DEMAND FORECASTING FOR RESTOCKING GOODS USING MACHINE LEARNING

BY

DILRUBA KHANOM DOLON

ID: 171-15-1503

MD. MEHRAB PATWARY

ID: 172-15-9719

MOHAMMAD JAKARIA ABEDIN

ID: 172-15-9674

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Computer Science and Engineering

Supervised By

Ms. Subhenur Latif
Assistant Professor
Department of CSE
Daffodil International University

Co-Supervised By
Mr. Masud Rabbani
Lecturer
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

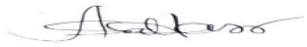
1 JUNE 2021

APPROVAL

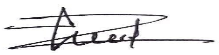
This project titled "**DEMAND FORECASTING FOR RESTOCKING GOODS USING MACHINE LEARNING**", submitted by **Dilruba Khanom Dolon**, **Md. Mehrab Patwary** and **Mohammad Jakaria Abedin** to the Department of Computer Science and Engineering, Daffodil International University, has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering (B.Sc) and approved as to its style and contents. The presentation has been held in 1st June 2021.

BOARD OF EXAMINERS

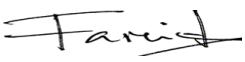
Dr. Touhid Bhuiyan — **Chairman**
Professor and Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University



Abdus Sattar — **Internal Examiner**
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University



Md. Jueal Mia — **Internal Examiner**
Senior Lecturer
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University



Dr. Dewan Md. Farid — **External Examiner**
Associate Professor
Department of Computer Science and Engineering
United International University

DECLARATION

We hereby declare that this thesis has been done by us under the supervision of **Ms. Subhenur Latif**, Assistant Professor, Department of CSE, and co-supervision of **Mr. Masud Rabbani**, Senior Lecturer, and **Department of CSE** Daffodil International University. We also declare that neither this thesis nor any part of this thesis has been submitted elsewhere for the award of any degree or diploma.

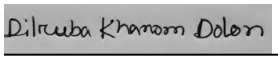
Supervised by:

Ms. Subhenur Latif
Assistant Professor
Department of CSE
Daffodil International University

Co-Supervised by:

Mr. Masud Rabbani
Lecturer
Department of CSE
Daffodil International University

Submitted by:



Dilruba Khanom Dolon
ID: 171-15-1503
Department of CSE
Daffodil International University

Md. Mehrab Patwary
ID: 172-15-9719
Department of CSE
Daffodil International University

Mohammad Jakaria Abedin
ID: 172-15-9674
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First of all, we want to render our gratitude to the Almighty Allah for the enormous blessing that makes us able to complete the final thesis successfully.

We are really grateful and express our earnest indebtedness to **Ms. Subhenur Latif,** Assistant Professor, Department of CSE Daffodil International University, Dhaka, Bangladesh. Profound Knowledge & intense interest of our supervisor in the field of "Machine Learning & Deep Learning" make our way very smooth to carry out this thesis. Her remarkable patience and dedication, scholarly guidance, continual encouragement, vigorous motivation, direct and fair supervision, constructive criticism, valuable advice, great endurance during reading many inferior drafts and correcting the work to make it unique paves the way of work very smooth and ended with a great result.

We would like to express our gratitude wholeheartedly to **Prof. Dr. Touhid Bhuiyan**, Professor, and Head, Department of CSE, for his kind help to finish our thesis and also to other faculty members and the staff of CSE department of Daffodil International University.

We would like to express thankfulness to the fellow student of Daffodil International University, who took part in this discussion during the completion of this work.

We would like to express our immense thanks to the kaggle website who provided us with the required raw data to make our work possible.

We would also like to thank the people who attend the survey done by us to collect the market real information.

Finally, we must acknowledge with due respect the constant support and passion of our parents and family members.

ABSTRACT

In the financial operations, many factors play a decisive role. Price and demand play an essential role among them, since they are the key determinants of the financial activities. Demand is not statically placed. In high-range prices, it is marked by unpredictable fluctuations. The principal determinant of market volatility is this form of fluctuation. We now have intelligent machines that can find the lessons from data in this age of artificial intelligence. Data insights can be obtained using machine learning techniques for prediction purposes. Prediction can be a successful way of eliminating market uncertainty. We try to find techniques for the machine learning in our work to help us predict the future demand for products at any business. Our work is based on the raw data from the website of Kaggle. Machine Learning has various prediction algorithms. We use gradient boosting, neural networking (MLP regression), linear regression, SVM, Decision Tree, regression random, forest regression to find the solution. In order to achieve the optimum accurately, we have compared the accuracy in terms of efficiency.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLE

CHAPTER 1
INTRODUCTION

1.1 Introduction

In Bangladesh super the number of super shop is increasing day by day. In the citizens of the western world and in Bangladesh, the habit of everyday shopping for the basics has shifted. By now, people have opted to go to the market once a week or once a month for essential goods. A large super-shop was built in the West for this purpose. Everything we can choose from the mega shop from live fish to smartphones. In Bangladesh, too, this is not impossible. In 2000, Bangladesh's first super shop was created [1]. Even in many South Asian countries, this industry is thriving. As can currently be said around 18% of Maldive customers, almost 7% of them are shopping from super shops in India, 1% in Pakistan and Nepal. 2% of people in Bangladesh shop from super stores. However, 98 percent of people still rely on conventional markets. In reality, a super shop operator has noticed, our government and municipal companies do not take enough initiatives or incentives to accelerate this region. Machines behave intelligently in this age of artificial intelligence. Einav et al. [1] define machines which are widely used for the purposes of prediction. The finance sector is the most exciting area for machine learning applications. For this purpose we use the machine learning approach (ML) to forecast onion prices for our collected demand data. We use the machine learning approach (ML) for this reason to forecast prices of onions for our collected demand information. Geron demonstrated that a number of tools such as Scikit-Learn, Tensorflow, Matplotlib, Pandas, Numpy are available for machine learning applications [2]. There are many feature selection algorithms that can be used to make a data set easy to use. Predicting demand is not a straightforward operation. We previously addressed the number of predictive factors. The forecasting demand can be subject to the abovementioned ML techniques. This is our key objective.

