ISSN No:-2456-2165

A Comparative Study of Various Machine Learning Models on Interval Data a Case Study of Maison Gil Ltd.

Emmanuel Byiringiro*[1], Wilson Musoni[2]
[1] Information communication and Technology (ICT), University of Kigali, Kigali, Rwanda.
[2] Information communication and Technology (ICT), Kepler College, Kigali, Rwanda.

Abstract:- A comparison of a performance of various machine learning models to predict the sales components is presented in this paper. The general aim of this thesis was to find a suitable machine learning model that can fit and well forecast the sales components before they happen and thus create the purchase orders before the product runs out of stock or sold. The dataset used in the thesis is from a product supplier consisting of product sales data for about a thousand assorted products on a time of three years. Firstly, a Literature review used to find suitable machine learning algorithms and then based on the results obtained, an experiment was performed to evaluate the performances of machine learning algorithms. Results from the literature review shown that regression algorithms namely Supports Vector Machine Regression, Ridge Regression, Gradient Boosting Regression, and Random Forest Regression are suitable algorithms and results from the experiment showed that gradient boosting has performed well than the other machine learning algorithms for the chosen dataset, The range of supervised and unsupervised algorithms is provided. The different models were compared using the performance metrics such as mean squared error (MSE), Mean Absolute Error (MAE) and Root Squared (R^2) called coefficient of determination as well based on the dataset's values against the predicted values. The results shows that the Gradient boosting have shown the highly fitting capability and as well with the highly accuracy compared to other Machine learning models. To conclude After the experimentation and the analysis, the Gradient boosting algorithm has been performed well when compared with the performances of the other algorithms and therefore, gradient boosting is chosen as the optimal algorithm for performing the sales forecasting of the sales components at Maison Gil Lt.

Keywords:- *Time Series Forecasting, Sales Forecasting, Mean Absolute Error, Mean Squared Error, Root Squared, Gradient boosting, Support Vector machine, Lasso Regressor, Linear regressor, Ridge regressor , Decision Tree Regressor , Random Forest Regressor .*

I. INTRODUCTION

Sales forecasts and predictions are especially important areas to always focus on. Efficient and optimal prediction methods are essential for all managers to maintain the efficiency of their selling strategies within their organizations. Performing this task manually can result in significant errors and poor decision-making management. Most importantly, it takes time. This is not desirable in this fast-moving world. Much of today economic growth is relying on the corporate sector which is mostly deals with the big sales and which is expected to produce enough products to meet aggregate demand and sales revenues. Coordinating with the market audience is the focus of the department. Therefore, it is important that the company was able to achieve this goal using a forecasting system. The forecasting process involves analyzing data from a variety of sources, including consumer behavior, market trends, and other factors. This analysis helps the companies as well to have effective financial resources management. The forecasting process can be used for a variety of purposes, such as forecasting future demand for a product or service or forecasting the number of products sold in each period.

➢ *Objectives*

This report examines whether machine learning techniques can accurately predict sales components, we choose to apply algorithms that are appropriate for forecasted sales components such as linear Regressor as the baseline model, Random Forest regressor, decision Tree regressor, Gradient boosting Regressor and compare their performance. reliable forecasting results could be extended to potentially improve automatic asset allocation in the sales portfolio or recommend a monthly – trading algorithm construction.

II. LITERATURE REVIEW

Numerous well-known strategies have been developed because of forecasting's significance in numerous industries. These techniques are typically referred to as statistical techniques, machine learning techniques, or hybrid techniques. Autoregressive (AR), Autoregressive Integrated Moving Average (ARIMA) (Gurnani et al., 2017). Exponential smoothing methods (Hoboken, 2015) are widely used in time series analysis, control, and forecasting. (Becker et al., 2017) and looked at the developments in forecasting during the last 25 years by looking through books on the subject, and as a result, the forecast has advanced.

Francis E.H. Tay and Lijuan Cao's research from 2001, focused on the use of a novel neural network technique called the support vector machine (SVM) in financial time series forecasting. They examined the viability of SVM in this context and proposed that SVMs achieve the best network

structure by implementing the structural risk minimization principle, which seeks to minimize an upper bound of the generalization error rather than the training error. Over time, this leads to a generalization performance that is superior to other neural network models. SVMs have been demonstrated the performance exceptionally well when used to tackle non-linear regression estimation problems.

Mariana Oliveira and Luis Torgo's study (2015). attempted with the ensembles with the goal of enhancing forecasting performance. and acknowledged ensembles as one of the most ambitious methods for tackling predictive challenges and conventional in utilizing variety among models to reduce the variance and bias components of the forecasting error. The summation of the most recent observed values was done by the authors using various embed sizes. Standard bagging and the standard ARIMA are compared in this research study, and successful results are obtained, demonstrating that the approach is promising and can be utilized as an option for forecasting the time series. RMSE is a tool for assessing performance (Mariana Oliveira and Luis Torgo, 2015).

Because of the data it can use and the money it can make, financial time series forecasting is invariably a practitioner's focal point. When forecasting financial time series, ensemble algorithms significantly improve base learners' performances. The four basic learner algorithms—SVR (support vector regression), BPNN (back-propagation neural network), RBFNN (radial basis function neural network), and locally weighted learning—were tested on the Shanghai Composite Index dataset by the authors (LWL). Comparing Random Subspace, Stacking, and Bagging reveals that the latter method, bagging, offers stable improvement for the selected dataset. The performance evaluation measures RMSE, and RAE are utilized in a comparison of ensemble methods for financial market prediction (A Comparison of Ensemble Methods in Financial Market Prediction | Request PDF, n.d.)

