Project Title:

Stock Market Prediction: A Study on Ensemble Techniques with Machine Learning and Deep Learning Models

Research Question:

How does ensemble methods such as stacking multiple machine learning models compare to individual models in stock market prediction.

Project Background:

Stock market is an essential part of economy. Investors face challenges due to the constantly changing and unpredictable nature of the financial market. To minimize risks and optimize investment approaches, it is essential to accurately predict financial returns in the volatile environment of stock markets. (Raipitam et al., 2023).
This project aims to see the effectiveness of ensemble methods such as stacking multiple machine learning models. In this project, we ensemble outputs of various models such as Random Forest, SVM, LSTM, and XGBoost. In this project, we use the share price data of BAE Systems Plc from 1st October 2019 to 1st October 2024. The final goal of the project is to compare the accuracy metrics of the ensemble approach and the individual models.

Objectives:

1) To do extensive background research on stock market predictions, ensemble methods, machine learning and deep learning models
2) To undertake the necessary preprocessing steps for the dataset.
3) To access the performance of individual models such as Random Forest, SVM, and LSTM in stock price prediction.
4) To design and access the performance of an ensemble model by combining predictions from individual models (Random Forest, SVM, LSTM, and XGBoost) using stacking.
5) To analyze and compare the accuracy of individual models with an ensemble approach.
6) To identify the benefits and limitations of ensemble methods for stock market prediction.

References:

Baswaraj, D., Natchadalingam, R., Rekha, R.S., Sahni, N., Divya, V. and Reddy, P.C.S., 2023, November. An Accurate Stock Prediction Using Ensemble Deep Learning Model. In *2023 International Conference on Research Methodologies in Knowledge Management, Artificial Intelligence and Telecommunication Engineering (RMKMATE)* (pp. 1-7). IEEE. https://ieeexplore.ieee.org/abstract/document/10369255

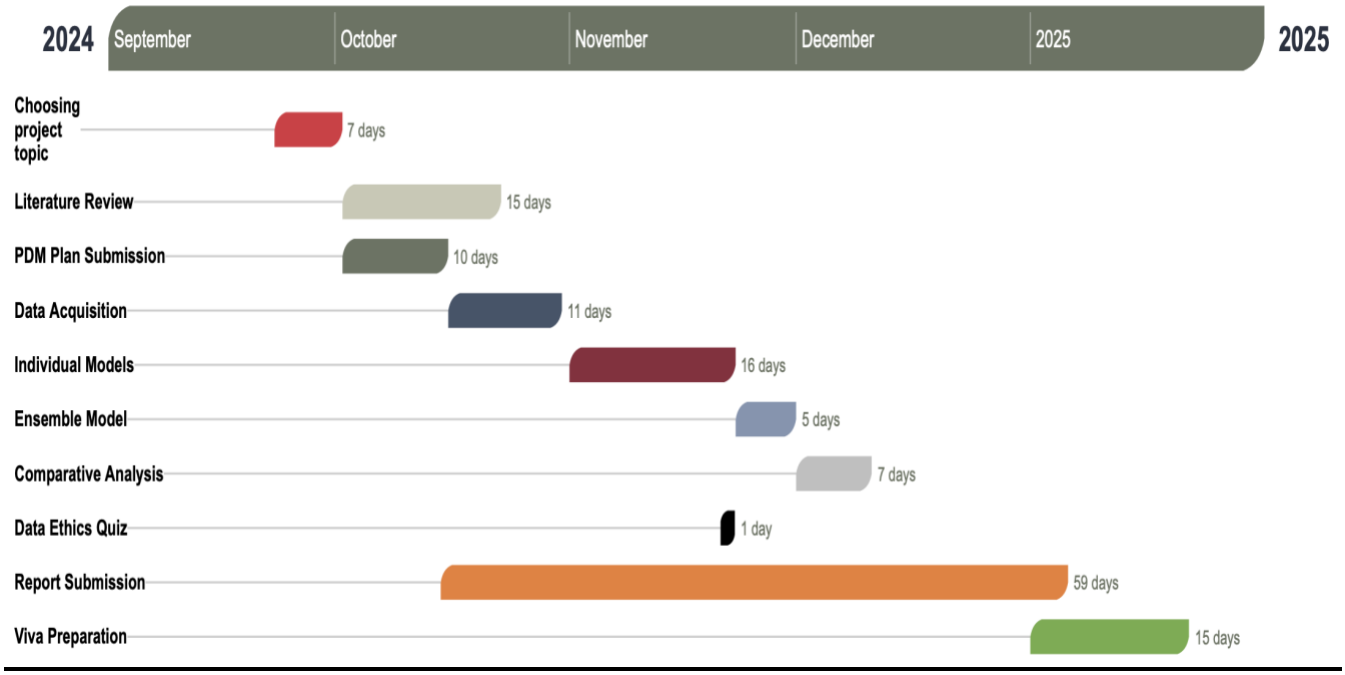
He, K., Yang, Q., Ji, L., Pan, J. and Zou, Y., 2023. Financial time series forecasting with the deep learning ensemble model. *Mathematics*, *11*(4), p.1054. https://www.mdpi.com/2227-7390/11/4/1054