1.2 Motivation

Market forecasting is the mechanism by which historical sales data are used to predict the expected consumer demand forecast. Demand Forecasting provides companies with an assessment of their consumers' expected volume of products and services. The demand forecast depends on the critical assumptions of the company such as revenue, profit margins, cash flow, investment, risk assessment, mitigation planning, etc. In this phenomenon, prediction can be fruitful. Predictions for the future prices of different products will apply to the financial sector. We conducted a primary review to find the best solution. Initially, we try to figure out whether these kinds of work have already been completed but have been done with little satisfaction. However, there was no significant work in the ML subfield, prediction, in the earlier era. Furthermore, we see the scope of this kind of work to be applied in the economic field. Customers conclude that the larger products are predictively designated for them. Form this discussion we can understand that demand forecasting system is very necessary for any shopping mall because if shop keeper know the future demand of any product. He can restock them with specific amount. As a result, money will be saved and product defected rate will be reduced. Form this point we were motivated. In Bangladesh there is no intelligence system that can predict future demand. It is another factor to do our work. The proposed model consists effective structures; and we named this structure as “Factors regression Model." History demand and demand variables are its inputs. The model results are demand for clean background and potential demand factors. The final result of the model is the final demand forecast and profit-maximizing activities. Business forecast is the summary of the basic forecast and the forecasts of operations. The model can be further tuned by opinions of experienced individuals who can modify any activities and recalculate the prevision model on the basis of the experts' proposed new parameters.

Finally, we will make a system that can predict future demand of any product of different store that can help the shop keeper to make maximum profit by using minimum capital.

1.3 Problem Definition

Different applications in the economic sector can be observed at this AI age. Prediction is used commonly to predict a product's price. This form of work is commonly used to analyze companies, make choices and analyze production. This may also be of assistance to the client. This is also unknown to the Bangladeshi economy. In our work, we see that in different times of year fluctuations in the onion market are extremely high. This incentive draws our attention. If we can predict demand of any product it can also help to reduce market instability of this product. For this reason, we analysis which parameter affect the demand of any goods. We found that time is important factor of this phenomenon. As demand is time series data we at first calculate correlation of time and demand of a specific product. We found that day, month, year, is important factor. Based on this knowledge we collected dataset and apply different machine learning algorithms. We collected our required dataset form kaggle website. The dataset contains different noise data. in preprocessing stage, we removed all the noise data. in our dataset time was date time format. So we need to detach them. And after splitting we got year month and days. We analyzed 2013 to 2017 daily data. after competition of data preprocessing we implemented different machine learning regression algorithms when we apply algorithm we use different hypothetical things. One of the most important is different size of test data. we choose best algorithm which is produced best r2_score. after selection of best algorithm, we evaluated this using real data that was not seen by our algorithm. In evaluation state our selected algorithm performed very nicely.

1.4 Research Questions

- How can the dataset be collected and made?
- Could the categories be described accurately?
- How do I classify the demand?
- Can the machine learning method accurately predict demand?
- Will the model work on the web?
- How can the people be helped by this?
- How this system can make market equilibrium position?

1.5 Research Methodology

The methodology of science literally refers to the practical "how" of any research work. In particular, a study is designed by a researcher in order to systematically ensure accurate and reliable outcomes that address the research goals and objectives. In this section we will talk about the data collection, data processing, data categorization, implementation of algorithms, Model training, Algorithm assessment, Use the model in a web application.

1.6 Research Objectives

- To anatomize how to classify or categorize the demand using some regression algorithm.
- Developing a model which can predict any product's future demand.
- To imagine an overview of various commodity prices.
- Develop a web application which displays the demand with machine learning and tools for web engineering.

1.7 Research Layout

Our study material has been published in relation to:

Chapter 1 Offer the study a summary. The initial analysis is an essential stage in this first segment. Furthermore, in this chapter we explain what inspired us to do this study. The problem definition is the most important part of this chapter. This segment also includes the research topic, the challenge.

Chapter 2 This consists of a context analysis and provides a brief overview of the work in this area. Related remarkable work with machine learning is here described.

Chapter 3 Gives a simple approach or process definition. How was the research performed in this segment addressed?

Chapter 4 Is about the performance measurement. The result of the analysis is contained in the chart.

Chapter 5 It's part of the study ending. This section describes the model's efficiency. This segment also provides an accuracy comparison. This section also includes the web

implementation portion of the model and the output. The segment ends with a view of the work's limits. The future work was also encoded.

1.8 Expected Outcome

• We will anticipate any product's potential demand.

• Commodity market volatility will be reduced.

• We've suggested a strategy for balancing the demand.

• A comprehensive web-based framework has been created to demonstrate the results of market research.

CHAPTER 2

BACKGROUND STUDY

2.1 Introduction

Machine learning for prediction has been achieved with many experiments. Prediction is one of Machine Learning's most widely used applications. There have been huge numbers of studies to forecast a future price of products so that the associated uncertainty can be resolved. These studies concentrated on issues and used various machine learning algorithms to resolve the problem. This chapter provides an overview of the work pertinent to the previously discussed subject by some experts.

2.2 Related Works

Prediction is one of the most frequently used machine learning frameworks. Price classification is an internationally well-examined area. We also covered a variety of similar works previously done in this section.

Gabralla et. al. [3]. experimented with machine learning models to predict the price of crude oil. This paper shows the dependence of the desired output on input material. Three algorithms were used for functions. The data set was randomly divided into training and testing data before forecasting and consisted of four classes. K star and SMO regression models were used for the prediction of IBL. Finally, for better prediction, the authors create an ensemble model. Data machine learning methods work successfully in numerical time series. It is therefore a strong stand for us to forecast prices of onions by means of machine learning.

Anandhi et al [4]. was responsible for the search and supply in Tamil Nadu of raw materials for pulpwood in India using the SVM algorithm. Pulpwood specifications and supplies were determined by collecting and analyzing a pulpwood dataset for fifteen years. The author found market supplies for the coming year after reviewing the data collection.

Supply is the pulpwood scarcity. In addition to SVM, more algorithms are included and the accuracy rate comparable.

In this regard the POS figures for customers in the restaurant in Japan have been utilized by Tanizaki et. al. [5]. very similar work. Different methods such as Bayesian, Boosted, Decision and Stepwise have been used for this work. The rate of data use was 40% to 100%. The highest accuracy of 91.7 percent is achieved by implementing Bayesian algorithms in the dataset where 100 percent data is used. In comparison, the neural network algorithm showed 98.17 percent accuracy, where the data use a rate of 30%.

Sinta et. al. [6]. have carried out another equivalent predictive work using the K-Nearest Neighbors Method ensemble in Indonesia using 14-year monthly price rice data. The authors of this paper addressed comparing K-Nearest Neighbors and K-Nearest Neighbor's ensemble and demonstrated that KNN is less effective because of time series data than the KNN ensemble. The authors have checked precision by using training data variations. After the forecast, prediction results were compared with the actual outcome and quite closely compared with the actual price. In comparison, the daily rice price figures on a dataset of about 730 observations have used several algorithms.

