Time Series Forecasting of Agricultural Products Sale Using Deep Learning

Conference Paper · July 2023

DOI: 10.1109/ICCCNT56998.2023.10307538

4 authors:

Md. Touhidur Rahman
Daffodil International University
1 PUBLICATION **3** CITATIONS
SEE PROFILE

Mst. Eshita Khatun
Louisiana State University
15 PUBLICATIONS **66** CITATIONS
SEE PROFILE

Archina Mahmuda Asha
1 PUBLICATION **3** CITATIONS
SEE PROFILE

Mohammad Jahangir Alam
Daffodil International University
15 PUBLICATIONS **109** CITATIONS
SEE PROFILE

2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT) | 979-8-3503-3509-5/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICCCNT56998.2023.10307538

Time Series Forecasting of Agricultural Products Sale Using Deep Learning

Md. Touhidur Rahman
Department of CSE
Daffodil International University
Dhaka, Bangladesh
touhidur15-10361@diu.edu.bd

Mst. Eshita Khatun
Department of CSE
Louisiana State University
Baton Rouge, LA 70803
mkhatu3@lsu.edu

Archina Mahmuda Asha
Department of CSE
Southeast University
Dhaka, Bangladesh
2017000000021@seu.edu.bd

Mohammad Jahangir Alam
Department of CSE
Daffodil International University
Dhaka, Bangladesh
jahangir.cse@diu.edu.bd

***Abstract*— Due to the massive production of data, the time series dataset is useful for time-based prediction as it holds time related information. Different forecasting techniques help to predict in the field of the stock market, sales, healthcare, banking, weather etc. The coming out of new companies has increased every year, and they have taken part in the competition with other existing organizations with their products and services. As the competition is growing day by day, sales analysis is a must for every organization to make a strong position in the competitive market. In that case, product and area-wise sales analysis can help a company to attain its aim by accomplishing its target. The success of a business mainly relies on the sales of its product. Predicting the sales may help the company to discern the amount of growth rate. Moreover, this research target is to estimate the production amount for their next manufacture. More precise predictions can give fame and make companies successful. Though it is a challenging task, with the help of machine learning algorithms this issue can be sorted out effectively. In this research, we have analyzed the product selling rate and used Multi Step LSTM for forecasting the sales of the agricultural product for the next month sales.**

Keywords*— Product Sale Forecasting, Agriculture Product, Time Series Models, LSTM*

I. INTRODUCTION

A time series is a gathering of observations of well-defined data items acquired through repeated measurements over time. A piece of data is tracked at an enhancement in time. To obtain significant statistics and other properties of data time series analysis plays an important role in in the business environment [1]. Time series forecasting is the method to examine time series data to make a prognostication using statistics and modeling. The applications of forecasting are seen in retail forecasting, finance forecasting, economic forecasting, business forecasting, weather forecasting, healthcare forecasting and so on. To prophesy future liveliness time series forecasting uses information concerning historical time stamped data and associated patterns.

In today's competitive era, every business has to keep it updated with current technology. Sales anticipating is crucially important for modern business to alleviate the cost and enhance customer service level [2] [3]. Sale is an important criterion for every company to make it successful in the competitive marketplace. Predicting sales can make every company one step further towards prosperity.

Forecasting the future always offers charms and fascination in persons though it is a venturesome task. It can be more glamorous in the case of predicting revenue by analyzing a company's sales [4]. The evolution of technology has made this task easier in recent times. Nowadays, a lot of retail companies are trying to improve their sales by analyzing and making predictions for future sales. In this regard, these companies take the assistance of modern technology. There are many general time series forecasting approaches that can be applied to anticipate agricultural product sales which are mainly based on time series [5].

In Computer Science, different data mining or machine learning algorithms can make this task effectively and make organizations successful in earning more revenue by forecasting their sales. With the help of machine learning tools, organizations can effectively do this task and earn more revenue. Machine learning approaches can analyze organizations' current sales including the demand of a specific product for a particular area and help organizations decide to focus on increasing supplies of that particular product for a specific region in the upcoming days so that the organizations can earn more revenue.

