IJCSE International Journal of Computer Sciences and Engineering Open Access
Survey Paper Vol.-7, Special Issue-14, May 2019 E-ISSN: 2347-2693

Machine Learning based Models used for Sales Prediction in Retail Shops: A Survey

Surendra Babu K N[1*], Mallikarjun M Kodabagi[2]

[1,2]School of Computing and Information Technology, REVA University, Bangalore, India

Corresponding Author: surendrababukn@reva.edu.in

DOI: https://doi.org/10.26438/ijcse/v7si14.516521 | Available online at: www.ijcseonline.org

Abstract. Globally retail industries are growing day by day, because retail industries give more profit in less time duration. In 2017 USD 23,460 billion was the value of global retail industry, it is expected to increase 5.3% during the forecast period (2018-2023) reaching to USD 31,880.8 by 2023 [11]. Every business revolves around one word – Profit! Every business man wants to increase the profit of his business, no one wants to loose. The best way to increase the profit is by extracting knowledge about the business and transforming that knowledge into right predictions. In retail sales business, prediction of future sales is very much essential to improve the business operation and to increase profit. Manually analysing large amount of data for predicting future sales may lead to less accurate results. Statistical techniques were used initially to forecast future sales, later Datamining techniques were inculcated into the process of prediction. Only Datamining techniques were not sufficient to accurately predict the sales, so Artificial Intelligence (AI) domain is chosen by software professionals for prediction. Machine Learning (ML) is an application of AI and Deep Learning (DL) is an upgradation of ML especially Artificial Neural Network (ANN). Various ML and DL prediction models are gaining more attention in recent days [12]. The models can be chosen based on the type of data that is being analysed and the response time of prediction models. This paper provides review of various prediction models used for sales prediction in retail industries based on data features and models.

Keywords: Data Mining, Deep Learning, Machine Learning and Retail sales prediction.

I. INTRODUCTION

The retailers are eager to know about sales happened in the retail shop, progress in terms of profit and aspects which can affect business. By inculcating certain steps, the situations that can lead to reduction of overall profit can be avoided. Knowing the future retail sales statistics, retailer can identify which / when / why product is sold more and which / when / why product is sold less. After identifying products, the retailer can decide which / when products to stock and which / when product not to stock. This process eventually leads to gain more profit. Manually predicting sales may not give accurate results. So, the prediction models are used and they are gaining significance in recent years. The retail sales prediction involves following steps: i) The past retail sales transaction data is processed by some procedures to extract required data set in a required format. ii) The necessary features are selected from the collected data set. iii) Then the data set is given to prediction models to predict future sales. Fig. 1 shows the steps involved in prediction process.

The prediction involves learning and then performing prediction. The available data is segmented into two sets, training set and testing set. The model is trained using training set of the data set so that it can learn about features present in the data set. Later the prediction model follows prediction algorithm for prediction. The prediction models are given with test set to predict and to know the accuracy of the prediction model.

In retail sales prediction, if the transaction data is available for 30 months, then 20 months data set may be used for training and remaining 10 months data may be used for testing. Three widely used learning techniques are: Supervised learning, unsupervised learning and Reinforcement learning. Fig.2 shows types of learning techniques used by predictive models. In supervised learning prediction is done on given set of samples, it searches for predefined pattern in data points assigned during training. In unsupervised learning, no predefined patterns specified using data points, patterns should be identified by the model based on some parameters (features). In reinforcement learning data points are analysed and decision is taken, then it will learn whether the decision is good or bad, and improves its decisions over time.

Following algorithms are widely accepted machine learning algorithms to construct predictive models. A)Regression: Simple linear regression, multiple linear regression and

Logistic regression. B)Classification: Decision Tree Classification, Rule based classification, Nearest neighbour classification, Naïve Bayes classification, Artificial Neural Networks, Support Vector Machines and Ensemble learning. C)Clustering: K-means clustering, Agglomerative clustering, Divisive clustering and Density based clustering.

