**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](#_Toc91842939)

[1.1 Purpose 4](#_Toc91842940)

[1.2 Glossary 4](#_Toc91842941)

[2. Context 5](#_Toc91842942)

[2.1. Business opportunities 5](#_Toc91842943)

[2.2. Vision 5](#_Toc91842944)

[2.3. Benefit analysis 5](#_Toc91842945)

[2.4. Requirements 6](#_Toc91842946)

[2.5. Functional epics solution 8](#_Toc91842947)

[2.5.1. Login 8](#_Toc91842948)

[2.5.2. View Home Page 9](#_Toc91842949)

[2.5.3. Favourite Lists 10](#_Toc91842950)

[2.5.4. View Specific Stock 11](#_Toc91842951)

[2.5.5. View Stock Price 12](#_Toc91842952)

[2.5.6. Reminder List 13](#_Toc91842953)

[2.5.7. User Profile 14](#_Toc91842954)

[2.6. Scope 14](#_Toc91842955)

[2.7. Outside scope 15](#_Toc91842956)

[2.8. Approval Criteria 15](#_Toc91842957)

[2.9. Technical Analysis 16](#_Toc91842958)

[2.10. Algorithm 17](#_Toc91842959)

[2.10.1 . LSTM (Long short-term memory) 17](#_Toc91842960)

[ How it works: 17](#_Toc91842961)

[ Advantages: 19](#_Toc91842962)

[ Disadvantages: 19](#_Toc91842963)

[2.10.2 . SVM (Support Vector Machine) 19](#_Toc91842964)

[ Advantages: 19](#_Toc91842965)

[ Disadvantages: 19](#_Toc91842966)

[2.10.3 . Linear Regression 22](#_Toc91842967)

[ Advantages: 22](#_Toc91842968)

[ Disadvantages: 22](#_Toc91842969)

[2.11. Roadmap 24](#_Toc91842970)

[3. References 24](#_Toc91842971)

# 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.

## 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. |

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

# 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.

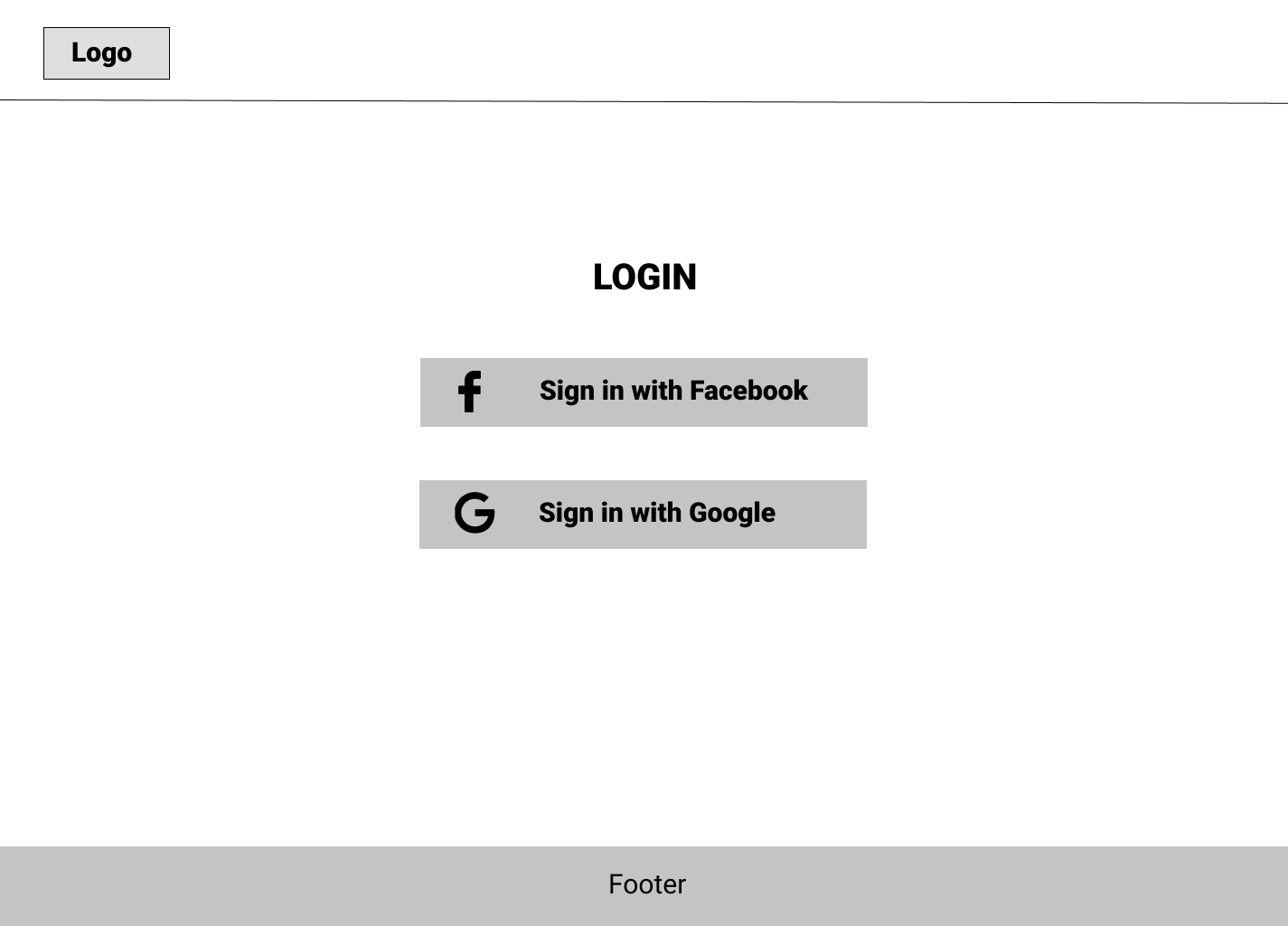
## Requirements

|  |  |  |
| --- | --- | --- |
| **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.