(Galadí et al., 2021) asserts that to address the challenge of capturing the non-stationarity property and to identify the precise movements, financial time series forecasting experiments were conducted utilizing an intelligent hybrid model. In this paper, support vector regression (SVR) and empirical mode decomposition (EMD) are both introduced, and a novel hybrid intelligent prediction technique is suggested. A non-linear, non-stationary signal is broken down into IMFs using EMD. In the experiment, the suggested algorithm separates the data into intrinsic modes (IMFs) by utilizing EMD and SVR with various kernel functions (L.I.N.F.-J, 2015). When compared to individual SVR results, the results demonstrate that the EMD-SVR model produces accurate results. Model performance is assessed using the metrics RMSE, MAE, and MAPE.

Because noise and non-stationarity are present, forecasting and modeling financial time series can be challenging. A nonlinear radial basis function neural network ensemble model using the datasets S and P index series and Nikkei 225 index series gathered from the DataStream was proposed by (Donglin Wang and Yajie Li, 2010). The experiment comprises four parts, the first of which involves bagging and boosting the data to separate training sets. The following stage involves feeding training sets into distinct RBF-NN models, which then yield different predictors based on the diversity principle. The third stage involves selecting appropriate neural network ensembles using the Partial Least Squares algorithm (PLS). In the last step, RBF-NN ensembles are forecasted using SVM-Regression. The suggested ensemble method is superior to several existing ensemble approaches, according to experimental results.

According to Mohammed Al-Gunaid et al., (2021) , the main goal of this article is to identify the variables that affect the amount of sowing crop sales and develop a method for the most accurate sales forecasting to support decision-making and improve the efficiency of agro-industrial company procedures. The methodology to forecasting sowing crop sales volumes was also covered in this research, along with the variables influencing sales and a comparison of methods for creating mathematical models. Forecasts are created using a neural network, random forests, and linear regression algorithms. We employ RMSE, ME, and MAPE as our evaluation metrics. Results have demonstrated that, when compared to neural networks, Random Forest has delivered better forecasts.

III. METHODOLOGY

The quantitative methods were briefly introduced and the appropriate methods for our research will be defined. In addition, secondary data will be applied. Regarding the analysis of collected data, the Data analysis will be done using the python with its data analysis libraries such as NumPy, With NumPy, a core Python library for scientific computing, you can create powerful three-dimensional array objects that may be utilized in experiments (Next, n.d.). You can also access a wide range of sophisticated mathematical operations on these arrays, Matplotlib a Python 2D plotting library called Matplotlib generates publication-quality figures in a range of physical formats and in cross-platform interactive settings. Seaborn, A matplotlib-based Python data visualization library is called Seaborn. It offers a sophisticated drawing tool for creating eye-catching and instructive statistical visuals. Sklearn, an open-source library that has effective data mining and data analysis features. The range of supervised and unsupervised algorithms is provided. The performance metrics to measure the accuracy model such as mean squared error (MSE), mean absolute error (MAE) were applied to find the accurate model. Finally, Root squared (R2) or the coefficient of determination have been applied for the fitting model testing.

A. Data Collection.

The yearly sales of shoes in manufacturing companies from the years 2018 to 2020 were taken from the Shoes manufacture and Sales database and utilized as the dataset for this thesis. There are three parts to the shoes: shoes for men, for women, and for children. And as well with their subcategories which are the casual and dress shoes. 156 weeks of data make up the final processed data. 156 weeks of

data are utilized to train models, while the remaining data is used for testing. Totals of 9994 observations and fifteen attributes was considered, and each observation represented a different row number.

B. Feature extraction

Feature Extraction is useful when you need to reduce the number of features required for processing without losing important or relevant information. Feature extraction can also reduce duplicate data in a data set (Jiang et al., 2017). The Total Unit Price and Quantity data set represents and creates the Total Sales (Total_Sales_RWF) of the data set. There was a performance correlation between Total Sales and Unit Price, and a negative correlation between Total_Sales_RWF and Quantity.

C. Data Analysis

Data analysis is the process of examining, cleaning, transforming, and modeling data with the goal of discovering useful information, drawing conclusions, and supporting decision-making. There are multiple aspects and approaches to data analysis, and different techniques were used under different names in different fields of business, science, and social sciences. Data mining is a specialized data analysis technique that focuses on modeling and knowledge discovery for prediction rather than purely descriptive purposes.

Sales forecasting were considered supervised machine learning. In supervised learning, you have an input variable (x) and an output variable (Y) and use an algorithm to learn a mapping function from input to output. Y = f(X), the goal is to fit a good mapping function so that given new input data (x), it can predict the output variable (y) for that data. The process of learning an algorithm from a training dataset is named supervised learning because it can be viewed as a teacher overseeing the learning process. we know the correct answer. The algorithm iteratively makes predictions on the training data. Learning stops when the algorithm reaches an acceptable level of performance.

