**A0100 - ANALYSIS REPORT**

.

.

.

.

.

.

.

.

.

.

.

.

.

.

.

.

|  |  |
| --- | --- |
| **Version:** | **1.2** |
| **Status:** | Draft |
| **Approver:** |  |
| **Author:** | Quách Hoàng Minh  Nguyễn Bảo Nguyên  Ngô Gia Hân  Nguyễn Vũ Anh Thư |



**Document history**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Version | Date | Author | Status | Remarks |
| 1.0 | 30-11-2021 | Quách Hoàng Minh  Nguyễn Bảo Nguyên  Ngô Gia Hân  Nguyễn Vũ Anh Thư | Draft |  |
| 1.1 | 23/12/2021 | Quách Hoàng Minh  Nguyễn Bảo Nguyên  Ngô Gia Hân  Nguyễn Vũ Anh Thư | Draft | Update section 2 |
| 1.2 | 31/12/2021 | Quách Hoàng Minh  Nguyễn Bảo Nguyên  Ngô Gia Hân  Nguyễn Vũ Anh Thư | Draft | Add section algorithm and roadmap |
|  |  |  |  |  |

**References**

|  |  |  |  |
| --- | --- | --- | --- |
| Reference | Title | Author | Version |
|  |  |  |  |

Table of contents

[1. Introduction 4](#_Toc92467698)

[1.1 Purpose 4](#_Toc92467699)

[1.2 Glossary 5](#_Toc92467700)

[2. Context 6](#_Toc92467701)

[2.1. Business opportunities 6](#_Toc92467702)

[2.2. Vision 6](#_Toc92467703)

[2.3. Benefit analysis 6](#_Toc92467704)

[2.4. Requirements 7](#_Toc92467705)

[2.5. Functional epics solution 9](#_Toc92467706)

[2.5.1. Login 9](#_Toc92467707)

[2.5.2. View Home Page 10](#_Toc92467708)

[2.5.3. Search Index 10](#_Toc92467709)

[2.5.4. Favourite Lists 11](#_Toc92467710)

[2.5.5. View Specific Stock 13](#_Toc92467711)

[2.5.6. Compare stocks 15](#_Toc92467712)

[2.5.7. View Stock Price 17](#_Toc92467713)

[2.5.8. Reminder List 17](#_Toc92467714)

[2.5.9. User Profile 19](#_Toc92467715)

[2.6. Scope 19](#_Toc92467716)

[2.7. Outside scope 19](#_Toc92467717)

[2.8. Approval Criteria 20](#_Toc92467718)

[2.9. Technical Analysis 21](#_Toc92467719)

[2.10. Algorithm 21](#_Toc92467720)

[2.10.1 LSTM (Long short-term memory) 23](#_Toc92467721)

[ How it works: 23](#_Toc92467722)

[2.11. Roadmap 26](#_Toc92467723)

[3. References 27](#_Toc92467724)

# Introduction

## Purpose

This documentation aims to record the agreement and decisions between customer and development team. In this analysis process, we would like to achieve the expectation between the team and the client about customers’ requirement’s solution.

This report serves as a consolidation of Trading Vision Project (TVP) original requirements, the development team's proposal and the dialogue about every single requirement and acceptance criteria throughout the project timeline.

The document is used as a reference for developing a more detailed specification and project plan in the O0500 deliverables. In the clarification phase multiple project deliverables are to be delivered. In the clarification report the deliverables listed below are embedded either as isolated sections or in their relevant context, e.g., the User Interface Prototype deliverable is split across relevant functional descriptions, while the High-level Software Architecture deliverable has its own dedicated section.

