PROJECT AND DATA MANAGEMENT PLAN

Project Plan

Title: Forecast Stock Market Trends and Portfolio Optimization Using Neural Networks and Machine Learning Algorithms.

Research question:

1) Which machine learning algorithms are the most efficient at forecasting stock values in the FTSE 100?
2) Does the use of technical indicators such as Moving Averages, RSI, MACD enhance the forecasting accuracy of machine learning models for stock returns?

Project Objectives:

1) Construct a machine learning model that utilizes Yahoo Finance data to forecast stock prices from different sectors and optimize stock investments.
2) Investigate the impact of different technical indicators on the capabilities which impact the prediction of machine learning models.

Summary of project and background:

The objective of this study is to assess the effectiveness of standard momentum investing techniques in comparison to advanced models based on machine learning in forecasting returns on the stock market and optimizing investment portfolios. The research will use historical stock price data from the FTSE 100 index using Yahoo Finance to deploy and assess several machine learning methods such as LSTM, CNN, Gradient Boosting Machines(GBM). The objective is to evaluate whether approach yields greater returns and superior risk-adjusted performance over a period. In addition, the research will investigate the use of technical indicators and ensemble learning methods to improve the accuracy of predictions and the performance of portfolios.

Technical indicators which are mathematically derived from past data on price, volume, or open interest. They are often used in technical analysis to forecast future price fluctuations. Technical indicators include moving averages, moving average convergence divergence (MACD), Relative-strength-index(RSI).

Reference List:

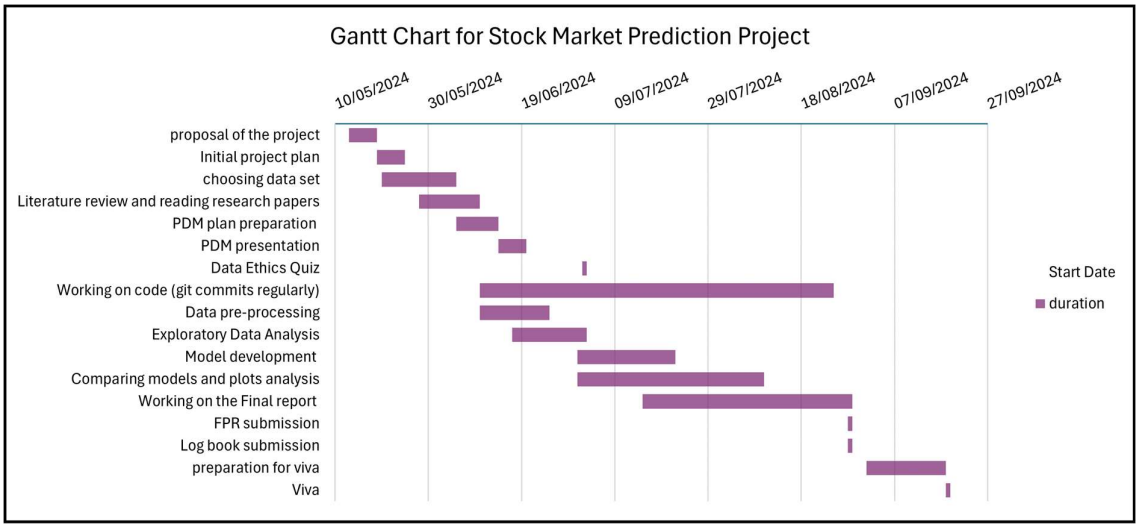
1) Y. K. Pardeshi and P. P. Kale, "Technical Analysis Indicators in Stock Market Using Machine Learning: A Comparative Analysis," *2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, Kharagpur, India, 2021, pp. 1-6, doi: 10.1109/ICCCNT51525.2021.9580172.
2) P. P. Kadu and G. R. Bamnote, "Comparative Study of Stock Price Prediction using Machine Learning," *2021 6th International Conference on Communication and Electronics Systems (ICCES)*, Coimbatre, India, 2021, pp. 1200-1204, doi: 10.1109/ICCES51350.2021.9489170.
3) B. Omar, B. Zineb, A. Cortés Jofré and D. González Cortés, "A Comparative Study of Machine Learning Algorithms for Financial Data Prediction," *2018 International Symposium on Advanced Electrical and Communication Technologies (ISAECT)*, Rabat, Morocco, 2018, pp. 1-5, doi: 10.1109/ISAECT.2018.8618774.

Task List and Project Timeline

Task list:

S/No	Task	Start Date	End Date
1	Proposal of the project	13/05/2024	19/05/2024
2	Initial project plan	19/05/2024	25/05/2024
3	Choosing data set	20/05/2024	05/06/2024
4	Literature review and reading research papers	28/05/2024	10/06/2024
5	PDM plan preparation	05/06/2024	14/06/2024
6	PDM presentation	14/06/2024	20/06/2024
7	Data Ethics Quiz	02/07/2024	02/07/2024
8	Working on code (with GitHub commits)	10/06/2024	25/08/2024
9	Data pre-processing	10/06/2024	25/06/2024
10	Exploratory Data Analysis	17/06/2024	03/07/2024
11	Model development	01/07/2024	22/07/2024
12	Comparing models and plots analysis	01/07/2024	10/08/2024
13	Working on the Final report	15/07/2024	29/08/2024
14	FPR submission	29/08/2024	29/08/2024
15	Logbook submission	29/08/2024	29/08/2024
16	Preparation for viva	01/09/2024	18/09/2024
17	Viva	19/09/2024	19/09/2024

Gantt Chart:



Data Management Plan

Data Collection:

- Data is collected from the Yahoo Finance website and downloaded into collab by yfinance API. the yfinance module is installed on Colab notebook and it enable to download share price data for specific stock by entering the ticker symbol directly for a given period and also access other information also.

Overview of the Dataset:

- The dataset will contain the data of 25-30 stocks from different sectors. The dataset will contain columns like: date, Open, High, Low, close, Adj close, volume.

Summary of data:

- Format would be CSV files.
- Records would be approximately 100K rows for all stocks.
- Data getting directly downloaded from yfinance API into colab.

Document control:

- Git is my version control system, which supports versioning and collaboration features. I will be committing data and files in a Git repository. Made a first push to the repository, including the dataset files and documentation, including a README file outlining its contents, properties, and preparation methods. My Github link is : https://github.com/Sasidhar0709/Project-Stock-market-Analysis

Metadata:

- My README file contains information on the dataset's attributes, kind of data, and source. This metadata offers a high-level summary of the dataset, including information about its properties, libraries, and functions.

Security and Storage:

- The code and the reports are backed up regularly. Code is uploaded to Git and saved in colab and the report and other documentation will be uploaded to OneDrive regularly. Data is shared with internally with the shared links.

Ethical Requirements:

- As the data is related to financial sector information, the data does not contain any personal information, and it meets the GDPR requirements and meets the UH ethical policies.