➢ *Python*

Analysis tools such as Python has been integrated. Python is an easy-to-learn and powerful programming language. It features an efficient high-level data structure and a simple and effective approach to object-oriented programming. Python's sophisticated syntax and dynamic typing, along with its interpreted nature, make it an ideal language for scripting and rapid application development in many domains on most platforms. A language excels at string processing. that is, string manipulation is still actively developed by a community of developers who list some languages with great string processing power and test whether they are object-oriented programming languages. (Programming Languages for Library and Textual Processing - Fosdick - 2005 - Bulletin of the Amer*ican Society for Information Science and Technology - Wiley Online Library*, n.d.).

➢ *Performance Metrics*

As predictive modeling handles two types of problems; classification and regression, there exit different metrics for each of them. The major difference between the evaluation of the two categories is that in regression problems there are continuous variables with which the model is dealing, and they can be used to calculate the error between actual values and the predicted ones. However, in classification problems there are correctly and incorrectly classified outputs which are used to compare the predicted to the actual values. It is important to use multiple metrics, as you cannot rely on a single number to decide on the best model and for comparisons between more than one algorithm or a more than one version of the same algorithm. This is basically because one metric can compensate for the limitation of the others. In this case four metrics will be used on each model to decide which one is the best fit and as well as the performance for the evaluation testing.

- *R-square:*

This metric is defined as the percentage of the response variable variation explained by a regression model. This variation is also called the Sum of Squares and it describes how data points have deviated from the mean which is itself a measurement of central tendency. R-squared always takes a value between 0 and 1 and the closer the value to 1 the better the model because it means that more variance is explained by the model (R-Squared Formula, Regression, and Interpretations, n.d.).

- *Mean Absolute Error (MAE):*

This is considered as one of the simplest, most intuitive, and easily interpretable metrics out there to evaluate models. In essence, MAE is a description of residuals magnitude, so it is as simple as the absolute difference between actual values and residuals (Pascual, 2018).

- *Mean Squared Error (MSE):*

The square root of the mean square error is known as the root mean square error (RMSE). The average squared difference between prediction and observation is its root. Less error indicates that the model is more accurate. For each one of the four models (Decision Tree model, Random Forest model, Extreme Gradient Boosting model, Support Vector Machine Regression model) and the ensemble model (Pascual, 2018).

D. Data Processing:

➢ *Cleaning the Data*

Cleaning data should be the first step in any Data Science (DS) or Machine Learning (ML) workflow. Without clean data it will be much harder time seeing the actual important parts in the exploration. Once the training of ML models begins, they will be unnecessarily more challenging to train. The main point is that to get the most out of a dataset, it should be clean. In the context of data science and machine learning, data cleansing means filtering and modifying data to make it easier to explore, understand, and model. Filter out parts you do not want or do not need so you don't have to view or edit them. Change it to the format you want so that you can effectively use the parts you want. The dataset used required the following changes to be considered clean

- Dropping of Rows with Null Values, check for incomplete (Missing values)
- Removal of all negative numbers and replacing them with absolute values
- Adjust the column Sales_Date into the right format of date Time format

➢ *Sample of dataset*

h_Location	Category	Sub_Category	Product_ID	Product_Type	Customer_ID	Sales_Date	Unit_Price	Quantity	Discount	Payment_Mode	Total_Sales_RWF
Nyarugenge	Men_Shoes	Casual_Shoes	MSH-03	Boots	JE-15475	1/1/2018	15280	1	0.0	Cash	15280
Nyarugenge	Men_Shoes	Casual_Shoes	MSH-02	Sneaker	KB-16585	1/1/2018	19900	2	0.0	Visa	39800
Kicukiro	Kids_Shoes	Casual_Shoes	KSH-02	Casual Shoes	CW-11905	1/1/2018	11230	3	0.8	Cash	24706
Nyarugenge	Men_Shoes	Casual_Shoes	MSH-06	Sandals	MY-18295	1/1/2018	14880	4	0.6	Cash	50592
Nyarugenge	Men_Shoes	Casual_Shoes	MSH-02	Sneaker	NF-18385	1/1/2018	15980	5	0.4	Cash	73508
Nyarugenge	Kids_Shoes	Casual_Shoes	KSH-02	Casual Shoes	RH-19510	1/1/2018	17460	5	0.8	Cash	73332
Musanze	Women_Shoes	Casual_Shoes	WSH-06	Biker Boots	SV-20365	1/1/2018	18460	4	0.3	Cash	68302
Nyarugenge	Men_Shoes	Casual_Shoes	MSH-04	Chealsea	EA-14035	1/1/2018	23000	1	0.0	Cash	23000

Fig 3.1 Sample of dataset

➢ *Handling the Missing Data*

The data we receive is rarely uniform. Data may be missing and should be handled so as not to impact machine learning model performance. To do this, we need to replace the missing data with the mean or median across the columns. Use sklearn for this. A preprocessing library containing a class called Imputer that helps maintain missing data.

Our object name is **imputer.** The Imputer class can take parameters like:

1**. missing_values**: Placeholders for missing values. All occurrences of missing_values are completed. You can give an integer or 'NaN' to find missing values.

2. Strategy: This is an imputation strategy. For Mean, replace missing values with the mean along the axis (column). Other strategies are median and most frequent.

3. axis: are assigned 0 or 1. 0's are allocated along the columns and 1's are allocated along the rows thereafter we fit the imputer object to our data.