The objectives of this research are follows:

- To explore companies current sales including every product and every region and to observe what type of products are on-demand in which region including different product sizes.
- To forecast the next month sales of the agricultural product.

For that we have collected a primary dataset of a company with different types of attributes to model the data training, testing and evaluating. The structure of this paper is following. In section I, we have described the introduction. Section II stated

the literature review based on time series forecasting related works. Section III is presented methodology of this experiment analysis. Results is elaborately discussed in section IV. Finally, in section V, We have concluded and accounted future scope.

II. Literature Review

Currently, many models are presented only for predicting upcoming sales rate of the product. Product sales rate analysis is a study which emphasizes evaluating business needs and finding solutions. In this research work, initial knowledge has been taken from the related paper works.

A lot of processes involve making prediction successful for a specific domain in machine learning such as data collection, dataset cleaning, model choosing, training, and final decision making. Machine learning is categorized into supervised, semi-supervised, unsupervised, and reinforcement [6]. Supervised learning covers classification algorithms, regression algorithms, and deep learning. An abundance of research papers has been published by the researchers by using machine ML techniques to analyze sales statements for a successful prediction strategy.

A study [7] prognosticate the performance of the Karachi Stock Exchange (KSE) in the marketplace. Different ML techniques have been used to analyze the data. Among all other algorithms, the Multi-Layer Perceptron technique responds best with 77% correct prediction, followed by RBF (63%), SVM (60%), and SLP (60%).

Another study [8] focused on the churn prediction problem. The researcher has worked with ten analytical algorithms. For this research, the dataset of customers of a Telecom company has been used containing 3333 records. With an accuracy of 96%, both random forest and ADA boost have obtained the highest performance. Multi-layer perceptron and SVM both have achieved 94% accuracy.

In [9] they have worked with supervised learning to forecast the success of a company based on Crunchbase data containing more than 60K companies' information. To perform their task, they have used the K-Nearest Neighbors (KNN) model and have achieved 44.45% of F1 score and 73.70% of accuracy. They have also explored Logistic Regression (LR) and Random Forests (RF) models. F1 score has been used as a metric to differentiate the performance of these three techniques.

A recent study [10] has used varied forecasting approaches. They have worked with a general dataset to foresee the sales of furniture analyzing their sales history. They have applied techniques such as Seasonal Autoregressive Integrated Moving Average, TES, Prophet, LSTM and CNN. Various accuracy measurement procedures differentiate the performances of these models. They proved that CNN and the Prophet models are performing better for forecasting the future sequence of seasonal objects.

In [11] a machine learning method has been used to analyze business intelligence (BI) systems. The Structural Support Vector Machines (SSVMs) method has been applied to carry out classification. On the training samples, more than 78% accuracy has been obtained for the SSVM. A group of researchers [12] has come up with a model for the prophecy in the business sector. This model can detect automated blunders and deception. A supervised learning approach based on RF has been used in this work. The dataset contains 1,517,400 financial statements from Dutch companies including 154 different attributes.

In [13] different data mining methods have been used to forecast sales by analyzing sales data. The dataset has been based on sales data from an e-fashion store. The machine learning algorithms that have been used in this work are the Generalized Linear Model, Decision Tree, and Gradient Boost Tree (GBT). GBT model has shown the best performance with 98% accuracy. On the other hand, GLM has obtained 64% accuracy whereas DT has gained 71% accuracy. In [14] a machine learning technique has been used to predict a company's success. Dataset has been collected from the Crunchbase database. With the help of three algorithms such as LR, SVM, and the GBC (Gradient Boosting Classifier) the task is performed. The Gradient Boosting Classifier offers the best upshot among all the algorithms.

Nowadays data mining techniques can assist a seller to forecast on product sale rate. In [15] the authors focused on the review analysis by using different machine learning algorithms to predict the most popular cosmetic brands to increase the product sale rate. Another study [16] has been found that they analyzed user reviews on the book selling so that they can analysis the factors which type of books has the high demand on market. A study [17] suggested a multi-stage deep learning technique for forecasting future business process incidents. Deep stacked autoencoders and feedforward multilayer neural networks are also included in this process. Different real-life datasets have been used in this research. Their suggested model attains a good outcome.

