

DEMAND AND SALES FORECASTING USING MACHINE LEARNING

BOKKA SATYA PRASAD¹

Department of Computer Science and System Engineering

Andhra University

Visakhapatnam, India

Abstract - Every aspect of life is changing as a result of machine learning, which has also significantly impacted situations in the real world. Machine learning has revolutionary applications in every industry, including education, healthcare, engineering, sales, entertainment, and transportation, to name just a few on an endless list. The traditional method of achieving sales and marketing objectives doesn't benefit businesses anymore because it doesn't take into account how people actually buy things. Advances in machine learning have led to significant changes in the field of sales and marketing. Due to these developments, it is now much easier to determine important factors like target market demographics and forecasting sales for the upcoming years. Informing the sales team's plans for a growth in their business. The purpose of this study is to suggest a factor for forecasting Big Mart Companies' upcoming sales while taking into account their past sales. Machine learning models like Linear Regression, K-Neighbors Regressor, XGBoost Regressor, are used to conduct an extensive study on sales prediction. Data characteristics used in the forecast include item weight, fat content, visibility, kind, and MRP as well as the year the outlet was established, the size of the outlet, and the type of outlet site.

Key Words: Standard Scaler, Label Encoder, Linear Regression, K-Neighbors Regressor, XGBoost Regressor

INTRODUCTION

Sales forecasting has always been a very important topic to focus on. All vendors must now anticipate well and optimally in order to maintain the effectiveness of marketing groups. Manually performing this work would be time-consuming, which is undesirable in today's fast-paced environment and could result in grave mistakes that would result in bad management of the firm. A significant portion of the world economy is dependent on the business sectors, which are literally expected to generate enough goods in the right amounts to satisfy demand. The primary objective of business sectors is market audience targeting. It is crucial that the business has been successful in achieving its goal by utilizing a system of forecasting. In order to make predictions, it is necessary to analyze data from a variety of sources, including market trends, customer behaviour, and other elements. The companies would benefit from this analysis by having better financial resource management. The forecasting method can be used for a variety of things, such as estimating future demand for the product or service and estimating how much of the product will be sold in a specific time frame. Here, machine

learning has a lot of potential for use. In the field of machine learning, computers are able to execute some jobs better than people. They are employed to carry out specific tasks in a methodical manner and produce improved outcomes for the advancement of the modern civilization. The foundation of machine learning is mathematics, which may be used to design various paradigms that are close to the ideal output. Machine learning has been shown to be beneficial in the instance of sales forecasting. It aids in more precise forecasting of upcoming sales.

In our article, we suggested machine learning techniques for data gathered from a grocery store's prior sales. Based on a few key characteristics identified from the available raw data, the goal is to predict the sales pattern and the quantities of the products to be sold. To fully understand the data, analysis and study of the acquired data have also been done. At each crucial stage of the marketing strategy, analysis would assist business organizations in arriving at a probable decision.

1. RELATED WORK

[1] 'Walmart's Sales Data Analysis - A Big Data Analytics Perspective'

This study analyses data gathered from a retail establishment and makes predictions about future management strategies for the establishment. The impact of numerous sequences of events, like the weather, holidays, and so on, can genuinely alter the state of various departments; therefore, it also analyses these effects and considers their impact on sales.

[2] 'Applying machine learning algorithms in sales prediction'

In this various separate machine learning algorithmic techniques are used to obtain superior, optimal results, which are then further analyzed for prediction task. It used an ensemble technique, four algorithms, etc. Different strategies have been used to implement feature selection.

[3] Using machine learning, the "Sales Prediction System" By using approaches like Clustering Models and measurements for sales forecasts, the goal of this article is to obtain accurate findings for predicting future sales or demands of a corporation. The potential of the algorithmic methods is appraised and utilized appropriately in subsequent studies.

[4] Using machine learning techniques, intelligent sales prediction is described in this study examines how judgments should be made based on experimental data and the knowledge gained via data visualization.

Techniques for data mining have been used. The Gradient Boost algorithm has demonstrated the highest degree of accuracy in predicting upcoming transactions.

[5] Using customer demographics at the store level, anticipate retail sales and recommend products.