The document is prerequisite for these documents

* A0140 - Functional Epics
* A0150 - User Interface Prototype
* O0500 - High-level Software Architecture
* P0100 - Project Roadmap

## Glossary

|  |  |
| --- | --- |
| **Stock Symbol** | Stock Symbol is a unique series of letters assigned to a security for trading purposes. |
| **Candlestick Chart** | Candlestick Chart used by traders to determine possible price movement based on past patterns, it shows four price points (open, close, high, and low) throughout the period the trader specified |
| **Price Ceiling** | Price Ceiling is the highest price at which an investor can place an order to buy or sell securities during the trading day. |
| **Price Floor** | Price Floor is the lowest level at which an investor can place an order to buy or sell securities during the trading day |
| **Highest Price** | Highest Price is the highest price exchanged in a session. |
| **Lowest Price** | Lowest Price is the lowest price exchanged in a session. |
| **Matching Order Price** | Matching Order Price is the process by which a securities exchange pairs one or more unsolicited buy orders to one or more sell orders to make trades. |
| **Reference Price** | Reference Price is the closing price at the previous most recent trading session. The reference price is taken as the basis for calculating the trading range of stocks in the session |
| **Volume** | Volumes measure the number of shares traded in a stock or contracts traded in futures or options. |
| **Session** | A trading session is a certain period of time in the stock market, at which time transactions and orders to buy and sell stocks will take place. |
| **UPCOM** | UPCoM (Unlisted Public Company Market) is the market at HNX for public companies not yet listed and was launched in June 2009 with 10 initial companies |
| **HOSE** | Ho Chi Minh Stock Exchange was formerly known as HCM Securities Trading Centre, established in 1998 under Decision No. 127/1998/QD-TTg of the Prime Minister |
| **HNX** | Hanoi Stock Exchange, formerly the Hanoi Securities Trading Centre located in Hanoi, Vietnam, was launched in March 2005 and handles auctions and trading of stocks and bonds. |
| **LSTM** | Long short-term memory |
| **SVM** | Support Vector Machine |
| **MAPE** | Mean absolute percentage error |

# Context

The customer has a budget, and they want to invest in the stock market. They want to know the information of many stocks, so that they can decide which stock to invest in.

## Business opportunities

Netcompany has a group of investors that wants to have a website to check the stock market’s information. For convenience in grasping market trends, customers want an application to view the stock market price and provide reliable news about the stock market. The website should also give the auto technology prediction in the coming week (5 days) without hiring stock forecasters. However, this will be a difficult problem for software developers because the stock market is complex and has many different price ranges. In addition, the stock price is always changing, and it fluctuates continuously from time to time.

The stock market is a volatile place. The cause of that fluctuation comes from many sources: market trends, investor sentiment, company's financial statements, etc. There are two ways to analyse securities: Fundamental analysis and technical analysis.

Fundamental analysis assumes that stock prices do not inevitably reflect the true intrinsic value of the underlying business and our solution for our customers lies in the second analytical method. Technical analysis generally assumes that a stock's price reflects all available information and that prices generally move according to trends. In other words, by analysing a stock's price history, you may be able to predict its future price behaviour. If you have ever seen someone trying to identify patterns in stock charts or discussing moving averages, that's a form of technical analysis (Frankel, 2021). As a result, we will use technical analysis as the main method in this project.

## Vision

When the project is completed, a stock market web application is delivered to the client. Customers can view stock price trends on HOSE, HNX and UPCOM stock exchanges. In addition, customers can use the predict function to predict prices in the next one week.

## Benefit analysis

Customers have capital and want to invest in stocks but are still confused because there are too many information sources. With this web application, customers will be able to easily track information about stocks on stock exchanges in Vietnam, and visually monitor chart, line, and column charts. In addition, customers can use the predict function as a source of reference, increasing their confidence in their investment.

