Electronic Assignment Cover sheet - Please complete out and attach as the first page of Assignment.

Student Declaration:

By submitting this assignment, I confirm that I am aware of DBS's policy regarding cheating, plagiarism and all other forms of academic impropriety. The coursework submitted is my own or my group's work, and all other sources consulted have been appropriately acknowledged. I am aware that in the case of doubt an investigation will be held.

Programme:
MSc in Business Analytics

Module:
B9BA107 Applied Research

Project Title:
Prediction And Visualization Of The FMCG Stocks Prices

Submitted to:

Dr. Paul Mc Evoy

Submitted by:

Konika Sankhe (10600575)

Abstract

This research attempts to understand the principle of Stock price forecasting and research which algorithms are the best suited to perform the evaluative task of predicting stock trends at the time of the Covid-19 and Ukraine-Russia Wars. Additionally, the stock trends visualizations will also be done.

Ten year's worth of historical stock price information will be made available for the fast-moving consumer goods (FMCG) sector by India's NSE and BSE. The years 2012 to 2022 will be chosen for the historical dataset. This work suggests the Deep Learning (DL)-based Long Short-Term Memory (LSTM) Algorithm and Auto-ARIMA. Additionally, the outcomes show that the Power BI platform can give an analysis of stock price patterns in an original and efficient manner. The research aims to forecast the churns for the upcoming years.

Keywords : Long Short-Term Memory, deep learning, ARIMA model, Power BI, Stock Analysis, Deep Learning, Stock Price prediction, Visualization

Contents

1. Introduction

Covid and War have imparted a priceless lesson for life. Countries' economies crumbled during the epidemic. An area of multidisciplinary scientific inquiry is stock price forecasting. In the area of time series prediction, it is a widely utilized application. Time series forecasting might assist investors in the financial sector to develop a solid plan and reduce the risk of current erroneous investments. The research suggests calculating stock prices by examining the historical price behavior of individual equities.

The system must be effective and reliable since investors incur risks when investing their money in stocks and may rely on it to offer them superior insights into various stocks and their patterns. The procedure only delivers the pro forma future price of the companies. The suggested approach is anticipated to take into account many aspects and components that might influence the performance and price of the stocks.

2. Literature Review

In this part, the results of various studies are gathered. Research articles that were read leading up to and during the development proposal event may be found in this section. The investigations helped in comprehending present strategies and potential strategies to best answer the given research issue.

The authors in their project "Stock Price Analysis and Prediction" calculated stock prices by examining the historical price behavior of individual equities. The researcher has used LSTM neural network to perform the time series [1]. The LSTM model has predicted the value almost closer to the actual value [1]. Yimian Yao in their research proposed the LSTM RNN-based stock price prediction model, which, after accuracy checking, can predict stock prices reasonably accurately. This allows stock speculators to have a broad understanding of the direction of the stock market and to buy or sell at the right time in order to maximize their profits [2]. As a result of the findings, investors may utilize the LSTM RNN approach in their investment process to anticipate stock prices with a fair degree of accuracy [2].

In order for stock speculators to have a broad understanding of the direction of the stock market and to buy or sell at the right time to maximize profit, this article presents the LSTM RNN-based share price, prediction model [3]. This model can make a reasonably accurate prediction of share value after an accuracy check. The results show that using the LSTM RNN method can make relatively accurate stock price predictions, and investors can use this method in the investment process to improve the investment success rate [3]. With an accuracy of 83.88 percent, the findings of the suggested model look rather good [3]. The findings unambiguously show that the most accurate model is the LSTM-based univariate model, which takes historical data from one week as input to predict the closing value of the Reliance Industries Limited time series for the next week [4]. They used LSTM, Auto-ARIMA, and Linear Regression and evaluated each one's precision to see which one predicted the outcome with the best accuracy. Extensive outcomes on several variables are provided for each deep learning model [4].