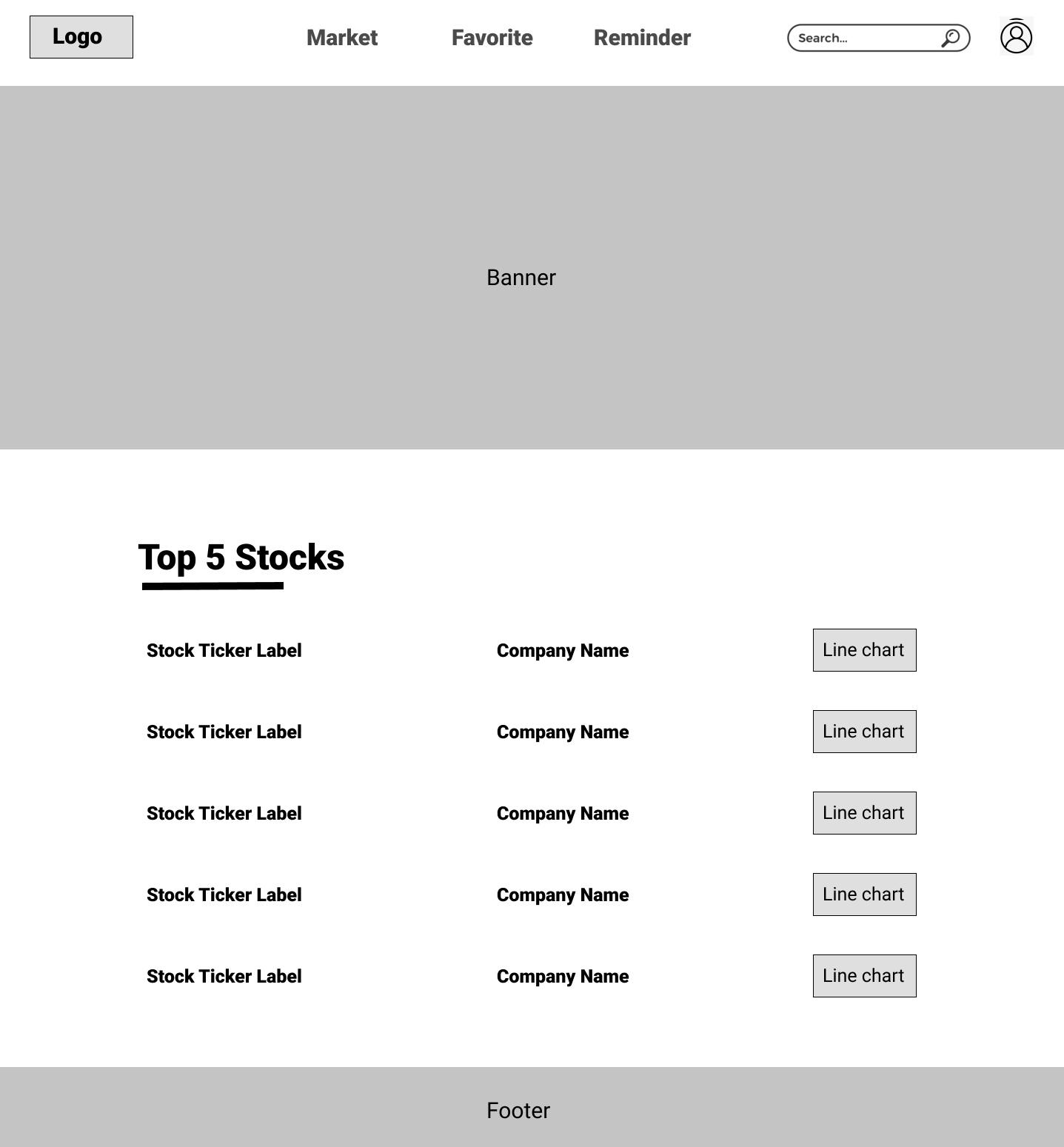
### Login



*Figure 1: 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.

