STOCK MARKET PREDICTION

COLLEGE OF COMPUTING AND INFORMATION SCIENCES

|  |  |
| --- | --- |
| NAME | REGISTRATION NO. |
| KALUUMA MAHAD | 17/U/1405/EVE |
| KINTU ARNOLD | 17/U/5188/EVE |
| sLUBWAMA ISAAC | 17/U/5739/PS |
| OLARA SAMUEL OBADIAH | 17/U/9634/EVE |

*Supervisor*

**Mr. Grace Kamulegeya**

System Requirements Specification

*The users of the data pipeline*

For our pipeline, there will be two users (actors in our case diagram) that is the system administrator and the investor who is the client in this case.

The system administrator uses the pipeline to:

1. Load data

The data cannot be used without being loaded so it is important to load the data by reading it in form of a csv file from its source.

1. Wrangle data

Wrangling data consists of four main categories that all have their different justifications which include;

Feature scaling which helps to scale the spaces in the values of the variables and put them in the appropriate range to be worked on.

Outlier removal helps to remove distant values from the other values in the dataset and therefore match the values also in the same range.

Data cleaning which will include the splitting of data into two sectors that are train and test so that the model is trained and tested appropriately using the algorithms like linear regression

Linear encoding in this case will not be necessary because the values in the data set are already in numeric representation as they are supposed to be.

1. Visualize the data

The administrator uses the pipeline to represent the data in visual and graphical formats that can help the investors analyse the data themselves by using the various tables and graphs.

1. Work algorithms and evaluate them

Predictive algorithms are used on both the train and test data that were split in order to derive a model that will be trained and tested using the values in the dataset.

Both the system administrator and the investor use the pipeline for deployment. Deployment is important to the administrator because he runs the whole system and also mobilizes all the algorithms to be used in the system. He is also responsible for the storage of the data of the system.

The investors use only the deployment part of the pipeline because they don’t need to know what happens in the background and therefore, they are focused on the foreground hence they use it to see the stocks that were run and predicted in order to find out the most valuable stocks to purchase.

Fig.1 on the next page shows the use case diagram for the pipeline

Figure 1.. Stock Prediction use case diagram.

For our use case diagram the wrangle data use case includes data sampling, data cleaning, visualization of data and removal of outliers. Data sampling is a statistical analysis technique which uses data points from our data sets for manipulation and analysis to identify patterns in a larger data set, for data cleaning there is removal of inaccurate records from our data set to ensure accuracy and for data visualization there is encoding of data as visual objects contained in graphics. Outlier removal deals with removal of extreme values that deviate from other data observations. This is all accomplished within the data wrangle use case

The deployment use case is interacted with all the actors that is system administrator and investor. The system administrator manages and controls the operations of the deployment use case, the investor on the other hand interacts with it as a web interface for registration on to the system that is account creation which he uses to monitor the current and predicted stock prices thus choosing a more viable stock for investment.