Sales-forecasting of Retail Stores using Machine Learning Techniques

Akshay Krishna, Akhilesh V, Animikh Aich, Chetana Hegde

RNS Institute of Technology, Bangalore – 560098

akshay.krishna154@gmail.com, chetanahegde@ieee.org

Abstract – **This is the age of the internet where the amount of data being generated is so huge that man alone is not able to process through the data. Many machine learning techniques hence have been discovered for this purpose. In this paper, we are trying to predict the sales of a retail store using different machine learning techniques and trying to determine the best algorithm suited to our particular problem statement. We have implemented normal regression techniques and as well as boosting techniques in our approach and have found that the boosting algorithms have better results than the regular regression algorithms.**

Keywords— **AdaBoost, Boosting Techniques, ElasticNet, Gradient Tree Boosting, Lasso Regression, Multiple Regression, Polynomial Regression, Ridge Regression, Supervised learning, XGBoost.**

I. Introduction

Sales forecasting can be defined as the prediction of upcoming sales based on the past sales occurred. Sales forecasting is of paramount importance for companies which are entering new markets or are adding new services, products or which are experiencing high growth. The main reason a company does a forecast is to balance marketing resources and sales against supply capacity planning.

Forecasting can help in answering some expository queries like “Do we have the right mix of price, promotion, and marketing in place to drive demand?”, “Do we have enough salespeople to get the volume of orders we have budgeted?”, “Do we have the essential demand-side resources in place?” and for these reasons, many of the companies allocate significant financial and human resources to perform this task genuinely, which requires large investment [3]. Manufactures organizations and business houses require an accurate and reliable forecast of sales data so that they don’t suffer from losses due to wrong or inaccurate prediction by the model. Companies mainly use sales forecasting to determine two things. First, to determine the current demand level of the service or product in the market. Second, to determine the future demand for a company’s services or products.

Forecasting can be used to predict sales revenue at product level, or at an individual business level, or at company level. In this paper we have concentrated on product level sales forecasting. Future sales plan aids in optimal utilization of facility, scheduling, conveyance and effective control of inventory. These, in turn, result in enhancement of clients’ satisfaction and also decrease in production cost. In the recent past, many investigations addressing the problem of sales forecasting have been reported [1]. Sales forecasts affect a company’s marketing plan directly. The marketing department is responsible for how clients and their customers interprets its services and products and compare it against its competitors and use the sales forecast to assess how marketing spending can increase sales and channel demand. It is important to develop effective sales forecasting models in order to generate accurate and robust forecasting results. In the business and economic environment, it is very important to accurately predict various kinds of economic variables such as Past Economic Performance, Current Global Conditions, Current Industry Conditions, Rate of Inflation, Internal Organizational Changes, Marketing Efforts, Seasonal Demands, etc. to develop proper strategies. On the contrary, inaccurate forecasts may lead to inventory shortage, unsatisfied customer demands, and product backlogs [2]. Due to these reasons, outmost importance is given to develop productive models in order to generate robust and accurate results.

In this paper we will be considering a variety of forecasting methods such as Multiple Regression, Polynomial Regression, Ridge Regression, Lasso Regression etc. along with various boosting algorithm like AdaBoost, Gradient Tree Boosting so as to get the maximum accuracy.

Multiple Regression is a statistical tool used to predict the output which is dependent on several other independent predictor or variable. It combines multiple factors to access how and to what extent they affect a certain outcome [4]. Polynomial Regression is an extension of simple linear regression. Where the model finds a non linear relationship between the independent variable x and the dependent variable y [5]. Here the model usually fits the variable y as the n^{th} degree of variable x. Ridge Regression is a way to create a model when the predictor variable has multi-colinearity [6]. Lasso is a regression method that performs both regularization and variable selection in order to enhance the prediction accuracy [7].

The AdaBoost is a boosting algorithm which is mainly used for improving the performance of the models. It utilizes the output of the other weak learners and combines the outputs of those algorithms into a weighted sum and finally arriving to the output. AdaBoost can significantly improve the learning accuracy no matter whether applied to manual data or real data [8]. The elastic net method overcomes the limitations of the Lasso. Gradient Tree Boosting is also a boosting algorithm which creates a model in the form of a group of weak predictors.

In this paper a wide range of forecasting methods are discussed because the combination of multiple forecasts can be used to increase the forecast accuracy.

II. RELATED WORK

Lots of work in the field of sale forecasting using machine learning has been proposed to date. A brief survey of the related work in the area of sales forecasting is presented in this section.