➢ *Removal of Negative Numbers and null values*

After attempting to plot 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 managing real world data. The best option would be to assume that these are indeed the correct Quantity and Unit Prices but incorrectly entered as negative values, therefore it is necessary to return the absolute numbers in 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	Unit_Price	Quantity
1/1/2018	15280	1
1/1/2018	19900	2
1/1/2018	11230	3
1/1/2018	14880	4
1/1/2018	15980	5
1/1/2018	17460	5
1/1/2018	18460	4
1/1/2018	23000	1

Fig 3.2 Sample of dataset with normal Date Format

Sales_Date	Unit_Price	Quantity
2018-01-01	15280	1
2018-01-01	19900	2
2018-01-01	11230	3
2018-01-01	14880	4
2018-01-01	15980	5
2018-01-01	17460	5
2018-01-01	18460	4
2018-01-01	23000	1

Fig 3.3 Sample of dataset after changes

➢ *Encoding categorical data*

Original Data

Team	Points
A	25
A	12
B	15
B	14
B	19
B	23
C	25
C	29

One-Hot Encoded Data

Team_A	Team_B	Team_C	Points
1	0	0	25
1	0	0	12
0	1	0	15
0	1	0	14
0	1	0	19
0	1	0	23
0	0	1	25
0	0	1	29

Fig 3.4 Converting Categorical data into dummy variables

➢ *Splitting the Data set into Training set and Test Set*

Now we divide our data into two sets, one for training our model called the **training set** and the other for evaluating the performance of our model called the **test set**. The split is generally 80/20 yet for this paper we applied 69/33 for the 33 of the testing and 69 for the training . To do this we import the "train_test_split" method of "sklearn.model_selection" library. **from sklearn.**model_selection **import** train_test_split .

Now to build our training and test sets, we will create 4 sets:

Then split the data into two **sets. One** for training **the model,** called the training **set,** and **another** for evaluating the performance of **the model,** called the test set. The split is **usually** 80/20. To do **this,** import the **"train_test_split"** method **from the "sklearn.model_selection"** library. **From** sklearn.model_selection import train_test_split . **To create the** training and test sets, we create **four** sets:

1. **X_train** (training part of the matrix of features),
2. **X_test** (test part of the matrix of features),
3. **Y_train** (training part of the dependent variables associated with the X train sets, and therefore also the same indices),
4. **Y_test** (test part of the dependent variables associated with the X test sets, and therefore also the same indices).

We will assign to them the test_train_split, which takes the parameters — arrays (X and Y), test_size (Specifies the ratio in which to split the data set).

```
X_train, X_test, Y_train, Y_test = train_test_split( X , Y , test_size = 0.2, random_state = 0)
```

➢ *Feature Scaling*

Most machine learning algorithms use the Euclidean distance between two data points in their calculations. Therefore, features with large magnitudes contribute more to distance calculations than features with small magnitudes. Standardization or z-score normalization is used to circumvent this feature. This is done by using "StandardScaler" class of "sklearn.preprocessing". from sklearn.preprocessing import StandardScaler sc_X = StandardScaler()

Further we will transform our X_test set while we will need to fit as well as transform our X_train set.
The transform function will transform all the data to a same standardized scale.

```
X_train = sc_X.fit_transform(X_train)
X_test = sc_X.transform(X_test)
```

E. Data Visualization

Data visualization is the graphical representation of information and data. By using visual elements such as charts, graphs, and maps, data visualization tools provide an accessible way to see and understand trends, outliers, and patterns in your data. In the world of big data, data visualization tools and technologies are essential for analyzing vast amounts of information and making data-driven decisions.

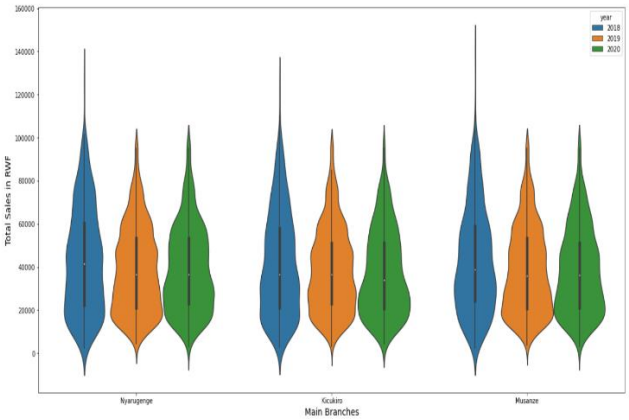

Fig 3.5 Branch_Location vs Total_sales_RWF Yearly Accordingly

IV. RESULT

A. Machine learning models

We split the dataset into two variables. X is the previously defined characteristic and y is the target value Total_Sales_RWF that we want to predict. Since this is a regression problem, we used regression techniques. The divisions of the tensile test are performed in relation to each other.