Stock price forecasts for "Predicting Stock Price using Data science technique" research were computed using logistic regression, random forest, decision trees, and naive Bayes theorem [5]. An evaluation of the proposed machine learning algorithm technique has been made with regard to identifying the confusion matrix, categorizing data based on priority, and comparing it with the best levels of precision, recall, and the F1 score. The genuine computation of the informational collection of the refreshed framework researchers needed was 67% utilizing a one-of-a-kind timberland calculation [5].

This paper suggests using leveraged Power BI framework to filter and visualize the data in a simple and digestible format which can assist to formulate a conclusion [6]. The findings show that the Power BI architecture can give an analysis of the cost of living and industry predominance in a creative and effective way. Therefore, future examinations for each individual year can aid in identifying the broad trends throughout time [6].

The above projects have used the LSTM and Auto ARIMA method to do time series forecasting however, they were not performed in the particular industry. The proposed study will use ARIMA and LSTM algorithms for stock price forecasting by analyzing NSE and BSE FMCG

industry’s stock prices. Consumer products that sell quickly and for a cheap price are known as fast-moving goods. Consumer packaged products are another name for these items.

Additionally, Power BI visualization be used to assess how the stock market evolved during COVID and the Ukraine-Russia War.

3. State of the Art

This part will provide the research question and the research objectives of the proposed study.

3.1 Research Questions

1. Which model will perform better for time series forecasting of FMCG stocks, LSTM or ARIMA?
2. How can Power BI visualization be used to assess how the stock market evolved during COVID and the Ukraine-Russia War?

3.2 Research Objective

The main goal of this project is to use the ARIMA (Auto Regressive Integrated Moving Average) model and LSTM Network (Long Short-Term Memory Network) which is an advanced RNN to perform a comparative analysis.

The NSE and BSE FMCG industry's stock prices dataset will be taken to apply these models. The research will show which model is outperforming the other. This will help to get an idea about the performance of these two models and which model can be used for better performance and accuracy as well as the time duration will be noted down.

The secondary goal of this proposed project is to use the business intelligence tool in order to perform the visualization of the FMCG industries. The visualization will be performed on particular industries to identify the differences between their stock markets and will be analysed to identify how the stock market has changed over the time period.

3.3 Hypothesis for Quantitative Research

A hypothesis is a claim that we are attempting to support or refute. It is used to indicate how variables are related to one another and whether or not this connection is important. It provides a prediction of the answers to research inquiry and is specific.

The following quantitative research hypothesis has been tested in the intended study:

The suggested study develops the premise that the stock price patterns in the two occurrences stated were different.

According to the second possibility, COVID-19 might have had a greater impact on trends than in wartime.

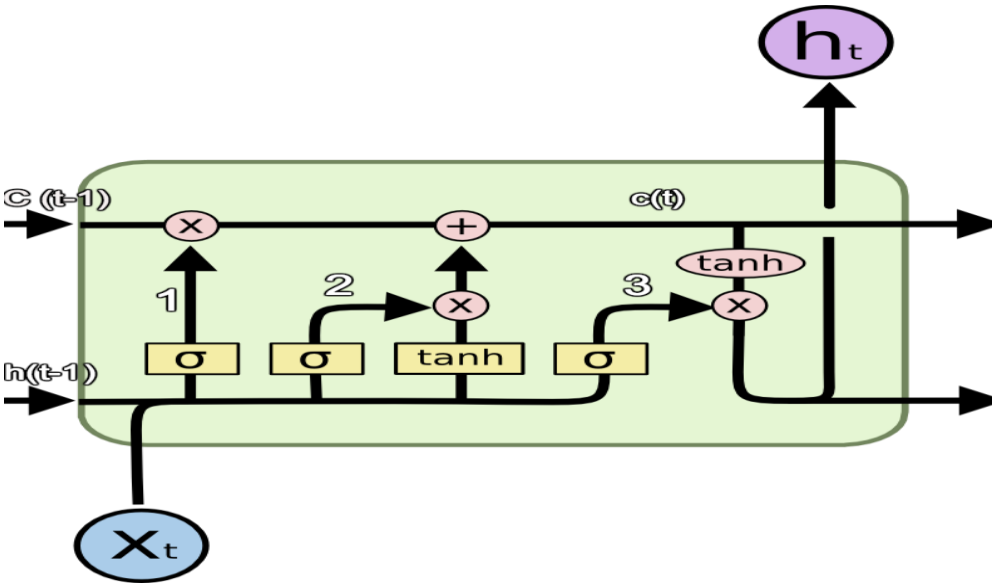
The third assumption is that LSTM would outperform ARIMA since a large amount of data will be gathered over a 10-year period. However, ARIMA can outperform it if the time series forecasting is done on a lower amount of data.

