

Project Phase III Report

On

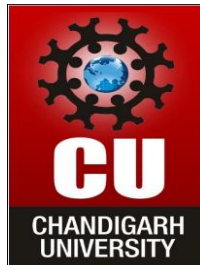
Stock Price Prediction Using Machine Learning

Submitted for the requirement of

Project course

BACHELOR OF ENGINEERING

COMPUTER SCIENCE & ENGINEERING



Submitted To:

Mr. Puneetparsad Singh

Submitted By:

Jaipreet Singh (20BCS1737)

Krishnkant Singh (20BCS1670)

Piyush Kumar Sharma (20BCS1688)

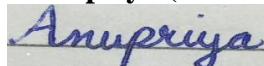
Yogesh Kumar (20BCS1702)

Pawan Kumar (20BCS1717)

Mentor: Mr. Puneetparsad Singh

Co Supervisor Signature

Ms. Anupriya (E-10436)



DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

CHANDIGARH UNIVERSITY, GHARUAN

June 2022

ABSTRACT

In this project we attempt to implement machine learning approach to predict stock prices. Machine learning is effectively implemented in forecasting stock prices. The objective is to predict the stock prices in order to make more informed and accurate investment decisions. We propose a stock price prediction system that integrates mathematical functions, machine learning, and other external factors for the purpose of achieving better stock prediction accuracy and issuing profitable trades. There are two types of stocks. You may know of intraday trading by the commonly used term "day trading." Intraday traders hold securities positions from at least one day to the next and often for several days to weeks or months. Long-Short Term Memory Models are very powerful in sequence prediction problems because they're able to store past information. This is important in our case because the previous price of a stock is crucial in predicting its future price. While predicting the actual price of a stock is an uphill climb, we can build a model that will predict whether the price will go up or down.

TABLE OF CONTENT

Sr no.	Topic	Page No.
1	Feature/characteristics identification	3
2	Constraints Identification	6
3	Analysis of features and finalization subject to constraints	7
4	Design selection	9

Features

- **Objective:** The objective is to predict the stock prices in order to make more informed and accurate investment decisions. We propose a stock price prediction system that integrates mathematical functions, machine learning, and other external factors for the purpose of achieving better stock prediction accuracy and issuing profitable trades.
- **Life Span:** Every project is invariably time-bound. At the time of planning, there will be the time phase of the project where the team can work independently on the project modules. By considering the length and time bounds, our project was divided as follows:
 1. A team was formed considering of five members and the project was assigned to the team. The team did a lot of research about the project and discussions were held related to the project's scope based on the deadline.
 2. After finalizing the scope of the project and all the things related to it, the project was divided into small actionable tasks and a roadmap was created to achieve the end result. Short measurable goals were established to reach the higher goal.
 3. After dividing the project into sub-parts, the actual implementation of project starts i.e. the true start of the project. During implementation, the process was monitored from time to time reviewing the team's efficiency and performance.
 4. When the project is completed, then retrospection of the project is done and the project is tested against different data to check its accuracy.
- **Required Funds:** Cost estimation is one of the essential factors for any project. So, calculating in advance the required funds for the project will be very impactful. But as our project is software-based and all the datasets are available on the internet free of cost, therefore no funding is necessary.
- **Life Cycle:**
 - (i) **Project Assignment Phase:** In the starting week of this semester, the project was assigned to us by our supervisor and a team was formed. In this phase, the idea for the project was explored and elaborated. The goal of this phase was to examine the feasibility of the project. In addition, decisions were made concerning who is to carry out the project and whether the project has an adequate base of support among those who are involved.

-
- (ii) **Project Definition Phase:** In this phase, the requirements that are associated with a project result are specified as clearly as possible. This involved identifying the expectations that all of the group members have with regard to the project result. Preconditions form the context within which the project were be conducted. All of the prerequisites were covered during this phase.
- (iii) **Execution Phase:** The actual implementation of the project started in this phase and it is currently in progress. The project was sub-divided into various parts among the team members and each one is working on it. The execution phase started in the March and it will approximately be finished in the first week of May.
- (iv) **Testing Phase:** The final outcome of the project in the Execution Phase will be tested against different inputs and efforts will be made to find and remove any type of bugs or errors in our project. This will approximately take a week at the most.
- **Team Spirit:** Team spirit is one of the most important principle to get the project finished on time. All of the group members, even though they are from different backgrounds, are able to get along with each other as everyone is aware of the fact that teamwork in managing a project leads to productivity and efficiency. Everyone is putting their strengths forward, holding themselves accountable, and they are not afraid to ask for help since they respect one another.
 - **Risk and Uncertainty:** Our project is based on forecasting the stock prices; therefore, risk and uncertainty is always associated with it. There is always the risk of market volatility, fake data and panic selling.
 - **Directions:** Projects are always performed by keeping in the mind the person who is going to use it due to which we have tried to create our project as user-friendly as possible. The user doesn't have to go through an extremely complex process to understand the proper way of using the software to get the desired output.
 - **Uniqueness:** This project uses LSTM model to predict stock prices instead of linear regression. Linear regression requires a series of assumptions to be made to be effective. But LSTM doesn't have any assumption like this. The reason they work so well is that LSTM can store past important information and forget the information that is not necessary.