## Requirements

|  |  |  |
| --- | --- | --- |
| **ID** | **Feature** | **Description** |
| 1 | Login/Logout | As a customer, I'm able to login/logout |
| 2 | Overview page | As a customer, after logged in successfully, I'm able to see the overview page:  + I can see the search button for search index        + I can see the list of capitalization-weighted indexes of all companies listed on the stock exchange and their values (date, point, volume)). E.g.: (these indices are just an example. We will choose them depend on what dataset we can find)              + VNIndex of HoSE              + VN30              + HNX-Index              + HNX30-Index              + UPCOM        + Favourite indices list:                + If it is unavailable, display empty                 + I can see the information of each stock index                           + date, price, volume                           + I can click the index to view the chart |
| 3 | Search index | As a customer, I'm able to search index  + I can search by index  + I can add index search result to favourite list  + I can click the index search result to view its chart |
| 4 | Index chart | As a customer, I'm able to view index chart:    + I can see the list of charts that I would like to view. They are line chart (by default) & candlestick chart    + I can view the index chart:             + line chart:                     + x-axis: time unit (date or month)                      + y-axis: price (VND)                   + prediction area                             + When I hover to the line, I can see the popup which will display a price      + I can choose to candlestick chart to view index      + I can see the bar chart which will display the volume of index by time unit in the bottom       + I can see the prediction volume area      + When I hover to the column, I can see the popup which will display a price      + I can do comparison by choosing other indices       + In the current chart, I can see one more line chart of the comparing index. |
| 5 | Prediction Index | Display prediction to the index chart |
| 6 | Setting reminder | As a customer, I can set reminder:  + I can choose the expectation point in the prediction area in a chart  + I can create reminders by adding titles, timing, and content.  + I can receive the reminder via email. |
| 7 | Manage user profile | As a customer, I'm able to view my profile:     + I can view my name from Facebook or Gmail  As a customer, I want to delete all of personal information and related actions when I do not continue using the system. |

## Functional epics solution

In this section, we will provide a prototype of the website, which will include these pages’ prototype: Homepage, Login. View stock price, view specific stock, favourite list, and reminder list. These prototypes will also have the descriptions of the pages and some basic information about the function that have on each page.