Many statistical methods such as regression, (ARIMA) Auto-Regressive Integrated Moving Average, (ARMA) Auto-Regressive Moving Average, have been used for creating multiple sales forecasting paradigms. But, sales forecasting is a sophisticated problem and is influenced by external as well as internal factors and there are two major drawbacks to the statistical approach as told by A. S. Weigend et al.,[9]. A hybrid seasonal quantile regression approach and (ARIMA) Auto-Regressive Integrated Moving Average approach for daily food sales forecasting were proposed by N. S. Arunraj et al.,[10] and also found that the performance of the individual model was comparatively lower than the hybrid model.

D. Fantazzini et al.,[11] proposed new multivariate models which is used in forecasting the car sales in Germany using Google data. In his paper, he stated that long-term forecast is of at most value for the car industry due to the lengthy period of time required for production and development process. X. Yua et al.,[12] used Support Vector Regression for magazine sales and newspaper forecasting. Support Vector Regression was used because it overcame the over-fitting problem and it also attained minimum structural risk rather than minimum empirical risk. E. Hadavandi et al.,[13] used an integration of Genetic Fuzzy Systems (GFS) and data clustering for the sales forecasting of the printed circuit board. In their paper by using K-means clustering created K clusters of all the records of the data. Then, all the clusters were fed into independent Genetic Fuzzy Systems (GFS) with the ability of database tuning and rule-based extraction.

Recognized work in the field of sale forecasting was carried out by P.A. Castillo et al.,[14], They performed sales forecasting on new published books in an editorial business management environment by applying computational methods. (ANN) Artificial Neural Networks are also used in the sales forecasting field. Fuzzy Neural Networks have been introduced with an objective to improve the prediction performance and also Radial Basis Function Neural Network (RBFN) is assumed to have a great potential for the prediction of sales [15, 16, 17].

The literature in this field shows that not much work has been done in swarm intelligence technique in effectively training the prediction models [18, 19]. The Genetic Algorithm (GA) is a potential candidate for training the ANN models. There are a lot many works in this field which has helped the organization to predict the future profit what they can make by investing at the proper place at right time. This paper is another contribution to this field.

III. ARCHITECTURE AND MODELING

Having the sales data of the retail store, the proposed work suggests the following various steps for predicting the sales of different categories available. The architectural diagram for the proposed algorithm is shown in Figure 1. The various steps involved are explained hereunder.

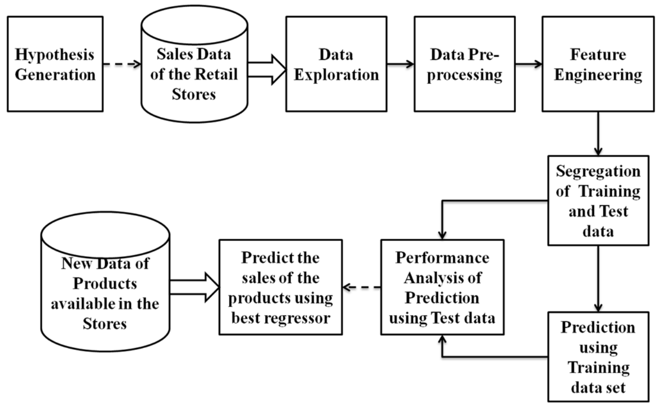


Figure 1.Architectural Diagram

A. *Hypothesis Generation*

This step is of primal importance in the process of data analysis. In this step various hypotheses are generated by analyzing the problem statement. The hypotheses are generated in such a way that they favor the needed outcome. Let us consider the problem statement-"The dataset which has been collected is of 2013 sales data for 1559 products across 10 stores in different cities. The aim is to build a predictive model and find out the sales of each product at a particular store".

So the major idea in this step is to identify the properties of a product and the stores, which can have an impact on sales. Since the forecasting is done on the basis of the products and stores, we can categorize the hypotheses as "Product Level Hypotheses" and "Store Level Hypotheses".

Few of the Product level hypotheses which can have an impact on sales are- Brand, Packaging, Display Area, Utility, Promotional Offers, Visibility in the Store, Advertising, etc, and the Store level hypotheses which can have impact are- City Type, Population Density, Store Capacity, Competitors, Location, Customer Behavior, Marketing, Ambiance, etc. For example, branded products have higher sales compared to other products because; the customer's trust in the brand is comparatively higher which leads to an increase in sales of that particular brand. Similarly, if we consider the stores being well preserved and are handled by humble and polite employees, they are expected to have higher sales because the walk-ins in the store would be higher. Therefore, we have come up with 15 hypotheses which give us a better understanding of the problem.

