

Presentation of Project and Data Management Plan

Project Plan:

The project plan includes the following tasks:

➤ **Project Title:** Stock market Prediction

➤ **Research Question:**

1. What is the impact of the global stock indexes on short-term fluctuations of the ISE? (indexes are featured columns)
2. Which ML model gives the best predictions accuracy with the low amount of data? (Data range from 500 to 1000 rows)
3. From dataset, what are the key features and data processing methods that improves the model performance or efficiency?
4. Challenges and limitations faced in forecasting the stock market using ML models.

➤ **Project Objectives:**

1. Finding dataset, collect data and preprocess data.
2. Feature Engineering (identifying the key features that impact the performance of the model prediction)
3. Exploratory Data Analysis (EDA)

➤ **Summary of project and background:**

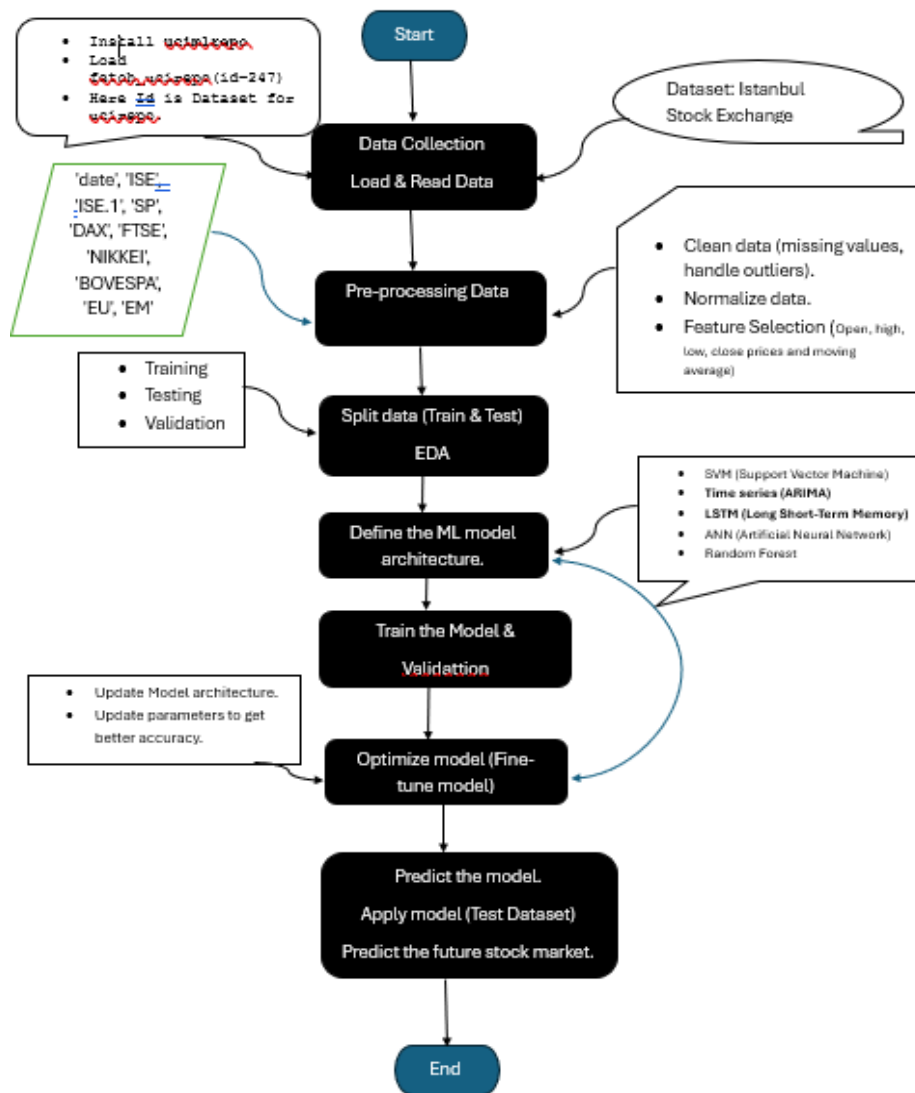
The project's main goal is to use the ML models to forecast the ISE performance. Based on the data chosen and relevant indices, the research project aims to analyse several models and identify which model delivers the most accurate predictions. Finding accurate predictions can help the investors, make smart decisions in maximizing returns and avoid risks in losing the money.

