

Analysis on the Prediction of Sales using Various Machine Learning Testing Algorithms

Dr. J.Sasi Kiran¹

Professor in CSE & Dean ,
Lords Institute of Engineering and
Technology(A), Hyderabad,
Telangana
sasikiranjangala@gmail.com

Dr.P S V Srinivasa Rao²

Professor in School of Engineering (CSE),
Malla Reddy University, Hyderabad,
Telangana
drpsvsrao@mallareddyuniversity.ac.in

Dr.P V R D Prasada Rao³

Professor, Dept of Computer Science
and Engineering
Koneru Lakshmaiah Education
Foundation, Vaddeswaram, AP, India
pvrdrasad@gmail.com

Dr. B.Sankara Babu⁴

Professor, Dept of CSE,
Gokaraju Rangaraju Institute of
Engineering and
Technology(Autonomous),
Hyderabad, Telangana
bsankarababu81@gmail.com

N.Divya⁵

Assistat Professor, Dept of CSE,
G Narayanamma Institute Of Technology
and Science, Hyderabad, Telangana
n.divya38@gmail.com

Abstract - The amount of data created is so large that we are unable to process it all on our own. For this goal, a variety of machine learning approaches have been developed. The objective is to perform prediction of sales for super markets using different regression and boosting approaches, and figuring out which algorithm is suitable for the job. Traditional sales and marketing goals do not assist organizations keep up with the pace of a competitive market because they are carried out without any knowledge about customers' purchasing behaviors. The advancements in machine learning have had a tremendous impact on the world of sales and marketing. Item weight, item fat content, item visibility, item kind, item MRP, outlet establishment year, outlet size, and outlet location type are all included in the prediction.

Key Words: Sales, linear regression, Random forest, XG Booster, Bayesian Regression

1. Introduction

Sales prediction is useful in a variety of industries since it aids in the improvement of a company's sales by allowing for future planning. Sales forecasting is the practice of analyzing a company's sales statistics from previous years to forecast the company's future short- and long-term sales performance.

This is one of the pillars of proper financial planning, Sales predication is a critical prerequisites for venture arranging and adju st choice making, permitting companiesto superior arrange their trade exercises. Accurate prediction of sales helps company to form their future plans and strategies. Sales forecasting is critical in guiding the

warehousing department's sales and marketing as they arrange warehouse locations [1]. Sales data, on the other hand, can better reflect future sales trends. Predicting sales is a common use of machine learning. Prediction of sales can be used to determine the impact of new product and the future plans of the product. Recent surveys showed machine learning had a great impact on sales prediction [2].

In the present world, due to rapid increase in usage of machine learning technology, many real world problems are solved using machine learning. Predicting the sales of the company is also a real world problem which helps company to boost their profits. So, the solution to this problem which is predicting sales of a company is done using machine learning algorithms [5]. There are many algorithms in machine learning which helps in predicting sales like linear regression, xg booster regression, and Random forest regression. We tried to get better accuracy and better sales prediction using these algorithms.Targeting the customers is the main focus of business sectors. Profitability is critical for any business.

Sales forecasting requires analysing data from a range of sources, such as market trends, economic indicators, and consumer behaviour.customer feedback and their responses and various other factors. Sales analysis helps company to decide whether to invest in a product or not [6]. This sales forecasting can be used for many purposes, such as predicting the future demand of the products or

services. We can also predict how the product will do in coming future. Machine learning is the improvement of a machine that's able to memorize and adjust without taking after express informational by utilizing calculations and measurable models to analyse and draw inferences from designs in information. So, we train algorithm using previous data to predict output of the new data. So, in this way we train machine to predict the sales of company by training it with previous data of sales of a company [7]. Linear regression is a machine learning approach that aids in the resolution of machine learning issues. Linear regression is a mathematical technique for predicting the value of a dependent variable (y) based on the value of an independent variable (x). A linear relationship between x(input) and y(output) is discovered as a result of this regression technique (output).

2. Literature Review

Kohli, Shreya, Gracia Tabitha Godwin, and Siddhaling Urolagin. "Sales Prediction Using Linear and KNN Regression." In *Advances in Machine Learning and Computational Intelligence*, pp.321-329. Springer, Singapore, 2021 [1].

The main concept of sellers is supply and for consumers it is demand. Predicting what consumers want accurately will help organizations improve sales. Deals Expectation is basically based on fore seeing deals of distinctive company outlets so that they can alter their future plans of the company based on the anticipated deals.

Behera, Gopal, and Neeta Nain. [3] "A comparative consider of huge shop deals forecast." In *Worldwide Conference on Computer Vision and Picture Preparing*, pp. 421-432. Springer, Singapore, 2019.

Deals expectation is concerned with evaluating future deals of companies such as general stores, basic supplies, eateries, pastry shops and baked goods.