Huang et. al [7]. used SVM to forecast the stock market. For stock market forecasting, the authors observed 676 pairs of data. The combined model achieved better results by applying different classification methodologies of 75%. Finally, a model has been proposed that SVM has integrated with other grading methods. We have improved precision in comparison to this work and have checked our accuracy with other master learning algorithms.

The relationship between prices and demand changes was established with Rafieisakhaei et. a. [8]. By creating a model with causal circuit and mathematical formulas, which mainly comprises oil price, oil demand and oil supply, the Authors pursued a digital solution. They suggested various forms of loops with the main component being global oil demand, oil prices and oil supply. Finding the demand/supply ratio of oil price by regression analysis on historical data prices. In comparison, our solution has been drawn by means of machine learning.

The rice production of Gandhi et al. was forecast with the aid of an SVM approach, similar work has been done on agriculture [9]. This paper mainly aimed at coping with the growing population in India with rice production. For data mining, SVM, SMO algorithm and uncertainty matrix methodologies used. The author used data mining to compile data by omitting unnecessary data before implementing algorithms. Subsequently, SVM was used to perform necessary functions for the production of the labeled training information. Then the result with the SMO algorithm was measured. Finally, it was measured using the precise confusion matrix of the model proposed.

The fluctuations in rice price in Bangladesh are defined by M. M. Hasan et. al.[10]. They attempted to forecast the future price of rice to reduce the rate of fluctuation. For this reason, standard machine classification algorithms such as KNN, Naïve Bayes, Decision Tree, SVM and Random Forest are used. Their higher accuracy rate by the decision tree algorithm was 98.17 percent. They divide rice prices in their work into three categories: High, medium and poor. They tried to forecast future rice prices by using the Decision Tree algorithm.

M. M. Hasan et al [11]. demonstrated the solution to market uncertainty by forecasting prices for onions. They used a machine learning algorithm for this purpose to predict the future price. Data of 2 years is used daily. Collection of information The implementation and assessment of the Data Analysis algorithm is its four key steps to complete entire work. They have been using the algorithms KNN, Naïve Bayes, Decision Tree. The highest level of precise accuracy is 99 percent thanks to Neural Network Algorithms.

2.3 Comparison of related work

Table 2.1 Comparison Table of related work

Related work	Percentage
Prediction of oil prices with ensemble machine learning,	76.94%
Encourage vector regression to predict pulpwood demand and supply.	85%
Restaurant demand forecasting using machine learning and analysis.	89%
Ensemble k-nearest approach to predict Indonesia's rice price.	90%
Forecast the course of bond market activity with a vector holder.	83%
Offer-and-demand analysis to forecast developments on the petroleum markets: A 2015 pricing case study.	75%
Prediction of rice crop yield in Indian by using vector support machines.	89.14%
Ascertaining the Fluctuation of Rice Price in Bangladesh Using Machine Learning Approach	98.17%
Solving Onion Market Instability by Forecasting Onion Price Using Machine Learning Approach	99%

From the discussion above, we found that price prediction work in Bangladesh was not noteworthy. Comparing the work in connection, we can see that our model has a larger data set and performed in certain categories more satisfactorily than others. In web-based applications, we can implement our knowledge.

2.4 Research Summary

The above study is carried out on a wide variety of research projects from a number of research teams and demonstrates what research was conducted in the field of market analysis. By analyzing, we have effective outcomes. While not enough resources are present, it is hoped that each sector can become more resourceful by adding buying information about various goods, after an individual day.

2.5 Challenges

The main challenges of our work was data collection. There is no source in data for demand in Bangladesh. We contacted different super shop but most of them they did not categories demand based list of their product. Another problem is during corona pandemic we did not visit more store. As a result, we collect data from kaggle website. And we tried to make a perfect model that can predict demand based on user dataset. So when we will implement our work we added special feature to change dataset of registered user. As a result, user can see future demand of product based on this dataset.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Introduction

Our working methodology includes a whole Six processing, Pre-processing, Dataset, Implementation Algorithm, Assessment steps. The chart of our work is presented in Figure 3.1

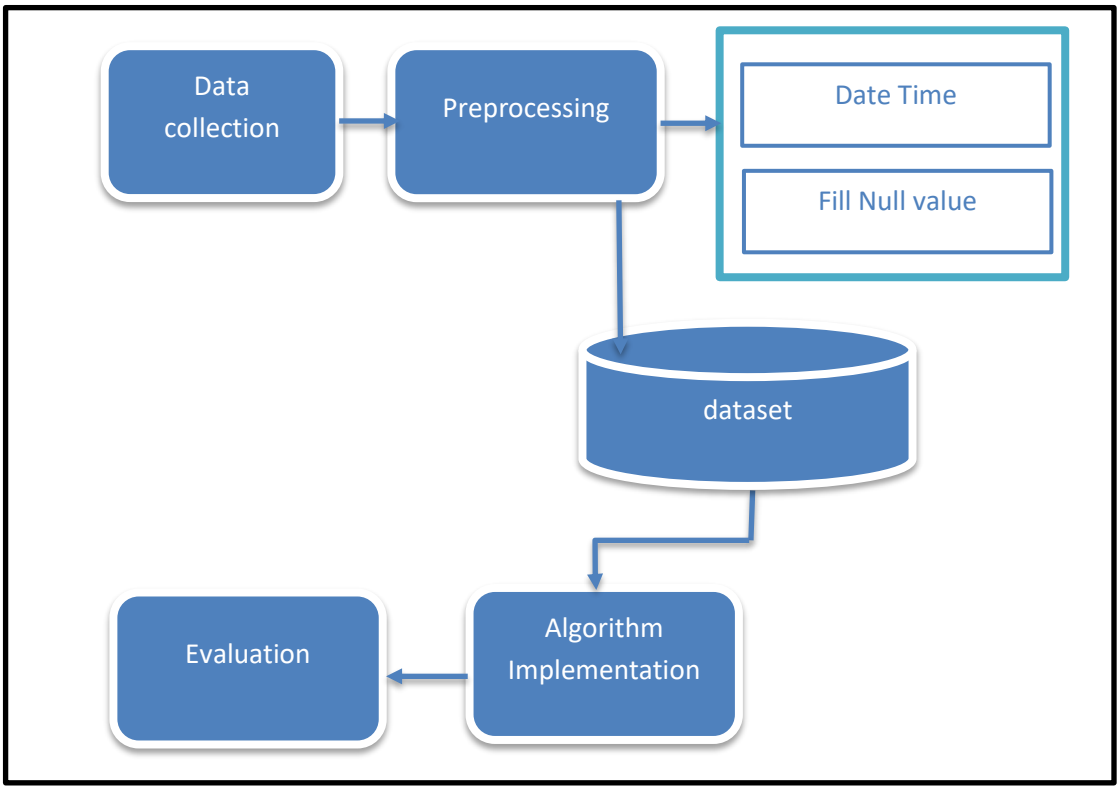


Figure 3.1: Methodology diagram

3.2 Data Collection

Day by day POS data is growing. This form of data is now used to analyze the industry. This study aims to make choices for the decision-maker. Due to the pandemic condition, we have accrued the necessary data from the Kaggle website. Day after day, we obtained

15,000 samples of 50 products. For more study, the data was not prepared, so the next move was taken to ready the data.