➢ *Machine Learning Models used:*

- Random Forest Regressor
- Lasso Regressor
- Gradient Boosting Regressor
- Decision Tree Regressor
- Ridge Regressor
- Support vector Machine
- Linear Regressor (Baseline Model)

➢ *The Process of Modeling the Data:*

ISSN No:-2456-2165

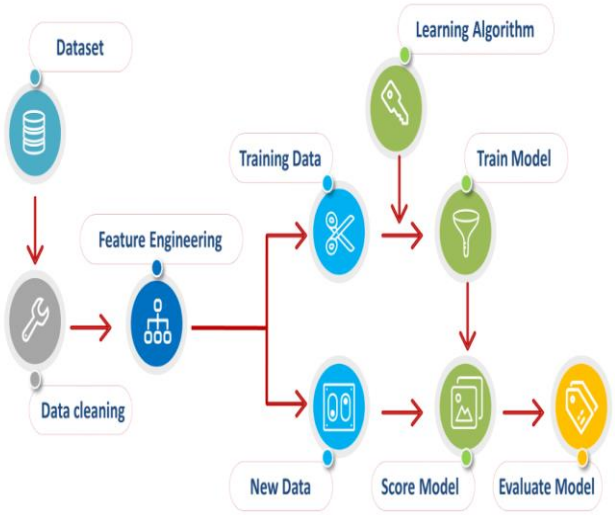

Fig 4.1 Process of Modeling

B. Linear Regression (baseline Model)

In statistics, linear regression is a linear approach for modeling the relationship between a scalar response (or dependent variable) and one or more explanatory variables (or independent variables). Linear regression was the first type of regression analysis that was rigorously studied and widely used in practical applications.

This is because a model that relates linearly to the unknown parameter is easier to fit than one that relates the parameter nonlinearly, and it is easier to determine the statistical properties of the resulting estimator.

Mean Absolute Error: 1451143912512108.2

Mean Squared Error: 3.182614187447832e+33

R^2 Score: -7.57388364480193e+24

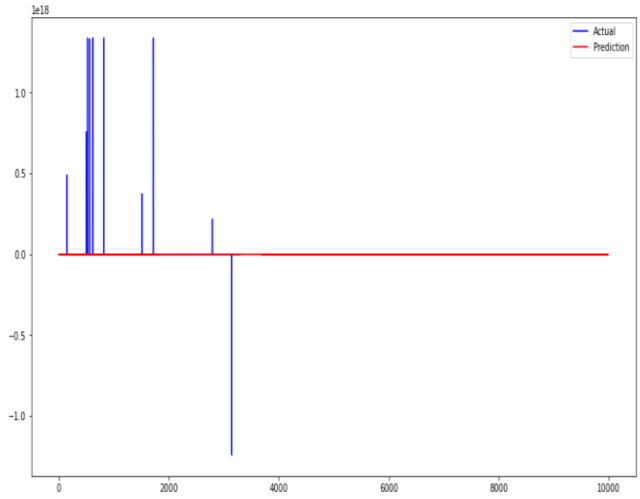

Fig 4.2 Linear Regression graph

C. Random Forest Regressor

Random Forest is a supervised learning algorithm that uses ensemble learning methods for classification and regression. It works by building several decision trees at training time and outputting the classes that are the modes of the classes (classification) or the mean predictions of each tree (regression).

➢ *Metrics Results Obtain with the Random Forest Regressor*

Mean Absolute Error: 2453.58

Mean Squared Error: 12776128.93

R^2 score: 0.97

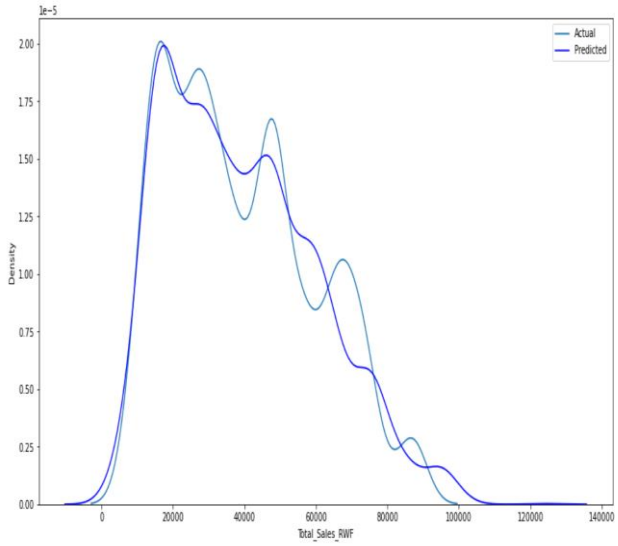

Fig 4.3 Random Forest Regressor graph

D. Lasso Regressor

In statistics and machine learning, Lasso (Least Absolute Shrinkage and Selection Operator, Lasso or LASSO) performs both variable selection and regularization to improve the predictive accuracy and interpretability of the resulting statistical models. It is a regression analysis method.

➢ *Metrics Results Obtain with the Lasso Regressor*

Mean Absolute Error: 3361.09

Mean Squared Error: 21687117.49

R^2 Score: 0.95

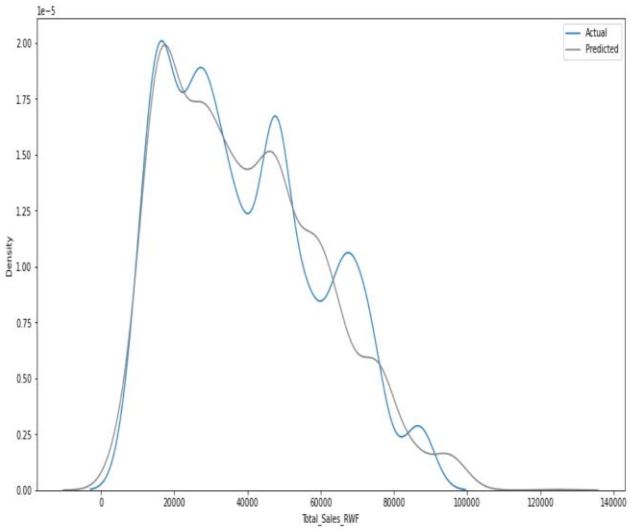

Fig 4.4 Lasso Regressor graph

E. Decision Tree Regressor

Decision tree regression observes object features, trains a model on the structure of the tree to predict future dates, and produces meaningful and continuous output. Continuous output means that the output/result is not discrete. i.e., it is only represented by a discrete known set of numbers or values.