-
- **Flexibility:** Change and project are synonymous. A project sees many changes throughout its life span. To make our predictions more accurate, we could've added a dropout layer in the neural network's layers which will help in avoiding overfitting but due to the time constraint, it was not possible.
 - **Sub-contracting:** Sub-contracting is a subset of every project and without which no project can be completed unless it is a proprietary firm or tiny in nature. The more complexity of a project the more will be the extent of contracting. Every project needs the help of an outsider consultant, engineer, or expert in that field. For the better implementation of our project, we were guided by our co-supervisor due to which we were able to resolve our project's problems easily.
 - **Cost:** If the quality of the project is to be changed there could be an impact on the cost of the project. The cost could increase if more resources are required to complete the project quicker. As our project is software related, no additional cost is required to increase efficiency of our project.

Constraints

- **Time:** Time constraints can be negotiated but can never be overcome completely. As we have to complete this project within a period of approximately 2.5 months due to which even though the project could have been made better but not within the given time limit.
- **Cost:** For project related to hardware, the budget may affect the efficiency of the project. However, our project is software-related, the cost of the project doesn't matter.
- **Scope:** The scope of our project is to predict the stock price of any company by using Machine Learning
- **Quality:** The prediction of stock prices is done by the LSTM model in Recurrent Neural Network (RNN). Although the project could've been made better by using dropout layers and activation functions but due to time constraint it was not done.
- **Benefits:** The accurate prediction of stock price will lead to more profit investors can make. Predicting how the stock market will move is one of the most challenging issues due to many factors that involved in the stock prediction, such as interest rates, politics, and economic growth that make the stock market volatile and very hard to predict accurately. The prediction of shares offers huge chances for profit and is a major motivation for research in this area; knowledge of stock movements by a fraction of a second can lead to high profits.
- **Risks:** The risk tolerance of the project usually can't be overcome. The first issue is that on internet fake data is also present which may cause huge loss to an investor. Another major issue is that stock market also depends on the people's feelings i.e. Bull and Bear market.

Analysis of Features and Finalization Subject to Constraints

- **Tiingo:** Tiingo is a financial data platform that makes high quality financial tools available to all. Tiingo has a REST and Real-Time Data API, which this library helps you to access. Presently, the API includes support for the following endpoints:
 - (i) **Stock Market Ticker Closing Prices + Metadata:** Data includes full distribution details and is validated using a proprietary EOD Price Engine.
 - (ii) **Curated news from top financial news sources + blogs:** Stories are tagged with topic tags and relevant stock tickers by Tiingo's algorithms
- **Normalization:** Normalization is a scaling technique in which values are shifted and rescaled so that they end up ranging between 0 and 1. It is also known as Min-Max scaling.

Here's the formula for normalization:

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}}$$

- **Train Test & Split:** The dataset is divided into two parts i.e. the training dataset and the testing dataset. The train-test split procedure is used to estimate the performance of machine learning algorithms when they are used to make predictions on data not used to train the model.
- **Long Short-Term Memory Network:** LSTMs are widely used for sequence prediction problems and have proven to be extremely effective. The reason they work so well is that LSTM can store past important information and forget the information that is not important. LSTM has three gates:
 - (i) The input gate: The input gate adds information to the cell state,
 - (ii) The forget gate: It removes the information that is no longer required by the model,
 - (iii) The output gate: Output Gate at LSTM selects the information to be shown as output.