Deals forecast makes a difference the company to play down the merchandise whose deals is to diminish and increment the stock of

The product that are attending to increment which is able lead to extend in deals of the company

Giering, Michael. [4] "Retail sales prediction and item recommendations using customer demographics at store level." *ACM SIGKDD Explorations Newsletter* 10, no. 2 (2008): 8489. This paper tells about retail sales prediction and a detailed review of product recommendation

system for retail stores. The relative importance of consumer demographic characteristics is estimated and integrated into the model to calculate the sales of each client type correctly. This data includes daily sales figures for 600 products at the store level and for a variety of non-overlapping customer groups. This information is beneficial to increasing sales and enhancing sales. A recommender system was built using a rapid online thin Singular Value Decomposition. When compared to a single aggregate model built for the entire dataset, it is discovered that clustering data across client types and demographics produces better results. The system implementation specifics are presented in this section, as well as practical concerns in real-world problems. Initial data from establishments that were evaluated for a year show an increase in sales. To confirm and demonstrate the scalability of this technique, a quick summary of how the primary procedures covered here were expanded to a considerably bigger data set is provided.

Bin, Okmyung. [11] "A prediction comparison of housing sales prices by parametric versus semi-parametric regressions." *Journal of Housing Economics* 13, no. 1 (2004): 68-84.

This consider gauges a hedonic cost work employing a semi-parametric relapse and compares the cost forecast execution with customary parametric models. This ponder utilizes a huge information set speaking to 2595 single-family private domestic deals between July 2000 and June 2002 from Pitt District, North Carolina. Information from Geographic Data Frameworks (GIS) are consolidated to account for location traits of the houses.

The comes about appear that the semi-parametric relapse out flanks the parametric partners in both in-sample and out-of-sample cost forecasts, showing that the semi-parametric demonstrate can be valuable for estimation and expectation of lodging deals costs.

Gopalakrishnan, T., Choudhary, R., & Prasad, S. (2018, December). Forecast of Deals Esteem in Online shopping utilizing Direct Relapse [8]. In *2018 4th Worldwide Conference on Computing Communication and Mechanization (ICCCA)* (pp. 1-6). IEEE.

The purpose of this paper is to examine a large

superstore's deals and forecast their future deals in order to differentiate them, expand their benefits, and make their brand truly outstanding and competitive according to showcase patterns while also ensuring client satisfaction. The Straight Relapse Calculation, which is a prominent calculation in the field of Machine Learning, is the method used for deal expectation. The deals data ranges from 2011 to 2013, with a prediction of data for 2014. The real-time information for the year 2014 is also taken at this stage, and the real information for the year 2014 is compared to the anticipated information to compute forecast precision. Normally, this is done in order for us to approve our results with the real ones. This, in turn, would support them in pursuing necessary actions (which would be evaluated later) for their advancement.

3. Problem Statement

To predict the sales of a company using machine learning algorithms for the better improvement of the company sales. Comparing different machine learning algorithms and understanding them to

produce better accuracy.

To predict the sales of a company which helps in boosting their sales and helps in better planning of future?

3.1 Objective

1. To understand the concept of sales of a company.
2. Sales forecasts to identify which item and product or service can be strengthened in future.
3. Sharing accurate forecasts with suppliers and partners to develop their business and also relationships.
4. To analyze methods to revenue maximization of sales.

4. Methodology

The proposed system implements the regression, classification algorithms to predict the sales. The proposed system implements boosting algorithms in addition for the sales forecasting. The proposed system may find out the best algorithm for sales prediction out of all algorithms implemented

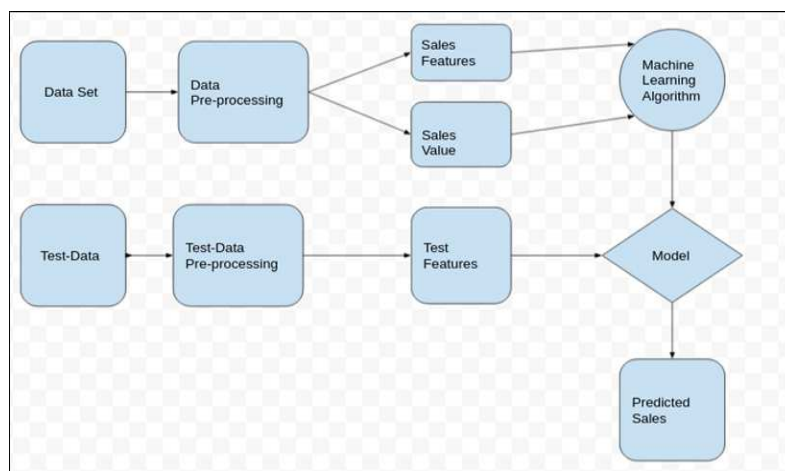


Figure 1: Model Architecture

5. Implementation

The sales prediction of a company may not be predicted as we just see the data. There is a lot of analysis that should be done to predict the sales initially, perform pre-processing on the data to understand key features which show effect on the sales. Import libraries like pandas, numpy, seaborn, matplotlib and warnings for smoother

execution. We use Anaconda jupyter to run the python model much smoother than the regular pc model Algorithms we used for better prediction are linear regression, XGBoost, random forest regression, Bayesian Regression is a popular and efficient supervised learning algorithm.