### Login

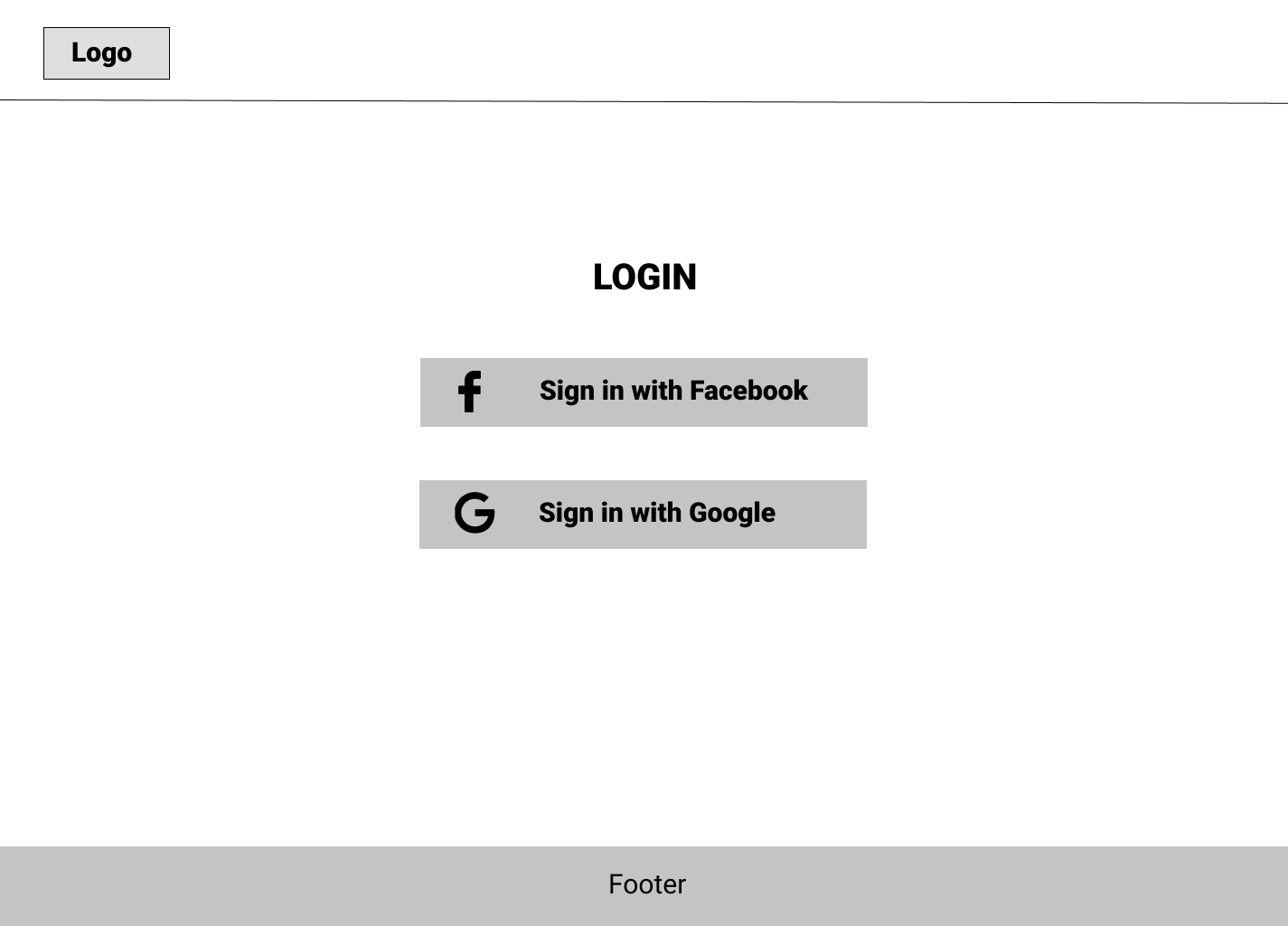


Figure : Login page

The login page will provide for users two ways to login to the page: Facebook or Google. If the user chooses to log in, a pop up will appear and this navigates the user to the login page of google or Facebook. The user will log in to this page.

### View Home Page

Table

Description automatically generated

Figure : Homepage

This is the homepage of the website; users can see the label and the company name including the chart of 5 stocks that are trending and have high volume on that day. We could search the index in the search bar on the header. The banner will show a wallpaper related to stock. There is a search function on the top right for the user to input the name of the index that they want to find.

### Search Index

When users click on the search bar, a popup will be displayed. Users can see all stocks belonging to each stock exchange or can write exact stock symbols to find. All stocks are always displayed in alphabetical order.

Graphical user interface, application

Description automatically generated

Figure : Search Index

### Favourite Lists

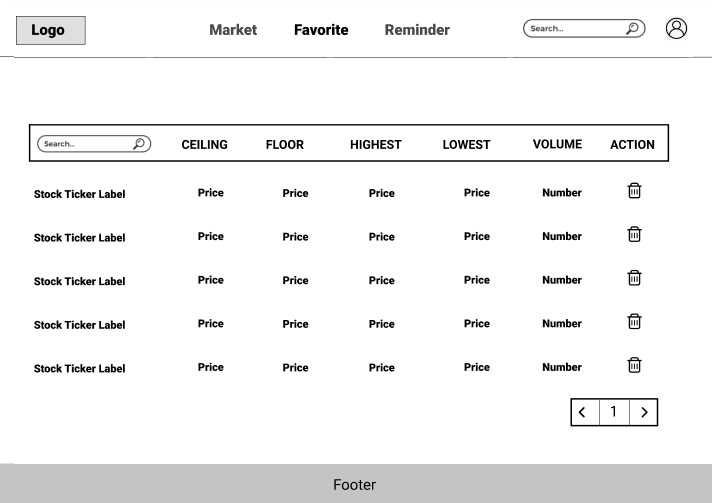


Figure : Favourite list

The favourite page, this page displays all the stock, which the user follows. It will have the search bar for searching the stock favourite and the information of all the stock users added to this list. These stocks also show the information including ceiling price, floor price, highest price, lowest price, volume. In case, user does not have any favourite stock, there is a message “**Add your favourite stock**”

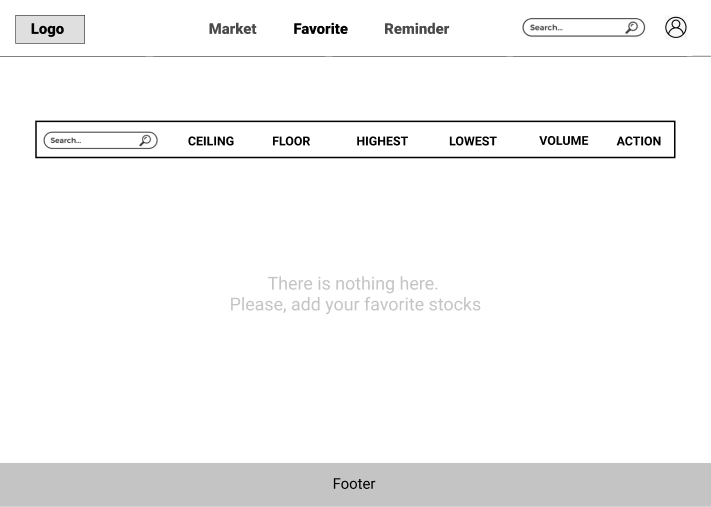


Figure : There is no stock in favourite list

Users can click the “**Bin**” icon to remove a stock from the Favourite List.