B. Data Exploration

When we consider a business problem, we try to achieve more accuracy by changing and implementing different models. But, after a certain point, we notice that we will be struggling to improve the accuracy of the model. To overcome such type of problems data exploration comes into the picture.

The first step in Data exploration is to look into the dataset and to discover the information regarding the available and the hypothesized data. The dataset under consideration has the following features as the variables as shown in Table 1. We can see that there are 6 features which are hypothesized and present in the dataset, 3 features present in the dataset but not hypothesized and 9 features hypothesized and not found in the data. This can be best depicted in the diagram as shown in Figure 2.

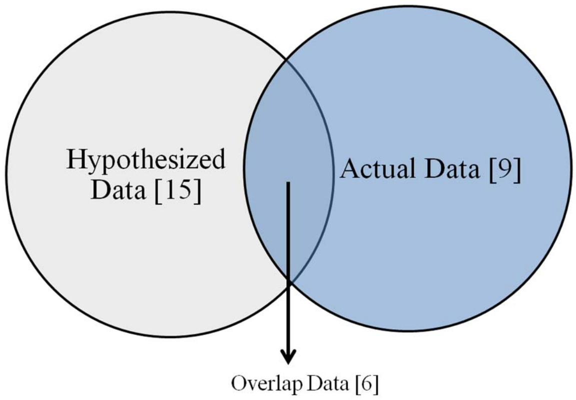


Figure 2.Venn Diagram

The dataset under consideration has some values missing in the columns "Outlet_Size" and "Item_Weight". The missing values of the data will be imputed in the data pre-processing section. The variables which are present in our dataset can be grouped as categorical variables and numerical variables. The general description of the numerical variables is given in Table 2. From the table, we can have two main observations.

1. Item_Visibility has the minimum value of 0. This value creates a contradiction because the value 0 indicates the product not being visible but being sold. Therefore, the visibility should be non-zero.
2. Outlet_Establishment_Year values vary from 1985 to 2009. These values here are converted to represent how old a particular store is in the pre-processing section.

By considering the categorical variables in the dataset we have 1559 unique products and 10 unique outlets. Exploring the dataset using the frequency of different categories, we have come to the following conclusions.

1. In the category Item_Fat_Content some of the "Low Fat" values are coded as "LF" and "low fat", also few of the "Regular" is coded as "regular".
2. In the Item_Type category, there are 16 unique categories present. These items vary from food to drinks, meat to canned food, household items to medicines, etc. Indicating that each store has a variety of products being sold.

Table 1.Data Description

Variable	Description
Item_Identifier	Product ID
Item_Weight	Weight of Product
Item_Fat_Content	Fat content in Product
Item_Visibility	% of the display area occupied
Item_Type	Category of the product
Item_MRP	Cost of the product
Outlet_Identifier	Store ID
Outlet_Establishment_Year	Year of establishment
Outlet_Size	Ground area covered by the store
Outlet_Location_Type	Type of city store located
Outlet_Type	Type of the store
Item_Outlet_Sales	Sales of product in a particular store

C. Data Pre-processing

This step typically imputes the missing values and handles the outliers present in the dataset. Missing values are found in Item_Weight and Outlet_Size columns of the dataset. Item_Weight is the numerical variable and Outlet_Size is the categorical variable.

Item_Weight is imputed by taking the average weight of that particular item in the dataset and replacing it with the missing values. Outlet_Size being a categorical variable, we cannot find the average but we consider the mode strategy to impute the missing values. Therefore, the missing values in the outlet size are found out by taking the mode of the sizes based on the outlet type.