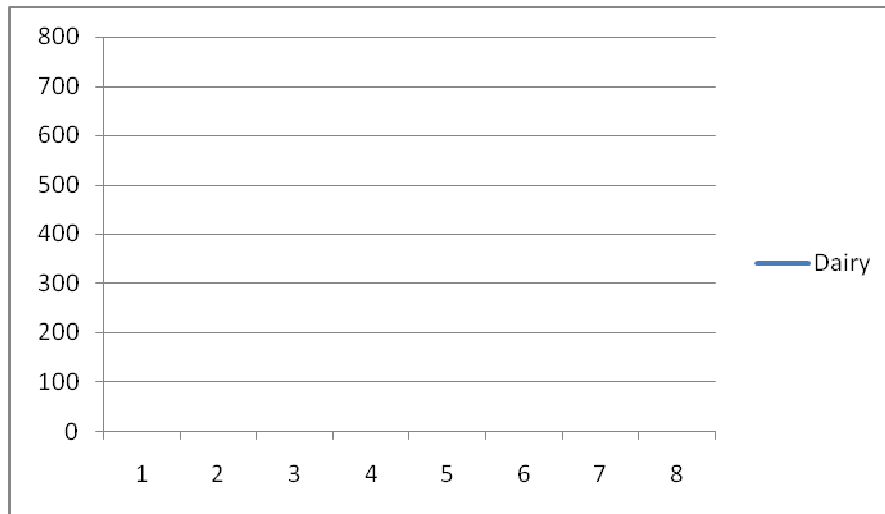


Figure 2: Item outlet sales and outline

Frail models irregular timberland could be an administered learning calculation which builds multiple decision trees and blend them together to induce a more exact and steady forecast for each class, Bayesian regression predicts membership probabilities, such as the likelihood that a given

record or data point belongs to that class. The most likely class is the one with the highest possibility. We then perform data visualization to better understand the data

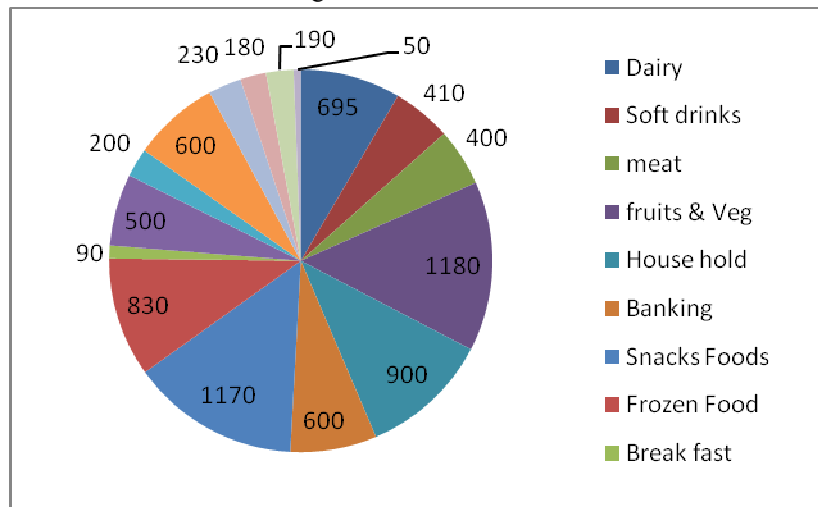


Figure 3:Item_type

Data pre-processing is done to better understand the dataset and by pre-processing we get a complete idea about the dataset like its parameters, attributes and hidden features can also be discovered using data preprocessing After complete understanding of the data set there is a need to check whether missing values are present in the data or not Missing values in data is filled by using mean and

mode Mean is used to fill the numerical missing values and mode is used to fill categorical data in the data set After filling missing values the next step is to encode categorical data Machine learning solves problems in a mathematical way using different mathematical methods.