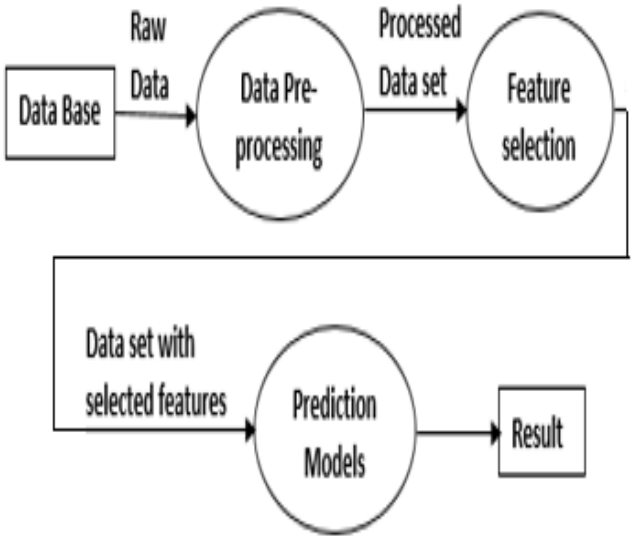

Fig. 1. Steps involved in prediction process

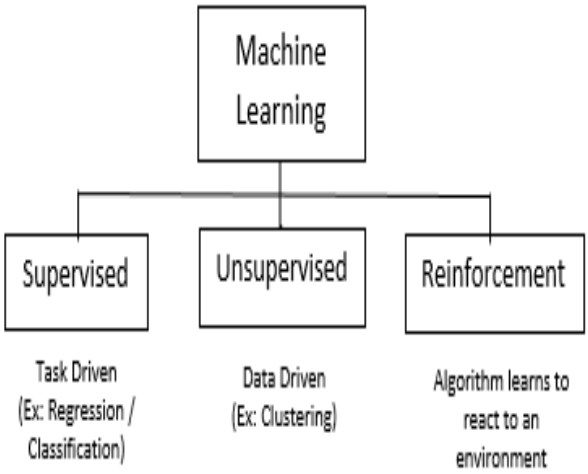

Fig. 2. Types of Learning Techniques

Regression

Regression is used when variable to predict is numerical (e.g. price of a house). For estimating relationships among variables some statistical methods are used in regression. In regression, two main variables are used called dependent variable and independent variable. The dependent variable is the output got from model and the output is calculated using input independent variables. It basically helps in understanding the relation between dependent variable with one or more independent variables. When one or more independent variable is varied then what is the change in dependent variable is observed and prediction is done.

Classification

In classification, variable to predict is part of one of the number of pre-defined category (class), which can be as simple as "yes" or "no". (e.g. taking decision whether to give loan or not to a new applicant). It is a technique where we categorize data into a given number of classes. Identifying class to which input data belongs to is the main goal of classification. It employs supervised machine learning method. Some of the classification models [14] used in machine learning are Logistic Regression, Naïve Bayes, Stochastic Gradiant Descent, K-Nearest Neighbors, Decision Tree, Random Forest, Support Vector Machines and one of the most known classification method is Artificial Neural Networks.

Clustering

To find groups of similar data objects from a large data then we employ a process called Clustering. Each identified group is called a cluster. A cluster consists similar data elements within cluster and dissimilar to data elements of other cluster. In clustering, variable to predict is part of one of the number of clusters generated by clustering algorithm. It is a Machine Learning technique that involves the grouping of data points. For a given a set of data points, clustering algorithm can be used to classify each data point into a specific cluster. Data points that are in the same cluster should have similar properties (features), while data points in different clusters should have highly dissimilar properties (features). The learning technique used by clustering is unsupervised learning. Clustering is a common technique for statistical data analysis used in many fields.

Prediction models can be constructed using any one machine learning algorithm or collection (fusion) of machine learning algorithms as specified in [13].

II. RELATED WORK

The prior researchers have discussed several predictive models to predict sales. The model proposed in [1] aims at prediction of food products sales and prediction of when season starts and ends by considering seasonal changes. It employs *ensemble learning approach with integration of classifiers dynamically* for improved identification of seasonal changes and changes in customer demands. The model is evaluated on data set from a Sligro Food Group N.V. Ensemble learning with original standard data set and with feature selection (OR/FS) produces high accuracy. It is observed that moving average performs quite well, but stable and accurate results can obtained by ensembles learning approach. The problem with ensemble approach is, it is not

good to use one global measure for all products in retail business. The work can be extended to estimate similarity of different products based on their sales data values. Further enhancements can be done for identifying relations between products in market basket data.