➢ *Metrics Results Obtain with the Decision Tree Regressor*

Mean Absolute Error: 2553.9

Mean Squared Error: 14885873.04

R^2 Score: 0.96

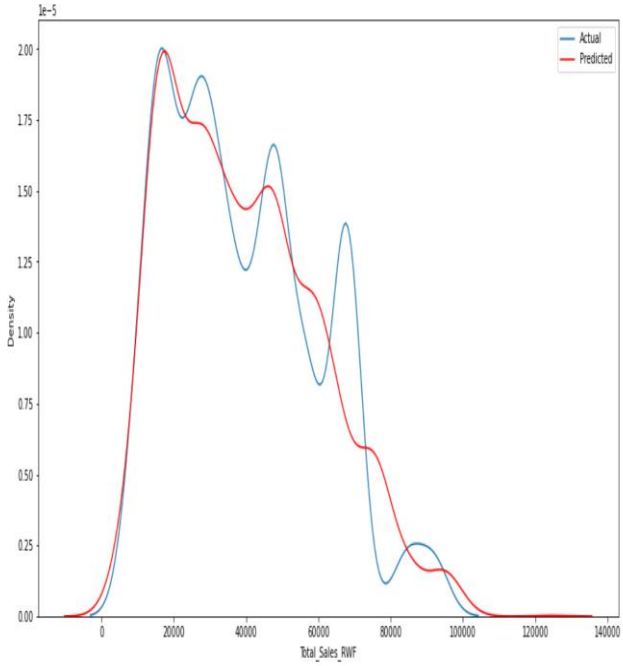

Fig 4.5 Decision Tree Regressor

F. Gradient Boosting Regressor

This is a popular boosting algorithm. In gradient boosting, each predictor corrects its predecessor's error. In contrast to Adaboost, the weights of the training instances are not tweaked, instead, each predictor is trained using the residual errors of predecessor as labels. There is a technique called the **Gradient Boosted Trees** whose base learner is CART (Classification and Regression Trees).

➢ *Metrics Results Obtain with Gradient Boosting Regressor*

Mean Absolute Error: 530.03

Mean Squared Error: 732126.21

R^2 Score: 1.0

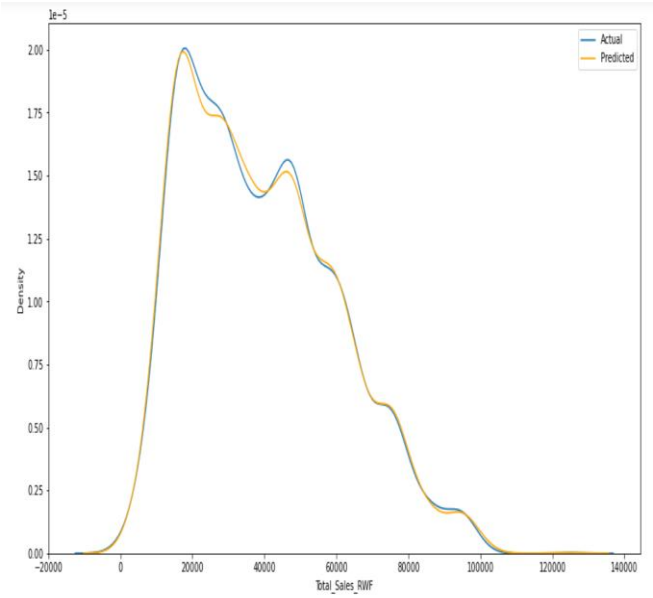

Fig 4.6 Gradient Boosting Regressor

G. Ridge Regressor

A Ridge regressor is a regularized version of a Linear Regressor. i.e to the original cost function of linear regressor we add a regularized term that forces the learning algorithm to fit the data and helps to keep the weights lower as possible. The regularized term has the parameter 'alpha' which controls the regularization of the model i.e helps in reducing the variance of the estimates. Cost Function for Ridge Regressor. Below is the cost function for the Ridge Regressor.

