International Journal of Business Analytics & Intelligence
9 (1 & 2) 2021, 41-46
http://publishingindia.com/ijbai/

A Novel Approach to Identify the Promotion of Sales in Supermarkets using Machine Learning Algorithms

G. Sujatha*, R. Sangeetha**

Abstract

Prediction of sales promotion is a significant method to attract customers and enable retailers in successful planning and forecasting. The core objectives of this paper are to introduce a novel interpretable machine language algorithm to enhance the sales in supermarkets based on the buying behaviour of the customers, and to find the factor of sensitivity that stimulates, satisfies, and builds loyalty among the customers. Decision tree, gradient boosted trees, and support vector machine learning algorithms will be mainly used to compare the accuracy level and to analyse the promotion of sales in supermarkets based on the buying behaviour of customers, and also to find factors influencing customer retention.

Keywords: Supermarket, Sales Promotion, Customers, Decision Tree, Gradient Boosted Tree, Support Vector

Introduction

The supermarket is the best platform for the customer to get everything in one place. Customers can easily compare the same product offered by different brands. A supermarket is a place where the customer can make a perfect purchase decision. The rapid growth of technology transforms the customer's buying behaviour from large retail grocery shopping. Supermarket shopkeepers are forced to promote beyond the online point of sales when they consider the role of technology in their business plans, such as the store's website, e-newsletters, social media advertisement, text messaging, and so on; however, the customer demands are high and continue to expand day by day.

This study aims at helping supermarkets predict and analyse their marketing demands and supply based on the buying behaviour of customers by using cognitive analysis. This helps us carry out our desired predictions that reclines within artificial intelligence. Artificial intelligence works like a human brain. This will be accomplished through trained data from existing historical data and it also visualises its patterns from it. This could be achieved through machine learning algorithms. Machine learning is a subset of artificial intelligence (AI), which creates applications that learn from data and predict their accuracy without being programmed to do so. In this paper, the researcher will analyse the customer behaviours and forecast sales using 3 machine learning algorithms, the decision tree, gradient boosted trees, and support vector machine algorithms. This paper intends to study the previous sales history, and whether the sales depend on factors such as location, population, brand, store capacity, store ambition, goods display, sales promotional techniques of the product, inventory goods moving time, and so on. Since the volume of the data set is large, it needs to be processed through machine learning algorithms to get an accurate prediction of the sales using the above mentioned algorithms of machine learning.

Review of Literature

This paper reveals how future sales of Big Mart Companies can be predicted, keeping in view the sales of previous years. It has predicted the pattern of sales and the quantities of the products to be sold based on some key features gathered from the raw data (Bajaj, Ray, Shedge, Vidhate & Shardoor, 2020). In this study, data collected from a retail store and prediction of the future strategies interrelated with store management is executed. Based on

* Associate Professor, PG and Research Department of Computer Science, Sri Meenakshi Government Arts College for Women (A), Madurai, Tamil Nadu, India. Email: sujisekar05@gmail.com
** Assistant Professor, Department of Computer Science, Lady Doak College, Madurai, Tamil Nadu, India. Email: rsangeetha@ldc.edu.in

events, such as the climatic conditions, holidays, and so on, it is found that these events can change the state of buying behaviour of the customer, which in turn affects the sales of the supermarket (Singh, 2017).. This paper shows that machine learning results in better demand predictability than the use of traditional models, and helps in improving the accuracy of demand forecasting (Tarallo, Akabane, Shimabukuro, Mello & Amancio, 2019). This paper studies the behaviour of the customer and predicts the sales volume (Zuo, Yada & Shawkat Ali, 2016). This study has identified the factors that influence the demand in the retail grocery store, using statistical software tools (Paták, Branská & Pecinova, 2015). The study is aimed at forecasting sales for a supermarket chain store using a machine learning algorithm, instead of using a traditional statistical method to forecast supermarket sales. It has left a lot of challenges unaddressed and mostly results in the creation of predictive models that perform poorly (Odegua, 2020). This article focuses on small-scale businesses, like a store, with day-to-day grocery movement, to predict the sales for the upcoming month using a machine learning algorithm (Ramesh, Mathew & Hemalatha, 2017).