3.4 Proposed Models (Resources)

Both ARIMA and LSTM will be used to check if they can be adapted to build prediction models for the Fast-moving consumer goods stock prices. Additionally, further research will include checking which one is the best of the two of them and providing an accurate prediction score.

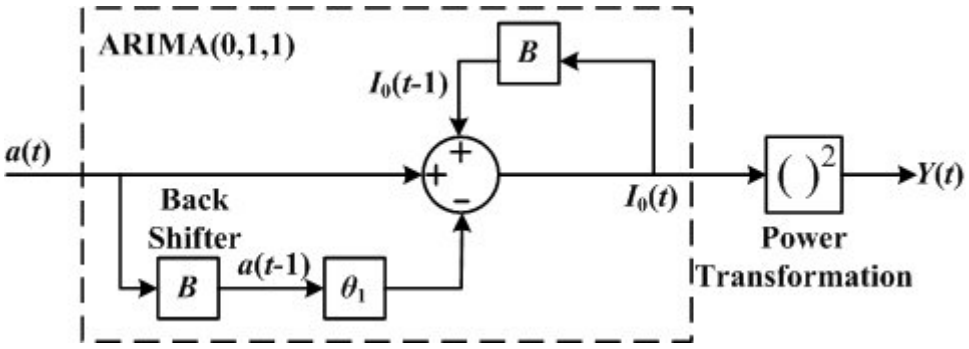
3.4.1 LSTM

Time series forecasting models may anticipate future values based on prior, sequential data by utilizing LSTM. This improves demand forecasters' accuracy, which helps the firm make better decisions. The Long Short-Term Memory Network (LSTMN), an improved RNN (sequential network), enables the permanent storage of data. A recurrent neural network, or RNN, is employed for permanent memory. Long-lasting dependencies cannot be remembered by RNNs. With LSTMs, long-term dependence issues are particularly avoided. LSTMs are widely employed and have shown to be quite effective for applications involving sequence prediction.



3.4.2 ARIMA

An ARIMA process underlies an ARIMA model for a time series in that the process that produced the observations is an ARIMA process. In cases where the series is steady, ARIMA may be preferable. Before using moving data, the Autoregressive Integrated Moving Average (ARIMA) Model converts it to stationary data. The most popular model for predicting linear time series data is this one. The ARIMA model is commonly employed in banking and is effective in forecasting movements in the share market. Even more so than the most widely used ANNs approaches, ARIMA models are recognized to be reliable and effective in financial time series forecasting, especially short-term prediction.