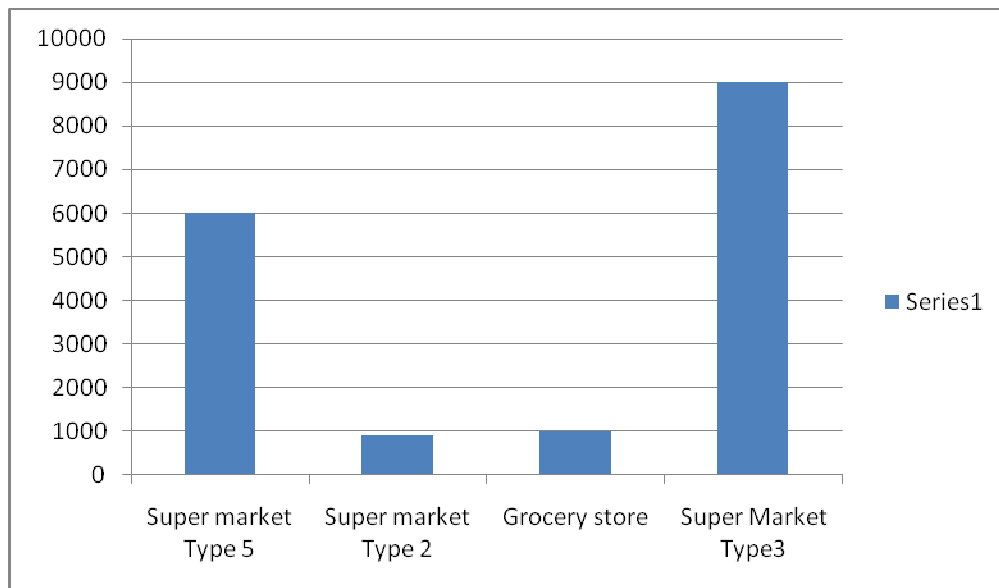


Figure 4:outlet_type

So, to solve any problem we must need more numerical data than categorical data. Label Encoding is a class imported from sklearn library to encode categorical data to numerical data for better results. Feature scaling is done on dataset to increase accuracy and to reduce errors. Feature scaling to the data is applied to standardize the independent features present in the data in a fixed range which increases accuracy and reduce errors. Splitting training and testing data. The data in the dataset is divided into two categories: training and testing. Models are trained using training data, while models are tested using testing data. Same dataset is used for training and testing for better accurate results.

6. Conclusion

Due to rapid increase in machine learning algorithms, we made a prediction model for prediction of sales and select the best model whose accuracy is high. We have done comparison between four machine learning algorithms using various attributes to predict the best model. Machine learning techniques including Linear Regression, XGBoost, and Random Forest, as well as different parameters like Root Mean Squared Error (RMSE), variance Score, and training and testing accuracies, all influence the precision of outcomes, are utilized to predict sales using the Bayesian algorithm.

References

1. Kohli, S., Godwin, G. T., & Urolagin, S. (2021). Sales Prediction Using Linear and KNN Regression. In *Advances in Machine Learning and Computational Intelligence* (pp. 321-329). Springer, Singapore.
2. Cheriyan, S., Ibrahim, S., Mohanan, S. and Treesa, S., 2018, August. Intelligent Sales Prediction Using Machine Learning Techniques. In *2018 International Conference on Computing, Electronics & Communications Engineering (ICCECE)* (pp. 53-58). IEEE.
3. Behera, Gopal, and Neeta Nain. "A comparative study of big mart sales prediction." In *International Conference on Computer Vision and Image Processing*, pp. 421-432. Springer, Singapore, 2019.
4. Giering, Michael. "Retail sales prediction and item recommendations using customer demographics at store level." *ACM SIGKDD Explorations Newsletter* 10, no. 2 (2008): 84-89.
5. <https://www.kaggle.com/devashish0507/big-mart-sales-prediction>
6. <https://github.com/aswintechguy/Machine-Learning-Projects/tree/master/Bigmart%20Sales%20Prediction%20Analysis%20-%20Regression>
7. <https://www.youtube.com/watch?v=CD58mco2XqA>
8. Gopalakrishnan, T., Choudhary, R., & Prasad, S. (2018, December). Prediction of Sales Value in Online shopping using Linear Regression. In *2018 4th International Conference on Computing Communication and Automation (ICCCA)* (pp. 1-

- 6).IEEE.
9. Ponce H, Miralles-Pechúan L, de Lourdes Martínez-Villaseñor M. Artificial hydrocarbon networks for online sales prediction. In Mexican international conference on artificial intelligence 2015 Oct 25 (pp. 498-508). Springer, Cham.
 10. Lin, Zhaojiang, Andrea Madotto, Genta Indra Winata, Zihan Liu, Yan Xu, Cong Gao, and Pascale Fung. "Learning to learn sales prediction with social media sentiment." In Proceedings of the First Workshop on Financial Technology and Natural Language Processing, pp. 47-53. 2019.
 11. Bin, Okmyung. "A prediction comparison of housing sales prices by parametric versus semi-parametric regressions." *Journal of Housing Economics* 13, no. 1 (2004): 68-84.