This paper describes a product recommendation system and a sales forecast system that were employed to the advantage of a set of retail outlets. Consumer demographic information has been utilized to precisely design each person's sales.

[6] "Building an intelligent sales forecast system using artificial neural networks and GAs."

In the study, deep neural network techniques are being used to predict their sales strategy for electronic components. To increase the system's effectiveness, other optimization methods are also applied, such as the Genetic Algorithm.

[7] "Bayesian learning for thousands of retailers' sales rate prediction"

In this study, it is demonstrated that numerous vendors would profit from the forecast of the single transaction's rate, demonstrating how the information gathered could be useful for the creation of a setup that would estimate a large number of outputs. The prediction makes use of neural network methodology. In order to gather insights, they have practiced Bayesian learning here.

[8] Combining Machine Learning and Data Mining for Effective User Profiling [6].

This study demonstrates how to use an automatic prototype to identify suspicious conduct. Several machine learning methodologies have been made in use for concluding this appropriate prototype. Here data mining and constructive induction techniques are merged to pull out the discrepancy found in the conducts of the owners of cell phones.

2 DATA VISUALIZATION

1.1 Heat map for determining correlation between the dataset attributes

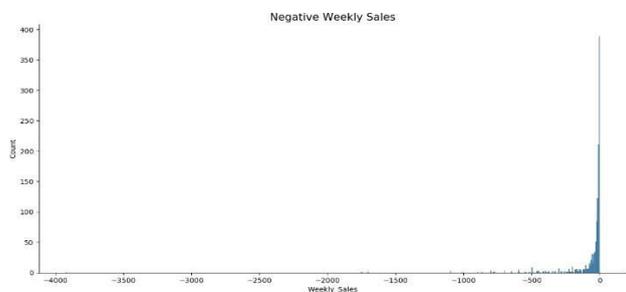


Fig-1: Negative Weekly Sales

Heat map, an element of the data visualization library called Seaborn, is a color encoded matrix which is used here to depict the correlation between target variable and the rest of the attributes.

Higher the intensity of the color of an attribute relative to the target variable, lower is the dependency of the target variable on the respective attribute. It is observed that target variable i.e., Item_Outlet_Sales is least dependent on Item_Visibility and most dependent on Item_MRP. Thus, higher the MRP of an item, lower will be the Item_Outlet_Sales.

1.2 Count Plots

The distribution of various Outlet Types i.e., Supermarket Type1, Supermarket Type2, Grocery Store, Supermarket Type3 is plotted. It is observed that maximum outlets are of Supermarket