D. Feature Engineering

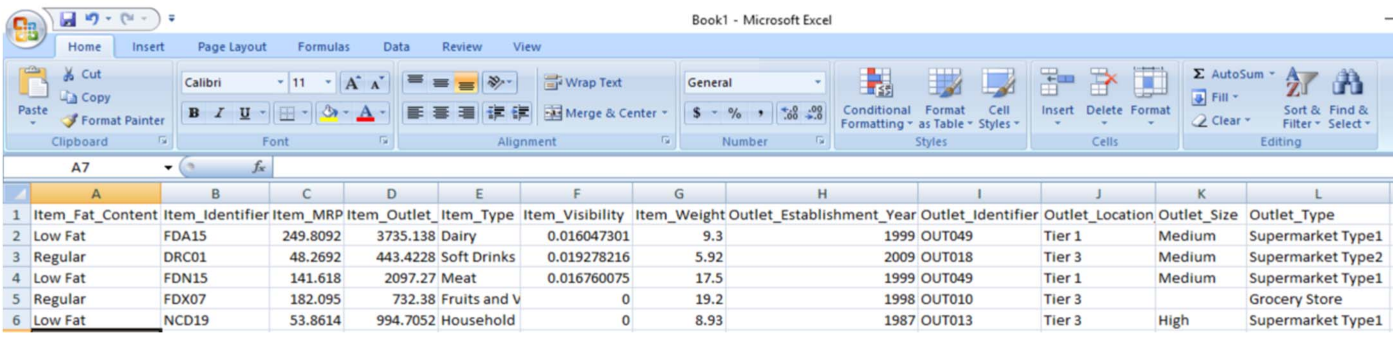
During the data exploration step, we identified a few of the nuances in the data. This step deals with those nuances and also is used for creating new variables out of existing variables so that our data will be ready to perform the analysis.

In the data exploration section, we saw that Item_Visibility has the minimum value of 0 which didn't make sense. To rectify this mistake we have considered this as a missing value and we have imputed it with the mean visibility of that product. In the hypothesis generation, we considered visibility as one of the store level hypotheses, which indicates higher the visibility of the product in the store, higher the sales. But, if we look at the visibility of a product in a particular store compared with the mean visibility of that product across all stores, we get the information telling how much importance was given to that product in that particular store compared to other stores. Therefore, we create a new variable called Item_Visibility_MeanRatio which hold these values.

We also saw in the data exploration section that Item_Type has 16 unique categories; if these categories are grouped under some common property they might be useful for performing the analysis. This common property can be found in Item_Identifier variable, which holds the Unique ID of each item. The value of this variable either starts with FD, DR or

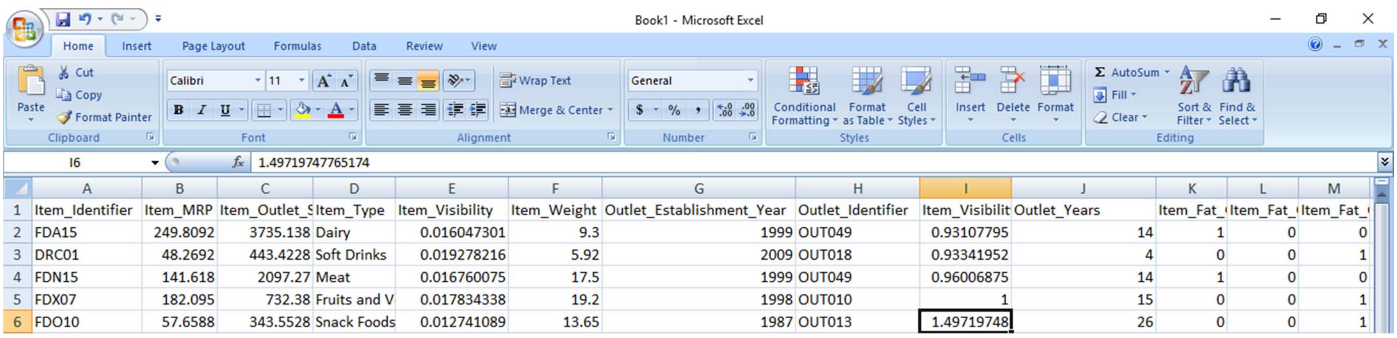
Table 2.Numerical Variable Description

	Item_MRP	Item_Visibility	Item_Weight	Outlet_Establishment_Year
count	14204.00	14204.00	11765.00	14204.00
mean	141.00	0.0659	12.79	1997.83
std	62.08	0.0514	4.65	8.37
min	31.29	0.0000	4.55	1985.00
25%	94.01	0.0270	8.71	1987.00
50%	142.24	0.0540	12.60	1999.00
75%	185.85	0.0940	16.75	2004.00
max	266.88	0.3283	21.35	2009.00



