# PROJECT PLAN

**Name: Susairaj Anthony**
**Student ID: 22071718**

**PROJECT TOPIC:**

**"Predicting Stock Market Trends Using Machine Learning Model (LSTM)"**

**RESEARCH QUESTIONS:**

- What impact does market volatility have on the stock price predictions made by the LSTM model?
- What effects do different financial indicators (e.g., RSI, moving averages, and trade volume) have on the precision of LSTM models used to predict the stock market?
- What are the limitations on LSTM models' ability to forecast changes in the stock market, and how may these limitations be overcome?
- How would data preprocessing techniques (e.g. normalization and handling missing values) influence the way LSTM models perform when making stock market predictions?

**PROJECT OBJECTIVES:**

- **Data Collection and Preprocessing:** Collect and preprocess historical stock market data from Yahoo Finance, handling missing values and normalizing for analysis and model training.
- **Data Exploration and Analysis:** Analyze data to identify patterns, trends, and correlations using statistical and visualization techniques for stock market insights.
- **Model Development and Implementation:** Develop and implement diverse machine learning models, including LSTM networks, for forecasting stock market movements, experimenting with configurations to optimize performance.
- **Model Evaluation:** Assess model performance using metrics like MAE, RMSE, and accuracy, comparing LSTM models with other algorithms to identify the most effective approach.
- **Insights and Recommendations:** Support traders and investors by providing practical insights and suggestions based on model forecasts. Give concise results that emphasize risks and investment opportunities based on anticipated market changes.

**SUMMARY OF PROJECT AND BACKGROUND:**

- This project utilizes Long Short-Term Memory (LSTM) neural networks to predict stock market trends.
- Recurrent neural networks (RNNs) of the LSTM network type are highly effective at detecting sequential dependencies across time.
- The LSTM model is fed historical data from Yahoo Finance, which includes trading volumes, price fluctuations, and other financial parameters.
- The goal of the LSTM network is to produce accurate predictions of future market movements by learning from historical patterns and trends.
- The LSTM model's performance is assessed by looking at its capacity to predict stock prices and give traders and investors insightful information that helps them make wise decisions in the financial markets.
- The goal is to develop a robust predictive model to provide valuable insights for traders and investors, aiding in more informed decision-making in financial markets.

**REFERENCE:**

- The data set retrieved from Yahoo Finance.
- Hochreiter, Sepp & Schmidhuber, Jürgen. (1997). Long Short-term Memory. Neural computation. 9. 1735-80. 10.1162/neco.1997.9.8.1735.
- Thanaki, J 2018, Machine Learning Solutions : Expert Techniques to Tackle Complex Machine Learning Problems Using Python, Packt Publishing, Limited, Birmingham. Available from:ProQuest Ebook Central. [11 June 2024].
- Atienza, R 2018, Advanced Deep Learning with Keras : Apply Deep Learning Techniques, Autoencoders, GANs, Variational Autoencoders, Deep Reinforcement, Learning, Policy Gradients, and More Packt Publishing, Limited, Birmingham. Available from: ProQuest Ebook Central. [11 June 2024].

# PROJECT TIMELINE: GANTT CHART

## Predicting Stock Market Trends Using Machine Learning Model (LSTM)

**Project Start Date: 5/14/2024**

PROJECT KEY: Lecture | Quiz | Submission | Supervisor Meeting | Coding Preparation | Viva | Report Preparation

