SALES FORECASTING USING PREDICTION ANALYTICS ALGORITHM

Dr. Vijayalakshmi. P[1] **B. Vanitha[2]**

ABSTRACT

Sales prediction plays a significant role in various firms that are engaged in merchandising, logistics, production and sales. This enables the users to allot resources efficiently for the estimation of sale revenue and set up an improved strategy for future growth of the firm. Machine-learning helps to find data accuracy in sales prediction. In this paper, forecasting the sales of a product is analyzed by using linear regression and K-means cluster approach that produces higher prognostic performance compared to the present predictive learning algorithms. A novel approach known as predictive analytic approach has been proposed. This approach is used to cluster the dataset based on consumable and non-consumable items. It selects the medium, high and low levels (central points) from the collected dataset. The performance of the proposed approach has been compared with linear regression and k-means cluster approaches. The experimental results prove that the proposed approach outperforms all other existing approaches.

Keywords: Machine-learning, sale prediction, clustering, accuracy of data.

1. INTRODUCTION

Sales prediction plays a big role in regulating business. The Knowledge of the past stock allows preparing for the need of the business and improves sales to succeed in the commercial environment. Linear regression is a supervised algorithm applied to find out the accuracy of sales in the market. K-means clustering is used to cluster the dataset based on the items that are sold. The proposed predictive analytic approach was organized in six phases, namely data collection, hypothesis definition, data exploration, data cleaning, data modeling and feature engineering. Data collection includes collecting data and dataset. The problem is analyzed using hypothesis definition. And the uniqueness of the data is explored. Cleaning is done to detect and correct the inaccuracies in the dataset. Data modeling is used to predict data using machine-learning techniques. Feature-engineering is used to import data from machine-learning algorithm, which predict sales.

2. BIG DATA AND DATA SCIENCE

Gopal Behera and Neeta Nain in 2019, proposed a comparison model using different machine-learning technique. Xgboost technique was used to predict sales in the market. Arpit Bansal, Mayur Sharma and Shalini Goel in 2017, used clustering technique to predict accurately and the central point using distance calculator. Their methodology improved the accuracy rate and clustering time of the data. Samaneh Beheshti-Kashi et al. in 2015, reviewed different methods to predict the strategy of sales and increase the demand of products. Bohdan M. Pavlyshenko in 2019, analysed the stacking model to predict the time series and validate data with precision.

Nikhil Sunil Elias and Seema Singh carried out a comparison of algorithms to prove hyper parameter values. An effective model to predict sales was the random forest method, which was used to analyze the comparison model. Gopalakrishnan T, Ritesh Choudhary and Sarada Prasad in 2018, implemented data visualization technique to increase the sale of the products and demand of the brand to increase profit using regression algorithm. Dr. Zainab Pirani et al., 2017, reviewed the database transaction of the product that analysed the increase in trends and sales pattern using regression algorithm.

[1]Assistant Professor, Dept. of CS, CA & IT
Karpagam Academy of Higher Education, Coimbatore.

[2]3rd Year MCA Student, Dept. of Computer Applicaiton
Karpagam Academy of Higher Education, Coimbatore.

3. DATA ACCURACY PREDICTION APPROACHES

Machine-Learning provides flexibility to seek knowledge without being explicitly programmed. It also concentrates on the computer programs that might access information. It used information to predict the choices supported by patterns. A machine-learning rule could be assessed by two ways, namely supervised machine- learning rule and unsupervised machine- learning rule. Supervised machine-learning rule was employed to research the better-known coaching dataset and turn out the prediction on output values. Unsupervised machine-learning rule showed that the system would display a performance to explain a hidden structure from untagged information. This rule explored the information and drew inferences from datasets to explain the hidden structures from untagged data. There are two existing approaches, namely such as linear regression and k-means clustering.

3.1 LINEAR REGRESSION APPROACH

Linear regression is the mathematical model that analyses the connection between a variable with the given set of independent variables.

The regression for the product sales can be calculated using the equation (1).

$$Y=mX+b \qquad \text{eqn.(1)}$$

- Y is the dependent variable, the sales of the item which is to be predicted for the accuracy.
- X is the independent variable, the dataset which is used to make predictions.
- m is the slope of the regression line which represents the effect that X has on Y.
- b is the line which crosses the y axis.