3.4.3 Power BI

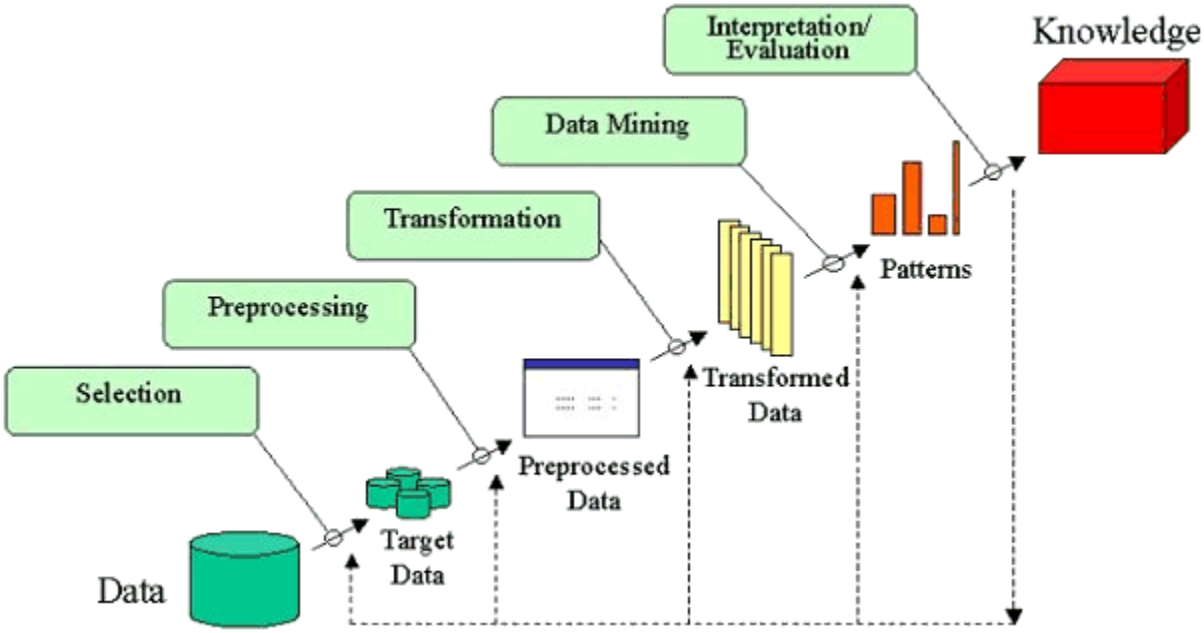
Microsoft's Power BI is a solution for business analytics that allows you to analyze and display data, draw conclusions from it, and share it with different organizational divisions. While Power BI is more suited to a wide audience that requires business information to supplement their analytics, Tableau is designed for data analysts. The cost per user is affordable but still more expensive than what you receive with Power BI if you already store a lot of data on worksheets and want to take the effort to extract your data from third-party tools to upload to Tableau. With easy access to straightforward visualizations, Power BI offers a robust backend data manipulation function. There isn't one in SSRS.

4. Research Methodology

This section will explain the specific methods or techniques that will be used to locate, pick, process, and evaluate data on a subject.

4.1 Time Series Forecasting and Prediction

The creation of this study will adhere to the Knowledge Discovery in Databases (KDD) approach. There are six steps in all, as shown in figure 1. Knowledge Discovery in Databases (KDD) is an ongoing process that enables the improvement of mining methods, the inclusion of fresh data, and the alteration of old data to provide more recent results.



The overall process of finding and interpreting patterns from data involves the following steps:

1. Data Selection
   The dataset will be picked up from the National Stock Exchange(NSE) and Bombay Stock Exchange (BSE) sites. It will be related to the Fast-moving consumer goods (FMCG) industry.
   The research will be performed on the last ten years of data, which will be from the year 2012 to 2022 in order to perform the time series forecasting.
   On the other hand, for visualization the different companies' data will be taken based on year, month, and days. Also, the data at the time of corona and the war will also be fetched from the site.