In this research, we have performed sales forecasting on agricultural product using deep learning techniques. The task involved forecasting the time series of the next month production rate of any company, in order to familiarize we have studied previously.

III. Methodology

The main goal of this research is to analyze the data mining techniques and deep learning for agricultural product sale forecasting and also analyzing the growth rate of the company.

A. Data Collection and Preparation

In our research, we have collected data from an agricultural company named "Win Agro" which is situated in Bogura district in northern part of Bangladesh. The company provides the products are basically based on the types of disease of the

agricultural field such as insect, fungus and weed etc. The company has a total of four zones. And we have collected all the zonal data from July 2022- December 2022. Initially, we have collected 1092 data. After processing the data, we finally got the desired dataset with a total of 832 data. Our dataset contains a total of 13 attributes such as month, product code, product name, packet size, cartoon size, cartoon price, packet price, cartoon quantity, packet quantity, value, region code, use & sell with 832 entries.

For any prediction-based research real and authentic data is a must. Primary data brings more accurate outcomes and is also responsible for making an efficient and powerful prediction model.

B. *Exploratory Analysis*

After Preprocessing the data, we have done statistical analysis to find out which product is the most selling products by size, by uses of the product and by some other criteria.

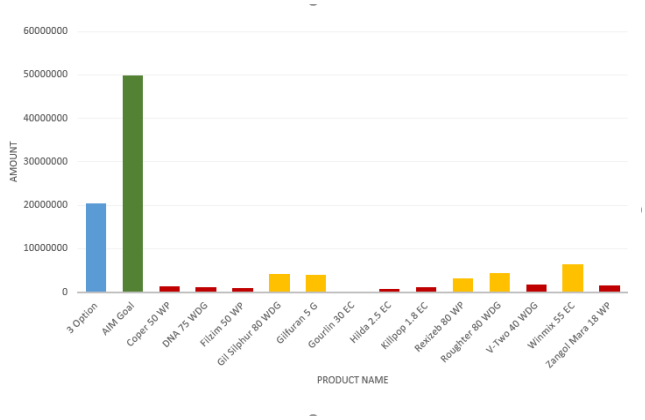

Figure 1. Statistical analysis of product wise sales

In figure 1, represents the sales amount of each product of the company that the product "AIM Goal" is the most selling product and "Gourlin 30" EC is the least selling product.

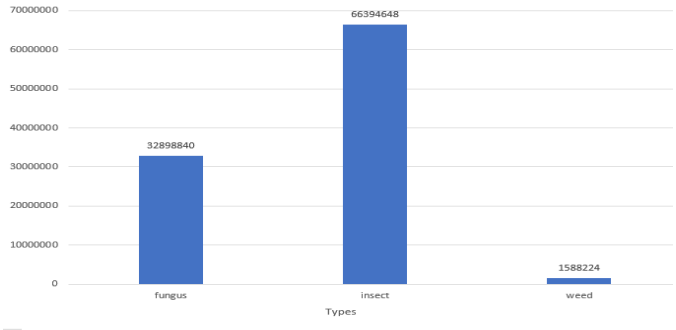

Figure 2. Sales product by type of use

In our dataset there are three types of disease related agricultural products based on uses such as fungus, insect and weed. Figure 2 represents the selling ration of those products. Findings conclude that most selling items are related to plant insects.