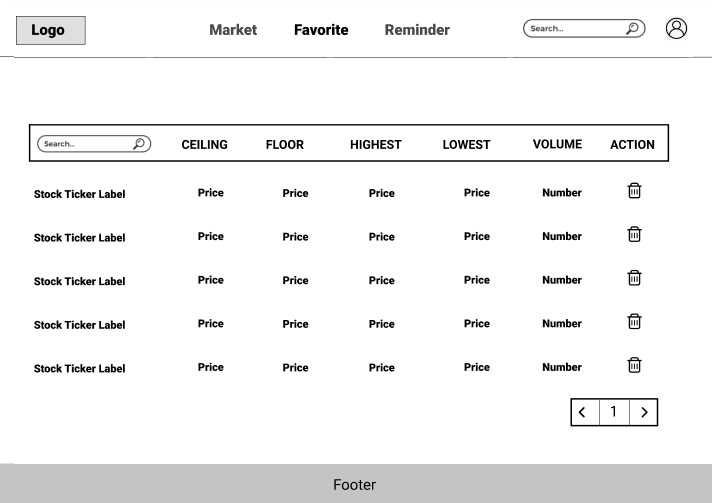
### View Home Page

**

*Figure 2: Homepage*

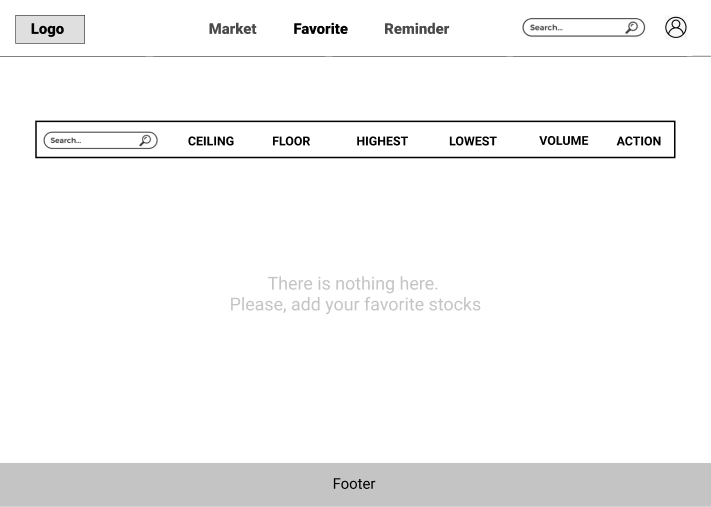
This is the homepage of the website; users can see the label and the company name including the chart of 5 stocks that are trending and have high volume on that day. We could search the index in the search bar on the header. The banner will show a wallpaper related to stock. There is a search function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

### Favourite Lists



*Figure 3: Favourite list*

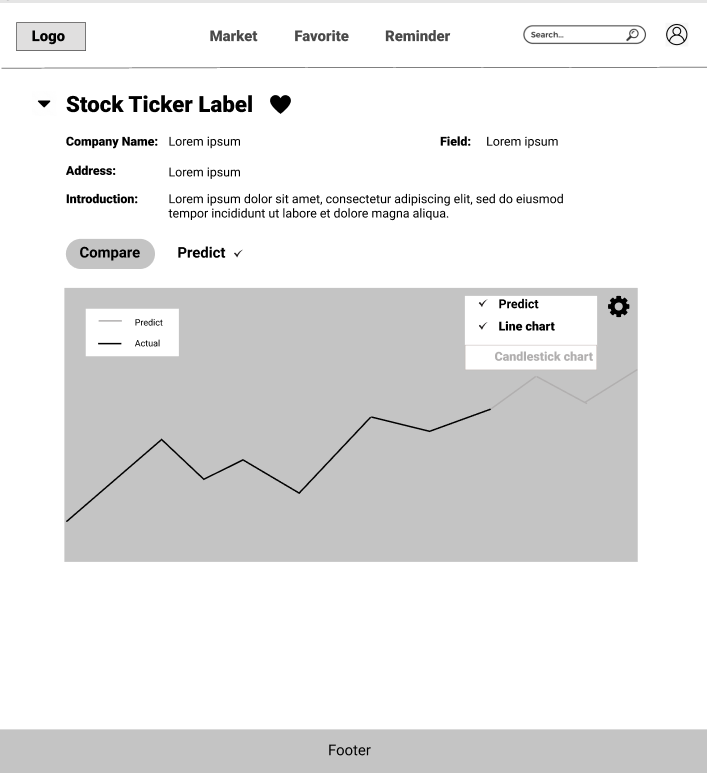
The favourite page, this page displays all the stock, which the user follows. It will have the search bar for searching the stock favourite and the information of all the stock users added to this list. These stocks also show the information including ceiling price, floor price, highest price, lowest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



*Figure 4: There is no stock in favourite list*

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

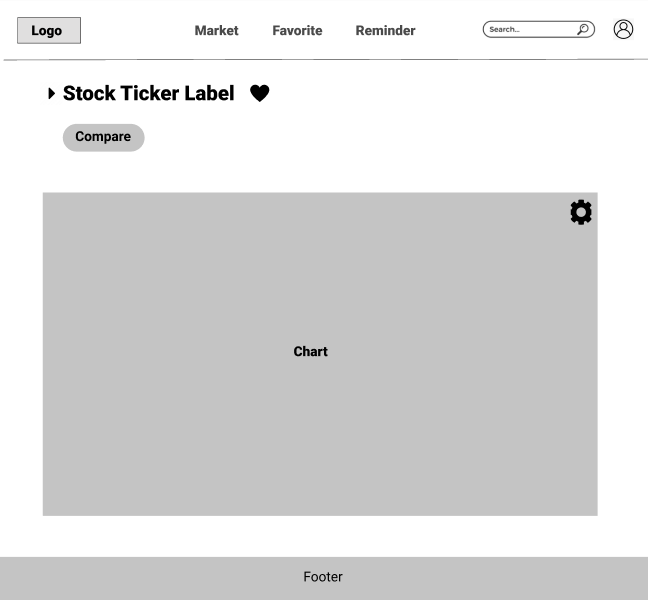
### View Specific Stock



*Figure 5: Specific Stock*

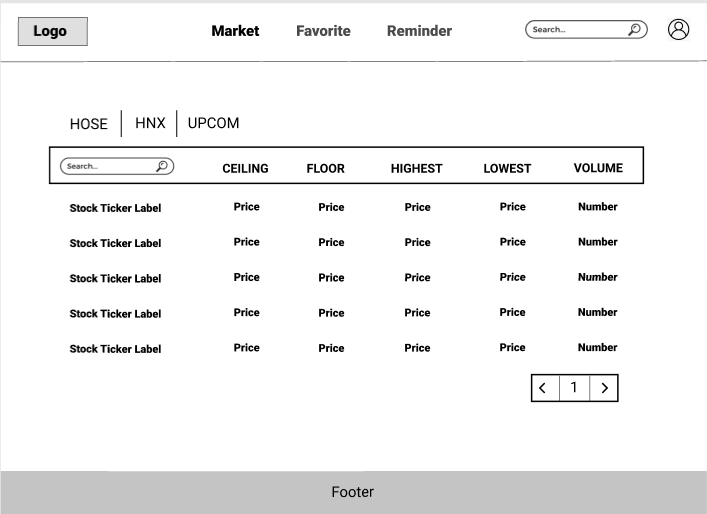
This is the stock detailed page, which will contain the stock’s sticker label and the detailed information of each stock. This will contain the name of the company, field, link to a webpage, a short description of the company, and some financial numbers. Also, this page will show the user the specific chart of that stock. These charts could be chosen to show a line graph or candlestick chart. On this page, they also can click on the “**Gear**” icon and choose “**Predict**” to view the prediction of stock trends made by us based on Artificial Intelligence.