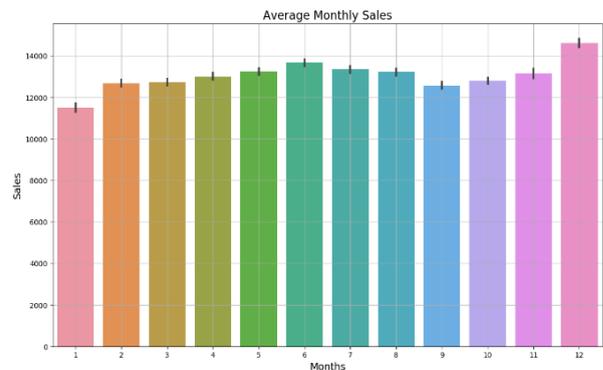


Fig-2: Average Monthly Sales

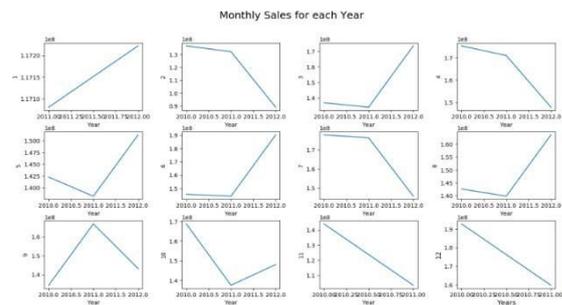


Fig-3: Monthly Sales for Each Year

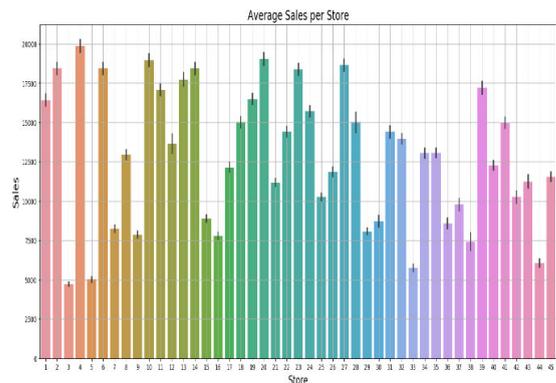


Fig-4: Average Sales Per Store

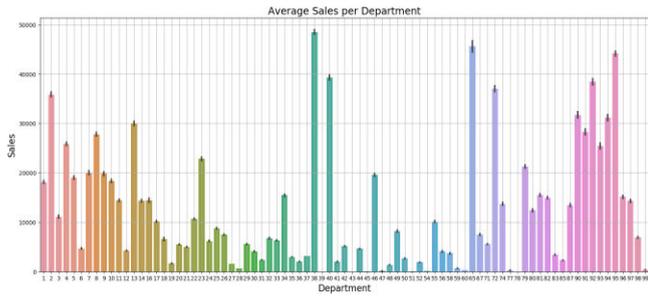


Fig-5: Average Sales Per Department

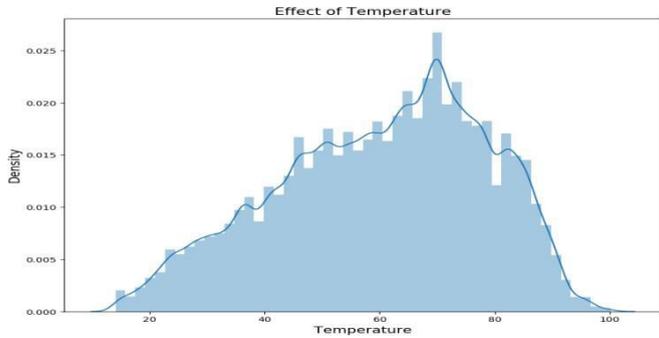


Fig-6: Effect on Temperature

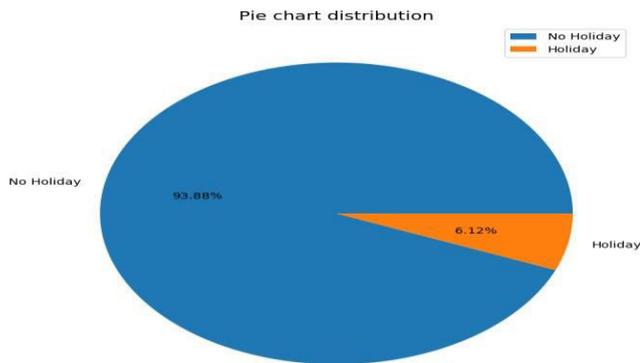


Fig-7: Pie Chart Distribution

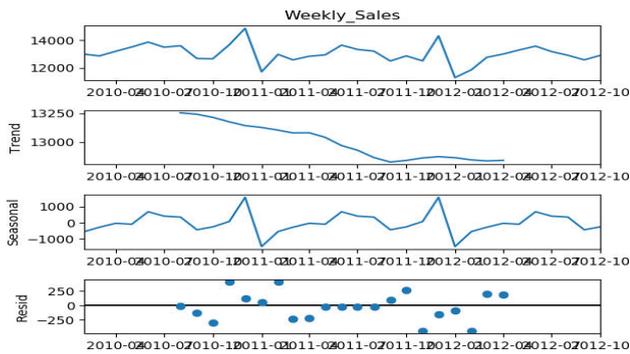


Fig-8 Weekly Sales

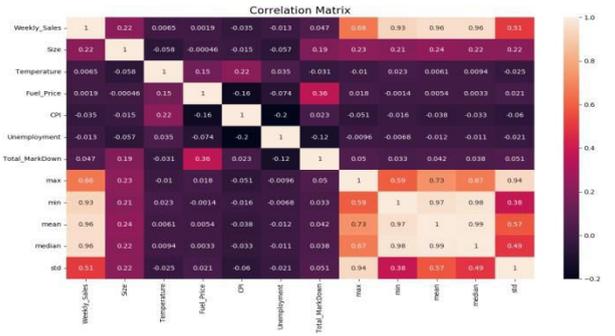


Fig-9 Correlation Matrix

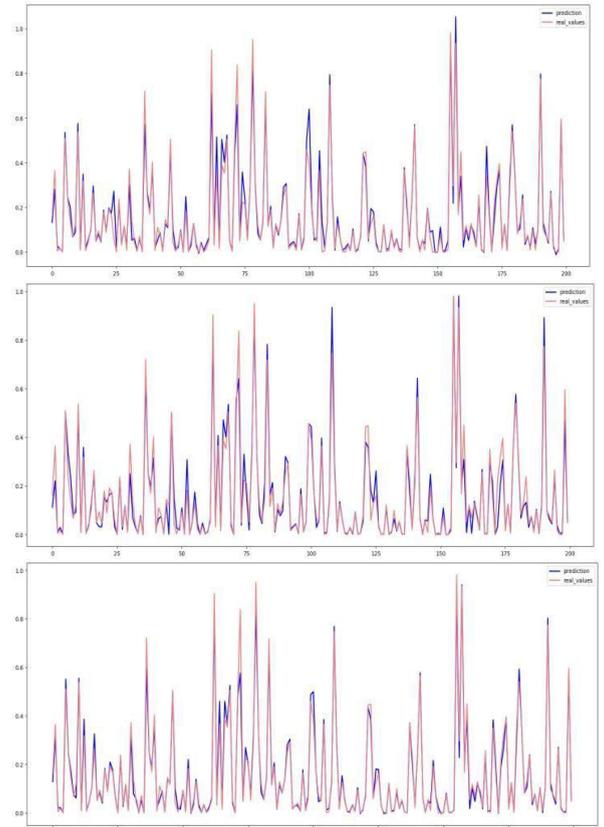


Fig-10: Trained Models

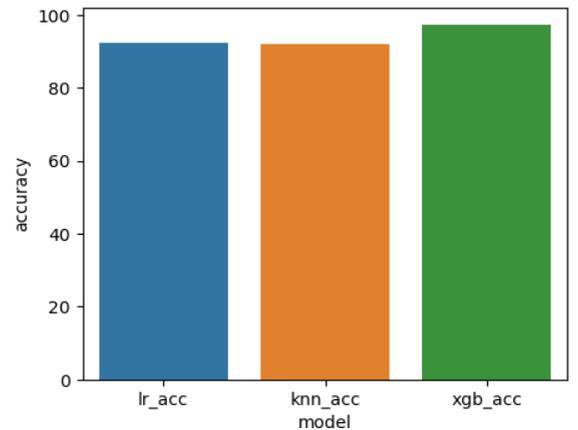


Fig-11: Accuracy of Different Machine Learning Models

2. DATA PREPROCESSING

In machine learning algorithm, data can't be used in its normal form as it is the as the way it is obtained, so the data needs to be devised before employing it in machine learning models. This technique is used to solve problems that are not yet known by the knowledge extractor. For example, if you have a problem with your car and want to know how much oil it takes to drive, you can ask the car's driver to tell you the exact amount of oil needed. This is called sales or demand of an item outlet.

Preprocessing Work:

The goal of preprocessing is to find out what kind of information the car needs before making any decisions about whether to use it or not. Proper formatted and cleaned data is essential for preprocessing. Data preprocessing consists of the following activities:

Dataset is imported by saving the dataset file as CSV file.

1) Importing the dataset: In our study we have used the dataset collected from a grocery store to check the futures include:

Item_identifier, Item_weight, Item_fat_content, Item_visibility, Item_type, Item_MRP, Outlet_identifier, Outlet_establishment_year, Outlet_size, Outlet_location_type, Outlet_type, Item_Outlet_sales.

2) Cleaning the Dataset: As a part of data cleaning, some columns are needed to be removed which do not contribute in attaining the final results of the algorithm. Here Item_identifier and Outlet_identifier is dropped.