Gantt Chart



Austen Tech

GANTT CHART



TASKS

FEBRUARY

MARCH

APRIL

MAY

Planning

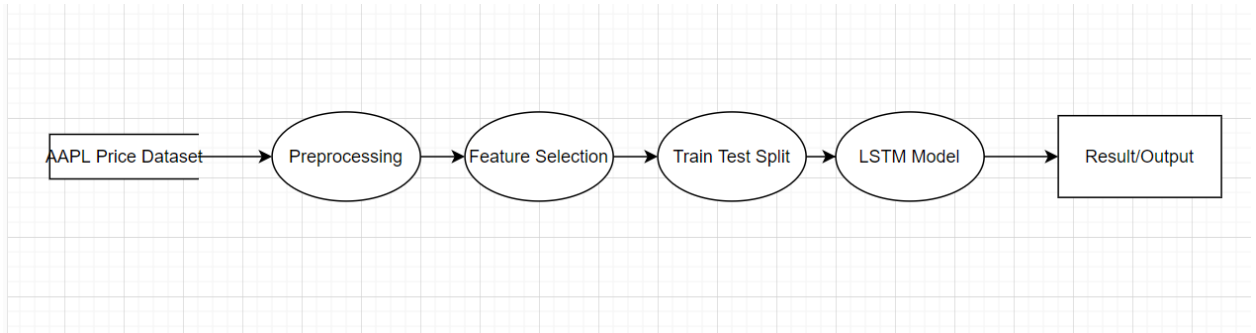
Research

Implementation

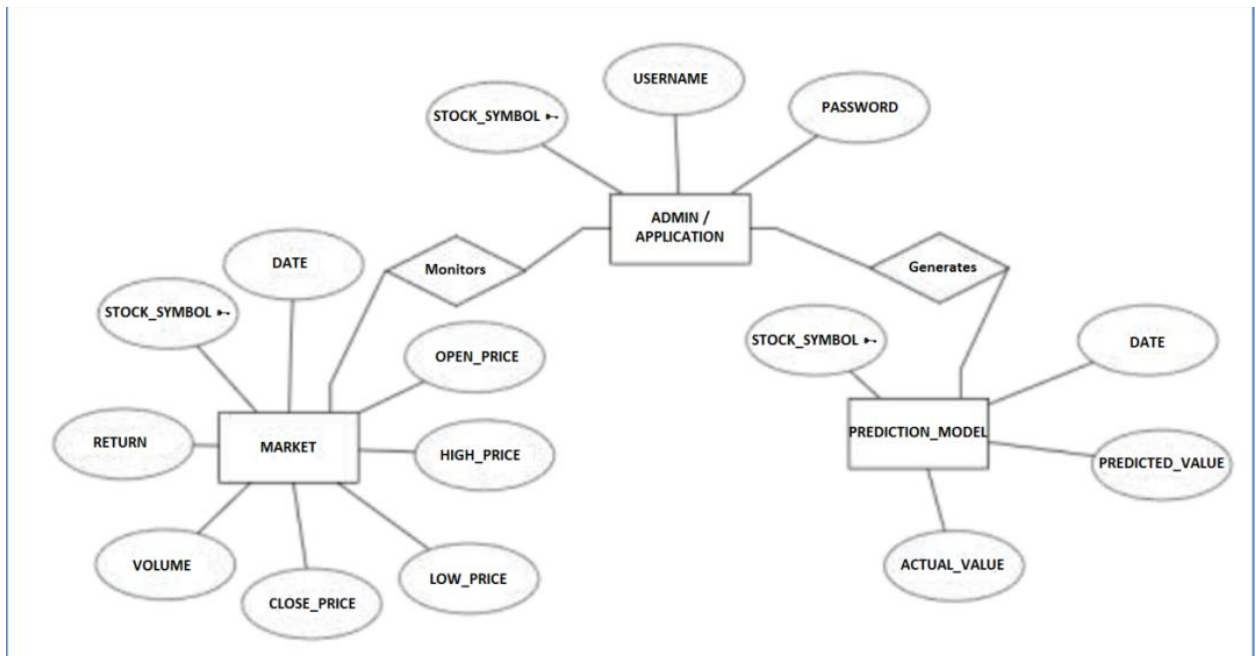
Testing

Design Selection