The model presented in [2] uses learning rule based on differential evolution (DE) to develop an efficient forecasting model. An Adoptive linear combiner (ALC) technique is used with its weights trained by DE. Two sets from www.forecast.org has been collected for retail sales prediction. Seasonally adjusted (SA) data set and not seasonally adjusted (NSA) data set. Both data sets are normalized by dividing each data by its maximum value in the data set for each feature. The input features are selected from the normalized data sets for the model. They are *first day of the month, mean and variance*. The available data set is split into two sets – training and testing sets. Using DE the ALC is trained and optimum weights are derived. Forecasting performance is tested by training model using various patterns. The Mean Absolute Percentage Error (MAPE) is calculated to analyze the performance of proposed DE model with Genetic Algorithm (GA) model. In advance, one month, three months and six months retail sales value is predicted by simulating the model. GA model is also simulated under the similar situations to compare the performance of the new model. The performance of DE model shows improved prediction accuracy when compared to GA model according to simulation study.

The model discussed in [3] aims to identify purchase pattern of customers to predict sales. Uses association rules to identify association between products. Purchase behavior of shoppers sometimes depends not only on products purchased but also on other products purchased with some dependency of purchased products. For 'N' number of total transactions i.e. rows and X & Y products in each row, an association rule specified as X => Y. The product dependency strength is calculated by using association rule in terms of *"support, confidence and lift"*. In this paper, the retail sales data of 11 items with 1,000 transactions along with customer profiles is considered. The 1,000 transactions are split into two sets based on period, first 500 transactions occurred in the period one (P-1) and remaining five hundred happened in period two (P-2). Various factors of changes in purchase behavior pattern is considered as given follows:

- Perished Pattern - appear in P-1, and not appeared in P-2.
- Added Pattern – not appear in P-1, and appeared in P-2.
- Emerging Pattern - Appears in both P-1 and P-2, but with various degrees of intensities provided by the strength of the association rule.
- Unexpected Pattern - Appears in both P-1 and P-2 but with different values in P-1 and P-2.

Association rules are used to measure the similarity in the context of change mining with conditional parts having products (items). Based on similarity in change mining the products to be stocked retail shop can be recommended so that it can increase sales intern increasing the profit.

The model specified in [4] aims to demand forecast of products in retail shop using Fuzzy neural network. Many attributes affect forecasting which leads to nonlinearity. Because of nonlinearity of attributes the traditional forecasting methods will not give a good demand forecasting. The demand forecast is done using fuzzy uncertainty Artificial Neural Network (ANN) and using traditional Holt-Winter's model. The performance of both the models are compared using Mean Absolute Percentage Error (MAPE). It is observed Holt-Winter's model has high error level than ANN. The statistics shows that MAPE is 29.1% for daily forecast period using Holt-Winter's model and MAPE is 10.1% for daily forecast period using fuzzy ANN.

The model in [5] uses Artificial Neural Network to forecast sales revenue of a Turkey's grocery retailing industries. The features used are marketing costs of grocery retailers, total profit and its competitor's total profit. Analysis is done on BIM, Migros and Carrefour retail industries. The neural network system is implemented using Excel with a NeuroXL add-on. Very strong forecasting is done for each retailer. The results of forecasted sales revenue on actual sales revenue rates for retailers are as follows: for BİM (94.87%), for Carrefour (89.35%), and for Migros (104.93%). These strong similarity rates are used by ANN to forecast retail sales. It is also used find factors that could affect sales revenue. In this research only grocery retailing industry is analyzed. Other industries can also be analyzed in the same manner. Further improvement can be done for country-wise comparison.

The model used in [6] aims to forecast sales of Walmart store using Hadoop Map Reduce. In map reduce data is given more importance than algorithms. So, a huge amount of data is required for processing. The three years of Walmart store data is considered with 45 stores having 99 departments in each store at various locations. The process includes three main steps, the first step collection of huge sales data, second is transferring data set into HDFS (Hadoop distributed file system) and third step is applying map reduce technique on the data sets. After Map Reduce, Hive will process the data, then dynamic partitioning and bucketing is done. The monthly sales are calculated. The mean feature among them is computed. Tableau data visualization tool is used to visualize sales data of Walmart store. For forecasting sales Holt-winters model is used.