3) Handling missing values: Missing data is something that needs to be manipulated so that there remains no discrepancy in the data to be fed into the model. Here there were some missing values in Item_weight and Outlet_size column. In case of Outlet_size, the entire row has been dropped with those of missing values and in case of Item_weight the missing spaces are filled with mean of all the other entries in the column.

4) Encoding categorical data: As machine learning concepts make use of mathematical models to solve problems, we need to make sure that we have enough numerical data to support our hypothesis in order to get the best out of it, as mathematics works well with numbers.

Label Encoder is a class imported from the sklearn library to do this task. We have encoded 'Item_fat_content', 'Item_type', 'Outlet_size', 'Outlet_location_type', 'Outlet_type', 'Outlet_establishment_year'.

5) Feature scaling: Feature scaling is a method in which we scale the data into an accurate and scalable size for the purpose of increasing accuracy and reducing error. It basically prevents the large variance of data points to be used in the algorithm and allows us to achieve better results. StandardScaler is a class imported from sklearn library. Here, standardization method has been used. We have standardized 'Item_weight' and 'Item_mrp'.

6) Extracting Independent and Dependent variables: Dependent variables are the targets or the output variables which needs to be finally evaluated and then compared against each other. Independent variables are the features or the input variables which can't be changed by any means and accordingly the targets are predicted. Independent variables include:

'Item_weight', 'Item_fat_content', 'Item_visibility', 'Item_type', 'Item_mrp', 'Outlet_establishment_year', 'Outlet_size', 'Outlet_location_type', 'Outlet_type'. Dependent variable include: 'Item_Outlet_sales'

7) Splitting the dataset into training and test dataset: To avoid overfitting, two separate datasets are not imported for train and test. So, splitting is done in a single dataset. The training dataset are the data we need to train the model on.

Test datasets are those ones which can be used to predict the outcome of a test.

3. ALGORIHTMS USED

3.1 Linear Regression

Linear Regression is the most commonly and widely used algorithm Machine Learning algorithm. It is used for establishing a linear relation between the target or dependent variable and the response or independent variables. The linear regression model is based upon the following equation:

$$y^{\wedge} = \theta_0 + \theta_1x_1 + \theta_2x_2 + \theta_3x_3 + \dots + \theta_nx_n$$

where, y^{\wedge} is the target variable, θ_0 is the intercept, $x_1, x_2, x_3, \dots, x_n$ are independent variables and $\theta_1, \theta_2, \theta_3, \dots, \theta_n$ are their respective coefficients.

The main aim of this algorithm is to find the best fit line to the target variable and the independent variables of the data. It is achieved by finding the most optimal values for all θ . With best fit it is meant that the predicted value should be very close to the actual values and have minimum error.

Error is the distance between the data points to the fitted regression line and generally can be calculated by using the following equation:

$$\text{Error} = y - y^{\wedge},$$

where, y is the actual value and y^{\wedge} is the predicted value.

3.2 K-Neighbors Regressor

KNN algorithm for Regression is a supervised learning approach. It predicts the target based on the similarity with other available cases. The similarity is calculated using the distance measure, with Euclidian distance being the most common approach.

Predictions are made by finding the K most similar instances i.e., the neighbors, of the testing point, from the entire dataset. KNN algorithm calculates the distance between mathematical values of these points using the Euclidean distance formula:

The value of K to be selected shouldn't be very small as it might result into noise in the data and in turn overfitting. The general solution is to reserve a part of data for testing the accuracy of the model. Then choose K=1, and then use the training part of modeling and calculate the accuracy of the prediction using all samples in the test set. Repeat this process increasing the K and choose K such that it is best for the model.

$$\sqrt{\sum_{i=1}^n (y_i - x_i)^2}$$

3.3 XGBoost Regressor

XGBoost also known as Extreme Gradient Boosting has been used in order to get an efficient model with high computational speed and efficacy. The formula makes predictions using the ensemble method that models the anticipated errors of some decision trees to optimize last predictions. Production of this model also reports the value of each feature's effects in determining the last building performance score prediction. This feature value indicates that outcome in absolute measures – each characteristic has on predicting school performance. XGBoost supports parallelization by creating decision trees in a parallel fashion. Distributed computing is another major property held by this algorithm as it can evaluate any large and complex model. It is an out-core-computation as it analyses huge and varied datasets. Handling of utilization of resources is done quite well by this calculative model. An extra model needs to be implemented at each step-in order to reduce the error.