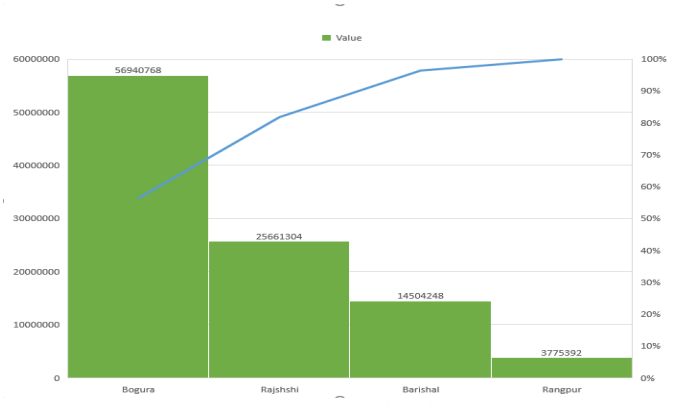

Figure 3. Different Zone of the selling product

Selling rate of the company in different zones shows at figure 3. Regression analysis findings is that the most selling Zone is Bogura which is quite higher than the other zones. The product selling company can focus on lower selling regions to increase the product sale.

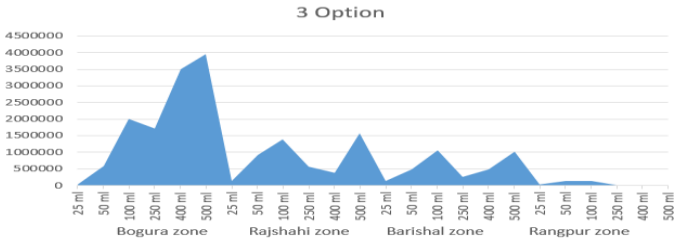

Figure 4. Packet size wise sales in several zones for "3 Option"

The company has a product with different pack sizes. From our dataset we have found that the most selling pack size for each product. Figure 4 shows the sales amount of the product named "3 Option" in different zones with the different size of the packet that the highest sales amount of the product is 500ml in the Bogura zone.

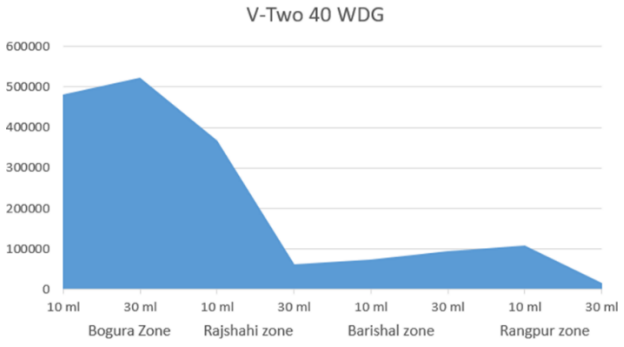

Figure 5. Packet size wise sales in several zones for V-Two 40 WDG

Figure 5 represents the sales of the product named "V-TWO 40 WDG" in different zones by packet size in amount. 100 ml pack size is the more demandable size of the product in the Bogura zone.

These statistical findings will help the company to make decision of the production of packet size, product types and choose the area for further developments.

IV. Experiment and Output Analysis

In analytical research, we have dropped unnecessary attributes at this section to boost up the performance. Our forecasting was developed with 10 attributes. After Feature selection, this study has generated the correlation matrix to represents how the features are related to each other. In figure 6 shows the visualization of the relation between features with heatmap. This study involves to predict of the production rate with multiple features of any product for the next month so we have implemented Long Short- Term Memory. LSTM is basically used for time series data processing and using multiple inputs rather than using only one feature.

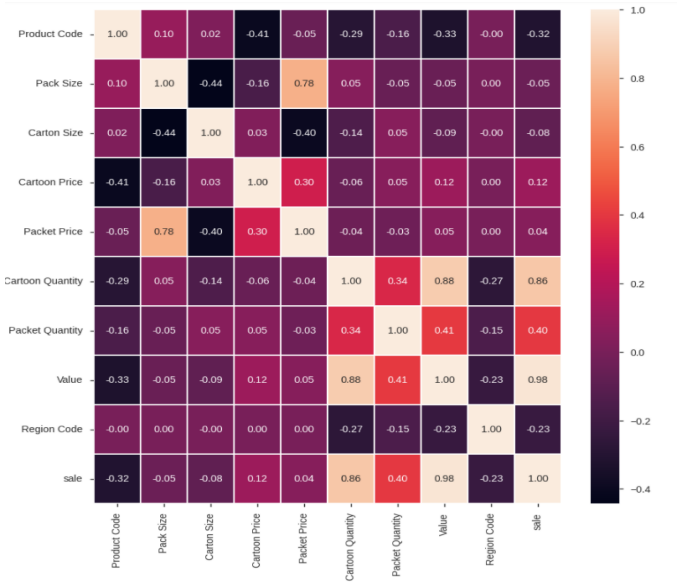

Figure 6. Correlation matrix with heatmap

Time series Forecasting: Time series forecasting used for analyzing historical data to get a predictable outcome. It analyzes the historical data using statistical methods and modeling then makes a prediction.