The prediction model in [7] aims to predict retail sales of a retail store using Deep Learning approach. Sales prediction model is constructed using point of sale (POS) data of retail

store. To this model if you give sales of a particular day then the changes in sales on the following day is predicted. Sale forecasting accuracy of 86% was got by using a deep learning model with L1 regularization. The retail store products have been precisely grouped into three categories. In this paper, the deep learning model is compared to a logistic regression model. Logistic regression is one of the method used for analysis of binary data. If number attributes are less then logistic regression gives best results, if number of attributes are increased then the accuracy of the result will go down. To increase the accuracy of linear regression when there is an increase in number of attributes then an appropriate attribute selection is required and it done through regularization or other methods. Deep learning is more accurate without any regularization techniques, but this model need some measures to gain very high accuracy. In future the model can be improved to predict the sales figures themselves. Sales prediction of the following day will always help for on-site needs. In future, scope of the prediction can be widened for even larger scale data. To perform this, it is important to know the necessary parameters with optimum values and detailed sales predictions.

Support Vector Machine (SVM) is the model used in [8] for predicting consumer purchasing in a super market from Japan. This model used RFID data to identify customer's in-store behavior. Using RFID data customer's purchase behavior is forecasted. This forecasting helps in identifying the future product sales. An RFID tag is attached with each shopping kart, when the customer uses this kart then the movements of the customer in the store is identified by this RFID tag. Using RFID tag, the data such as how much time the customer spent on selecting a particular product, type of product etc are identified to forecast customer's purchase behavior. The proposed SVM model is compared with Linear Regression and Bayes Classifier. It is observed that the performance of SVM is better than other two models.

The unique model specified in [9] is based on Deep learning and Empirical Mode Decomposition. The method uses 'divide and conquer' technique. The retail time series data is converted into sub-sequence (signals) and suitable deep learning model is developed for each sub-sequence using valid feature. The model used multiple sets of data such as i) Retail sales data ii) Holiday data iii) Weather data iv) Environmental data. This model is compared with KNN, GBDT(Gradient Boosting) and XGBoost algorithms. It is proved that the performance of the proposed model is better than KNN, GBDT and XGBoost algorithms. The model focused on solving problem of single-step prediction, future they have planned for working on multi-step prediction problem.

SPV model in[10] proposed to predict sale state of commodity. The model mainly concentrates on profit of the retail shop. It has used three main factors i) Season Ratio (S) ii) Profit Ratio (P) and iii) Sales Volume (V) forming SPV commodity sale model. The model uses ID3 algorithm for classification, K-means clustering algorithm for clustering and Markov Model for prediction. It is observed that there is 10% increase in performance of prediction of SPV model when compared to old season analysis model.

III. COMPARATIVE ANALYSIS

Table.1 shows a comparative analysis of predictive models used for forecasting / predicting retail sales:

CHALLENGES

The features selected and type of prediction algorithm used will judge the performance of prediction model. Following are challenging features which has to be considered to get very high prediction accuracy:

Life span of product: Manufacturing date, expiry date, temperature under which the products are preserved.
Seasonality: Seasons, Holidays, Weekends, Special Events, Births & Deaths, Political Events, and Calendar effects; on what day of the week the holiday occurs.
New Trends: Fashions and Styles, Age of customer and Publicity
Weather: Temperature, rain
Competition: Internal and External

CONCLUSION

By this survey it is concluded that the prediction algorithms are playing a key role in increasing profit of retail shops by providing information about products to be stocked and / or products not be stocked. This paper shed light on some machine learning prediction algorithms used for forecasting or predicting retail sales. This paper mainly reviews predictive algorithms based on feature extraction techniques, types of features used and the methods to increase the accuracy of the model. The study also shows that deep learning models can give accurate results when trained with proper features with sufficient data set.

REFERENCES

[1]. Patrick Meulstee and Mykola Pachenizkiy, "Food Sales Prediction: 'If Only It Knew What We Know'", 2008 IEEE International Conference on Data Mining Workshops.
[2]. Ritanjali Majhi, Ganapati Panda, Babita Majhi, S. K. Panigrahi and Manoj Ku. Mishra, "Forecasting of Retail Sales Data Using Differential Evolution", *2009 World Congress on Nature & Biologically Inspired Computing (NaBIC 2009).*
[3]. Pradip Kumar Bala, "Mining Changes in Purchase Behavior in Retail Sale with Products as Conditional Part", 2010 IEEE.