XGBoost objective function at iteration t is:

$$L(t) = \sum_{i=1}^n L(y_{out,i}, y_{out1_i^{(t-1)}}) + f_t(x_i) + g(f_t)$$

where, y_{out} = real value known from the training dataset, and the summation part could be said as $f(x + dx)$ where $x = y_{out1_i^{(t-1)}}$

We need to take the Taylor approximation. Let's take the simplest linear approximation of $f(x)$ as:

$$f(x) = f(b) + f'(b)(x-b) \quad dx = f_t(x_i)$$

where, $f(x)$ is the loss function L, while b is the previous step (t-1) predicted value and dx is the new learner we need to add in step t.

Second Order Taylor approximation is:

$$f(x) = f(b) + f'(b)(x-b) + 0.5f''(b)(x-b)^2$$

$$L(t) = \sum_{i=1}^n [L(y_{out,i}, y_{out1_i^{(t-1)}}) + h_i f_t(x_i) + 0.5k_i f_t^2(x_i)] + g(f_t)$$

If we remove the constant parts, we have the following simplified objective to minimize at step t,

$$L1(t) = \sum_{i=1}^n [h_i f_t(x_i) + 0.5k_i f_t^2(x_i)] + g(f_t)$$

4. RESULTS

Machine Learning algorithms such as Linear Regression, K-Nearest Neighbors algorithm, XGBoost algorithm and Random Forest algorithm have been used to predict the sales of various outlets of the Big Mart. Various parameters such as Root Mean Squared Error (RMSE), Variance Score, Training and Testing Accuracies.

5. CONCLUSION

With traditional methods not being of much help to the business organizations in revenue growth, use of Machine Learning approaches proves to be an important aspect for shaping business strategies keeping into consideration the purchase patterns of the consumers. Prediction of sales with respect to various factors including the sales of previous years helps businesses adopt suitable strategies for increasing sales and set their foot undaunted in the competitive world.

6. REFERENCES

1. Singh Manpreet, Bhawick Ghutla, Reuben Lilo Jnr, AesaanFS Mohammed, and Mahmood A. Rashid. "Walmart's Sales Data Analysis-A Big Data Analytics Perspective." In *2017 4th Asia-Pacific World Congress on Computer Science and Engineering (APWC on CSE)*, pp. 114-119. IEEE, 2017.
2. Sekban, Judi. "Applying machine learning algorithms in sales prediction." (2019).
3. Panjwani, Mansi, Rahul Ramrakhiani, Hitesh Jumrani, Krishna Zanwar, and Rupali Hande. *Sales Prediction System Using Machine Learning*. No. 3243. EasyChair, 2020.
4. Cheriyan, Sunitha, Shaniba Ibrahim, Saju Mohanan, and Susan Tressa. "Intelligent Sales Prediction Using Machine Learning Techniques." In *2018 International Conference on Computing, Electronics & Communications Engineering (iCCECE)*, pp. 53-58. IEEE, 2018.
5. Giering, Michael. "Retail sales prediction and item recommendations using customer demographics at store level." *ACM SIGKDD Explorations Newsletter* 10, no. 2

- (2008):84-89.
6. Baba, Norio, and Hidetsugu Suto. "Utilization of artificial neural networks and GAs for constructing an intelligent sales prediction system." In *Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks. IJCNN 2000. Neural Computing: New Challenges and Perspectives for the New Millennium*, vol. 6, pp. 565-570. IEEE, 2000.
 7. Ragg, Thomas, Wolfram Menzel, Walter Baum, and Michael Wigbers. "Bayesian learning for sales rate prediction for thousands of retailers." *Neurocomputing* 43, no. 1-4 (2002): 127-144.
 8. Fawcett, Tom, and Foster J. Provost. "Combining Data Mining and Machine Learning for Effective User Profiling." In *KDD*, pp. 8-13. 1996.