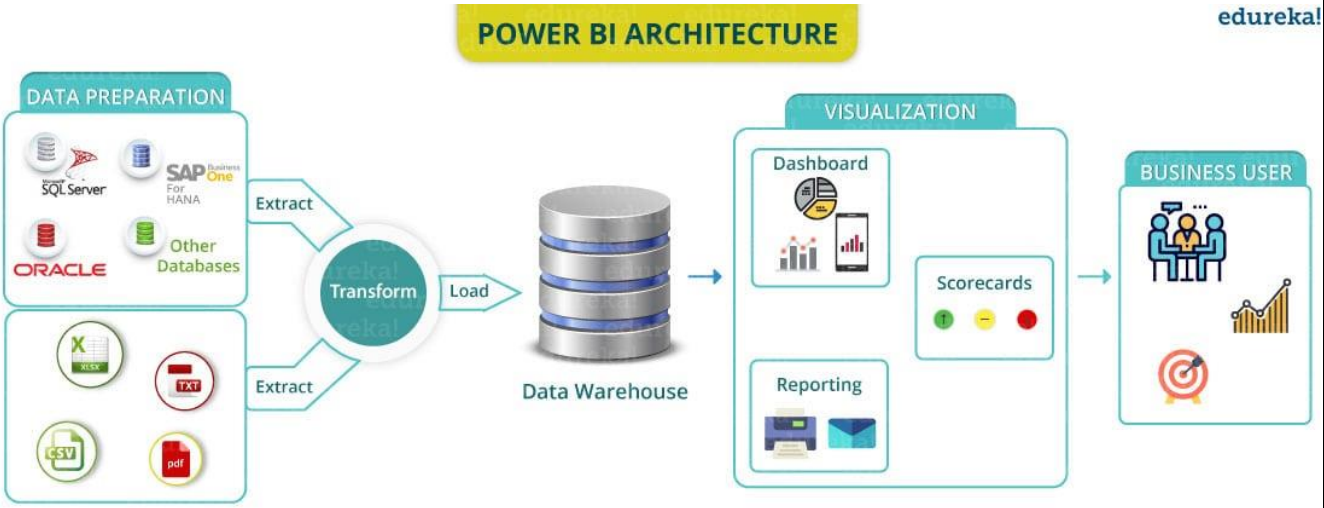
2. Data Cleaning / Preprocessing
   The practice of correcting or deleting inaccurate, damaged, improperly formatted, duplicate, or missing data from a dataset is known as data cleaning. Only the pertinent entries must be kept once the data has been filtered to eliminate all irrelevant elements. There are two methods to accomplish it: either manually eliminating each and every

entry, or using a Python script. To prevent human mistakes, the first technique is typically employed when datasets are quite big.

3. Data Transform
   The process of changing data from one format or structure to another is known as data transformation. When translating and mapping data from one "raw" data form into another format for warehousing and analysis, transformation activities are also known as data wrangling or data munging.

4. Data Mining
   The proposed study will be performed based on the LSTM and ARIMA time series forecasting methodology. This will be used in order to predict the future prices of the FMCG stock market. Additionally, the Power BI will be used in order to visualize the last 5 years of data which were before Covid, During Covid, and at the time of the Ukraine-Russia war.

4.2 Methodology of Power BI visualization



Data Preparation –

Some processing is necessary on the raw data. The raw data is transformed through a number of processing or purification steps, including the elimination of duplicated numbers, etc. After processing the data, we apply pertinent business principles to it in order to change it into what we need for our business. The data warehouses receive the modified data and load it. With this, the entire ETL procedure is finished.

Data Transform –

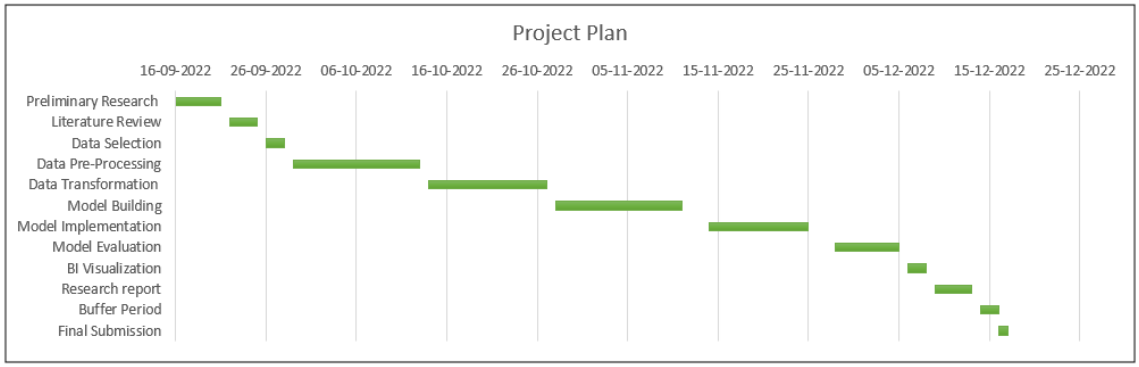
Once the data has been prepared it will be extracted in to Power BI and loaded in to the data warehouse. Relational environments called data warehouses are used to analyze data, particularly historical data. Data warehouses will be used to identify trends and connections in their data over time.