### View Specific Stock

Graphical user interface, application

Description automatically generated

Figure : Specific Stock

This is the stock detailed page, which will contain the stock’s sticker label and the detailed information of each stock. This will contain the name of the company, field, link to a webpage, a short description of the company, and some financial numbers. Also, this page will show the user the specific chart of that stock. These charts could be chosen to show a line graph or candlestick chart. On this page, the line chart also shows the prediction of stock trends predicted by a selected algorithm. The graph showing the prediction will be colored differently from the actual line.

Users also can view candlestick charts by clicking on “**Candlestick chart”**.

Graphical user interface

Description automatically generated

Figure : Candlestick chart

In order to add a stock to Favorite list, users can click on the “**Heart**” icon, then the information of that stock will be added into the Favorite page.

Users can click the “**Add**” icon to add reminders on the prediction line. They also can hover at any point on the line chart to view specific date and price.

Graphical user interface, application

Description automatically generatedGraphical user interface, application

Description automatically generated

Figure : Before and After adding favourite list.

Graphical user interface

Description automatically generated

Figure : User also can hide the company information.

### Compare stocks

When the Compare button is clicked, this popup will show up. You can search like what you do in the search function.

Graphical user interface, application

Description automatically generated

Figure : Choose appropriate stock to compare.

Then, you can choose which stock you want to compare with. There will exist two-line charts, each of them will have different colours for the actual part and prediction part. If you want to close the compare function, click the Compare button again, and choose the stock that you chose before one more time.

Graphical user interface, application

Description automatically generated

Figure : Compare two stocks

Graphical user interface, application

Description automatically generated

Figure : Click "Tick" icon to turn off comparison

### View Stock Price

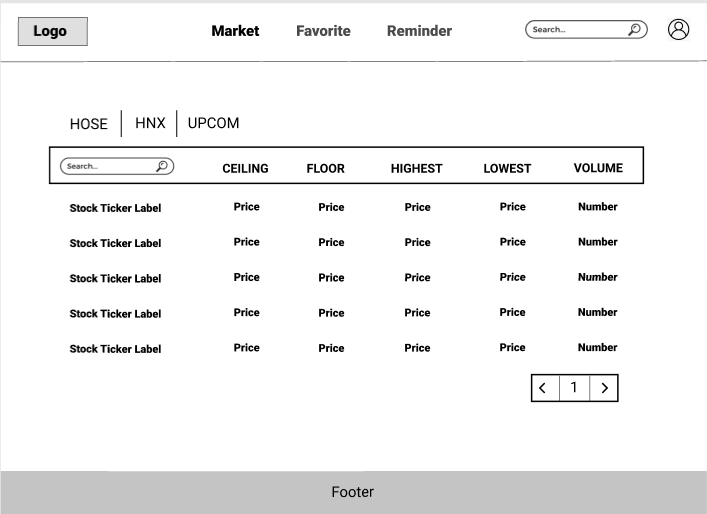


Figure : View Stock Page

This is the market page which will have the navigator header contain page navigation and search stock bar like the other page. In the main content, there will be a tab for users to choose their exchange market. Their option could be UPCOM, HNX or HOSE, for each exchange, there will be a list of stocks which include the information about ceiling, floor, highest, lowest and volume of every stock. Also, this will include a search bar for customer search for the ticket of stock they want to find.

### Reminder List

Table

Description automatically generated

Figure : Reminder list

The final wireframe design will be the reminder page. This page will contain the reminder that users have set before with the detailed time they want to notify and the stock or content in each reminder. Users could also edit or remove the reminders in this page.

When users click “**Add**” button, a popup will be displayed to add reminder. Users can click on the “Plus” icon in order to add their own reminder.