Raipitam, S.K., Kumar, S., Dhanani, T., Bilgaiyan, S. and Gourisaria, M.K., 2023, September. Comparative study on stock market prediction using generic CNN-LSTM and ensemble learning. In *2023 International Conference on Network, Multimedia and Information Technology (NMITCON)*(pp. 1-6). IEEE. https://ieeexplore.ieee.org/document/10275849

Song, H. and Choi, H., 2023. Forecasting stock market indices using the recurrent neural network based hybrid models: CNN-LSTM, GRU-CNN, and ensemble models. *Applied Sciences*, *13*(7), p.4644. https://www.mdpi.com/2076-3417/13/7/4644

Task List & Project Timeline:

Tasks	Description	Start Date	End Date
Choosing Project	Selecting a topic for the final project and thinking about a research direction	23-09-2024	01-10-2024
Literature Review	Reviewing research papers on the chosen topic and selecting a few papers according to our direction	02-10-2024	22-10-2024
PDM Plan Submission	Prepare project overview, project timeline, and data management plan	02-10-2024	15-10-2024
Data Acquisition	Gather and pre-process the required dataset for the project	22-10-2024	30-10-2024
Individual Models Development	Develop and test Random Forest, SVM, and LSTM models for Stock Prediction	01-11-2024	22-11-2024
Ensemble Model Development	Develop and test the ensemble of the three individual models and XGBoost on Stock data	23-11-2024	30-11-2024
Comparative Analysis	Compare prediction accuracy of individual and ensemble models	01-12-2024	10-12-2024
Data Ethics Quiz	Quiz on data ethics	21-11-2024	21-11-2024
Report	Writing Background, Methodology, Results, Limitations, and Conclusion	15-10-2024	31-12-2024
Final Project Report Submission	Submit FPR	01-01-2025	06-01-2025
Viva Preparation	Viva Preparation	01-01-2025	21-01-2025
Viva	Viva Schedule	07-01-2025	21-01-2025



Data Management Plan:

Overview of dataset

The dataset consists of historical stock prices of BAE Systems Plc (BA.L). BAE Systems Plc is a defence and aerospace company headquartered in the United Kingdom. The data includes daily prices such as open, high, low, close, adjusted close, and volume. The dataset includes stock prices from 1st October 2019 to 1st October 2024.
The data is sourced from Yahoo Finance through the Python 'yfinance' library. Yahoo Finance collects financial data from stock exchanges globally to provide public access to financial data for research, investment analysis, and historical tracking.

Data Collection

The stock price data is collected using the 'yfinance' Python library. The data is obtained through an API request to Yahoo Finance. Yahoo Finance BA.L is the official website for stock prices of BAE Systems Plc.

Metadata

The dataset will be downloaded as a pandas dataframe, which can be saved as a CSV file for further analysis. The data size is under 1MB. The dataset consists of daily stock prices, with approximately 1250 records for the five years.

Document Control

The GitHub repository stores the project's dataset, report, and code. I intend to commit weekly.
GitHub link

ReadMe File

The ReadMe file in the GitHub repository will include an overview of the project, a few lines of code, and references used in the project.

Security and Storage

The report and code will be backed up weekly. The data will be stored in GitHub, OneDrive, and on my laptop. The GitHub repository will be made private, with access only to staff and markers.

Ethical Requirements

The dataset does not contain any personal data as it contains only stock market data. So, our dataset comes under GDPR guidelines.
As it contains only stocks, the dataset confirms UH's ethical policies.
We have permission to use the data from Yahoo Finance as it is publicly available and free to use for research purposes.
Yahoo Finance provides financial data transparently and ethically. The data is collected from stock exchanges.