| | | | May 13, 2024 | May 20, 2024 | May 27, 2024 | June 3, 2024 | June 10, 2024 | June 17, 2024 | June 24, 2024 | July 1, 2024 | July 8, 2024 | July 15, 2024 | July 22, 2024 | July 29, 2024 | August 5, 2024 | August 12, 2024 | August 19, 2024 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| TASK | DESCRIPTION | TASK DATES | Wk 1 | Wk 2 | Wk 3 | Wk 4 | Wk 5 | Wk 6 | Wk 7 | Wk 8 | Wk 9 | Wk 10 | Wk 11 | Wk 12 | Wk 13 | Wk 14 | Wk 15 |
| Lecture - 1 | Module overview and choosing a project | 14/May/24 | Lecture - 1 | | | | | | | | | | | | | | |
| Lecture - 2 | Project plan and literature search | 21/May/24 | | Lecture - 2 | | | | | | | | | | | | | |
| choosing a project | Agree topics and dataset with supervisors by email | 22/May/24 | | Topic: Stock Market | | | | | | | | | | | | | |
| Supervisor Meeting - 1 | Discussed the Project Topics | 21/May/24 | | | Initial Plan | | | | | | | | | | | | |
| Lecture - 3 | Data Ethics and PDM Preparation | 3/Jun/24 | | | | Lecture - 3 | | | | | | | | | | | |
| Supervisor Meeting - 2 | PDM Presentation | 13/Jun/24 | | | | PDM Preparation | PDM Presentation | | | | | | | | | | |
| Lecture - 4 | Data Ethics | 17/Jun/24 | | | | | | Lecture - 4 | | | | | | | | | |
| Supervisor Meeting - 2 | Literature and Coding Discussion | 21/Jun/24 | | | | | | Literature preparation | | | | | | | | | |
| Quiz | Ethics Quiz | 2/Jul/24 | | | | | | | | 02 July 2024 | | | | | | | |
| Literature Review | Research on LSTM-based stock market analysis | 1 Jul - 10 Jul | | | | | | | | Literature Review | | | | | | | |
| Lecture - 5 | Results and Preparation of FPR and viva | 15/Jul/24 | | | | | | | | | | Lecture - 5 | | | | | |
| Coding | Investigate featuring | 17 Jun - 30 Jun | | | | | | Investigate featuring | | | | | | | | | |
| | Data Preprocessing | 1 Jul - 14 Jul | | | | | | | | Data Preprocessing | | | | Mock Viva | Mock Viva | | |
| | Dataset Analysis with Model | 1 Jul - 28 Jul | | | | | | | | Dataset Analysis with LSTM Model | | | | | | | |
| | Data Visualization | 15 Jul - 28 Jul | | | | | | | | | | Data Visualization | | | | | |
| | Prediction Evaluation | 22 Jul - 31 Jul | | | | | | | | | | | Prediction | | | | |
| Final Report Preparation | Write a Summary | 29 Jul - 7 Aug | | | | | | | | | | | | Summary | | | |
| | Describe the Methodology and Background | 29 Jul - 7 Aug | | | | | | | | | | | | Methodology and Background | | | |
| | Compose the Outcomes | 8 Aug - 14 Aug | | | | | | | | | | | | | Results | | |
| | Add a Conclusion | 15 Aug - 18 Aug | | | | | | | | | | | | | | Conclusion | |
| | Report Submission | 19 Aug - 23 Aug | | | | | | | | | | | | | | | Submit FPR |
| Viva | Viva Preparation | 1 Sep - 10 Sep | | | | | | | | | | | | | | | |
| | Viva Meeting | 11 Sep - 19 Sep | | | | | | | | | | | | | | | |

**Gantt Chart Link:** Data Management Project Plan 1.xlsx

# DATA MANAGEMENT PLAN

**DATA COLLECTION:**

The data for this project will be sourced from Yahoo Finance.

**OVERVIEW OF THE DATASET:**

- Yahoo Finance provides a comprehensive collection of financial market data, including historical stock prices, trading volumes, and market indices.
- The data spans various financial instruments like stocks, indices, and commodities.
- The data provided was collected by Yahoo Finance from various reliable financial exchanges and sources.

**SUMMARY OF DATA:**

- The dataset is available in CSV format.
- The number of records for Barclays PLC (BARC.L) over a ten-year period, assuming daily trading data, would be almost 8,000.
- The data for Barclays stock over ten years, along with files for processing and analysis, is expected to be around 100 KB.

**ETHICAL STANDARDS:**

The Yahoo Finance data set consists mainly of financial data and does not include any personal data.

- Given the impersonal type of the data and the fact that it does not include any personal information, adherence to General Data Protection Regulation for data processing and storage is crucial.
- The data usage adheres to the ethical policy of the University of Hertfordshire. It will be evaluated to ensure it meets with the standards.
- We can utilize Yahoo Finance data for non-commercial, personal purposes. Terms of use and license agreement need to be reviewed frequently to ensure ongoing compliance.

**DOCUMENT CONTROL:**

- **File naming:** As a reason of clarity and accessibility, a standard file naming convention will be followed. This will be the format: **Stock Market Analysis_22071718** (Topic and Student ID).
- **Version control:** Version control will be maintained by using GitHub to manage revisions to code files and data.
- **GitHub URL:** The project repository can be accessed at GitHub Link

**METADATA:**

- **User Documents:** The GitHub repository will include a thorough README.md file that defines the dataset, illustrates how to use the data, along with step-by-step instructions for executing the code.

**SECURITY AND STORAGE:**

- **Secure Storage:** The privacy and security of data will be maintained using encrypted, secure storage devices. This includes the use of secure cloud storage services and encrypted Google Drives.
- **Data sharing:** Only authorized personnel will have restricted access to the GitHub repository.
- **Backups:** For prevention of data loss, regular backups will be planned. To provide reliability, daily backups will be created and stored on both local and cloud storage systems. Moreover, autosave will be implemented to ensure the completeness and consistency of backups.