$$J(\Theta) = \frac{1}{m}(X\Theta - Y)^2 + \alpha\frac{1}{2}(\Theta)^2$$

➢ *Metrics Results Obtain with Ridge Regressor*

Mean Absolute Error: 3575.78

Mean Squared Error: 24472465.37

R^2 Score: 0.94

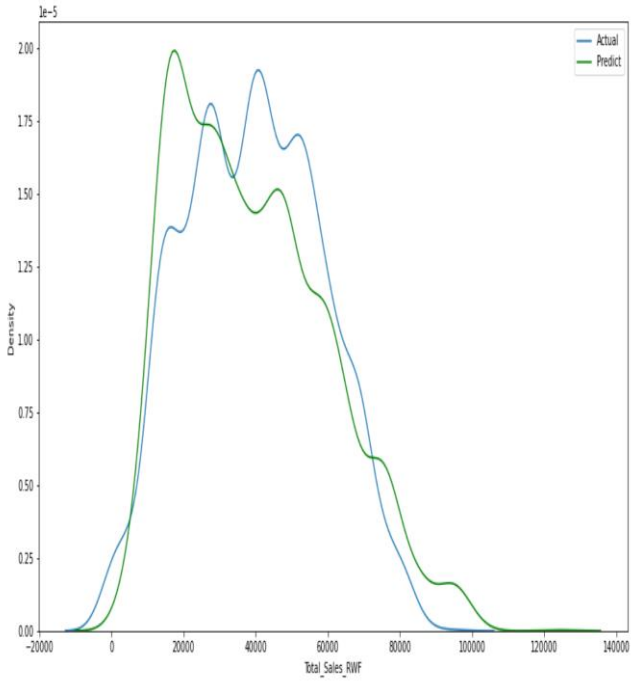

Fig 4.7 Ridge regressor

H. Support Vector Machine

Support Vector Machine Regression or simply Support Vector Regression is one of the margin maximization algorithms; it tries to find the best hyper plane with the largest margin; biggest distance between boundary/separation line and closest data points to it. Data points that lie in the margin of separation are called support vectors from where the algorithm takes its name. It also depends on epsilon-intensive loss function, and thus unlike other algorithms where all data points in the training set are considered, SVR focuses on those points that are epsilon-deviated from the hyper plane. The more data points in the hyper plane the better the model the more it is robust to outliers and the better can be generalized.

➢ *Metrics Results Obtain with Support Vector Machine*

Mean Absolute Error: 3033.52
Mean Squared Error: 91442912.1
R^2 Score: 0.79

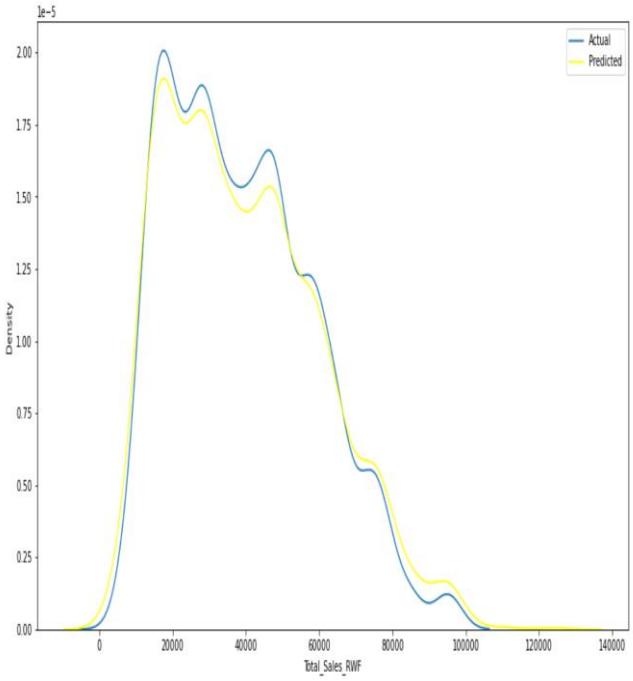

Fig 4.8 Support Vector Machine

I. Overall Results of the Models Towards their Metrics

Table 1. Result of all models that has been tested

Models (Algorithms)	Mean Absolute Error (MAE)	Mean Squared Error (MSE)	Root Square Error (R^2)
Ridge Regression	3575.78	24472465.37	0.94
Support Vector Machine	3033.52	91442912.1	0.79
Random Forest Regression	2453.58	12776128.93	0.97
Gradient Boosting Regressor	530.03	732126.21	1
Decision Tree regressor	2553.9	14885873.04	0.96
Lasso Regression	3361.09	21687117.49	0.95
Linear Regressor	1451143912512108.2	3.182614187447832e+33	-7.57388364480193e+24

J. Findings

This post uses the GBR model, which stands for Gradient Boosting Regressor, a type of machine learning. It relies on the intuition that the best possible next model combined with the previous model minimizes the overall prediction error. If a slight change in a case's prediction causes no change in error, the case's next target outcome is zero.

ISSN No:-2456-2165

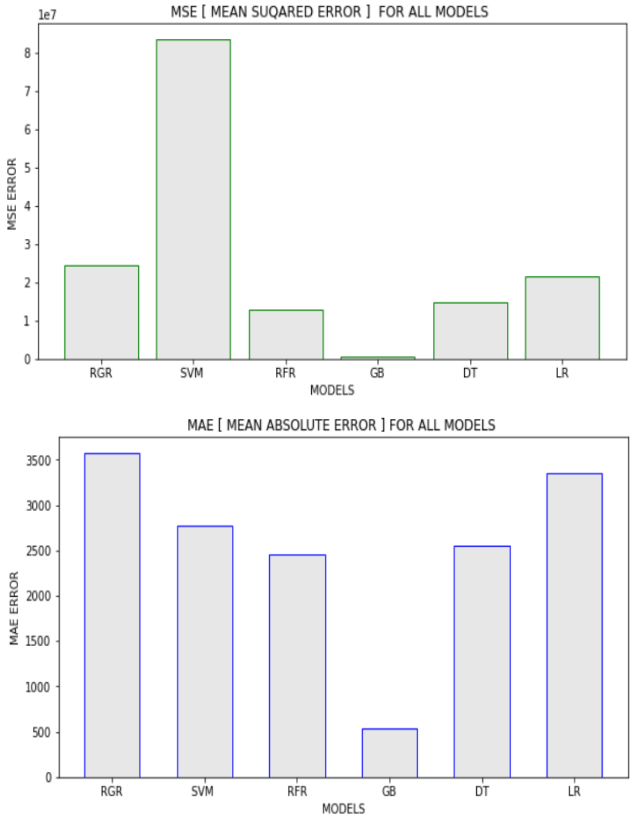

Fig 4.9 MAE For All Models

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 are assigned a class or label.
- GBR method is used to forecast the sales components 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 is predicted given set of inputs.