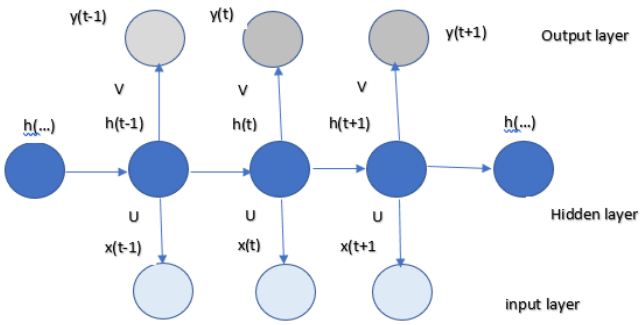

Figure 7. Time series architecture

In figure 7, there is three layers which are hidden layer, input layer and output layer. For applying the time series forecasting this study is used the historical data as the input layer.

Long Short-Term Memory: LSTM provides the solution for short term memory of RNN. After executing the many hidden layers, the RNN results may disappear due to the multiplication [4]. LSTM has three gates including input, forget and output gate. LSTM can be explained by following equations.

$$i_t = \sigma(w_i[h_{t-1}, x_t] + b_i)$$
$$f_t = \sigma(w_f[h_{t-1}, x_t] + b_f)$$
$$o_t = \sigma(w_o[h_{t-1}, x_t] + b_o) \quad (1)$$

Here,
i_t, f_t, o_t denotes input, forget and output gate
σ = sigmoid function
w_x = weight for the respective gate (x) neurons
h_{t-1} = output of the previous LSTM block at (timestamp $t-1$)
x_t = input at current time stamp
b_x = biases for the respective gates (x)

Input gate extracts the new information to preserve the new cell state. Some information from the cell state throws away by the forget gate and finally output gate is used to make active to the final output of LSTM block at timestamp t.

In this research, we have split the dataset into train data and test data then applied Multi Step LSTM for time series forecasting the sales rate of the product.

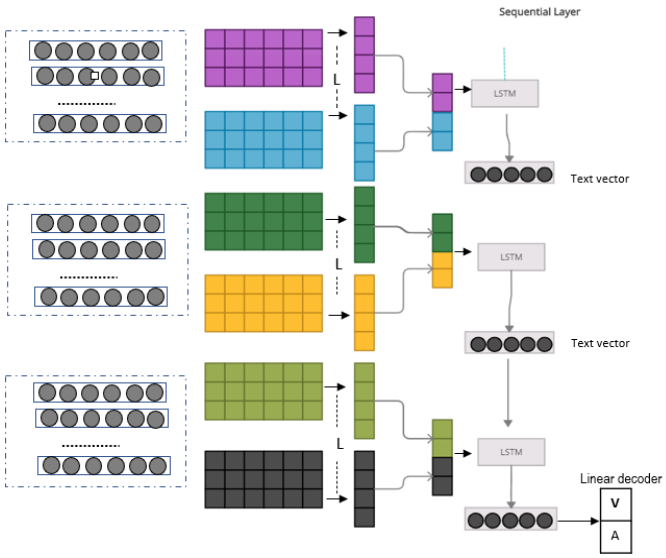

Figure 8. Long Short-Term Memory layer

In figure 8 illustrates the process visualization of LSTM where input data is used as word vector then by applying the LSTM it will process towards the architecture for getting the text vector by sequential layer.

```
Xtrain, Ytrain = X[0:int(X.shape[0] * (1 - test_share))], Y[0:int(X.shape[0] * (1 - test_share))]
Xval, Yval = X[int(X.shape[0] * (1 - test_share)):], Y[int(X.shape[0] * (1 - test_share)):]
Shape of training data: (1, 1, 4)
Shape of the target data: (1, 1)
Shape of validation data: (1, 1, 4)
Shape of the validation target data: (1, 1)
```