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



*Figure 6: User can hide the information of specific stock*

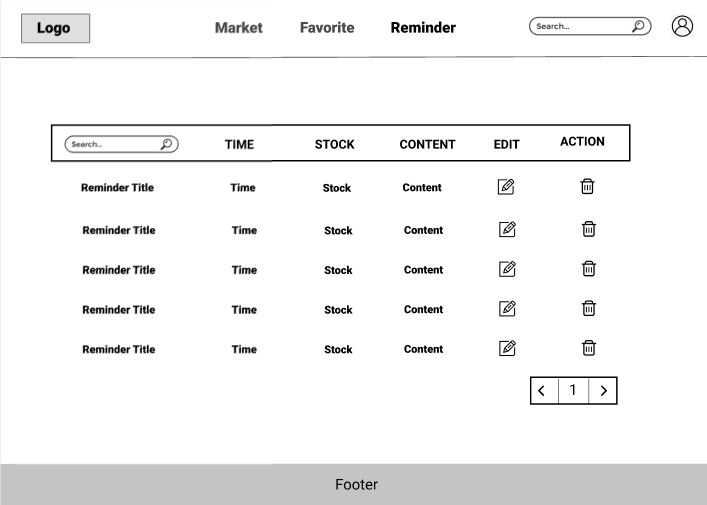
### View Stock Price



*Figure 7: 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.

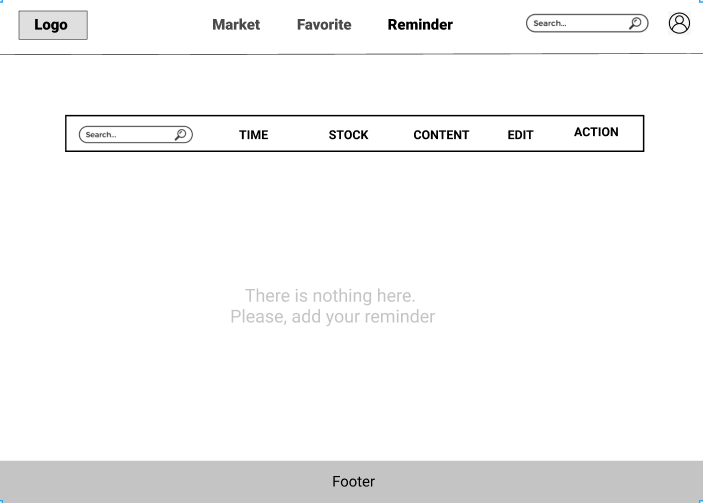
### Reminder List



*Figure 8: Reminder list*

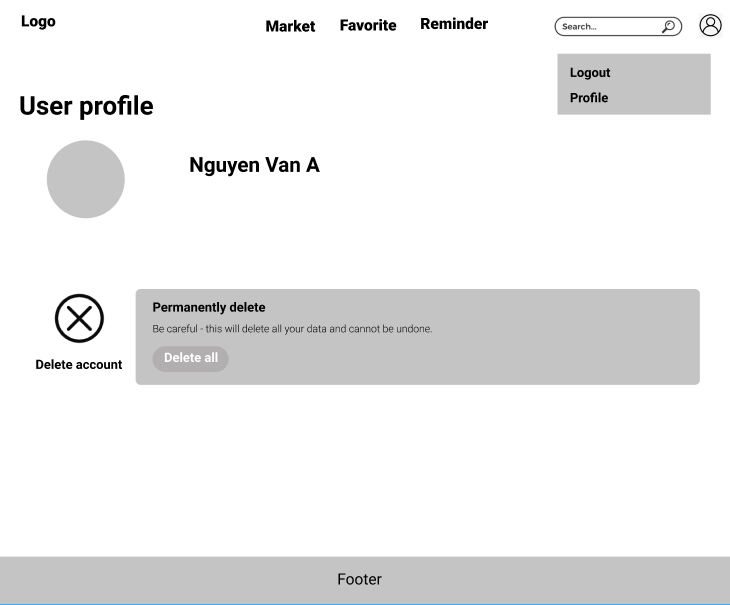
The final wireframe design will be the reminder page. This page will contain the reminder that users have set before with the detailed time they want to notify and the stock or content in each reminder. Users could also edit or remove the reminders in this page.

In case, user does not have any reminder, a message “**Add your reminder**” will be displayed.



*Figure 9: There is no reminder in the reminder list*

### User Profile



*Figure 10: 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

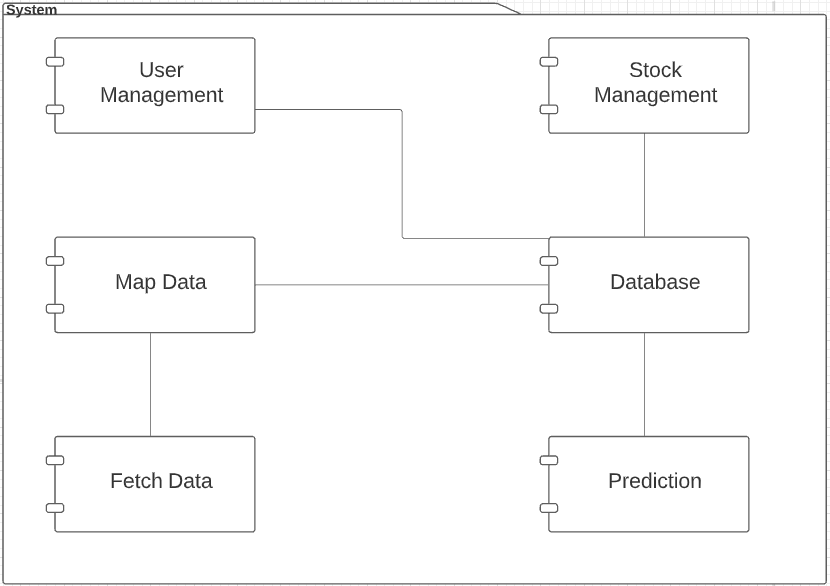
Below features are not included in this project:

* 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



*Figure 11: Component Diagram*

The component diagram above shows the high-level view of the technical architecture of our system. Five components play their role in the system respectively below:

* **Fetch Dat**a fetches data from other third parties (VCBS - Vietcombank Securities, CafeF). The data includes stock information (High, Low, Ceiling, Floor, Match, Volume, Time, and Stock Ticker) and company information (Full company name, field, address, website, brief information, EPS, P/E, book value, shares outstanding, market capitalization, total assets, ROA, ROE). Data will be collected every one minute.
* **Map Data** maps data from other sources to system databases appropriately.
* **Database** is the place storing all system data, and response when requested.
* **User Management** manages all users’ necessary information, reminder list, favourite list.
* **Stock Management** manages all stocks and their company’s information, generates charts to display for users.
* **Prediction** predicts stock price for upcoming week based on a selected algorithm.

### 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.

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.

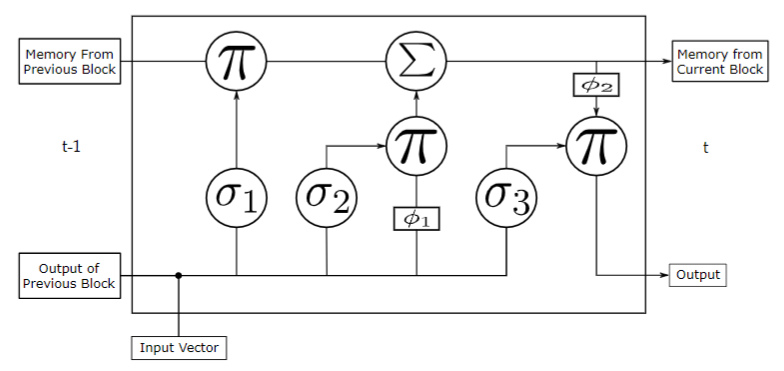
Support Vector Machine (SVM) is an algorithm belonging to the Supervised Learning group and used in classification. This algorithm will try to maximise the margin (M) to find the best hyperplane and divide data, the data points in two boundaries is called support vector because we could use them to support finding hyperplane.  
  
Linear regression uses the relationship between the data-points to draw a straight line through all them. This line can be used to predict future values.

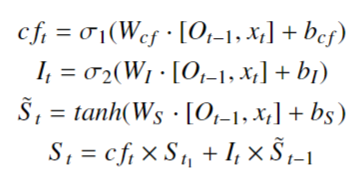
We will choose the LSTM algorithm in our project because after some demo code in the data we will use about Vietnam stock market, LSTM has the best performance in prediction.

### . LSTM (Long short-term memory)

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

#### How it works:

****

****

*Figure 12: Repeating module LSTM*

The symbol π and Σ represent element wise multiplication and addition respectively. The concatenation operation is represented by symbol (•) bullet. The fundamental component of LSTMs is the cell state, a line running from Memory from Previous Block (S t−1) to Memory from Current Block (S t). It allows the information to flow straight down the line. Networks can decide the amount of previous information to flow. It is controlled through the first layer (σ1). The operation performed by this layer is given in (2). The new information to be stored in the cell state is computed using two network layers. A sigmoid layer (σ2) that decides values to update (It) (see (3)) and tanh layer ϕ1 that evolves a vector of new candidate values (S˜t) as shown in (4). The combination of to be added in the state. Finally, cell state is updated using (5) (Jitendra Kumara, 2018).



*Figure 13: LSTM future prediction*

The graph below shows the comparison between prediction and actual price. The RMSE score is 1.37