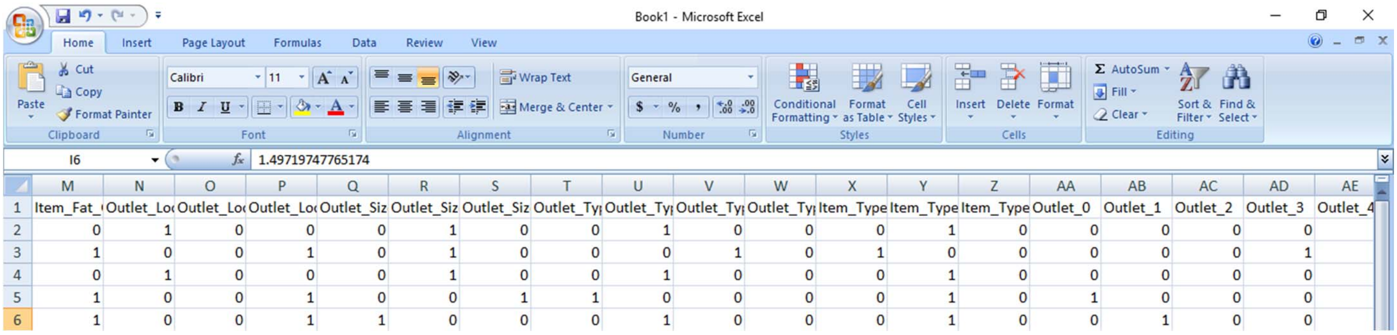
	A	B	C	D	E	F	G	H	I	J	K	L
1	Item_Fat_Content	Item_Identifier	Item_MRP	Item_Outlet_	Item_Type	Item_Visibility	Item_Weight	Outlet_Establishment_Year	Outlet_Identifier	Outlet_Location	Outlet_Size	Outlet_Type
2	Low Fat	FDA15	249.8092	3735.138	Dairy	0.016047301	9.3	1999	OUT049	Tier 1	Medium	Supermarket Type1
3	Regular	DRC01	48.2692	443.4228	Soft Drinks	0.019278216	5.92	2009	OUT018	Tier 3	Medium	Supermarket Type2
4	Low Fat	FDN15	141.618	2097.27	Meat	0.016760075	17.5	1999	OUT049	Tier 1	Medium	Supermarket Type1
5	Regular	FDX07	182.095	732.38	Fruits and V	0	19.2	1998	OUT010	Tier 3		Grocery Store
6	Low Fat	NCD19	53.8614	994.7052	Household	0	8.93	1987	OUT013	Tier 3	High	Supermarket Type1

Figure 3.Sample Data before Feature Engineering



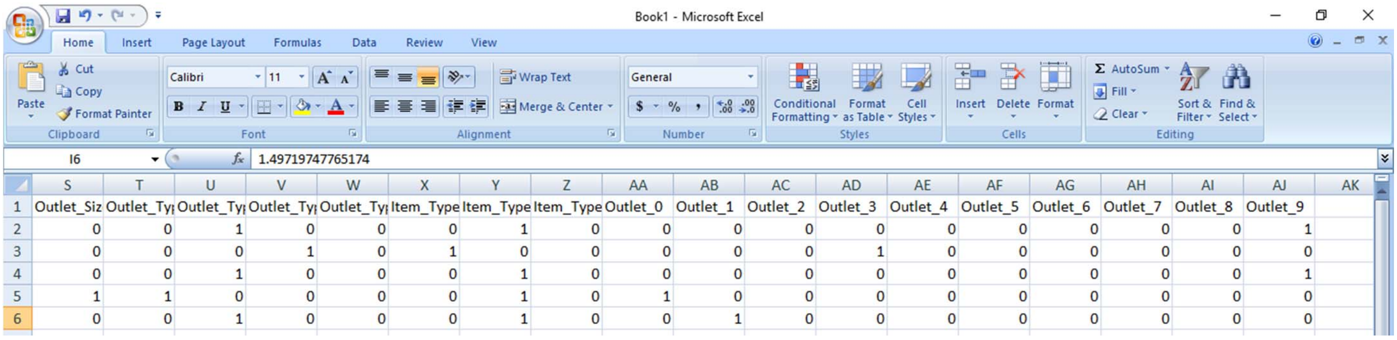
	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Item_Identifier	Item_MRP	Item_Outlet_S	Item_Type	Item_Visibility	Item_Weight	Outlet_Establishment_Year	Outlet_Identifier	Item_Visibilit	Outlet_Years	Item_Fat_(	Item_Fat_(	Item_Fat_(
2	FDA15	249.8092	3735.138	Dairy	0.016047301	9.3	1999	OUT049	0.93107795	14	1	0	0
3	DRC01	48.2692	443.4228	Soft Drinks	0.019278216	5.92	2009	OUT018	0.93341952	4	0	0	1
4	FDN15	141.618	2097.27	Meat	0.016760075	17.5	1999	OUT049	0.96006875	14	1	0	0
5	FDX07	182.095	732.38	Fruits and V	0.017834338	19.2	1998	OUT010	1	15	0	0	1
6	FDO10	57.6588	343.5528	Snack Foods	0.012741089	13.65	1987	OUT013	1.49719748	26	0	0	1

Figure 4a.Sample Data after Feature Engineering



	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE
1	Item_Fat_	Outlet_Lo	Outlet_Lo	Outlet_Lo	Outlet_Siz	Outlet_Siz	Outlet_Siz	Outlet_Ty	Outlet_Ty	Outlet_Ty	Outlet_Ty	Item_Type	Item_Type	Item_Type	Outlet_0	Outlet_1	Outlet_2	Outlet_3	Outlet_4
2	0	1	0	0	0	1	0	0	1	0	0	0	1	0	0	0	0	0	
3	1	0	0	1	0	1	0	0	0	1	0	1	0	0	0	0	0	1	
4	0	1	0	0	0	1	0	0	1	0	0	0	1	0	0	0	0	0	
5	1	0	0	1	0	0	1	1	0	0	0	0	1	0	1	0	0	0	
6	1	0	0	1	1	0	0	0	1	0	0	0	1	0	0	1	0	0	

Figure 4b.Sample Data after Feature Engineering



	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK
1	Outlet_Siz	Outlet_Ty	Outlet_Ty	Outlet_Ty	Outlet_Ty	Item_Type	Item_Type	Item_Type	Outlet_0	Outlet_1	Outlet_2	Outlet_3	Outlet_4	Outlet_5	Outlet_6	Outlet_7	Outlet_8	Outlet_9	
2	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	
3	0	0	0	1	0	1	0	0	0	0	0	1	0	0	0	0	0	0	
4	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	
5	1	1	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	
6	0	0	1	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	

Figure 4c.Sample Data after Feature Engineering

NC, which can be considered Food, Drinks, and Non-Consumables. Therefore, by using these two columns Item_Type and Item_Identifier we create a new column called as Item_Type_Combined which categorizes each product as Food, Drinks or Non-Consumables.

The variable Outlet_Establishment_Year holds the year when the store was established. We create a new variable out of this which indicates the number of years the store was in operation. This resulted in values ranging from 4 to 28 years.

In Item_Fat_Content variable we saw that there was a difference in the representation of categories. We correct this typo by replacing LF and low fat with Low Fat, reg with Regular. But, after changing the categories we see that the Non-Consumables are also having certain fat content as Low Fat or Regular. This is also modified in such a way that the Non-Consumables are categorized as Non-Edible in Item_Fat_Content.

As we know that machine learning algorithms take numerical values as input, we encode the categorical variables into numerical variables by using LabelEncoder and OneHotEncoder of sklearn's pre-processing module.

A sample data, before and after Feature Engineering is depicted in Figure 3 and Figure 4 respectively.

E. Segregation of Training and Testing Data

A major step in machine learning is to feed some amount of data into the algorithm and to train it to understand the pattern of data. Once the algorithm learns the pattern, one has to feed another dataset to check the level of understanding done by the algorithm. It is a practice to divide the available data into two subsets in the ratio of 4:1 for the purpose of training and testing. But, most of the times we may need to readjust this ratio for getting better performance. Moreover, the ratio may vary from one regression algorithm to the other.

F. Model Building for Prediction

After the dataset is split into training and testing sets, the training set is fed into the algorithm so that it can learn how to predict the values. Various regression algorithms like Multiple Regression, Polynomial Regression, Ridge Regression, Lasso Regression, etc. have been applied. Along with these boosting algorithms like AdaBoost, XGBoost has also been applied to the dataset for increasing the accuracy.

Multiple Regression is an extension of simple linear regression. It is used when we want to predict the value of the dependent variable based on the value of two or more independent variables. In general, multiple linear regression procedures will estimate the value of the variable based on the following equation –

$$y = a_0 + a_1x_1 + a_2x_2 + \cdots + a_nx_n$$

Here, *a* represents the coefficients.

Polynomial Regression is the form of regression analysis, in which the relationship between the independent variable and the dependent variable is modeled as an n^{th} degree polynomial. This is a non-linear regression model and polynomial regression is the extension of simple linear regression with the order being equal to 1. The polynomial regression procedures will estimate the value of the variable based on the following equation –

$$y = a_0 + a_1x + a_2x^2 + a_3x^3 + \cdots + a_nx^n$$

LASSO regression in machine learning stands for Least Absolute Shrinkage and Selection Operator. It performs both variable selection and regularization in order to enhance the