Figure 9. Train and test with shape of the data

After applying the model with epoch 20 we got the validation loss and the training loss which is shown in figure 10. Figure 11 states the sample of epoch steps with training and validation loss.

```
Epoch 1/20
1/1 [==============================] - 1s 1s/step - loss: 0.9478 - val_loss: 1.0990
Epoch 2/20
1/1 [==============================] - 0s 29ms/step - loss: 0.9432 - val_loss: 1.0972
Epoch 3/20
1/1 [==============================] - 0s 29ms/step - loss: 0.9386 - val_loss: 1.0953
Epoch 4/20
1/1 [==============================] - 0s 35ms/step - loss: 0.9340 - val_loss: 1.0935
Epoch 5/20
1/1 [==============================] - 0s 28ms/step - loss: 0.9294 - val_loss: 1.0917
Epoch 6/20
1/1 [==============================] - 0s 32ms/step - loss: 0.9248 - val_loss: 1.0899
Epoch 7/20
1/1 [==============================] - 0s 35ms/step - loss: 0.9202 - val_loss: 1.0881
Epoch 8/20
1/1 [==============================] - 0s 28ms/step - loss: 0.9156 - val_loss: 1.0862
```

Figure 10. Sample of epoch steps with training loss and validation loss

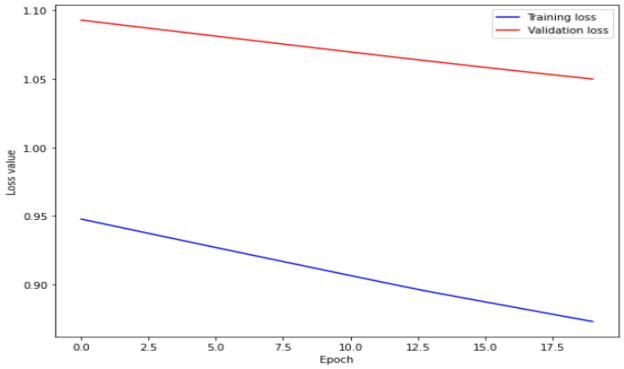

Figure 11. Training loss and validation loss

After that we got the forecasting value based on all the samples in the validation set which was 0.04831757. And the final average absolute error was 8339.66 which is showed at figure 12.

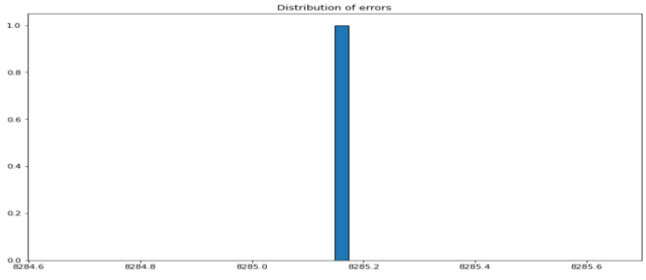

Figure 12. Final average absolute error

And the median absolute error was 8399.66. After applying the Multi Step LSTM in this research dataset we have got the prediction best result for the next month production rate.

V. Conclusion

Machine learning algorithms are now able to change technology in a way which was never imagined before. Way back in the 90's who could have thought that machines or computers would be able to learn by itself. But we are living in an era where machines are ruling everywhere and almost in every sector. In this research, we have applied deep learning model Multi Step LSTM to forecast next month manufacture of the agriculture product sale on the basis of product size, pack size, uses of the product and some other criteria.

For further implementation of the study, we have the desire to predict what types of products can be on-demand in the future and specific zones. Agricultural product makes the plantation fast which will help to reduce global warming.

References

[1] Nunnari, Giuseppe, and Valeria Nunnari. "Forecasting monthly sales retail time series: a case study." In 2017 IEEE 19th conference on business informatics (CBI), vol. 1, pp. 1-6. IEEE, 2017.