This research paper focuses on the benefits of machine learning in sales forecasting for short shelf-life and highly-perishable products, as it accesses the accuracy level of traditional statistical techniques. The result of this study reveals that it improves inventory balancing throughout the chain, reducing stockout rates at points of sale, and increased availability and profitability (Tarallo, Akabane, Shimabukuro, Mello & Amancio, 2019). This review paper analyses the existing machine learning algorithms for food sales prediction and describes the important design decisions of data analysts working on food sales predictions, such as the temporal granularity of sales data, the input variables to use for predicting sales and the representation of the sales output variable, forecasting food sales prediction, and evaluating accuracy using machine learning techniques (Tsoumakas, 2018). In this paper, the researcher investigated forecasting sales for a supermarket chain store called 'Chukwudi Supermarkets', with three machine learning algorithms (K-Nearest Neighbor, Gradient Boosted, and Random Forest). The study revealed that the random forest algorithm performs better than the other 2 models; gradient boosted models easily overfit the dataset and that k-nearest neighbor, even though fast, performs poorest among the 3. Further, the important variables that helped in better sales forecast were supermarket type (grocery store), product price, and supermarket opening year (Odegua, 2020).

Methodology

Algorithm Used

Decision Tree Algorithm

A decision tree is a supervised learning technique. Classification and regression problems are carried out by these algorithms. Generally, this is used for classification problems. This is a tree-structured classifier, where internal nodes are used to identify the features of a dataset, branches denote the decision rules, and each leaf node represents the outcome. There are 2 nodes, the decision node and the leaf node. The decision node has multiple branches, whereas the leaf node is the output and does not have any branches. Tests or decisions are performed based on the given data set features. Based on the conditions, graphical output is generated for all possible solutions to a particular problem. It is very similar to a tree; it starts with the root node, which explores further branches. CART (classification and regression tree) algorithm is used to construct a tree. This is simply to ask the questions with yes/no answer structures; based on the answer, the decision tree splits the tree into sub-trees. Selecting the best attributes for the root node and sub-node is a major issue in decision tree implementation. This is resolved through the attribute selection measure or ASM. Using this technique, one can identify the best attributes. There are two techniques in ASM, information gain and Gini index.

Information Gain

It is used to calculate how much information is provided to a class. The formula to predict information gain is

Information Gain = Entropy(S) − [(Weighted Avg) * Entropy(each feature)

Entropy: Entropy is a metric to find the impurity of the given attribute. It mentions randomness in data.

Formula to calculate entropy:

Entropy(S) = −P(yes) log2 P(yes) − P(no) log2 P(no)

Where,

- S = Total number of samples
- P(yes) = probability of yes
- P(no) = probability of no

Gini Index

The Gini index determines the impurity or purity in the decision tree constructed by implementing the CART (Classification and Regression Tree) algorithm. The following formula is used to calculate the Gini index.

$$\text{Gini Index} = 1 - \sum jPj2$$

In this paper, it is proposed to use a decision tree algorithm for the following 2 reasons.

- It is easy to understand the decision tree logic.
- It can effectively interpret human thinking while making decisions.

In this study, classification and regression, given a sample, is {(x n, y n)} N n = 1.

IF-THEN rule is applied and transformed into trees.

Classification: class label y ∈ {1, ..., K} (or proportion of each class p1, ..., pK).

Regression: numeric value y ∈ R (average of the outputs for the leaf's instances).

The predicted output for an instance x is obtained by a path from the root to a leaf. In the best case (balanced tree) for binary trees, the path has a long log of 2L if there are L leaves (Carreira-Perpinan, 2016).

Gradient Boosted Trees

Gradient boosted trees usually outperform the random forest. It builds the model step by step, and it generalises them by allowing optimisation of an arbitrary differentiable loss function.

A gradient boosting machine (GBM) compiles all the predictions which are generated from multiple decision trees to build the final predictions. That is, all the weak learners of the gradient boosting machine are decision trees.

In gradient boosted algorithms, every node has different features. So, these algorithms can capture different predictions from the data.

Further, each new tree will be considered the errors or mistakes done by the previous trees. So, each successive decision tree is made by the errors of the previous trees. This is the way the trees in a gradient boosting machine algorithm are constructed and minimise the overall prediction errors.

In this study, the mean rooted squared error is a loss function.

$$\text{loss} = (1/2) \times (y - y')2$$

where, y is the actual value and y' is the prediction. Each prediction descents in gradient boosting, as the partial derivative of loss function according to the prediction.

∂ loss /∂ y' = ∂((1/2) x (y − f(x))2)/∂y' = 2. (1/2). (y − y'). ∂(−y')/∂y' =2. (1/2). (y − y'). (−1) = y' − y

In this paper, α is the learning rate and set 1

y' = y' − α. (∂ loss /∂ y')

−α. (∂ loss /∂ y') = −α. (y' − y) =α. (y − y') = y − y'

Using this we can predict the new gradient boosted decision tree.

Support Vector Machine (SVM)

Support vector machines (SVMs) are a set of supervised learning methods used for classification, regression, and outlier's detection. The main advantages of SVM are their effectiveness in high dimensional spaces and in all other cases where the number of dimensions is greater than the number of samples (Ray, 2017).

SVMs are unique in nature compared to other classification algorithms, because it chooses the decision boundary by calculating the mean distance of the given data points of all the classes. This is called the maximum margin classifier or the maximum margin hyperplane in SVM algorithms. Then the support vector line is found,

which predicts the accuracy level of the data. This line is called the margin. Making a straight line between the 2 classes is known as a simple linear SVM.

Data Preprocessing

The goal of preprocessing is to find out what kind of information the prediction of sales needs before making any decisions. The data set is chosen from a dataset kaggle, called 'Supermarket sales'. To find the increase in sales, the individual frequent product type is analysed. This is done to predict the managerial decisions of the supermarket stores in the future.

The data consist of numerous supermarket variables like opening gross income, cost of goods sales (Cogs), total, quantity, gender, product line, customer type, and so on.

The data set contains a sample of 1,000 instances with 17 variables. For this study, 7 variables were taken.

RapidMiner tool is used for data preparation, visualisation, model building, and analysis of sales prediction and managerial decisions.

Table 1: Database Structure

Data Feature	*Variable Type*
Gross profit	Numeric
Cost of goods sales (Cogs)	Numeric
Total	Numeric
Quantity	Numeric
Gender	String
Product line	String
Customer type	String

From this variable, the gross profit is predicted based on cogs, total sales, quantity, gender, product type, and customer type. This study has revealed that the cogs and total variables are in perfect correlation, and quantity has a high correlation.

Experimental Work

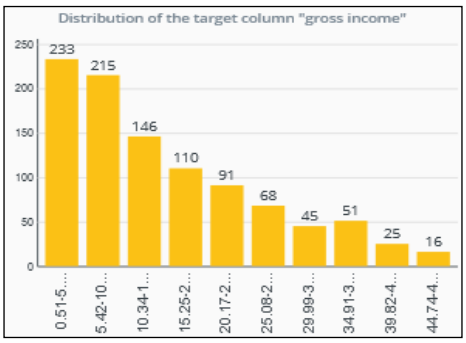


Fig. 1: Distribution of the Target Column "Gross Income"

Table 2: Model Comparison

Model	*Root Mean Squared Error*	*Average Absolute Error*	*Average Relative Error*	*Squared Correlation (R^2)*
Decision Tree	0.163	0.092	0.9	1
Gradient Boosted Trees	0.274	0.083	0.68	1
Support Vector Machine	0.008	0.006	0.06	1

Table 1 shows the database structure and variables used in this study.

Fig. 1 shows that the distribution of the target column grosses profit. Gross profit goes in the y-axis and various product types, like food and beverages, health and beauty products, and so on; supermarket shopkeepers will take an accurate purchase decision.

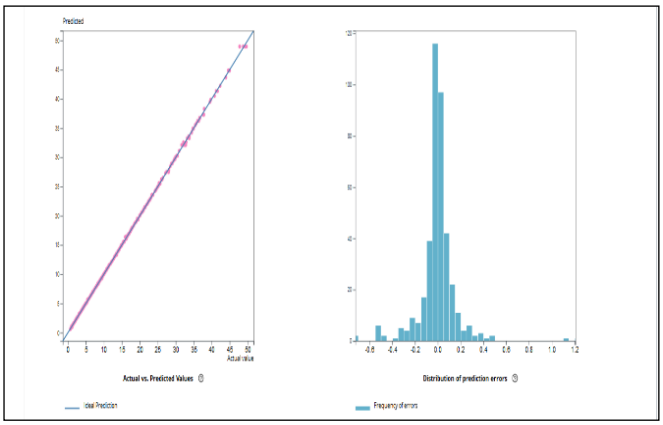
Table 2 interprets root mean squared error, average absolute error, and average relative error of decision tree as 0.163, 0.092, and 0.9, respectively. Gradient boosted trees values were 0.274, 0.083, and 0.68, and support vector machine values were 0.008, 0.006, and 0.06, respectively. This study found that SVM shows lower errors compared to other algorithms. However, squared correlation (R^2) is

1 for all the 3 algorithms mentioned.

Decision Tree

This model accomplished an average absolute error of 0.092. This means if gross income is predicted as 25.14, the real value is likely to be between 25.048000000000002 and 25.232.

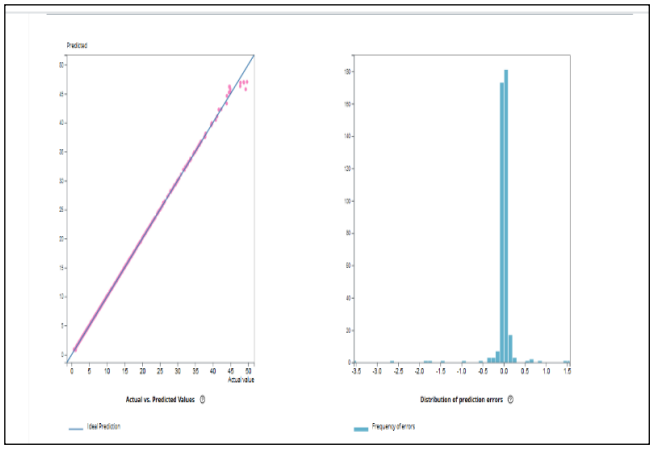
The R^2 statistic is 1. Generally, this is seen as a better value.



Gradient Boosted Trees

This model accomplished an average absolute error of 0.083. This means if you predict gross income 25.14 as a value, the real value is likely to be between 25.057000000000002 and 25.223.

The R^2 statistic is 1. Generally, this is seen as a better value.



Support Vector Machine

This model accomplished an average absolute error of 0.006. This means if you predict gross income 25.14 as a value, the real value is likely to be between 25.134 and 25.146.

The R^2 statistic is 1. Generally, this is seen as the best value.

Total number of support vectors: 1000

Bias (offset): 21.210

w[Gender] = −7.878

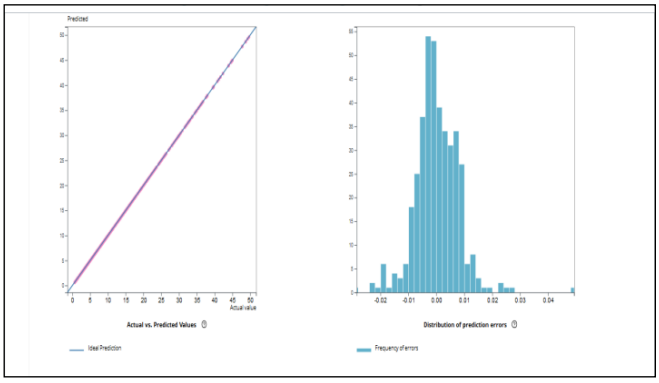
w[Product line] = −17.844

w[Customer type] = −2.887

w[Total] = 2029.165

w[cogs] = 2029.165

w[Quantity] = 333.829



Results and Discussions

From the results above, we see that the support vector machine shows below-average error compared to the decision tree and gradient boosted trees. In addition, the models in this study were trained on a data set, which might be a contributing factor to the errors. The performance of the gradient boosted trees was unforeseen, since multiple studies have shown its usage in sales forecasting. The gradient boosted trees show an unexpectedly higher distribution of prediction error compared to the other models. Although, the result could be an effect of the parameter settings used since there are numerous ways

to choose initial parameters. As the result shows, support vector machine algorithm gives better sales predictions than the other 2 algorithms. The supermarket store type seems to sell more and generally have higher sales than the other types of stores. There are several factors involved in sales forecasting, such as customer buying attitude, customer satisfaction towards the product, discounts, and offers. Sales predictions and profit are not only focused on the above-said factors, but also have to consider the seasonal product sales.

Conclusion

From this study, a supermarket shopkeeper can easily take effective market and management decisions with more accuracy without time consumption. In addition, this study will help them retain their customers by analysing customer behaviours. Supermarket owners can earn the trust of their customers and can concentrate on promoting sales and attracting new customers. This sales forecasting method will help the shopkeepers explore an alternative method for them to optimise their services to the consumers and enable efficient inventory management to control the available stock. Shortages or excess stock might be reduced, demand forecasting can be done on time, and relationship can be maintained with vendors. Further, supermarket economic benefits in the future can be predicted without any complicated statistical work.

References

Bajaj, P., Ray, R., Shedge, S., Vidhate, S., & Shardoor, N. (2020). Sales prediction using machine learning algorithms. *International Research Journal of Engineering and Technology (IRJET), 7*(6), 3619-3625.

Carreira-Perpinan, M. A. (2016, November 28). *Merced, Introduction to Machine Learning.* EECS, University of California.

Odegua, R. (2020, January). Applied machine learning for supermarket sales prediction.

Paták, M., Branská, L., & Pecinova, Z. (2015, August). *Demand forecasting in retail grocery stores in the Czech Republic.* 2nd International Multidisciplinary Scientific Conference on Social Sciences and Arts SGEM2015.

Ramesh, K., Mathew, J. J., & Hemalatha, N. (2017). Machine learning approach for the predictive analysis of sales in grocery store. *International Journal of Latest Trends in Engineering and Technology,* Special Issue, 95-98. e-ISSN: 2278-621X.

Ray, S. (2017, September 13). Understanding support vector, machine (SVM) algorithm from examples (along with code).

Singh, M. (2017). *Walmart's sales data analysis – A big data analytics perspective.* 2017 4th Asia-Pacific World Congress on Computer Science and Engineering (APWC on CSE).

Tarallo, E., Akabane, G. K., Shimabukuro, C. I., Mello, J., & Amancio, D. (2019). Machine learning in predicting demand for fast-moving consumer goods: An exploratory research. IFAC-PapersOnLine, 52(13), 737-742.

Tarallo, E. A., Akabane, G., Shimabukuro, C. I., Mello, J., & Amancio, D. (2019, August). *Machine learning in predicting demand for fast-moving consumer goods: An exploratory research.* Elsevier, 9th IFAC Conference on Manufacturing Modeling, Management and Control MIM 2019, Berlin, Germany.

Tsoumakas, G. (2018, June 14). A survey of machine learning techniques for food sales prediction. *Artificial Intelligence Review, 52*(1), 441-447. Retrieved from https://doi.org/10.1007/s10462-018-9637-z

Zuo, Y., Yada, K., & Shawkat Ali, A. B. M. (2016, December). Prediction of consumer purchasing in a grocery store using machine learning techniques. *IEEE, 3rd Asia-Pacific World Congress on Computer Science and Engineering (APWC on CSE).*