Table 3.Simulation Results of Various Algorithms

Algorithms	Training to Testing Ratio					
	80:20		75:25		70:30	
	RMSE	R^2	RMSE	R^2	RMSE	R^2
Linear Regression	1129.49	0.56	1146.35	0.56	1152.45	0.56
Polynomial Regression	1118.65	0.57	1133.43	0.57	1130.57	0.58
Lasso Regression	1129.27	0.56	1146.34	0.56	1152.70	0.56
Ridge Regression	1129.48	0.56	1146.36	0.56	1152.47	0.56
AdaBoost	1290.42	0.43	1203.23	0.51	1350.72	0.40
GradientBoost	1088.64	0.59	1103.9	0.59	1107.71	0.59

prediction accuracy and interpretability. The major objective of LASSO regression is to solve –

$$\min_{\beta_0,\beta}\left\{\frac{1}{N}(y - \beta_0 - x_i{}^T\beta)^2\right\}$$

Where β is subjected to the following constraint and t is the free parameter which determines the amount of regularization.

$$\sum_{j=1}^{p}|\beta_j| \leq t$$

AdaBoost stands for Adaptive Boosting. This algorithm is mainly used in improving the performance of the model created. In this algorithm the final output is obtained by taking the weighted sum of the outputs of the weak learners. In this algorithm the output is manipulated in such a way that, the algorithm favors the instances which were mispredicted by the weak learners. Hence making the algorithm adaptive.

AdaBoost refers to a particular method of training a boosted model. Where the boosted model is in the form –

$$F_T(x) = \sum_{t=1}^{T} f_t(x)$$

In the above formula f_t represents a weak learner and x represents an object which is as an input to the weak learner. The weak learner f_t returns the predicted value of the object. $h(x_i)$ is the hypothesis generated by each weak learner for every data sample in the training set. A coefficient α_t is assigned for every weak learner $h(x_i)$, which is selected at iteration t such that the sum of training error E_t is minimized.

$$E_t = \sum_i E|F_{t-1}(x_i) + \alpha_t h(x_i)|$$

G. Performance Analysis of Prediction

The performance of any regression algorithm is computed by feeding in the test data in the training model. This procedure gives us the idea how well a model has learned the pattern present in the dataset and is able to predict the values of the new data. The performance of any given regression model can be computed using the Root Mean Square Error (RMSE) or the R^2 error.

RMSE is given by the following formula –

$$RMSE = \sqrt{\frac{\sum_i^n (y_i - a_i)^2}{n}}$$

Where y_{ou} are the predicted value and a_i is the value given in the dataset.

The R^2 error is given by the following formula –

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}}$$

Where SS_{res} is called a residual sum of squares and is given by –

$$SS_{res} = \sum_i (y_i - a_i)^2$$

and, SS_{tot} is called the total sum of squares which is given by –

$$SS_{tot} = \sum_i (a_i - \bar{a})^2$$

Where $\bar{a}$ is the mean of the observed data which is given by –

$$\bar{a} = \frac{1}{n}\sum_{i=1}^{n} a_i$$

IV. IMPLEMENTATION AND PERFORMANCE ANALYSIS.

The proposed algorithm is implemented using Python 3.6 and sklearn 0.19.1. The sales forecast dataset form Kaggle [20] is been pre-processed with the steps mentioned above to predict the sales. The dataset consists of 8523 entries and has 12 features. Different models are created and simulated based on this dataset containing 8523 entries. The simulation result is shown in Table 3.

V. CONCLUSIONS

In this paper, we compare the performance of different algorithms on store sales dataset and analyze the algorithm with the best performance. Here we have observed that the AdaBoost algorithm has the highest RMSE value of 1350.72 and the algorithm with the least RMSE value is GradientBoost having 1088.64. The algorithm in terms of highest R^2 is GradientBoost with the value of 0.59 and the algorithm with the least R^2 is AdaBoost with the value of 0.40. Hence, by the obtained results we can see that the GradientBoost algorithm is the best predictor for the considered dataset having the least RMSE value of 1088.64 and the highest R^2 value of 0.59. We can also conclude that without proper hyperparameter tuning the AdaBoost algorithm won't be able to perform as expected and the performance deteriorates.

VI. REFERENCES

[1] Ching Wu Chu and Guoqiang Peter Zhang, “A comparative study of linear and nonlinear models for aggregate retails sales forecasting”, Int. Journal Production Economics, vol. 86, pp. 217-231, 2003.

[2] Giuseppe Nunnari, Valeria Nunnari, “Forecasting Monthly Sales Retail Time Series: A Case Study”, Proc. of IEEE Conf. on Business Informatics (CBI), July 2017.