The X variable consists of item MRP, item appearance, item weight, item location, item fat. The Y variable consists of sales value to predict the accuracy. The mathematical calculations are done for the dataset such as mean, standard deviation, count, minimum and maximum.

3.2 K-MEANS CLUSTERING APPROACH

The K-means approach is used to group similar data points into a cluster. This algorithm is an iterative algorithm and used to split the data and fit them only in one group. K-means is used to cluster the data items using the scatter diagram. The Euclidean distance is used to calculate the intervals between the clusters and plot the clusters along with centroids in the graph. The mean value and the centroids are calculated. The following is the k-means algorithm:

Step 1: Determine the number of clusters from the dataset.

Step 2: Randomly choose the distinct data points and assign every information point to the cluster.

Step 3: Measure the interval between data points and choose the cluster.

Step 4: Add data points to the neighboring cluster.

Step 5: Compute the mean value, including the latest point of the primary cluster.

Step 6: Repeat the steps until the optimum data point is reached.

4. THE PROPOSED PREDICTIVE ANALYTIC ALGORITHM

The proposed predictive analytic approach is divided into six phases namely data collection, hypothesis definition, data exploration, data cleaning, data modeling and feature engineering. In the first phase, dataset of the products is collected from the relevant sources. The dataset is loaded into framework and the datasets are combined. In the second phase, the hypothesis is used to analyze the problems in the dataset. The data are statistically related to the two variables. In the third phase, exploration of data is used to find the uniqueness of the data and drop unnecessary data from the dataset. In the fourth phase, the inaccurate data is detected and corrected. Data-cleaning is used to check and fill the missing values in the dataset. In the fifth phase, predictive analytical approach is used to improve data accuracy and form data clusters. In the sixth phase, feature engineering is used to import the data from machine-learning algorithm to predict sales.
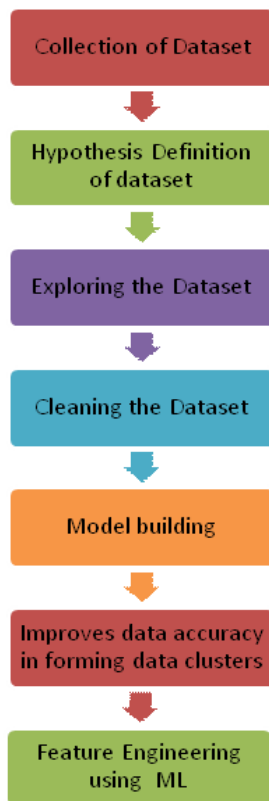
Figure 1: The Dataflow of
Predictive Analytical Approach

A dataset contains a variety of information that consists of features such as Product identifier, Product appearance, Product element, Warehouse kind, Product MRP, Warehouse symbol, Product Weight, Warehouse size, Manufacture Year, Warehouse Location, and Product Sales. Data exploration is employed to analyze the information to create true analysis. The assembling of dataset method is used to examine the uniqueness of the dataset. The inaccurate dataset is detected and renamed. Therefore, this method helps to remove any duplication of attributes. Linear regression and K-means algorithm are used for prediction of sales. The method of using the import information in machine- learning algorithms is to predict the correct sales. Linear regression method is used to get the correlation between the input and output values. The mathematical calculations are done for the data such as mean, standard deviation and-count, minimum and maximum. The accuracy rate is predicted and

it performs cross validation such as mean squared error. K-means clustering is used to cluster the data items using the scatter diagram. The Euclidean distance is used to calculate the intervals between the clusters and plot the clusters along with centroids in the graph. The mean value and the centroids are calculated. This approach is enforced using anaconda Spyder framework.

5. EXPERIMENTAL RESULT

The proposed predictive analytical approach is done using the Anaconda Spyder framework. The big mart dataset has been loaded in anaconda framework. The dataset is being combined into one file and it describes numerical calculation such as count, mean, standard deviation, minimum and maximum of dataset as shown in figure 2. The dataset has been checked to find out the mislaid values. The model building is used to predict the accuracy of the sales using linear regression method. The accuracy rate of sales is predicted and the cross validation is performed as shown in figure 3.


Figure 2. Statistical Summary

Figure 3: Accuracy rate

Figure 4 shows the impact of a product visibility over customer demand. The less visible items (scattered) are all sold in large quantities compared to more visible items for products of daily use.
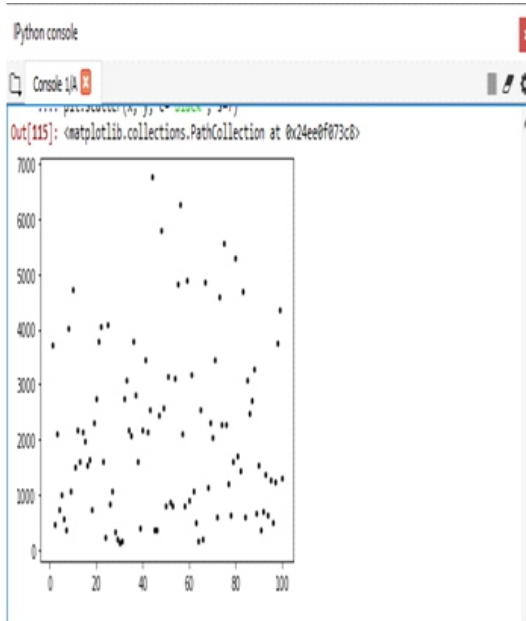
Figure 4. Impact of sales item Vs Customer Demand

Figure 5 shows that the dataset is clustered based on consumable and non-consumable items. It selects the medium-high and low -levels- from the dataset.
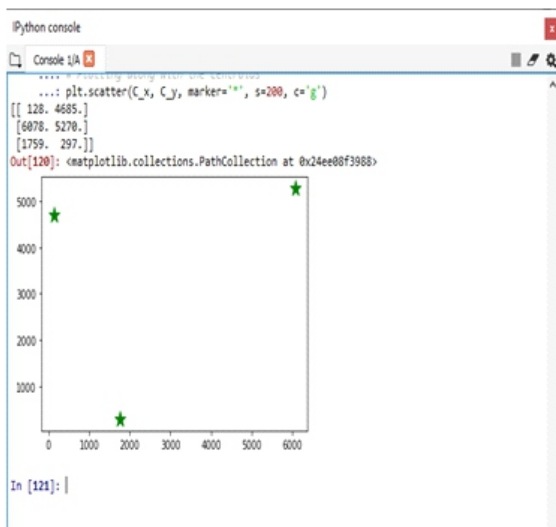
Figure 5. Clusters are Grouped

6. RESULT AND CONCLUSION

In day to day life every shopping mall would love to predict customer demand to avoid the shortfall of sales in every season. The accuracy of prediction in sales depends on the sale of a product and the customer demands. The profit of sales is directly proportional to the accuracy of sales prediction. The research area in predictive learning algorithm helps in deciding the accuracy of sales and user demand. In this paper the experimental result shows that the proposed technique outperforms the existing models in sales prediction.

Reference:

[1]   Gopal Behera and Neeta Nain "A Comparative Study of Big Mart Sales Prediction", 2019.

[2]   Arpit Bansal, Mayur Sharma and Shalini Goel "Improved K-mean Clustering Algorithm for Prediction Analysis using Classification Technique in Data Mining", 2017.

[3]   Samaneh Beheshti-Kashi et al "A survey on retail sales forecasting and prediction in fashion markets", 2014.

[4]   Bohdan M. Pavlyshenko "Machine-Learning Models for Sales Time Series Forecasting", 2019.

[5]   Nikhil Sunil Elias and Seema Singh "Forecasting of walmart sales using machine-learning algorithms".

[6]   Gopalakrishnan T, Ritesh Choudhary and Sarada Prasad "Prediction of Sales Value in online shopping using Linear Regression", 2018.

[7]   Dr. Zainab Pirani et al "Analysis and Optimization of Online Sales of Products", 2017.

[8]   Elcio Tarallo et.al "Machine-learning in Predicting Demand for Fast-Moving Consumer Goods: An Exploratory Research", 2019.

[9]   Wenjie Huang et.al " A Novel Trigger Model for Sales Prediction with Data Mining Techniques", 2015.

[10]   Ashwini Rekha. Banjanagari, Vijaykumar. B "Retail Giant Sales Forecasting using Machine learning", 2019.

[11] Robert Fildes, Shaohui Ma, Stephan Kolassa, "Retail forecasting: research and practice", 2018.