[2] Nunnari, Giuseppe, and Valeria Nunnari. "Forecasting monthly sales retail time series: a case study." In 2017 IEEE 19th conference on business informatics (CBI), vol. 1, pp. 1-6. IEEE, 2017.

[3] Pavlyshenko, Bohdan M. "Machine-learning models for sales time series forecasting." Data 4, no. 1 (2019): 15.

[4] Ensafi, Yasaman, Saman Hassanzadeh Amin, Guoqing Zhang, and Bharat Shah. "Time-series forecasting of seasonal items sales using machine learning–A comparative analysis." International Journal of Information Management Data Insights 2, no. 1 (2022): 100058.

[5] Yoo, Tae-Woong, and Il-Seok Oh. "Time series forecasting of agricultural products' sales volumes based on seasonal long short-term memory." Applied sciences 10, no. 22 (2020): 8169.

[6] Custódio, João Pedro Gil, Carlos J. Costa, and João Paulo Carvalho. "Success prediction of leads–A machine learning approach." In 2020 15th Iberian Conference on Information Systems and Technologies (CISTI), pp. 1-6. IEEE, 2020.

[7] Usmani, Mehak, Syed Hasan Adil, Kamran Raza, and Syed Saad Azhar Ali. "Stock market prediction using machine learning techniques." In 2016 3rd international conference on computer and information sciences (ICCOINS), pp. 322-327. IEEE, 2016.

[8] Ereiz, Zoran. "Predicting default loans using machine learning (OptiML)." In 2019 27th Telecommunications Forum (TELFOR), pp. 1-4. IEEE, 2019.

[9] Afolabi, Ibukun, T. Cordelia Ifunaya, Funmilayo G. Ojo, and Chinonye Moses. "A model for business success prediction using machine learning algorithms." In Journal of Physics: Conference Series, vol. 1299, no. 1, p. 012050. IOP Publishing, 2019.

[10] Ensafi, Yasaman, Saman Hassanzadeh Amin, Guoqing Zhang, and Bharat Shah. "Time-series forecasting of seasonal items sales using machine learning–A comparative analysis." International Journal of Information Management Data Insights 2, no. 1 (2022): 100058.

[11] Neha, M. (2020). Prediction Analysis of Stock Price using Machine Learning Approach. International Journal for Research in Applied Science and Engineering Technology, 8(8), 198–207,August, 2020.

[12] Kessler, Travis, and John Hunter Mack. "ECNet: Large scale machine learning projects for fuel property prediction." Journal of Open Source Software 2, no. 17 (2017): 401.

[13] Cheriyan, Sunitha, Shaniba Ibrahim, Saju Mohanan, and Susan Treesa. "Intelligent sales prediction using machine learning techniques." In 2018 International Conference on Computing, Electronics & Communications Engineering (iCCECE), pp. 53-58. IEEE, 2018.

[14] Żbikowski, Kamil, and Piotr Antosiuk. "A machine learning, bias-free approach for predicting business success using Crunchbase data." Information Processing & Management 58, no. 4 (2021): 102555.

[15] Rabeya, Tapasy, Eshita Khatun, Sheak Rashed Haider Noori, Sharmin Akter, and Israt Jahan. "Bengali Review Analysis for Predicting Popular

Cosmetic Brand Using Machine Learning Classifiers." In Inventive Communication and Computational Technologies: Proceedings of ICICCT 2022, pp. 251-260. Singapore: Springer Nature Singapore, 2022.

[16] Khatun, Mst Eshita, and Tapasy Rabeya. "A Machine Learning Approach for Sentiment Analysis of Book Reviews in Bangla Language." In 2022 6th International Conference on Trends in Electronics and Informatics (ICOEI), pp. 1178-1182. IEEE, 2022.

[17] Mehdiyev, Nijat, Joerg Evermann, and Peter Fettke. "A multi-stage deep learning approach for business process event prediction." In 2017 IEEE 19th conference on business informatics (CBI), vol. 1, pp. 119-128. IEEE, 2017.