[3] https://halobi.com/blog/sales-forecasting-five-uses/. [Accessed: Oct. 3, 2018]

[4] Zone-Ching Lin, Wen-Jang Wu, “Multiple LinearRegression Analysis of the Overlay Accuracy Model Zone”, IEEE Trans. on Semiconductor Manufacturing, vol. 12, no. 2, pp. 229 – 237, May 1999.

[5] O. Ajao Isaac, A. Abdullahi Adedeji, I. Raji Ismail, “Polynomial Regression Model of Making Cost Prediction In Mixed Cost Analysis”, Int. Journal on Mathematical Theory and Modeling, vol. 2, no. 2, pp. 14 – 23, 2012.

[6] C. Saunders, A. Gammerman and V. Vovk, “Ridge Regression Learning Algorithm in Dual Variables”, Proc. of Int. Conf. on Machine Learning, pp. 515 – 521, July 1998.IEEE TRANSACTIONS ON INFORMATION THEORY, VOL. 56, NO. 7, JULY 2010 3561.

[7] ”Robust Regression and Lasso”. Huan Xu, Constantine Caramanis, Member, IEEE, and Shie Mannor, Senior Member, IEEE. 2015 International Conference on Industrial Informatics-Computing Technology, Intelligent Technology, Industrial Information Integration.”An improved Adaboost algorithm based on uncertain functions”.Shu Xinqing School of Automation Wuhan University of Technology.Wuhan, China Wang Pan School of the Automation Wuhan University of Technology Wuhan, China.

[8] Xinqing Shu, Pan Wang, “An Improved Adaboost Algorithm based on Uncertain Functions”, Proc. of Int. Conf. on Industrial Informatics – Computing Technology, Intelligent Technology, Industrial Information Integration, Dec. 2015.

[9] A. S. Weigend and N. A. Gershenfeld, “Time series prediction: Forecasting the future and understanding the past”, Addison-Wesley, 1994.

[10] N. S. Arunraj, D. Ahrens, A hybrid seasonal autoregressive integrated moving average and quantile regression for daily food sales forecasting, Int. J. Production Economics 170 (2015) 321-335.

[11] D. Fantazzini, Z. Toktamysova, Forecasting German car sales using Google data and multivariate models, Int. J. Production Economics 170 (2015) 97-135.

[12] X. Yua, Z. Qi, Y. Zhao, Support Vector Regression for Newspaper/Magazine Sales Forecasting, Procedia Computer Science 17 (2013) 1055–1062.

[13] E. Hadavandi, H. Shavandi, A. Ghanbari, An improved sales forecasting approach by the integration of genetic fuzzy systems and data clustering: a Case study of the printed circuit board, Expert Systems with Applications 38 (2011) 9392–9399.

[14] P. A. Castillo, A. Mora, H. Faris, J.J. Merelo, P. Garcia-Sanchez, A.J. Fernandez-Ares, P. De las Cuevas, M.I. Garcia-Arenas, Applying computational intelligence methods for predicting the sales of newly published books in a real editorial business management environment, Knowledge-Based Systems 115 (2017) 133-151.

[15] R. Majhi, G. Panda and G. Sahoo, “Development and performance evaluation of FLANN based model for forecasting of stock markets”.Expert Systems with Applications, vol. 36, issue 3, part 2, pp. 6800-6808, April 2009.

[16] Pei Chann Chang and Yen-Wen Wang, “Fuzzy Delphi and back propagation model for sales forecasting in PCB industry”, Expert systems with applications, vol. 30,pp. 715-726, 2006.

[17] R. J. Kuo, Tung Lai HU and Zhen Yao Chen “application of radial basis function neural networks for sales forecasting”, Proc. of Int. Asian Conference on Informatics in control, automation, and robotics, pp. 325- 328, 2009.

[18] R. Majhi, G. Panda, G. Sahoo, and A. Panda, “On the development of Improved Adaptive Models for Efficient Prediction of Stock Indices using Clonal-PSO (CPSO) and PSO Techniques”, International Journal of Business Forecasting and Market Intelligence, vol. 1, no. 1, pp.50-67, 2008.

[19] R. Majhi, G. Panda, B. Majhi and G. Sahoo, “Efficient prediction of stock market indices using adaptive bacterial foraging optimization (ABFO) and BFO based techniques”, Expert Systems with Applications, Elsevier, vol. 36, issue 6, pp. 10097-10104, August 2009.

[20] https://www.kaggle.com/brijbhushannanda1979/bigmart-sales-data. [Accessed: Jun. 28, 2018].