*Figure 14: LSTM testing and training comparison*

Source: <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>

#### 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.

#### 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 (Support Vector Machine)

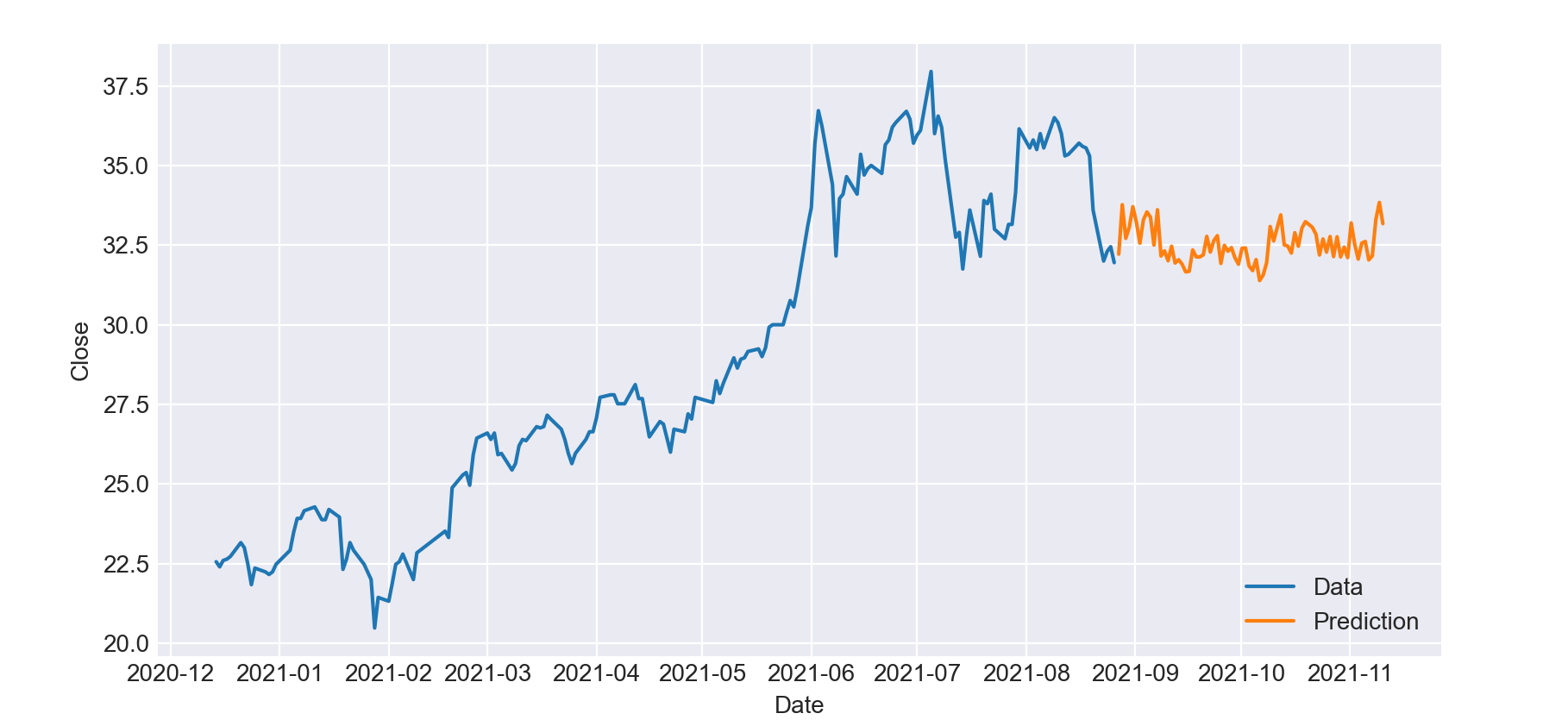
* **Input**: Close price and date of s specific stock
* **Output**: predicted price
* **Accuracy:** 29%
* **RMSE**: 0.05

#### Advantages:

* 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

#### Disadvantages:

* 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.

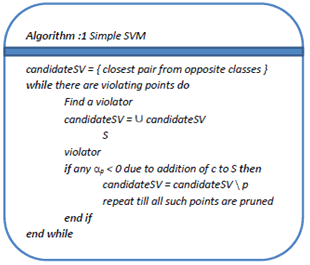
****

*Figure 15: Forecast for ticker ACB*



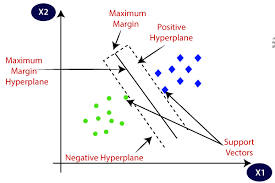
*Figure 16: Testing and training comparison of stock ACB*

### 



*Figure 17: SVM pseudo*

SVM chooses the extreme points/vectors that help in creating the hyperplane. These extreme cases are called support vectors, and hence the algorithm is termed as Support Vector Machine. Consider the below diagram in which there are two different categories that are classified using a decision boundary or hyperplane. (Support Vector Machine Algorithm, n.d.)



*Figure 18: Working of Support Vector Machine*

### . Linear Regression

* **Input**: Close price and date of s specific stock
* **Output**: predicted price
* **Accuracy: 7%**

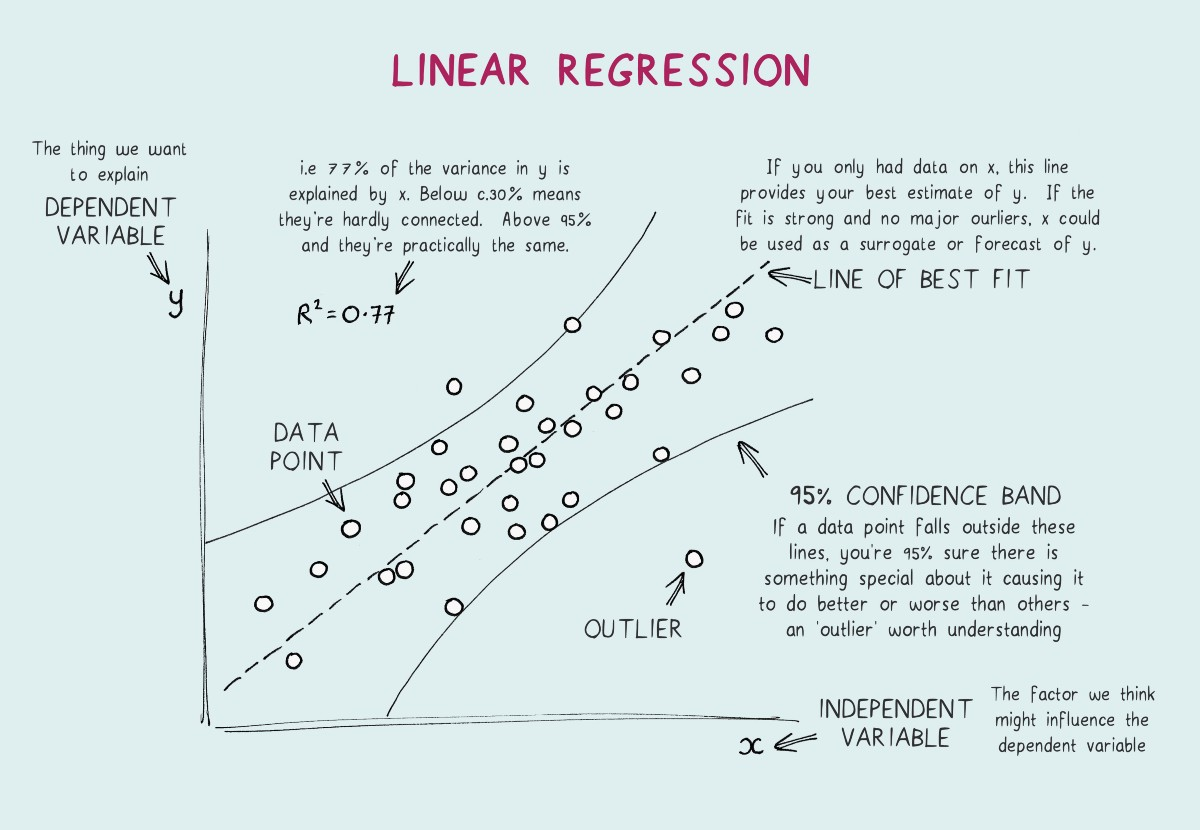
#### Advantages:

* Simple implementation
* Performance on linearly separable datasets
* Overfitting can be reduced by regularisation

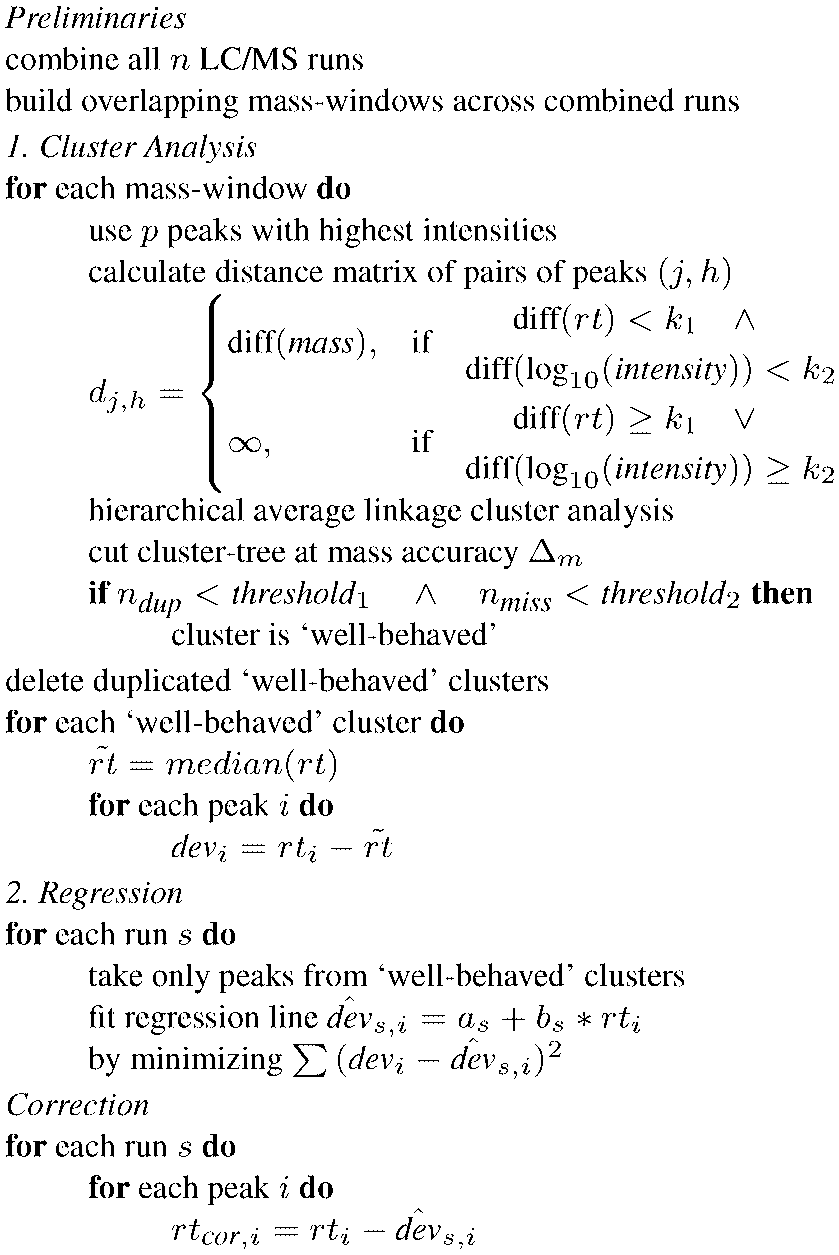
#### Disadvantages:

* Prone to underfitting
* Sensitive to outliers
* Linear Regression assumes that the data is independent