3.2.1 Date time:

Our initial collected dataset contains date as a date time format (2013/11/3). At first we need to separate year, month and day, because we worked with daily data. For this purpose, we used panda's library.

3.2.2 Fill null value:

Our initially collected dataset contains null sometimes. As a sensitive data we full this null value by 0. Because Mean or other method is not best for this purpose. And for this operation we also used panda's library.

3.3 Dataset

Our dataset contains total six features. Year, Month, Day, Store, Items, and sales. Sales is our dependent variables. That is depends on year month day and store. The sample of our dataset is given bellow: Our dataset contains about 15000 row. Table3.1 represents the sample of our dataset. Here sales represent the demand of any items that depends of year, month, day, and store. When we analyzed our dataset we noticed that there is lots of hidden information. All of this analysis is given in statistical analysis part.

Table 3.1 Dataset sample

year	month	day	store	item	sales
2014	9	29	4	11	41
2014	11	2	1	16	22
2014	3	16	2	40	42
2013	3	1	10	21	30
2016	9	29	8	36	121
2016	6	3	9	36	97
2016	9	11	10	12	96

3.5 Features

The Features is an observable property or feature of a phenomenon found in machine learning and pattern recognition. Choosing descriptive, discrimination and independent features is a vital step in the identification, grouping, and regression of efficient algorithms. After analyzing the data, we make the features of the work.

Table 3.2 Description of features

SL No.	Name of Attributes	Description
1	Year	We selected at the year of 2013 to 2017 data.
2	Month	Price is also depending on month. As a result, we selected month as our attribute.
3	Day	We work with daily data. That means daily demand. So we selected daily demand of a month as an attribute.
4	Store	After collecting data, we noticed that demand depends on Store. The demand of same product depends on Store number. So we selected 10 store data as our attribute.
5	item	Each store contains different item and different item's demand is different. So we selected 50 item demand of different 10 store.
6	Sales	Sales is our dependent variable. This variable is depending on all the attribute like Year, Month, Day, Store, Item. This is our predicted variable.

3.7 Algorithm Implementation

In this section we described the process of algorithm implementation. For doing this process we have to complete the previous process to make the required dataset. When dataset is prepared we made the price into three categories named high, mid, low. Then we proceed to algorithm implementation. As our work is to predict the onion price we try to find the best techniques to find the future price. We found classification algorithms best for our work. We applied five different ML algorithms name Gradient Boosting, Random

Forest, Decision Tree, SVM and Neural Networks gradually and measure the result and compared with each other to find the optimal algorithm to predict the price more efficiently. The following table 3.2 shows the result with accuracy rate. Among them NN got the highest accuracy.

Table 3.3 Parameter usages

Algorithms	Details
Gradient Boosting	random_state = 42, n_estimators=200,
Decision Tree	random_state= 0
Random Forest	Kernel = linear
Neural Network	random_state=1, max_iter=500, activaton = ‘relu’
Linear Regression	random_state= 0
Lasso Regression	random_state= 0, alpha=1.0

Table 3.3 illustrates the parameters and the various things that we applied for implementing the chosen algorithms.

3.8 Evaluation

We then continue with the execution of an algorithm when dataset is prepared. As we are working on predicting the market, we are trying to identify the right strategies to meet potential demand. For our job, we find the best regression algorithms. In order to help forecast the market, we progressively implemented six different ML algorithms, Gradient Boosting, Random Forest, Decision Tree, SVM and Neural Networks, and calculated the effects and compared them.

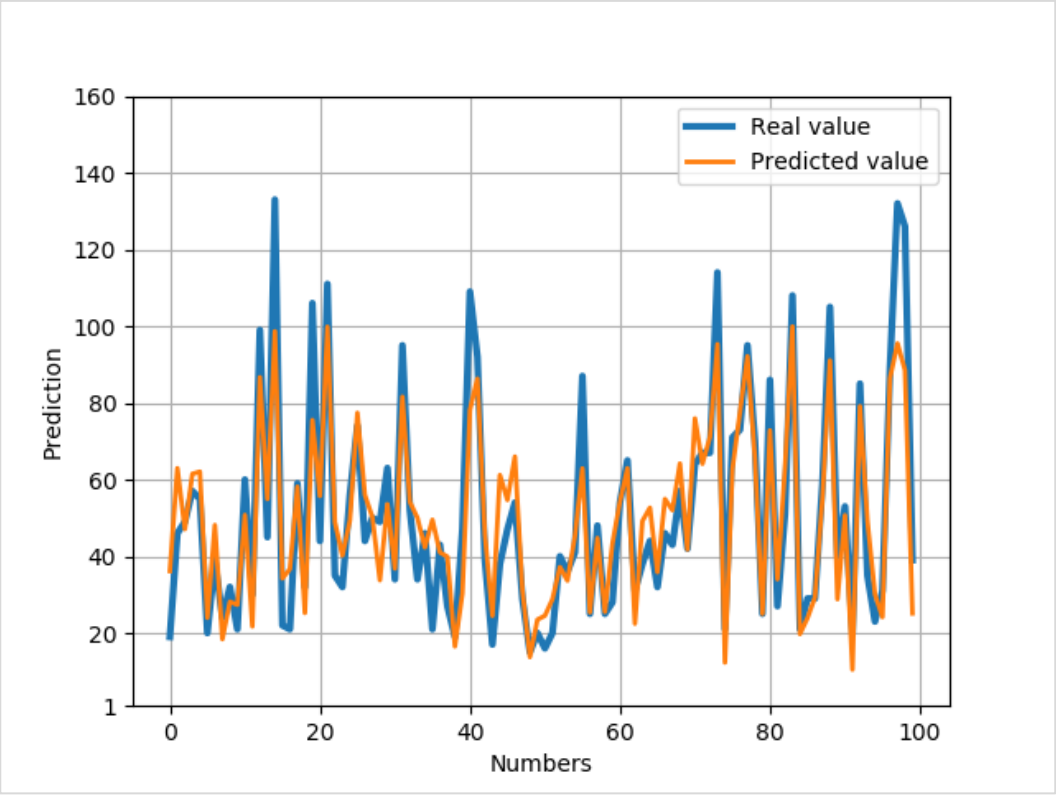


Figure 3.2: Real and predicted comparison

Figure 3.1 represents the real and predicted demand of randomly chosen 50 items. We validate about 100 unknown real data to see how much accurate our model generate prediction. In figure 3.1 blue color represent the real demand and orange color represents the predicted demand. From this graph we can see that our model predicts so accurate that it overlaps the original line most of the time.

So we can say that our model produced very good performance with real data.

CHAPTER 4

RESULT ANALYSIS

4.1 Introduction:

This chapter 4 focuses primarily on descriptive interpretation of the test data and study outcomes. When we analyze the issue, what is the results analysis? What is it? The portion of consequences should be planned to narrate findings without understanding or analysis, and recommendations for the discussion paper. The findings will be reported and the analysis revealed. This part of the study discusses whether the outcomes have been accomplished.

4.2 Experimental Result

For finding best fit dataset we applied a hypothetical factor, we used 30% to 70% test data usage rate. By this process we tried to find out which percentage is best for our model. All of our analysis is given bellow:

Table 4.1 For 30% Data Usage

Parameter	Algorithms					
	NN	*DT*	*RF*	*GB*	*Linear*	*Lasso*
MAE	23.29	13.29	9.75	9.50	22.86	22.87
MSE	817.37	325.40	167.70	157.45	819.80	920.70
RMSE	8.82	18.03	12.94	12.54	82.63	28.64
R2_Score	0.02	0.61	0.80	0.80	0.04	0.04

Here, for 30% data usage rate for 6 algorithms, we can see the best value for R2 score is 0.80 and produced by Random Forest and Gradient Bosting. But for MSE (Mean Squared Error) they produced different score. 157.45 is the less MSE score that is produced by GB algorithm. So, the Gradient Boosting Regression is the best algorithm for 30% data usage rate.

Table 4.2. For 40% Data Usage

Parameter	Algorithms					
	NN	*DT*	*RF*	*GB*	*Linear*	*Lasso*
MAE	22.80	13.33	9.80	9.83	22.81	22.83
MSE	830.60	328.586	168.89	175.75	815.68	816.77
RMSE	28.82	18.12	12.99	13.25	28.56	28.57
R2_Score	0.02	0.61	0.80	0.79	0.04	0.04

4.1.2 Table for a data usage rate of 40%, we can see that Random Forest Regression gives the best MSE (mean squared error) of the R2 value 168,89, and Random Forest of the R2. The Random Forest Regression is thus the best algorithm for a 40 percent data utilization rate.

Table 4.3. For 50% Data Usage

Parameter	Algorithms					
	NN	*DT*	*RF*	*GB*	*Linear*	*Lasso*
MAE	22.95	13.18	9.96	9.84	22.58	22.60
MSE	805.85	312.29	174.16	173.49	797.35	798.27
RMSE	28.38	17.67	13.19	13.16	28.23	28.35
R2_Score	0.03	0.62	0.79	0.79	0.04	0.04

The rate of data use of 6 algorithms is here 50 percent. The R2 score is 0.79 and is generated by random forest and bosting gradient. However, they provided separate score for MSE (Mean Squared Error). The lower MSE score produced by the GB algorithm is 173.49. The Gradient Boosting is therefore the best algorithm for 50% data use.

Table 4.4. For 60% Data Usage

Parameter	Algorithms					
	NN	*DT*	*RF*	*GB*	*Linear*	*Lasso*
MAE	22.72	14.00	10.21	9.81	22.54	22.56
MSE	828.47	357.48	181.48	172.07	792.79	794.04
RMSE	28.78	18.90	13.47	13.11	28.15	28.17
R2_Score	0.002	0.56	0.78	0.79	0.04	0.04

The rate of data use of 6 algorithms is here 60 percent. The R2 score is 0.79 and is generated Gradient Boosting. The lower MSE score produced by the GB algorithm is 172.07. The Gradient Boosting is therefore the best algorithm for 60% data use. In this part Random forest also produced very good performance. The r2 score is 0.78 that is very near to GB

Table 4.5. For 70% Data Usage

Parameter	Algorithms					
	NN	*DT*	*RF*	*GB*	*Linear*	*Lasso*
MAE	22.42	14.65	10.55	9.80	22.54	22.56
MSE	802.63	397.08	194.15	171.06	790.57	791.81
RMSE	28.33	19.92	13.93	13.07	28.11	28.13
R2_Score	0.03	0.52	0.76	0.79	0.04	0.04

In this case, the data consumption rate of 6 algorithms is 70%. The score of R2 is 0.79 and the gradient increases. The GB algorithm's lowest MSE score is 171.06. Therefore, gradient boosting is the perfect algorithm for the use of 70%. Random forest has also done very well in this section. The ranking for r2 is 0.76.

There are six different types of regression algorithms. The neural network, the decision tree, the Random Forest, the Gradient Boosting, the Linear and Lasso. Almost every algorithm has a good forecast, but Gradient Boosting's values are the best for both MSE and R2 scores. We therefore choose to use the algorithm of gradient boosting for prediction.

4.3 Statistical Analysis

In this part we analyzed different important information of our dataset. we tried to illustrated the correlation between different attribute.

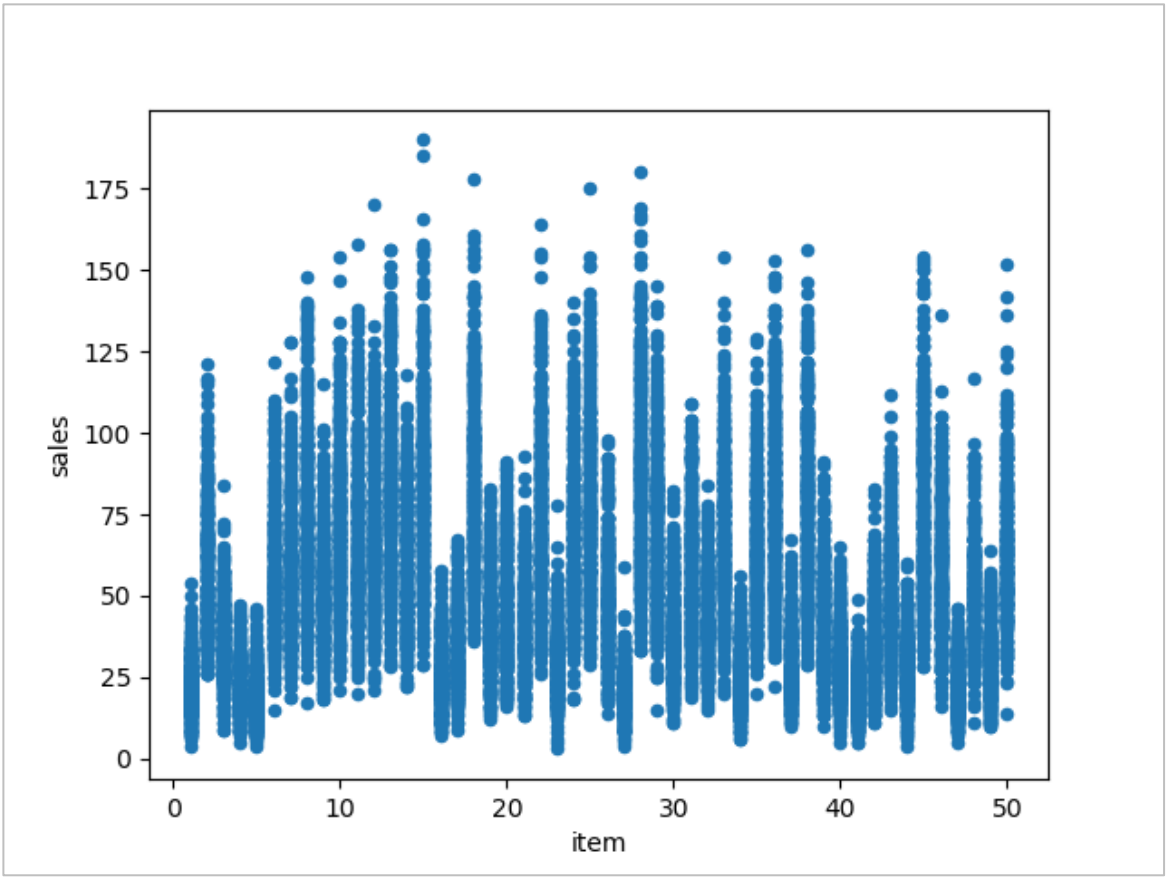


Figure. 4.1 Item vs Sales

Figure 4.1 represents the item vs sales. We worked with 50 types of items. From whole dataset we tried to find out most saleable items. This analysis helps a seller to restock more

item which sells rate is high. For example, in this graph 15 number item's sales rate is high. That means the demand of this product is high.

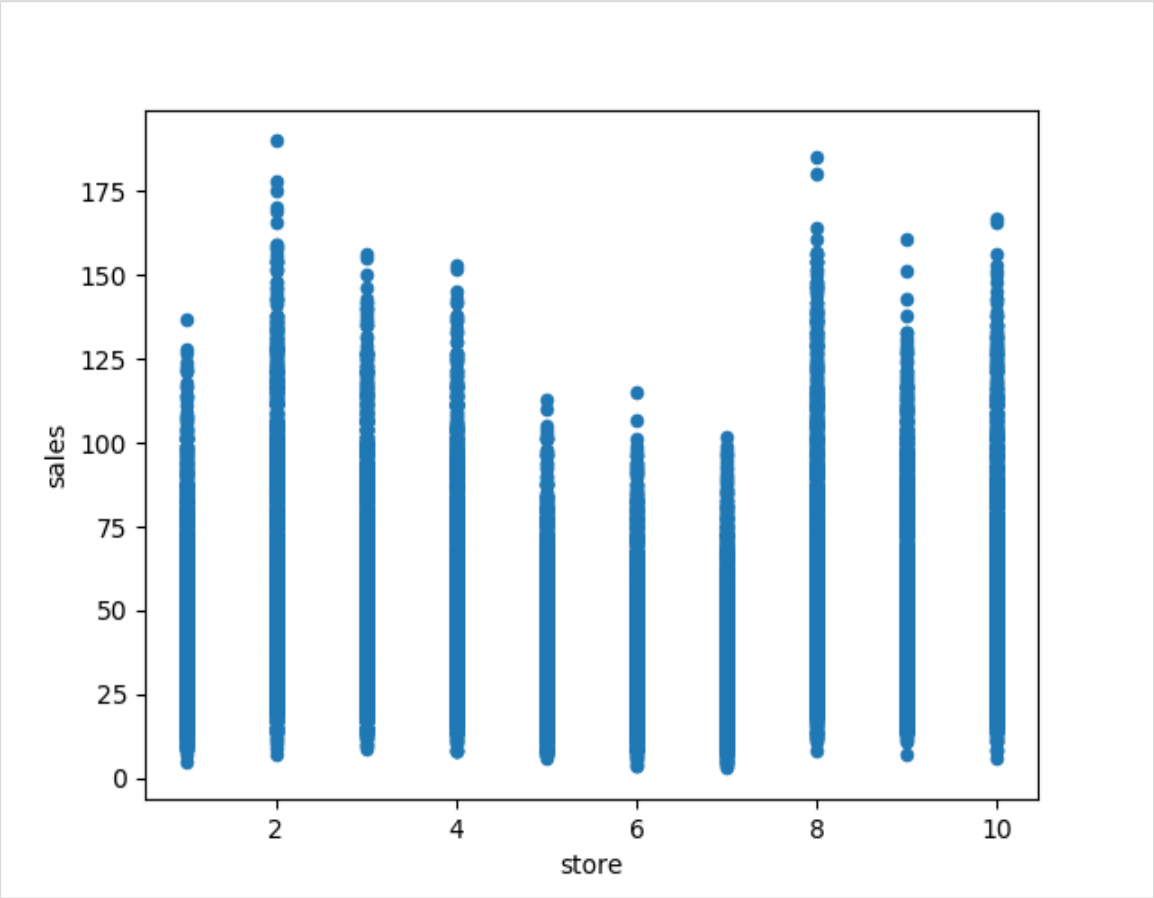


Figure. 4.2 Store vs Sells graph

Figure 4.2 represents the store vs sells. Form this analysis we tried to find out most sellable store. We work with 10 different store, And the common items of this store. From this graph we can see that the highest sellable store are number 2 and number 8. And lowest sellable store is number 7. Number 5, 6,7 store's sells is comparatively lower than other store. This graph helps the shop keeper to find out lagging of his service. And by following other store service quality they can improve the sales rate.

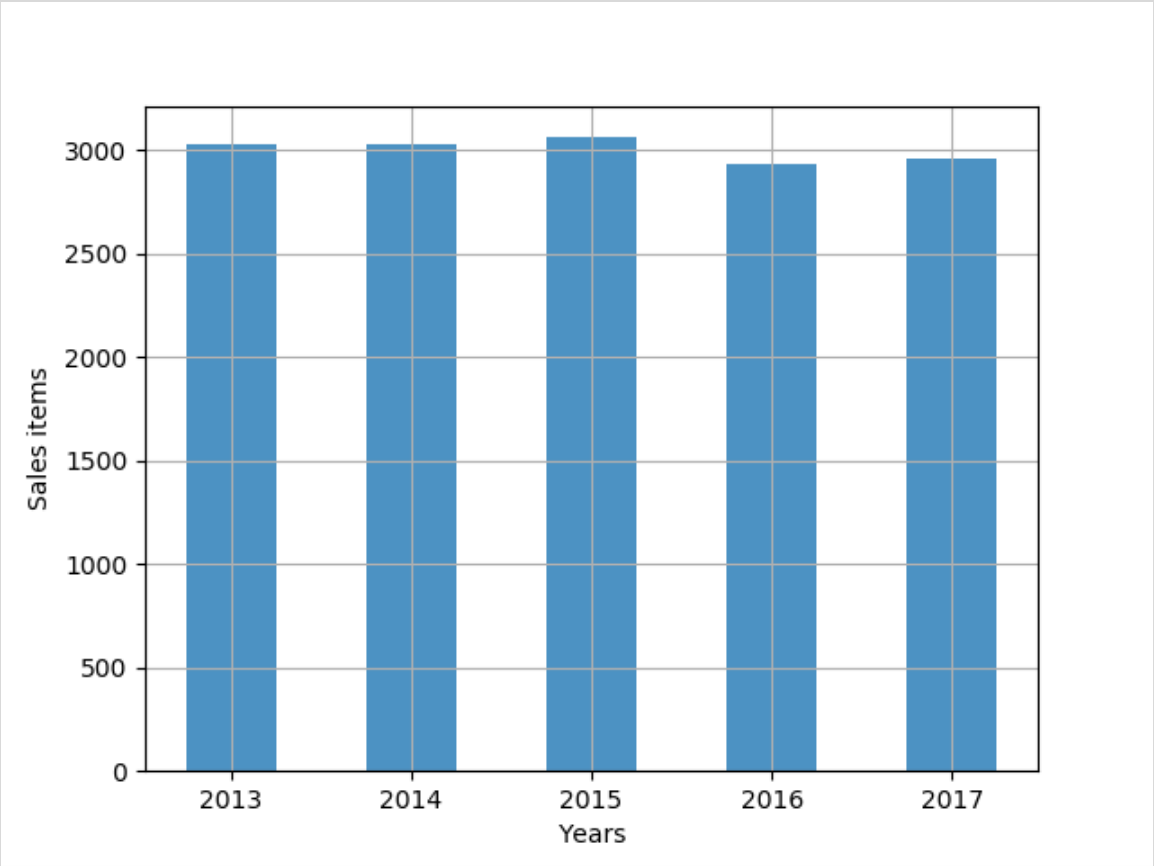


Figure. 4.3 Years vs Sales items

Figure 4.3 represents the relation between years and sales items. This analysis is done by combining all 10 selected stores. Form this graph we can see that the highest sales is happened in the year of 2015. The less sell rate is happened in the year of 2016. The year 2013 to 2015 sales rate is comparatively higher than the year 2016 and 2017. Form this graph we can say that when time is increasing the sales rate is decreasing of this 10 store.

CHAPTER 5

SUMMARY, CONCLUSION AND FUTURE WORK

5.1 Summary of the Study

Much research has been performed in the field of machine learning, but the amount of such research is very low in Bangladesh. Though work with the predictive style is a popular term for machine learning, Bangladeshi products are still unknown. This style of analysis is being used lately as such a work results produce a drastic change in our machine life. We have some fascinating real-life applications to learn from this kind of analysis. But little research in the field of Bangladesh's economy is being done. However, we hope that many scholars in this field have researched from various countries.

5.2 Conclusion

Every algorithm functions closely in our work. We researched the performance of six regression Machine learning algorithms and found the best algorithm for a better prediction of 50 different products.

In this work we used dataset from kaggle website. Our dataset was un organized. So at first we need to organized them. For this purpose, us used different data modification library. After preprocessing we apply our model. And find best model which is produced best r2 score. Finally, we have a model from which we can determine the future demand of 50 items. This project helps the shop keeper to reduce their lost. Because customer can see the future demand of any product as a result they can restock product which future demand is high.

5.3 Recommendations

There are a few remarkable suggestions for this:

- To improve data collection reliability, to achieve better results from the inquiry.
- Better results would also yield a better result in the processing of data.

● Make a more complicated algorithm like LSTM more accurate on our dataset.

5.4 Future Work

The following are the potential guidelines for the progress of this work:

- ❑ In our data collection in the future, we shall apply more sophisticated algorithms
- ❑ By adding other Bangladeshi stores, we will increase our data.
- ❑ We would also try to balance linear programming in the supply chain production/demand supply chain with the ML approach.
- ❑ Our work will forecast the optimum level of demand of all financial commodities based on the related results.
- ❑ We are aiming to create a product that demonstrates and informs customers of the future market for the product in a dynamical manner.

REFERENCE

[1] Super Shops Super Growth, available at << http://businessoutlookbd.com />> last access on 2 June 2020 at 11:10 am.

[2] A. Géron, Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. O'Reilly Media, 2019.

[3] L. A. Gabralla, R. Jammazi, and A. Abraham, "Oil price prediction using ensemble machine learning," in 2013 International Conference on Computing, Electrical and Electronic Engineering (ICCEEE), 2013, pp. 674-679: IEEE.

[4] V. Anandhi, R. M. J. I. j. o. f. c. Chezian, and communication, "Support vector regression to forecast the demand and supply of pulpwood," vol. 2, no. 3, p. 266, 2013.

[5] T. Tanizaki, T. Hoshino, T. Shimmura, and T. J. P. C. Takenaka, "Demand forecasting in restaurants using machine learning and statistical analysis," vol. 79, pp. 679-683, 2019.

[6] D. Sinta, H. Wijayanto, and B. J. A. M. S. Sartono, "Ensemble k-nearest neighbors method to predict rice price in Indonesia," vol. 8, no. 160, pp. 7993-8005, 2014.

[7] W. Huang, Y. Nakamori, S.-Y. J. C. Wang, and o. research, "Forecasting stock market movement direction with support vector machine," vol. 32, no. 10, pp. 2513-2522, 2005.

[8] M. Rafieisakhaei, B. Barazandeh, and M. Tarrahi, "Analysis of supply and demand dynamics to predict oil market trends: A case study of 2015 price data," in SPE/IAEE Hydrocarbon Economics and Evaluation Symposium, 2016: Society of Petroleum Engineers.

[9] N. Gandhi, L. J. Armstrong, O. Petkar, and A. K. Tripathy, "Rice crop yield prediction in India using support vector machines," in 2016 13th International Joint Conference on Computer Science and Software Engineering (JCSSE), 2016, pp. 1-5: IEEE

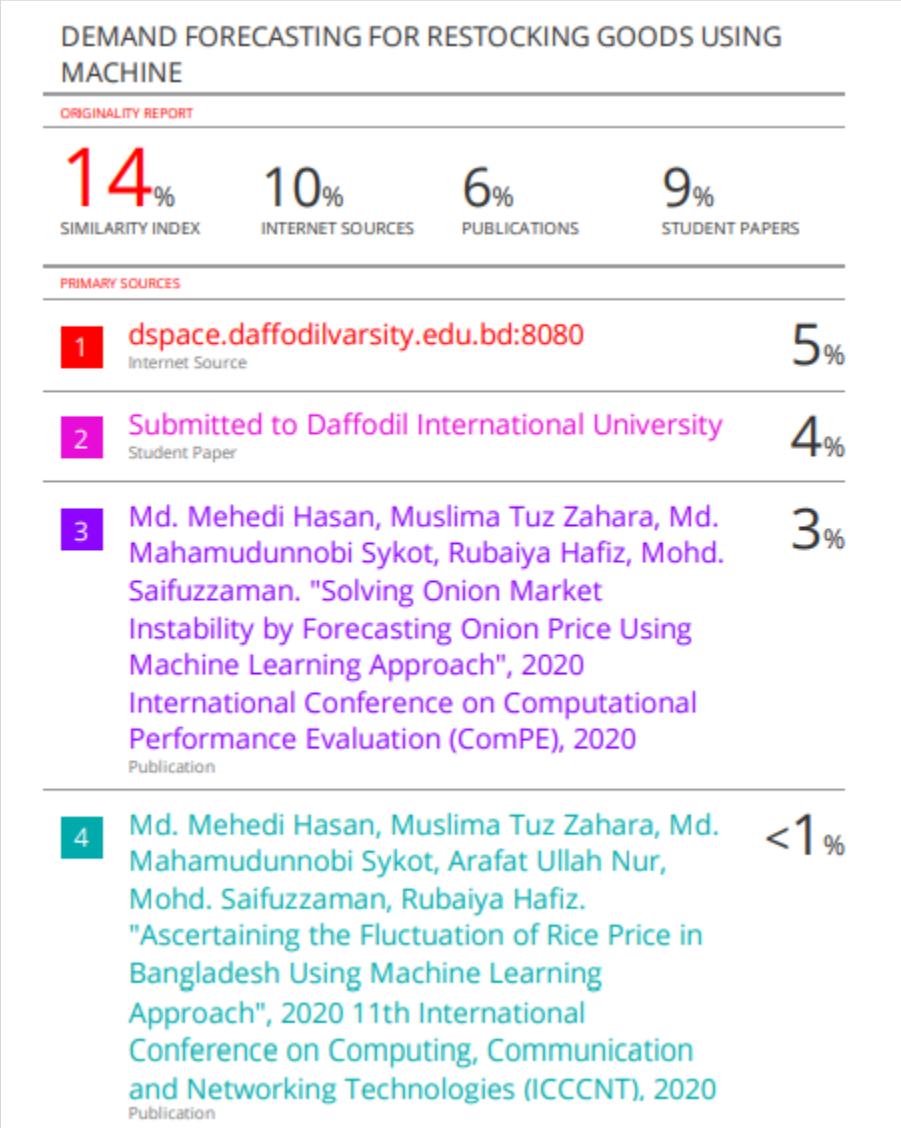
[10] M. M. Hasan, M. T. Zahara, M. M. Sykot, A. U. Nur, M. Saifuzzaman and R. Hafiz, "Ascertaining the Fluctuation of Rice Price in Bangladesh Using Machine Learning Approach," 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kharagpur, India, 2020, pp. 1-5, doi: 10.1109/ICCCNT49239.2020.9225468.

[11] M. M. Hasan, M. Tuz Zahara, M. M. Sykot, R. Hafiz and M. Saifuzzaman, "Solving Onion Market Instability by Forecasting Onion Price Using Machine Learning Approach," 2020 International Conference on Computational Performance Evaluation (ComPE), Shillong, India, 2020, pp. 777-780, doi: 10.1109/ComPE49325.2020.9200033.

APPENDIX

First, the methodology for our analysis to undertake the analysis we face so many challenges was described. In addition, in this area, little work has already been completed. This wasn't typical work. We couldn't get that much help from anywhere. Another problem was the gathering of data, an immense task for us. Because at first it was very difficult to find out dataset of our local super shop. We tried to collect but due to corona pandemic it was very difficult for us. As a result, we collected our dataset from kaggle website. After collecting dataset set we preprocessed it and then we applied algorithms. In near future we tried to build an intelligence system that can also predict our Bangladeshi products demand.

PLAGIARISM REPORT



DEMAND FORECASTING FOR RESTOCKING GOODS USING MACHINE

ORIGINALITY REPORT

14%	10%	6%	9%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080 Internet Source	5%
2	Submitted to Daffodil International University Student Paper	4%
3	Md. Mehedi Hasan, Muslima Tuz Zahara, Md. Mahamudunnobi Sykot, Rubaiya Hafiz, Mohd. Saifuzzaman. "Solving Onion Market Instability by Forecasting Onion Price Using Machine Learning Approach", 2020 International Conference on Computational Performance Evaluation (ComPE), 2020 Publication	3%
4	Md. Mehedi Hasan, Muslima Tuz Zahara, Md. Mahamudunnobi Sykot, Arafat Ullah Nur, Mohd. Saifuzzaman, Rubaiya Hafiz. "Ascertaining the Fluctuation of Rice Price in Bangladesh Using Machine Learning Approach", 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2020 Publication	<1%