Visualization –

The act of turning raw data into graphical or pictorial representations, such as charts, graphs, diagrams, photos, and films that illustrate the stock market's developments during COVID-19 and the Ukraine-Russian War, is known as data visualization. This will make it possible to learn from it. so that it would be simpler to produce reports and conduct a rapid examination of the data.

4.3 Research Timeline

Task	Start Date	End Date	Duration
Preliminary Research	16-09-2022	21-09-2022	5
Literature Review	22-09-2022	25-09-2022	3
Data Selection	26-09-2022	28-09-2022	2
Data Pre-Processing	29-09-2022	13-10-2022	14
Data Transformation	14-10-2022	27-10-2022	13
Model Building	28-10-2022	11-11-2022	14
Model Implementation	14-11-2022	25-11-2022	11
Model Evaluation	28-11-2022	05-12-2022	7
BI Visualization	06-12-2022	08-12-2022	2
Research report	09-12-2022	13-12-2022	4
Buffer Period	14-12-2022	15-12-2022	2
Final Submission	16-12-2022	16-12-2022	1



5. Conclusion

Stock market forecasting seeks to forecast future changes in a financial exchange's stock value. Investors will be able to earn more if share price movements can be predicted accurately. The main question is then how effective and potent these newly developed procedures are in comparison to conventional techniques. The proposed study will use the Long Short-Term Memory (LSTM) Neural Network and AutoRegressive Integrated Moving Average. This study will contrast the performance of ARIMA and LSTM as exemplary forecasting methods for time series data.

This study will develop a methodology for analyzing the influence of COVID 19 and the conflict between Ukraine and Russia on FMCG stock prices. This study will offer a visualization that users may use to examine the potential effects on the economy of different sectors. The visualization will be limited to 5 years to analyze the trends before the covid-19, at the time of covid-19, and at the time of war, which means the data will be analyzed from the year 2017 to 2022.

Deep learning may be used to develop a number of additional prediction issues in the fields of finance and economics. By using similar approaches on additional issues and datasets with varying amounts of features, the authors want to examine the progress brought about by deep learning.

6. Bibliography

[1] C. a. S. D. a. S. B. a. S. N. B. Vora, "Stock Price Analysis and Prediction}," *2021 International Conference on Communication information and Computing Technology (ICCICT),* 2021.

[2] Y. Yao, "Data Analysis on the Computer Intelligent Stock Prediction Model Based on LSTM RNN and Algorithm Optimization," *2022 IEEE International Conference on Electrical Engineering, Big Data and Algorithms (EEBDA),* 2022.

[3] P. S. a. G. A. a. K. Y. a. A. G. K. Sisodia, "Stock Market Analysis and Prediction for Nifty50 using LSTM Deep Learning Approach," *2022 2nd International Conference on Innovative Practices in Technology and Management (ICIPTM),* 2022.

[4] M. a. U. J. a. P. K. a. D. R. Hirey, "Analysis of Stock Price Prediction using Machine Learning Algorithms," *2022 International Conference for Advancement in Technology (ICONAT,* 2022.

[5] M. a. B. K. a. J. J. a. I. G. Roobini, "Predicting Stock Price using Data science technique," *2022 Second International Conference on Artificial Intelligence and Smart Energy (ICAIS),* 2022.

[6] W. a. X. H. a. N. M. a. S. W. a. W. M. Hu, "Power BI for Impacts Analysis on Cost of Living Caused by Industry Prevalence in Smart Cities," *2019 3rd International Conference on Smart Grid and Smart Cities (ICSGSC),* 2019.

[7] V. Padghan, "Power BI Tutorial," 07 July 2022. [Online]. Available: https://www.edureka.co/blog/power-bi-tutorial/.

[8] N. Pyvovar, "Data Science project management methodologies," [Online]. Available: https://medium.datadriveninvestor.com/data-science-project-management-methodologies-f6913c6b29eb.

[9] N. Sinha, "Towards Data Science," [Online]. Available: https://towardsdatascience.com/understanding-lstm-and-its-quick-implementation-in-keras-for-sentiment-analysis-af410fd85b47.