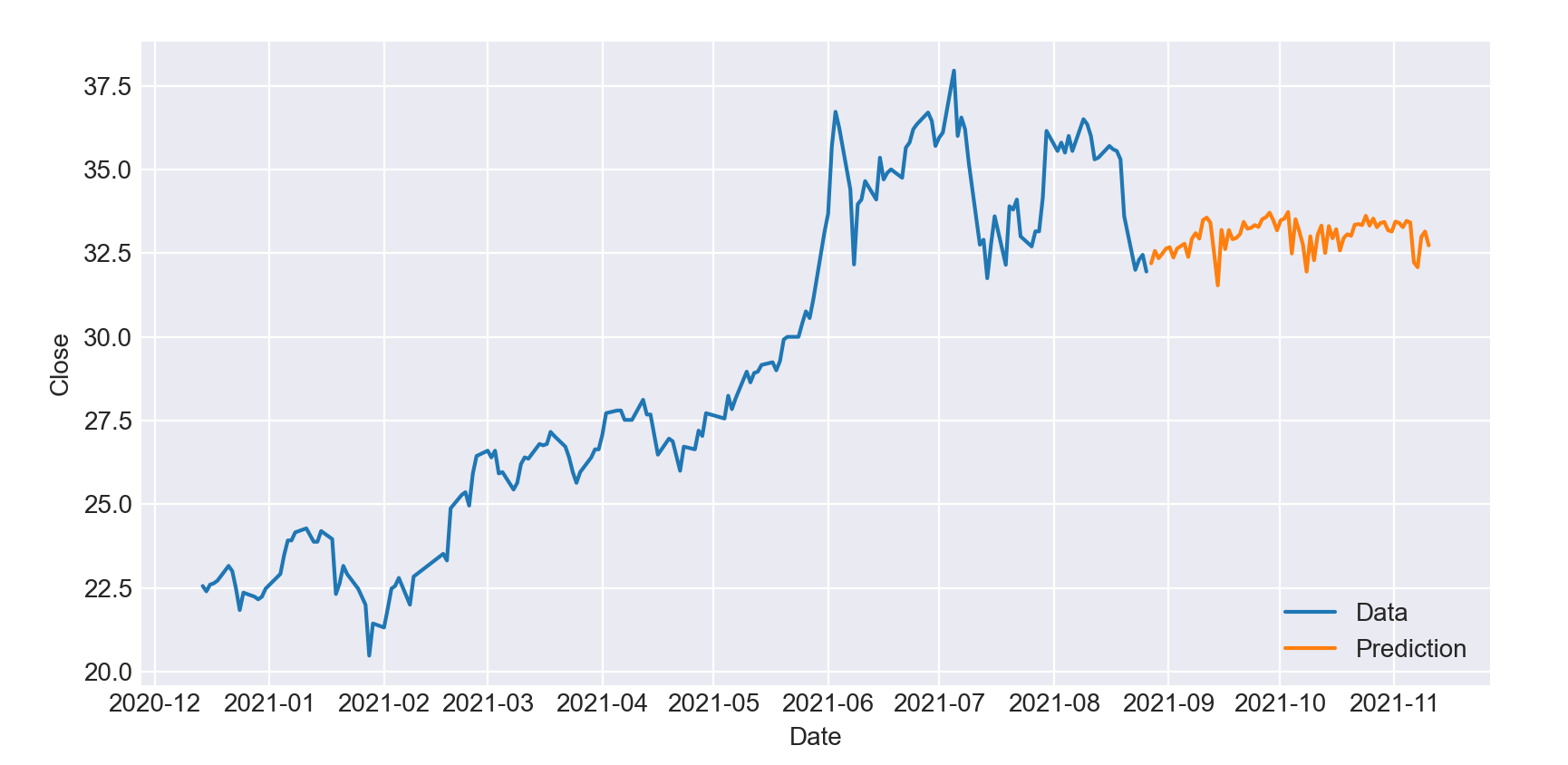
Regression analysis is a statistical methodology that allows us to determine the strength and relationship of two variables. Regression is not limited to two variables; we could have 2 or more variables showing a relationship. The results from the regression help in predicting an unknown value depending 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. (Wong, 2020)



*Figure 19: Working of Linear Regression*



*Figure 20: Linear Regression Pseudo*

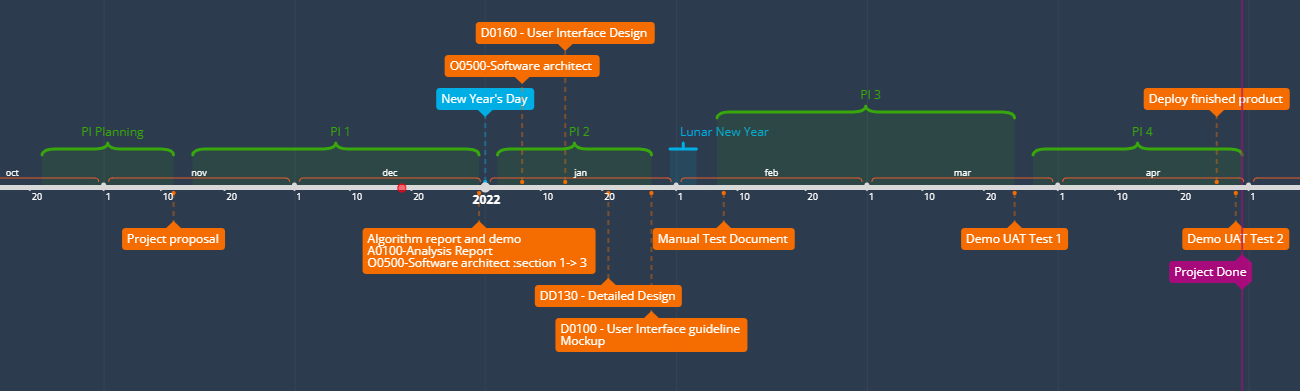
**

*Figure 21: Forecast for ticker ACB*

### 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 22: 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

Jitendra Kumara, R. G. (2018). *Long Short Term Memory Recurrent Neural Network (LSTM-RNN) Based Workload Forecasting Model For Cloud Datacenters*. Retrieved from Science Direct : https://www.sciencedirect.com/science/article/pii/S1877050917328557

*Support Vector Machine Algorithm*. (n.d.). Retrieved from Javatpoint Logo: https://www.javatpoint.com/machine-learning-support-vector-machine-algorithm

Wong, J. (2020, Nov 28). *Linear Regression Explained*. Retrieved from Towards Datascience: https://towardsdatascience.com/linear-regression-explained-1b36f97b7572