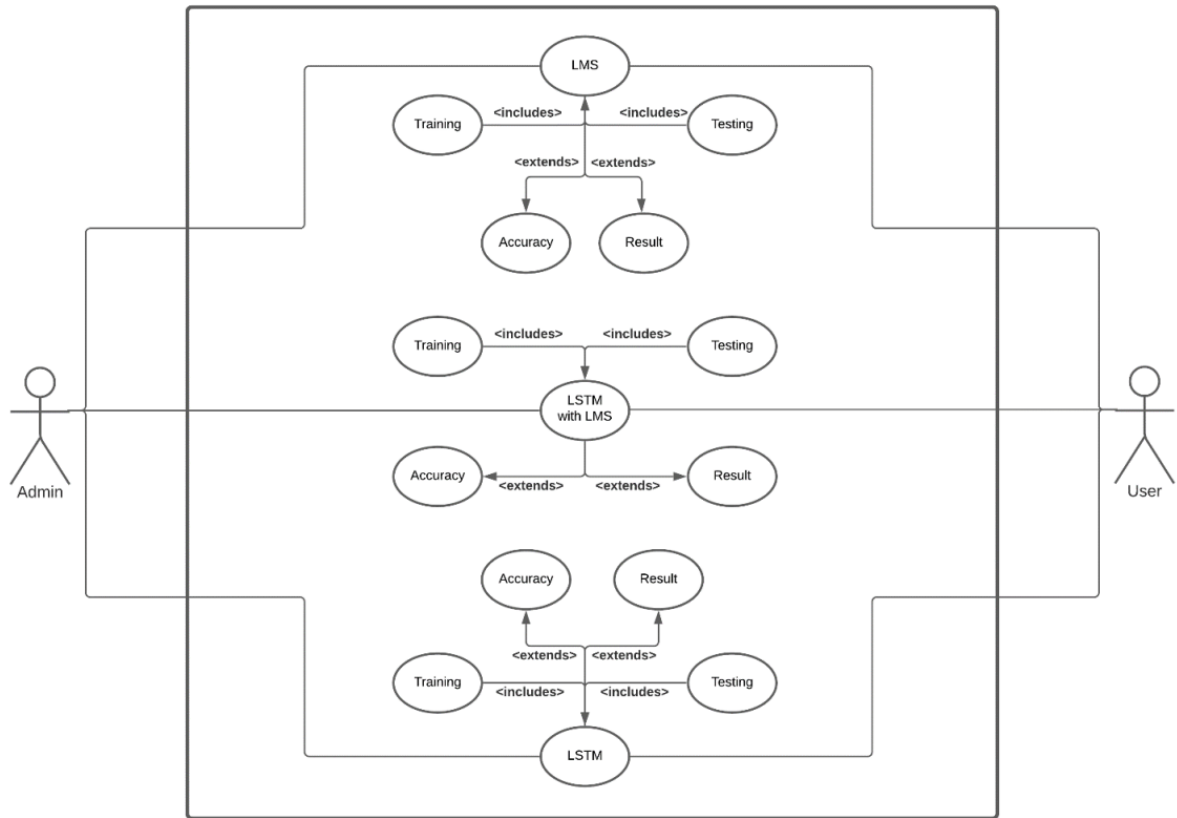
1. Data Flow Diagram:



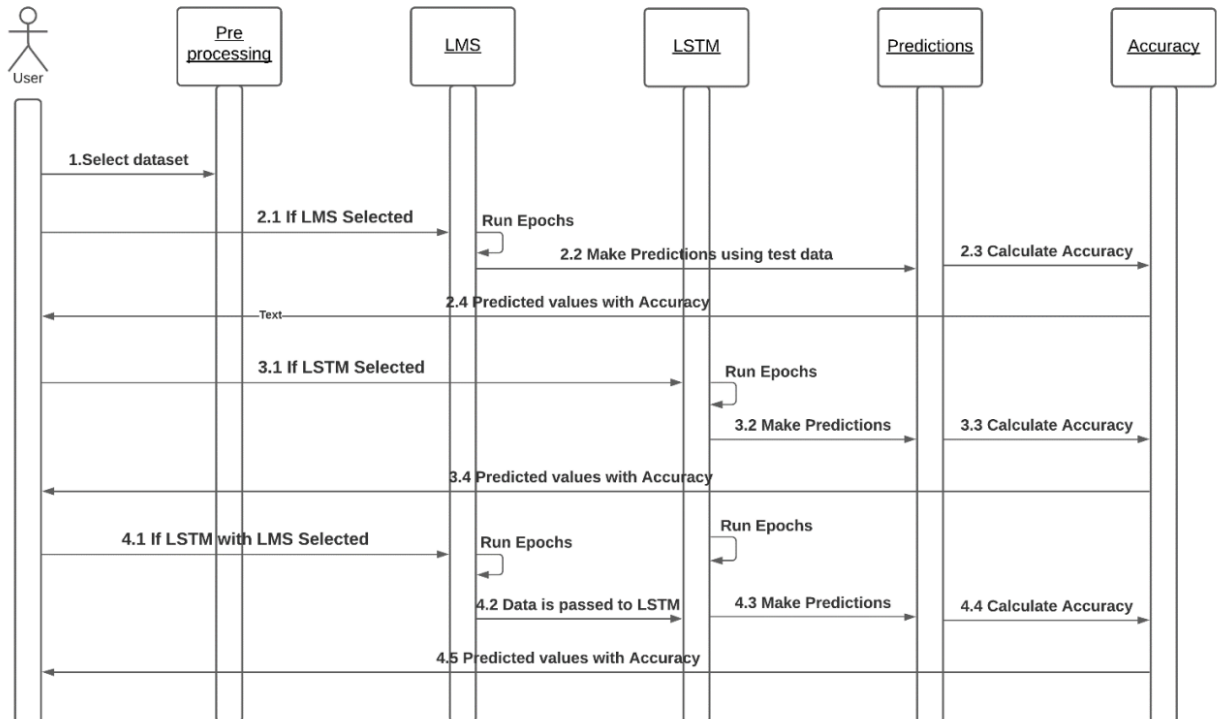
2. Entity Relation Diagram:



3. Use Case Diagram:



4. Sequence Diagram:



5. Collaboration Diagram:

