Data Science and Machine Learning: A Survey on the Future Revenue Predictions and the Amount of Product Sales

LWANGA Derrick Master's, Supervisor by Dr.Musoni Wilson Ph.D.
[1] Information Communication and Technology (ICT), University of Kigali, Kigali, Rwanda.
Subject Area: Machine Learning.

Abstract:- Human beings have always been fascinated by the future. Humans have been inspired to innovate by their desire to explore the future and learn about the unknown. Revenue are estimation normally depends on the sales of existing products. The sales forecast is one of the vital objectives in the business plan for any company. Sales forecasting is the process of determining the future revenue by using the prediction of the amount products sales. Success of the business usually depends on the amount of sales. Sales forecasting or revenue estimation is very important in the way that helps the company to determine the vital products or services which in return helps to reduce the costs of investing in non-profitable products. We needed to have a lot of information to develop sales predictions for a Bangalore Mart 250 supermarket each product previous sales records as a result, we acquired two year sales data from Bangalore Mart. The results of this research study was achieved through use of Machine Learning Models which include Linear Regression Random, Forest Regressor Lasso, Regressor Gradient, Boosting Regressor, Decision Tree Regressor and Ridge Regressor. **Keywords: Machine learning, Mean Absolute Error, Mean Squared Error, Root Squared Error, Python, One Hot Encoding, Ridge Regressor, Lasso Regressor, Random Forest Regressor, Gradient boosting Regressor, Decision Tree Regressor. Sales Forecasting**

I. INTRODUCTION

The rise of new technologies has positively affected our everyday lives especially in terms of business where a lot of business solutions have been provided by information technology. Now it is almost impossible to run a business without any information technology solutions. Even our day today lives are dependent Information Technology. One of Rwanda's Vision pillars is private-sector led development, and for this to happen successfully it is important to integrate. ICT into the private sector and help small to medium business owners run their businesses moreefficiently and this is where the sales forecasting system could come in handy, to help people who might not have the accounting or business background to understand the importance of predicting future expenses and sales to help better stock products, better plan the direction of thebusiness in terms of advertising as well, as well as general planning for the financial year. If supermarket owners know which products will likely sell more during the course of a month or the year than they can plan marketing campaigns and materials efficiently? There are countless advantages to the integration of this sales forecasting system that could possibly be used in otherbusiness outside of just local Supermarkets

II. METHODOLOGY

➢ *Data Analysis*

Analysis of data is a process of inspecting, cleaning, transforming, and modeling data with the goal of discovering useful information, suggesting conclusions, and supporting decision making. Data analysis has multiple facets and approaches, encompassing diverse techniques under a variety of names, in different business, science, and social science domains. Data mining is a particular data analysis technique that focuses on modeling and knowledge discovery for predictive rather than purely descriptive purposes.

Sales Forecasting is considered Supervised Machine Learning. Supervised learning is where youhave input variables (x) and an output variable (Y) and you use an algorithm to learn the mapping function from the input to the output. Y = f(X), The goal is to approximate the mappingfunction so well that when you have new input data (x) that you can predict the output variables (Y) for that data. It is called supervised learning because the process of an algorithm learning from the training dataset can be thought of as a teacher supervising the learning process. We know the correct answers, the algorithm iteratively makes predictions on the training data and iscorrected by the teacher. Learning stops when the algorithm achieves an acceptable level of performance

➢ *Cleaning the Data*

Cleaning data should be the first step in any Data Analysis activity. Without clean data it'll be much harder time seeing the actual important parts in the exploration. Once the training of ML models begins, they'll be unnecessarily more challenging to train. The main point is that to get the most out of a dataset, it should be clean.

ISSN No:-2456-2165

In the context of data science and machine learning, data cleaning means filtering and modifying the data such that it is easier to explore, understand, and model. *Filtering out* the parts you don't want or need so that you don't need to look at or process them. *Modifying* the parts, you *do need* but aren't in the format you need them to be in so that you can properly use them.

The dataset used needed the following changes to be considered clean:

1. Dropping of Rows with Null Values including NaN and null rows

2. Removal of all negative numbers and replacing them with absolute values

3. One Hot Encoding

➢ *Null Values*

Most data science algorithms do not tolerate nulls (missing values). So, one must do something toeliminate them, before or while analyzing a data set. There are many techniques for handling nulls. Which techniques are appropriate for a given variable can depend strongly on the algorithms you intend to use, as well as statistical patterns in the raw data, in particular, the missing values, and the randomness of the locations of the missing values. Moreover, different techniques may be appropriate for different variables, in a given data set. Sometimes it is useful to apply several techniques to a single variable. Finally, note that corrupt values are generally treated as nulls. The figure below shows the rowsin the dataset used in this project and the number of missing values in each row.

```
Train:

Invoice_ID             0
Branch                 0
Customer_Type          0
Gender                 0
Product_Type           0
Unit_Price             0
Quantity_Sold          0
Total sales in RWF     0
Sales_Date             0
Time                   0
Payment_Mode           0
dtype: int64

 Invoice_ID             0.0
Branch                 0.0
Customer_Type          0.0
Gender                 0.0
Product_Type           0.0
Unit_Price             0.0
Quantity_Sold          0.0
Total sales in RWF     0.0
Sales_Date             0.0
Time                   0.0
Payment_Mode           0.0
dtype: float64
```

Fig 1 : Null Rows dataset

It is clear that The value 0 (all bits at zero) is a typical value used in memory to denote null. It means thatthere is no missing data which can impact the implementation of the algorithm. So the best option is to remove all the rows with Null Values as this is a large dataset and removing these rows will not have too much of an impact on the algorithm implemented.

➢ *Removal of Negative Numbers*

After attempting to plot a few graphs, it was clear that within the data set there were negative values in the Unit Price and Quantity column which makes no sense because a customer cannot buy a negative amount of Stock in terms of quantity or pay for a product with a negative Unit Price. So the natural assumption is that this must have been errors which is expected when handling real world data.

The best option would be to assume that these are indeed the correct Quantity and Unit Pricesbut incorrectly entered as negative values, therefore it is necessary to return the absolute numbersin each row to get rid of all the negative values.

➢ *Breaking down Sales_Date*

It is important to make sure all columns are in the correct format to be able to be processed by the chosen algorithm. So the first step was to transform Sales_Date into the right date format.

Sales_Date includes the day, month, year, and time the purchase was made by a customer. To properly see how each of these factors affect sales, it is necessary to turn each of these factors into individual columns. Below is a picture of the final result after the mentioned operations.

➢ *Sample of dataset after changes*

	Product_ID	Mass	Product_Class	Product_Type	Unit_Price	Quantity sold	Branch_Id	Sales_Date	Branch_Size	Branch_Location	Branch_Type	Total sales in RWF
0	KLMS17	12.090	Less fat	Cheese	3922.00444	11.0	S009	1/3/2019	Level2	Gasabo	Simba_Branch1	43142.04884
1	CHKS22	7.696	Normal	Sodas	757.82644	12.0	S008	1/3/2019	Level2	Nyarugenge	Simba_Branch2	9093.91728
2	BSLS12	22.750	Less fat	Meat	2223.40260	13.0	S009	1/3/2019	Level2	Gasabo	Simba_Branch1	28904.23380
3	DHDS18	24.960	Normal	Fruits line	2858.89150	14.0	S001	1/3/2019	Level2	Nyarugenge	Simba_Branch4	40024.48100
4	YXLS01	11.609	Less fat	kitchen items	845.62398	15.0	S003	1/3/2019	Level1	Nyarugenge	Simba_Branch1	12684.35970

Fig 2 Sample of dataset after changes

➢ *Data Visualization*

In this part, we use Google colab software to show the time series plots of products for more comprehension. The horizontal axis indicates months and the vertical one presents the amount of sales. As the figure shows, there are considerable trend in all of them. For instance, in sales, we can see strong nonlinearity and upward trend, it also shows a weak seasonality, and its mean increases as the number ofmonths' increases. It means that sales figures are not fluctuated over fixed mean. The variance is also altering. Thus, this series is not stationary and it would need differencing or even de-trending to become stationary. Data visualization is the graphical representation of information and data. By using visual elements like charts, graphs, and maps, data visualization tools provide an accessible way to see and understand trends, outliers, and patterns in data. In the world of Big Data, data visualization tools and technologies are essential to analyze massive amounts of information and make data-driven decisions.

Data visualization should help one answer the following questions:

This is the dataset analysis with all the different data visualizations that can help the Bangalore Mart 250 Supermarket owner understand everything about their sales. The user will be able to interact with the dataand put different filters to see different data aspects of the data set.

Because of the variability of the unique values of the numeric columns a scatter plot with the target valuewill be of use.

- *Categorical Data Set*

	Invoice_ID	Branch	Customer_Type	Gender	Product_Type	Unit_Price	Quantity_Sold	Total sales in RWF	Sales_Date	Time	Payment_Mode
990	3448131	A	Normal	Female	Food and beverages	56.56	5	296.9400	3/22/2019	19:06	Credit card
991	3451633	B	Normal	Female	Sports and travel	76.60	10	804.3000	1/24/2019	18:10	Ewallet
992	3455135	A	Normal	Male	Electronic accessories	58.03	2	121.8630	3/10/2019	20:46	Ewallet
993	3458637	B	Normal	Male	Fashion accessories	17.49	10	183.6450	2/22/2019	18:35	Ewallet
994	3462139	C	Member	Female	Electronic accessories	60.95	1	63.9975	2/18/2019	11:40	Ewallet
995	3465641	C	Normal	Male	Health and beauty	40.35	1	42.3675	1/29/2019	13:46	Ewallet
996	3469143	B	Normal	Female	Home and lifestyle	97.38	10	1022.4900	3/2/2019	17:16	Ewallet
997	3472645	A	Member	Male	Food and beverages	31.84	1	33.4320	2/9/2019	13:22	Cash
998	3476147	A	Normal	Male	Home and lifestyle	65.82	1	69.1110	2/22/2019	15:33	Cash
999	3479649	A	Member	Female	Fashion accessories	88.34	7	649.2990	2/18/2019	13:28	Cash

Fig 2: Data Set

III. RESULT

- *Machine Learning Models*

First of all, we will divide our dataset into two variables X as the features we defined earlier and y as salesthe target value we want to product. This is a regression problem so we will use Regression methods.

Machine Learning Models we will use:

- Linear Regression.
- Random Forest Regressor.
- Lasso Regressor.
- Gradient Boosting Regressor.
- Decision Tree Regressor.
- Ridge Regressor.

The process of Modeling the Data:

- Importing the Model.
- Fitting the Model.
- Predicting Product Sales.
- Regression metrics

Score Metrics for Regression: Mean Absolute Error(MAE)- Mean of the absolute value of errors(absolute distance from true value)

Mean Squared Error(MSE)- Mean of the squared value of errors (squared distance from true value)R^2 (Coefficient of determination)- Regression score function.

- *Linear Regression*

In statistics, linear regression is a linear approach to modelling the relationship between a scalar responseand one or more explanatory variables. Linear regression was the first type of regression analysis to be studied rigorously and to be used extensively in practical applications. This is because models which depend linearly on their unknown parameters are easier to fit then models which are non-linearly related to their parameters and because the statistical properties of the resulting estimators are easier to determine. Mean Absolute Error: 4260281900475389.0 Mean Squared Error: 2.0816586633209133e+32 R^2 Score: -3.373208798734109e+27

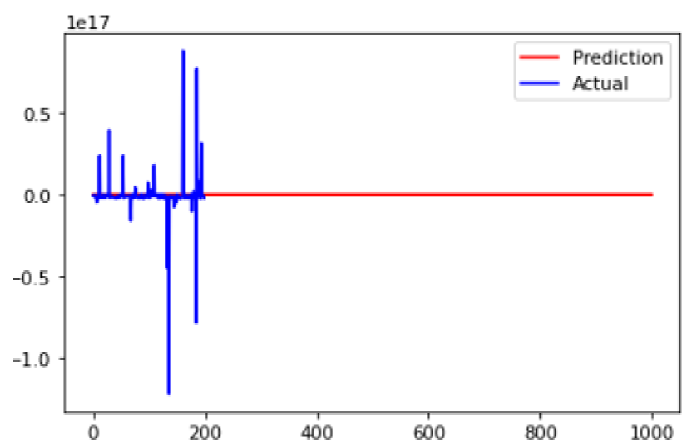

Fig 3: Linear Regressor

y_test	prediction
183.645	-7.31462E+14
450.1035	-9.3192E+14
252.042	-1.41685E+15
139.923	3.41732E+13
231.2415	-1.92433E+15
271.2885	-5.22176E+14
184.107	-2.46188E+15
33.3585	-5.03604E+15
295.6905	-6.49986E+14
180.6	-7.43964E+14
132.762	-1.52962E+15
262.458	-7.10132E+14
53.34	2.35E+16
94.1745	-2.55962E+15
269.934	-1.57097E+15
722.232	-2.40549E+15
888.405	-2.37166E+15
319.788	-4.05644E+14
749.49	-1.80404E+15
189.0945	80

Table 1: Predicted Table of Linear Regressor

➢ *Random Forest Regressor*

Random forest is a supervised Learning algorithm which uses ensemble learning method for classification and regression. It operates by constructing a multitude of decision trees at training timeand outputting the class that is the mode of the classes(Classification) or more prediction (regression)of the individual trees.

Mean Absolute Error: 81.15 Mean Squared Error: 10721.4R^2 Score: 0.8263

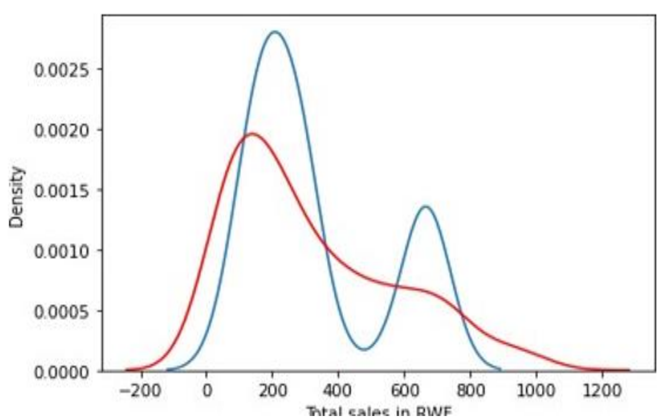

Fig 4: Random Forest Regressor

y_test	prediction
183.645	295.8001357
450.1035	295.8001357
252.042	204.3502547
139.923	269.8305605
231.2415	236.7240066
271.2885	295.8001357
184.107	171.6434557
33.3585	159.332204
295.6905	295.8001357
180.6	173.9326495
132.762	167.9326819
262.458	269.8305605
53.34	175.9844042
94.1745	171.6434557
269.934	248.509019
722.232	675.813486
888.405	675.813486
319.788	295.8001357
749.49	674.6874227
189.0945	295.8001357
737.7615	675.813486
344.4	295.8001357

Table 2: Predicted Table of Random Forest Regressor

➢ *Gradient Boosting Regressor*

Gradient Boosting Regressor is trained with the dataset by using 5-fold time series cross-validation approach where 80% of the data was used for training and 20% of the data was used as the test set and the performances have been measured by using the metrics, MAE, R_2 and MEAN SQUAREDERROR. The following are the results obtained:

Mean Absolute Error: 8.93 Mean Squared Error: 175.28R^2 Score: 0.9972

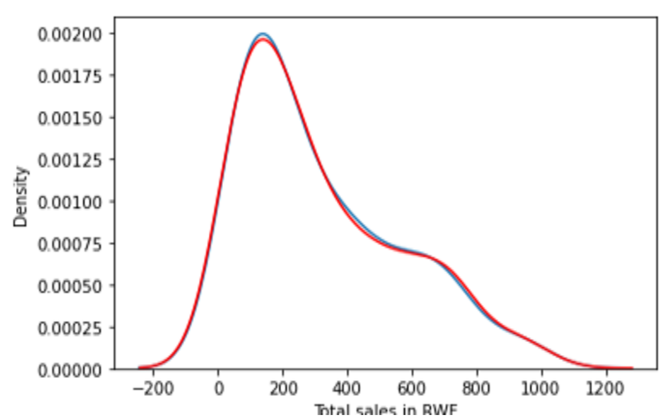

Fig 5: Gradient Boosting Regression Prediction

ISSN No:-2456-2165

y_test	prediction
183.645	182.4677865
450.1035	436.5291926
252.042	249.6448266
139.923	141.9073197
231.2415	239.5117948
271.2885	308.1323586
184.107	181.1521226
33.3585	41.96693354
295.6905	337.3053427
180.6	161.1490238
132.762	135.6827636
262.458	273.3479838
53.34	60.03432395
94.1745	97.09529761
269.934	277.1089912
722.232	729.3015136
888.405	886.5206752
319.788	321.5106092
749.49	716.8551388
189.0945	187.3079236
737.7615	755.0012901
344.4	347.1860577

Table 3: Predicted Table of Gradient Boosting Regression

y_test	prediction
183.645	354.7662893
450.1035	475.4590819
252.042	256.8883256
139.923	165.7192469
231.2415	284.155282
271.2885	316.7210893
184.107	287.0973663
33.3585	-91.10331079
295.6905	325.6735729
180.6	155.0040476
132.762	136.1970526
262.458	260.1413328
53.34	30.81342935
94.1745	282.7019613
269.934	285.5906715
722.232	637.8594768
888.405	746.8190943
319.788	409.3889213
749.49	660.8275414
189.0945	209.7935015
737.7615	653.8934932
344.4	447.5744014

Table 4: Predicted Table of Lasso Regressor

➢ *Lasso Regressor*

Mean Absolute Error: 64.05 Mean Squared Error: 7462.48 R^2 Score: 0.8791

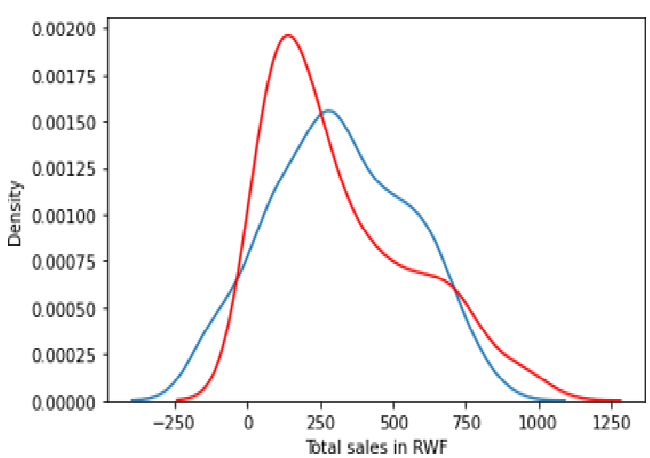

Fig 6: Lasso Regressor Prediction

➢ *Decision Tree Regressor*

In the shape of a tree structure, a decision tree constructs regression or classification models. It breaksdown a dataset into smaller and smaller chunks over time while also building a decision tree.

Mean Absolute Error: 76.21 Mean Squared Error: 9428.49R^2 Score: 0.8472

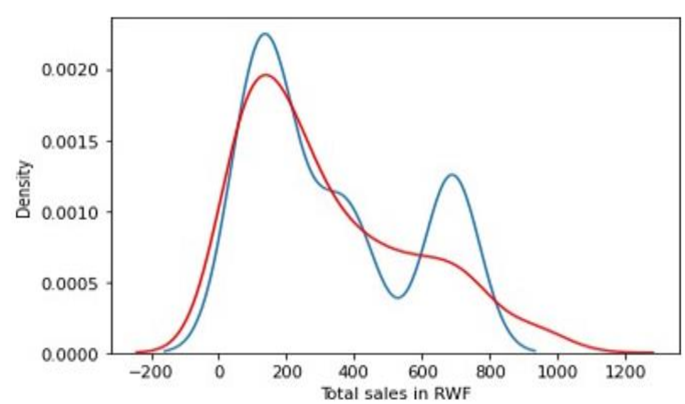

Fig 7: Decision Tree Regressor Prediction

y_test	prediction
183.645	194.2774324
450.1035	402.8178416
252.042	337.1478716
139.923	194.2774324
231.2415	337.1478716
271.2885	402.8178416
184.107	84.17655556
33.3585	147.2812895
295.6905	402.8178416
180.6	147.2812895
132.762	84.17655556
262.458	402.8178416
53.34	147.2812895
94.1745	84.17655556
269.934	337.1478716
722.232	688.902375
888.405	688.902375
319.788	194.2774324
749.49	688.902375
189.0945	194.2774324
737.7615	688.902375
344.4	194.2774324

Table 5: Predicted Table of Decision Tree Regressor

➢ *Ridge Regressor*

Ridge Regressor is trained with the dataset by using 5-fold time series cross-validation approach where 80% of the data was used for training and 20% of the data was used as the test set and the performances have been measured by using the metrics MAE and MSE. The following are the results obtained by the Ridge Regressor: Ridge Regressor is trained with the dataset by using 5-foldtime series cross-validation approach where 80% of the data was used for training and 20% of the data was used as the test set and the performances have been measured by using the metrics MAE and MSE. The following are the results obtained by the Ridge Regressor:

Mean Absolute Error: 87.43 Mean Squared Error: 13140.15R^2 Score: 0.7871

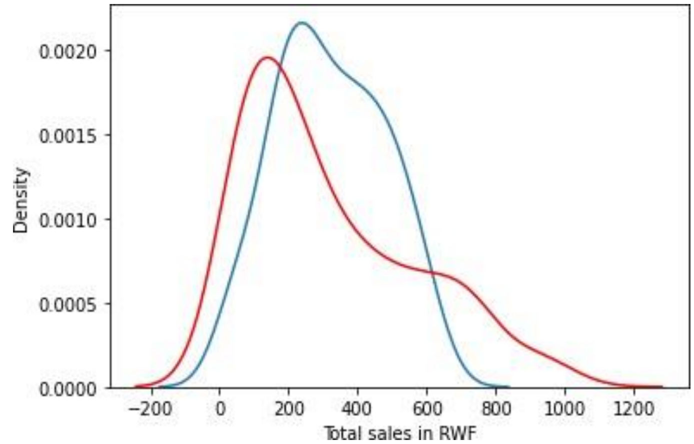

Fig 8: Ridge Regressor Prediction

y_test	prediction
183.645	386.4439384
450.1035	448.3073834
252.042	268.4584596
139.923	235.2635007
231.2415	271.9585774
271.2885	330.4895218
184.107	277.0186945
33.3585	63.32596952
295.6905	339.5273646
180.6	199.9733564
132.762	203.3193579
262.458	287.9754674
53.34	162.1518404
94.1745	224.1153392
269.934	278.6980996
722.232	539.4290807
888.405	621.5984824
319.788	403.6693712
749.49	573.3217527
189.0945	357.0893307
737.7615	547.1450267
344.4	463.1480609

Table 6: Predicted Table of Ridge Regressor

➢ *Comparative study of the Fitting Models basing on Metrics*

- *Performance evaluation results*

	models	MAE	MSE	R^2
5	Gradient Boosting Regressor	8.940000e+00	1.753500e+02	9.972000e-01
2	Lasso Regressor	6.405000e+01	7.462480e+03	8.791000e-01
3	Decision Tree Regressor	7.621000e+01	9.428490e+03	8.472000e-01
1	Random Forest Regressor	8.115000e+01	1.072140e+04	8.263000e-01
4	Ridge Regressor	8.743000e+01	1.314015e+04	7.871000e-01
0	Linear Regression	4.260282e+15	2.081659e+32	-3.373209e+27

Table 7: Performance Evaluation Results

From the table above, Gradient Boosting Regressor performed well with both the metrics MAE and MSE, Gradient Boosting Regressor has least error in forecasting the Sales Prediction using Data components when compared to the Ridge Regressor, Random Forest Regressor and Decision Tree Regressor. The Linear Regressor machine demonstrated the worst performance with the highest errorin both metrics.

Noting that, Interpreting the MAE can be easier than interpreting the MSE. Say that you have a MAE of 10. This means that, on average, the MAE is 10 away from the predicted value. In any case, **the closer the value of the MAE is to 0, the better fitting.** That said, the interpretation of the MAE is completely dependent on the data. In some cases, a MAE of 10 can be incredibly good, while in others it can mean that the model is a complete failure The interpretation of the MAE depends on the range of the values andas well the acceptability of error, for simple example, in the example of a MAE of 10, if the values rangedfrom 10,000 to 100,000 a MAE of 10 would be great. However, if the values ranged from 0 through 20, a MAE would be terrible.

This Article makes use of the GBR model stands for Gradient Boosting Regressor is a typeof machine learning boosting. It relies on the intuition that the best possible next model, when combined with previous models, minimizes the overall prediction error. If a small change in the prediction for a case causes no change in error, then next target outcome of the case is zero.

The GBR model is well suited for this article because of the following reasons:

- GBR is suitable for this article because it classified prediction problems where inputs areassigned a class or label
- GBR method is used to forecast the sales revenue of upcoming period. According to results there are high similarities between forecasted and actual data.
- GBR is suitable for regression prediction problems where a real-valued quantity ispredicted given set of inputs

IV. CONCLUSION

The main objective of this article was to create a management decision making tool, which would help supermarkets predict sales for future time periods in order to better plan their stock purchases, marketing campaigns and budgets for their financial year. A supermarket was used as the case study. Interviews and observation were the data collection techniques used which led to the following findings. There is a system in use which does not make any future predictions but instead works hand in hand with the stock management system to alert the management whenever the quantity of stock in store is too low, and therefore more stock needs to bepurchased. After analyzing the existing system, it was now time to design a prediction algorithm using python Google colab. The accuracy of the prediction model is rather high and can be taken to market as is, several real life tests can be done on several supermarkets to test how well the predictions fair against real sales of the rime period the prediction model was meant to forecast. The cleaning process and modeling process require a data scientist to handle all these processes and therefore data scientists must always be at the help of this article to get things to run smoothly. In conclusion the desired results were achieved and the Sales Forecasting System was implemented and is running as expected.

REFERENCES

[1]. (Shubham S. Jaiswal and Meenakshi Garg, Volume: 07 Issue: 09 | Sep 2020). https://irjet.com/archives/V7/i9/IRJET-V7I9382.pdf
[2]. Bowerman, B. L., and O'Connell, R. T., (1986), Time Series Forecasting: Unified Concepts and Computer Implementation, Boston, MA, USA, PWS Publishing Co., Translated by Reza Shiva, (1th edn)in 1996, Iran.
[3]. Box, G.E., Jenkins, G.M., (1976), Time Series Analysis, San Francisco, CA. Holden-Day. Box, G. E. P.,Jenkins, G. M., and Reinsel, G. C. (1994), Time series analysis: Forecasting and control (3rd ed.), Englewood Cliffs, NJ: Prentice-Hall.
[4]. Dhond, A., Gupta, A., and Vadhavkar, S., (2000), "Data Mining Techniques for Optimizing Inventories for Electronic Commerce", In Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining, Pages: 480 – 486, Boston, Massachusetts, United States.

ISSN No:-2456-2165

[5]. Doganis, P., Alexandridis, A., Patrinos, P., and Sarimveis, H., (2006), "Time series sales forecasting for short shelf-life food products based on artificial neural networks and evolutionary computing", Journal of Food Engineering, 75:2, 196–204.

[6]. Gupta, A., Maranas, C.D., and McDonald, C.M., (2000), "Mid-term supply chain planning under demand uncertainty: customer demand satisfaction and inventory management", Computers and Chemical Engineering, 24: 12, 2613–2621.

[7]. A Feasibility Study of Setting-Up New Production Line – Either Partly Outsource a Process or Fully Produce In-House url: http://www.diva-portal.org/smash/get/diva2:121377/FULLTEXT01.pdf (Visited2nd May 2022)

[8]. SALES FORECASTING MANAGEMENT Improvement of new product forecasting process in the Swedish company Heliospectra url: http://www.diva-portal.se/smash/get/diva2:968684/FULLTEXT01.pdf (Visited 6th May 2022)

[9]. Sales prediction using machine learning url:https://ui.adsabs.harvard.edu/abs/2022AIPC.2444d0001P/abstract (Visited 15th May 2022)

[10]. Beginner's guide to sales forecasting methodology url:https://www.anaplan.com/blog/sales-forecasting-guide/ (Visited 20th May 2022)

[11]. Increasing sales forecast accuracy with technique adoption in the forecasting process url: https://www.diva-portal.org/smash/get/diva2:722754/FULLTEXT01.pdf (Visited 26th April 2022)

[12]. Chuning Deng and Yongji Liu, (2021). https://www.hindawi.com/journals/wcmc/2021/9969357.

[13]. Sales forecasts for existing consumer products and services url: https://www.sciencedirect.com/science/article/abs/pii/S0169207000000583 (Visited 2nd May 2022)

[14]. Evaluating predictive count data distributions in retail sales forecasting url:ttps://www.sciencedirect.com/science/article/abs/pii/S0169207016000315 (Visited 6th May 2022)

[15]. https://www.sciencedirect.com/science/article/pii/S0377221721006500. (Thanos E.GoltsosaAris A.SyntetosaChristoph H.GlockbGeorgeIoannouc, Volume 299, Issue 2, 1 June 2022, Pages 397-419)

[16]. Doganis, P., Alexandridis, A., Patrinos, P., and Sarimveis, H., (2006), "Time series sales forecasting for short shelf-life food products based on artificial neural networks and evolutionary computing", Journal of Food Engineering, 75:2, 196–204.

[17]. Sales Prediction using Machine Learning url: https://www.scribd.com/document/549715095/009-Praveen-NCDST503 (Visited 16th May 2022) Sales Forecasting in Small and Medium-Sized Enterprisesurl: url:https://www.theseus.fi/bitstream/handle/10024/106191/Haataja_Timo.pdf?sequence=1&isAllowed=y (Visited 10th May 2022)

[18]. The definitive guide to sales forecasting methodologies url: https://www.zendesk.com/blog/5-essential- sales-forecasting-techniques/ (Visited 10th May 2022)

[19]. Kuo, R.J., Wu., P. and Wang, C.P., (2002), "An intelligent sales forecasting system through integration of artificial neural networks and fuzzy neural networks with fuzzy weight elimination", Neural Networks, 15: 7, 909–925.

[20]. Shanmugasundaram, J., Nagendra Prasad, M. V., Vadhavkar, S., Gupta, A., (2002) "Use of Recurrent Neural Networks for Strategic Data Mining of Sales Information", MIT, Sloan School of Management, Working Paper, No. 4347-02.

[21]. Two Crows (1999), Introduction to Data Mining and Knowledge Discovery, (3rd edn), U.S.A., by TwoCrows Corporation.

[22]. Yip, D.H.F., Hines, E.L. and Yu, W.W.H., (1997), Application of Artificial Neural Networks in Sales Forecasting, IEEE International Conference on Neural Networks, Houston, TX, USA., 4, 2121-2124,

[23]. Simple Linear Regression Analysis url:http://home.iitk.ac.in/~shalab/regression/Chapter2-Regression- Simple Linear Regression Analysis.pdf (Visited 6/06/2022)

[24]. Long Short-Term Memory Networks with Python. en-US. url: https:// machinelearningmastery.com/lstms-with-python/ (visited on 25th August 2022).