V. CONCLUSION

Today, sales forecasting is an integral part of every industry, especially seasonal companies. Accurately assessing the future can help companies create more jobs, so knowing which models can produce better results is important. In this study, we examined the ability of several machine learning models to predict sales components and compared their comparisons to have the best fitting models means the Model accuracy.

After performing sales forecasting with various methods, the following conclusions can be drawn from this study. First, most regression methods have performed better than the classical forecasting methods. In addition, the results have indicated the superiority of gradient boosting over the other models. Moreover. In addition, the findings have highlighted the potential usefulness of regression models. Considering the purpose of this method, which is producing quick and accurate forecasts, this model has performed surprisingly good, and better than the other classical methods. Furthermore, the findings have indicated that using the holiday factor improves the performance of Prophet model.

In summary, four models which are Decision Tree Reg ressor, Random Forest Regressor, Gradient boosting model and Linear Regressor which was the baseline on this paper c an forecast the sales of Shoes Sales in the dataset within less than 9994 units of the real sales. It should be mentioned that the Shoe sales range from around 420 to over 13500. Theref ore, a model with MSE of 681475.74 and a MAE of 536.52 with the R_2 of 1.0 can be classified as an acceptable model.

The results of this task will vary depending on the data set selected. It is possible to apply the same model to other seasonal time series. However, we need to tune the parameters to find the best predictive model. As a future study, it would be worthwhile to evaluate the mentioned forecasting methods for other seasonal factors and compare the performance of the applied models. In addition, this study has other directions for future research. To improve your results, you can evaluate complex LSTM and CNN models. Another application is multivariate time series forecasting. Moreover, the development of hybrid models representing a combination of traditional and modern forecasting methods is of value and potential for future research.

REFERENCES

[1]. *A Comparison of Ensemble Methods in Financial Market Prediction | Request PDF*. (n.d.). Retrieved November 11, 2022, from https://www.researchgate.net/publication/261204601_A_Comparison_of_Ensemble_Methods_in_Financial_Market_Prediction

[2]. Al-Gunaid, M., Salygina, I., Shcherbakov, M., Trubitsin, V., & Groumpos, P. (2021). Forecasting potential yields under uncertainty using fuzzy cognitive maps. *Agriculture & Food Security*, *10*. https://doi.org/10.1186/s40066-021-00314-9

[3]. Becker, R. A., Dreier, D. A., Manibusan, M. K., Cox, L. A. (Tony), Simon, T. W., & Bus, J. S. (2017). How well can carcinogenicity be predicted by high throughput "characteristics of carcinogens" mechanistic data? *Regulatory Toxicology and Pharmacology*, *90*, 185–196. https://doi.org/10.1016/j.yrtph.2017.08.021

[4]. Galadí, J. A., Silva Pereira, S., Sanz Perl, Y., Kringelbach, M. L., Gayte, I., Laufs, H., Tagliazucchi, E., Langa, J. A., & Deco, G. (2021). Capturing the non-stationarity of whole-brain dynamics underlying human brain states. *NeuroImage*, *244*, 118551. https://doi.org/10.1016/j.neuroimage.2021.118551

[5]. Gurnani, M., Korke, Y., Shah, P., Udmale, S., Sambhe, V., & Bhirud, S. (2017). Forecasting of sales by using fusion of machine learning techniques. *2017 International Conference on Data Management, Analytics and Innovation (ICDMAI)*, 93–101.

[6]. Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., Wang, Y., Dong, Q., Shen, H., & Wang, Y. (2017). Artificial intelligence in healthcare: Past, present and future. *Stroke and Vascular Neurology*, *2*(4), 230–243. https://doi.org/10.1136/svn-2017-000101

[7]. Next, S. T. X. (n.d.). *40 Most Popular Python Scientific Libraries*. Retrieved November 11, 2022, from https://www.stxnext.com/blog/most-popular-python-scientific-libraries/
[8]. Pascual, C. (2018, September 26). *Tutorial: Understanding Regression Error Metrics in Python*. Dataquest. https://www.dataquest.io/blog/understanding-regression-error-metrics/
[9]. *Programming languages for library and textual processing—Fosdick—2005—Bulletin of the American Society for Information Science and Technology—Wiley Online Library*. (n.d.). Retrieved November 11, 2022, from https://asistdl.onlinelibrary.wiley.com/doi/full/10.1002/bult.2005.1720310607
[10]. *R-Squared Formula, Regression, and Interpretations*. (n.d.). Investopedia. Retrieved November 11, 2022, from https://www.investopedia.com/terms/r/r-squared.asp