Graphical user interface

Description automatically generated

Figure : A popup to add reminder

In case, user does not have any reminder, a message “**Add your reminder**” will be displayed. Users can click on the “Plus” icon in order to add their own reminder.

Graphical user interface, text, application

Description automatically generated

Figure : There is no reminder in the reminder list

### User Profile

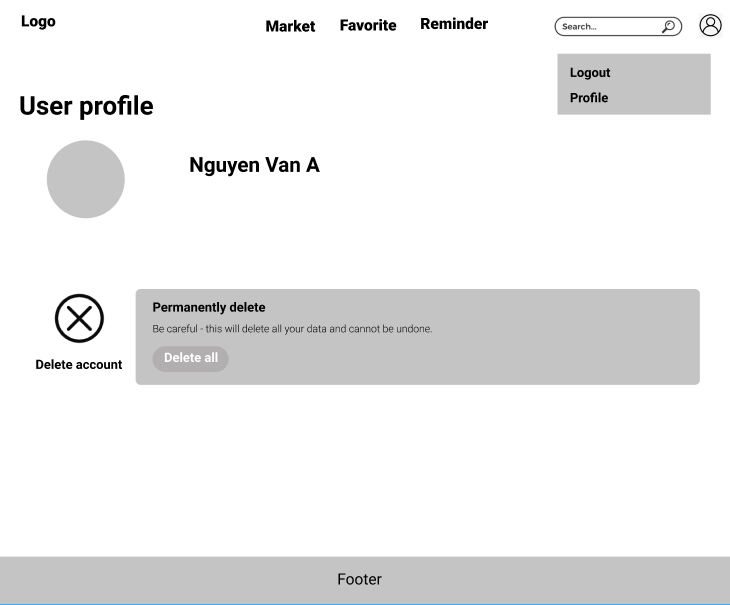


Figure : User Profile page

Users can use the dropdown menu from their account icon in the header to logout or access to the profile and delete account. This is the page for users to manage their account. Users can choose to delete all if they want to delete all information of users on the website.

### Scope

* Costs must not exceed the estimated budget.
* The final product must be released within 6 months.
* Must have a UAT version for customers to try before releasing the final version.
* The final product must satisfy the Minimum Viable Product (MVP).
* There are documents and resources that prove the predictions are reliable.

### Outside scope

Below features are not included in this project:

* Allow users to directly buy stocks on the website
* Have a box chat between all users of the page
* Have an admin page.
* Support real-time stock information

### Approval Criteria

* Web app that shows charts representing stock market prices as well as with the table of the stock information.
* The algorithm can predict about 55-60% correctly.
* Successfully deploying web apps and functions to web browsers.
* Price chart viewing, and price prediction functions can work well.
* Have an account to meet the account protection requirement of users.
* It is possible to switch back and forth between different types of shares

### Technical Analysis

Diagram

Description automatically generated

Figure : Component Diagram

System gets data from two sources: CafeF for stock price and VCBS for company information. The retrieved data will be mapped to the database of system MongoDB in suitable schema. After that, those data will be pre-processed to prepare for the prediction phase afterwards. Finally, on the monitor of the users shows the chart according to the output of the prediction.

### Algorithm

Taking everything into consideration, we saw from the Journal Articles of various authors that our problem could be solved by one of three algorithms. These algorithms are Long short-term memory (LSTM), Support Vector Machine (SVM) or Linear Regression. After evaluating the performance of the three algorithms on the Vietnamese stock market, we decided to choose the LSTM algorithm for our project. First, LSTM has the minimum MAPE value when compared with 2 other algorithms. Moreover, for linear regression, it is proved that it is only effective for a certain period of time in the old regimes. As the regime shift happens in the financial industry, the model became less effective. (Zhichao Zou, 2020). In addition, LSTM is well-known for its application in time-series forecasting due to its ability to have several lags of unknown duration between important events in a time series (KUMAR, 2020).

In order to evaluate the algorithms’ performance, we will use mean absolute percentage error (MAPE), which is a method of evaluation for LSTM model (Yulian Wen, 2020). MAPE means the percentage of difference between predicted values and actual values. For example, MAPE value of 11.5% means that the average difference between the predicted value and the actual value is 11.5%.

