or more sophisticated modelling approaches. This part of the task allows you to tailor the project to your interests and demonstrate deeper technical engagement.
- **Reporting (5 marks)**
Submit well-documented code (e.g., via GitHub) along with a report that explains your overall approach, rationale for design decisions, key results, and any challenges encountered. Include visualizations where appropriate to support your analysis.

Due date: 11:59pm Sunday 2 November 2025

Assessment Criteria:

You can get up to 30 marks for successfully completing all requirements of Task C.7.