➤ **Reference list:**

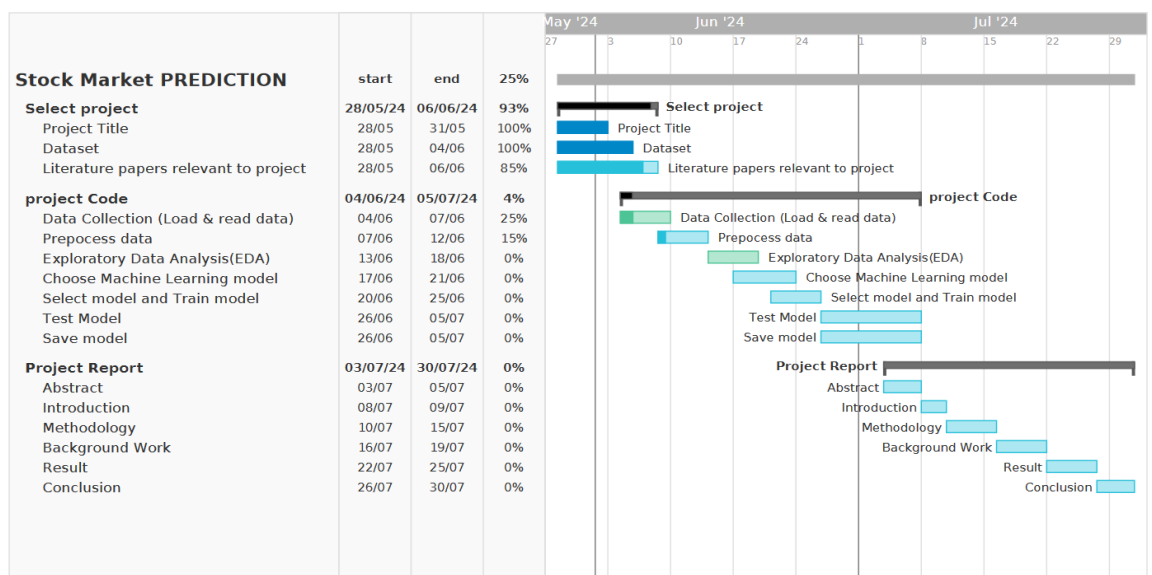
1. Chatterjee, A., Bhowmick, H. and Sen, J., 2021. Stock Price Prediction Using Time Series, Econometric, Machine Learning, and Deep Learning Models. arXiv. Available at: <https://arxiv.org/pdf/2111.01137>
2. Phuoc, T., Anh, K., T., P., Tam, H., P., Nguyen, C., V., 2024. Applying machine learning algorithms to predict the stock price trend in the stock market – The case of Vietnam. Nature. Available at: <https://www.nature.com/articles/s41599-024-02807-x>
3. Sonkavde, G., Dharrao, D., S., Bongale, A., M., Deokate, S., T., Doreswamy, D., Bhat, S., K., 2023. Forecasting Stock Market Prices Using Machine Learning and Deep Learning Models: A Systematic Review, Performance Analysis and Discussion of Implications. Mdpi. Available at: <https://www.mdpi.com/2227-7072/11/3/94>

Task List and Project Timeline:

➤ **Task list**



➤ Project Timeline in Gantt Chart



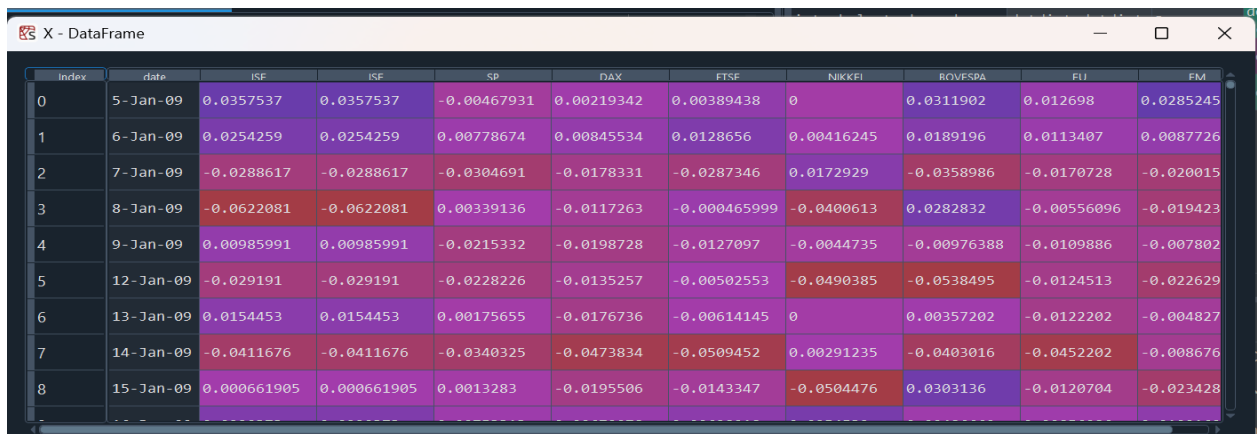
Data Management Plan:

➤ Overview of the dataset & Data collection:

I found this dataset from **University of California Irvine (UCI)** machine learning repository. The dataset contains stock values of different countries along with the "Istanbul stock market" (ISE). Here ISE is the target value, where its stock will be dependent on the other countries stock market.

Dataset-Link: <https://archive.ics.uci.edu/dataset/247/istanbul+stock+exchange>

Below is the sample dataset.



index	date	ISE	ISE.1	SP	DAX	FTSE	NIKKEI	BOVESPA	EU	EM
0	5-Jan-09	0.0357537	0.0357537	-0.00467931	0.00219342	0.00389438	0	0.0311902	0.012698	0.0285245
1	6-Jan-09	0.0254259	0.0254259	0.00778674	0.00845534	0.0128656	0.00416245	0.0189196	0.0113407	0.0087726
2	7-Jan-09	-0.0288617	-0.0288617	-0.0304691	-0.0178331	-0.0287346	0.0172929	-0.0358986	-0.0170728	-0.020015
3	8-Jan-09	-0.0622081	-0.0622081	0.00339136	-0.0117263	-0.000465999	-0.0400613	0.0282832	-0.00556096	-0.019423
4	9-Jan-09	0.00985991	0.00985991	-0.0215332	-0.0198728	-0.0127097	-0.0044735	-0.00976388	-0.0109886	-0.007802
5	12-Jan-09	-0.029191	-0.029191	-0.0228226	-0.0135257	-0.00502553	-0.0490385	-0.0538495	-0.0124513	-0.022629
6	13-Jan-09	0.0154453	0.0154453	0.00175655	-0.0176736	-0.00614145	0	0.00357202	-0.0122202	-0.004827
7	14-Jan-09	-0.0411676	-0.0411676	-0.0340325	-0.0473834	-0.0509452	0.00291235	-0.0403016	-0.0452202	-0.008676
8	15-Jan-09	0.000661905	0.000661905	0.0013283	-0.0195506	-0.0143347	-0.0504476	0.0303136	-0.0120704	-0.023428

➤ Summary of data:

- The dataset has 10 columns and 536 rows, as you can see the sample data on the above Image.
- Here I'm providing the Dataset Index ['date', 'ISE', 'ISE.1', 'SP', 'DAX', 'FTSE', 'NIKKEI', 'BOVESPA', 'EU', 'EM'] as show in the above image.

➤ Document control:

GitHub-Link: https://github.com/RaghavendhraRao/Stock-market-prediction-ISE-_DSFP

File Name: Stock-market-prediction-ISE-_DSFP

- **Metadata:** A ReadMe file is created for the project in the GitHub, which helps the staff & markers to install and run the code.

➤ Security and storage:

As the code is hosted on the GitHub, all the commits and backups can be seen and to ensure version control and storage. Dataset can be installed via pip, and anyone with the repository link can access the code.

➤ Ethical requirements:

Chosen dataset satisfies the GDPR (General Data Protection Regulation) and satisfies the UH ethical policies. All the datasets from the UCI used by the machine learning community to work on there research projects which is legal and ethical to use in the current project.