The lower the value for MAPE, the better a model is able to predict values. For example, a model with a MAPE of 5% is more accurate than a model with a MAPE of 10%.

Table : Comparison between 3 algorithms

|  |  |  |  |
| --- | --- | --- | --- |
|  | **LSTM** | **SVM** | **Linear Regression** |
| **Description** | The Long Short-Term Memory is an extended version of Recurrent Neural Network (RNN). While RNN is well-known in solving sequential data problems, its structure is still simple so the ability to connect layers over a large distance is not good. Basically, it could not remember information of far data, so the first elements of input normally do not have much effect on the output prediction. The reason for this is the RNN effect by vanishing gradient. To solve the problem of RNN, LSTM was designed to work by remembering the related information that is important for the prediction and removing all the other data | Support Vector Machine (SVM) is an algorithm belonging to the Supervised Learning group and used in classification. This algorithm will try to maximize the margin (M) to find the best hyperplane and divide data , the data points in two boundaries is called support ᴠeᴄtor because we could use them to support finding hyperplane. | Linear regression uses the relationship between the data-points to draw a straight line through all of them. This line can be used to predict future values. Regression analysis is a statistical methodology that allows us to determine the strength and relationship of two variables. Regression is not limited to two variables; we could have 2 or more variables showing a relationship. The results from the regression help in predicting an unknown value depending on the relationship with the predicting variables. When there is a single input variable, the regression is referred to as Simple Linear Regression. We use the single variable (independent) to model a linear relationship with the target variable (dependent). We do this by fitting a model to describe the relationship. |
| **Input** | The close price and date | The close price and date | The close price and date |
| **Output** | The predicted price with future date | The predicted price | The predicted price |
| **Advantages** | - Can process inputs of any length.  - Can use their internal memory for processing the arbitrary series of inputs which is not the case with feedforward neural networks  - Even if the input size is larger, the model size does not increase | - SVM works relatively well when there is a clear margin of separation between classes.  - SVM is more effective in high dimensional spaces.  - SVM is effective in cases where the number of dimensions is greater than the number of samples.  - SVM is relatively memory efficient | - Simple implementation - Performance on linearly separable datasets - Overfitting can be reduced by regularization |
| **Disadvantages** | - They require a lot of resources and time to get trained  - LSTMs are prone to overfitting and it is difficult to apply the dropout algorithm to curb this issue | - SVM algorithm is not suitable for large data sets.  - SVM does not perform very well when the data set has more noise i.e. target classes are overlapping.  - In cases where the number of features for each data point exceeds the number of training data samples, the SVM will underperform.  - As the support vector classifier works by putting data points, above and below the classifying hyperplane there is no probabilistic explanation for the classification. | - Prone to underfitting - Sensitive to outliers - Linear Regression assumes that the data is independent |

Table : Comparison MAPE between 3 algorithms

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **ACB (5 days)** | **ACB (15 days)** | **VNM (5 days)** | **VNM (15 days)** | **ACB training and testing** | **VNM training and testing** |
| **LSTM** | 1.35 | 2.87 | 0.6 | 1.42 | 5.95 | 14.48 |
| **SVM** | 3.2 | 4.7 | 12.3 | 39 | 9,8 | 88 |
| **Linear Regression** | 2.3 | 2.4 | 10.5 | 16 | 8.3 | 64 |

### LSTM (Long short-term memory)

* **Input:** Close price and date of s specific stock
* **Output:** Predicted price with future date

#### How it works:

In order to implement LSTM, we are going to create a sequential model with multiple layers. We will add LSTM layers to our model followed by a dense layer that predicts the future stock price. We imported the Sequential class from keras.models library and Dense, LSTM from Keras.layers library from Python for the training and testing process.

Table : MAPE of LSTM on different predicting days

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **ACB (5 days)** | **ACB (15 days)** | **VNM (5 days)** | **VNM (15 days)** | **ACB training and testing** | **VNM training and testing** |
| **MAPE** | 1.35 | 2.87 | 0.6 | 1.42 | 5.95 | 14.48 |

In this LSTM algorithm analysis, we choose 2 stocks ACB (appeared in Hose lately on January 9, 2020) which has a smaller data compared to VNM (appeared on January 19, 2006). For the ACB, the training data is from December 14, 2020 to March 1, 2021 with the length of 50, the testing set is from March 2, 2021 to December 10, 2021 with the length of 199. For the VNM, the training set is from January 19, 2006, to April 2, 2009 (length: 793) and testing set is from April 7, 2009, to December 10, 2021 (length: 3169). The prediction starts from December 13, 2021, for both ACB and VNM.

When the size of the data is large (like VNM), it shows that the MAPE is smaller (1.35 and 0.6). Moreover, because our LSTM algorithm’s performance significantly decreases when forecasting more than 1 week, so our scope for predicting stock’s price is only 5 days.

Chart, histogram

Description automatically generated

Figure : ACB training and testing

Chart, histogram

Description automatically generated

Figure : VNM training and testing

Chart, line chart

Description automatically generated

Figure : ACB 5 days training

Chart, line chart

Description automatically generated

Figure : VNM 5 days predicting

Chart, line chart

Description automatically generated

Figure : ACB 15 days predicting

Chart, line chart

Description automatically generated

Figure : VNM 15 days predicting

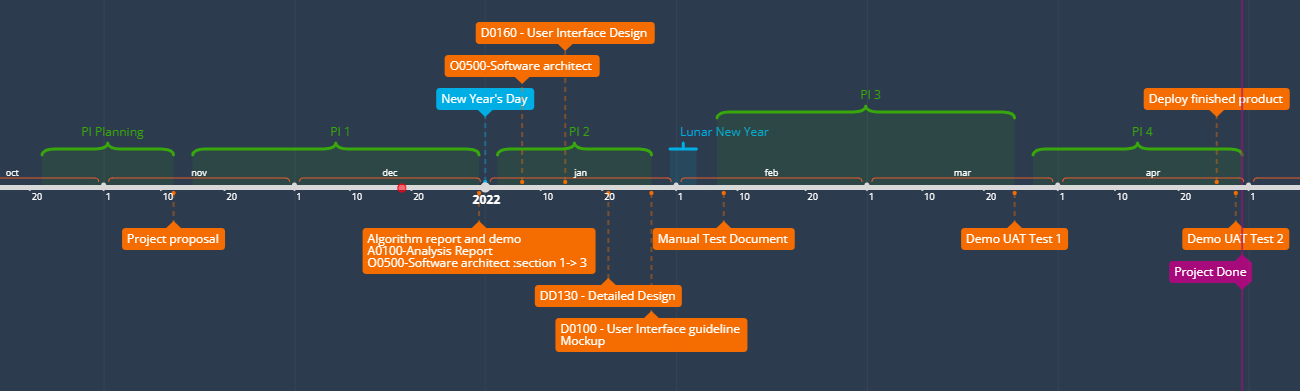
### Conclusion

After evaluating LSTM, Linear Regression and SVM, it is clear that LSTM gives the best performance among the three algorithms (lowest MAPE). Moreover, while Linear Regression can be prone to regime shift in the financial industry, LSTM outperforms SVM in general due to its ability to learn selectively and can remember or forget the required historical data (Sai Krishna Lakshminarayanan, John McCrae). In conclusion, because of its advantages, the LSTM algorithm will be chosen and applied for predicting stock price in the project.

### Roadmap

The project roadmap is established based on priorities of functionality from TVP. The roadmap is still high-level but has placed each theme of functionality into a release.

As established in Project Dependencies, the project is constrained by several dependencies, which is handled by scheduling functionality with dependencies in releases where time permits completion of the dependencies. Due to this, some functionality is planned later than their original priority would demand, in order to accommodate for the dependency. All project and system dependencies have been marked as milestones:



*Figure 23: Roadmap*

The result is the following releases:

* **Release 1**: Demo for UAT test 1
* **Release 2**: Finish product
* **Release 3:** Demo for UAT test 2

# References

* <https://towardsdatascience.com/time-series-forecasting-with-recurrent-neural-networks-74674e289816>
* <https://www.educba.com/recurrent-neural-networks-rnn/>
* <https://link.springer.com/chapter/10.1007/978-3-030-16145-3_17>
* <http://ceur-ws.org/Vol-2563/aics_41.pdf>