[4]. M. Ahsan Akhtar Hasin, Shuvo Ghosh, and Mahmud A. Shareef, "An ANN Approach to Demand Forecasting in Retail Trade in Bangladesh", 2011 *International Journal of Trade, Economics and Finance, Vol. 2, No. 2.*
[5]. Dilek Penpece and Orhan Emre Elma, "Predicting Sales Revenue by Using Artificial Neural Network in Grocery Retailing Industry: A Case Study in Turkey", 2014 *International Journal of Trade, Economics and Finance, Vol. 5, No. 5.*
[6]. Anita S Harsoor. "Forecast of sales of Wallmart Sotres using Big Data Applications", 2015 IJRET.
[7]. Yuta Kaneko and Katsutoshi Yada, "A Deep Learning Approach for the Prediction of Retail Store Sales", 2016 IEEE *16th International Conference on Data Mining Workshops.*
[8]. Yi Zuo, Katsutoshi Yada and A B M Shawkat Ali, "Prediction of Consumer Purchasing in a Grocery Store Using Machine Learning Techniques", 2016 IEEE *3rd Asia-Pacific World Congress on Computer Science and Engineering.*
[9]. Shucheng Mou, Yang Ji and Chujie Tian, "Retail Time Series Prediction Based on EMD and Deep Learning", 2018 IEEE Proceedings of IC-NIDC.
[10]. Jing Zhang and Juan Li, "Retail Commodity Sale Forecast Model Based on Data Mining", 2016 IEEE *International Conference on Intelligent Networking and Collaborative Systems.*
[11]. https://www.businesswire.com/news/home/20190102005213/en/Global-31.88-Billion-Retail-Industry-Analysis-Outlook
[12]. Venkatesh Shankar, "How Artifi cial Intelligence (AI) Is Reshaping Retailing", Elsevier Publications, Invited Commentary / Journal of Retailing 94 (4, 2018).
[13]. Mohit Gurnani, Yogesh Korkey, Prachi Shahz, Sandeep Udmalex, Vijay Sambhe and Sunil Bhirudk, "Forecasting of sales by using fusion of Machine Learning techniques", 2017 IEEE, International Conference on Data Management, Analytics and Innovation (ICDMAI), pp 93-101.
[14]. Rana Alaa El-Deen Ahmed, M.Elemam.Shehab, Shereen Morsy, Nermeen Mekawie, "Performance study of classification algorithms for consumer online shopping attitudes and behavior using data mining" 2015 IEEE, Fifth International Conference on Communication Systems and Network Technologies, pp 1344-1349.
[15]. Vishal Shrivastava, Prem narayan Arya, "A Study of Various Clustering Algorithms on Retail Sales Data", International Journal of Computing, Communications and Networking -2012, Vol.1 No.2. pp 68-74.

Table.1 comparative analysis of predictive models

Reference Paper	**Data Set**	**Data Pre-processing/ Feature Extraction**	**Prediction Model**	**Results / Accuracy**
[1]	Data from retail company Sligro Food Group N. V.	*Data pre-processing:* i)Symbolic Aggregate Approximation (SAX) ii)Agglomerative hierarchical clustering (AHC) *Feature* Extraction*/Features:* i)Filter-based individual feature selection ii)More than 100 features are being used (eg. Products, purchases, Holidays, school holidays, weather properties etc.)	i)Ensemble learning ii)24 heterogeneous classifiers are used such as C4.5, decision tree, k-nearest neighbour, two rule learner, libSVM, regression etc.	60% more accuracy in predicting food sales compared to currently used baseline technique (Moving Average and Simple regression)
[2]	Data is collected from www.forecasts.org	*Data pre-processing:* Normalization of data is done by division of each data value by maximum data value in each set. *Features used are:* i) First day of the month after normalization and ii) The mean and variance calculated for this month	i)Differential Evolution ii)Adoptive linear combiner	Improved prediction performance compared to standard Genetic Algorithm
[3]	Data from retail shop with 11 items are considered.	*Features used are:* Customer's shopping record and Customer demographic information	Association Rule mining	- - - - -

[4]	Data from a super market in the city of Dhaka is used.	*Following are influencing features:* i. Weekend ii. Holiday iii. Festival period day iv. Promotional activity details v. Item availability vi. Range of price. vii. Rate of Consumption of products viii. Brand value. ix. Climatic conditions x. Seasonality.	Fuzzy neural network	MAPE is reduced from 29.1% to 10.1 % when compared Fuzzy Neural Network with Holt-Winter's model.
[5]	Data from BIM, Migros and Carrefour retail industries are used	Features used are marketing costs, total profit, Competitor's total profit and sales revenue.	Feed forward ANN	90% accuracy.
[6]	Walmart store sales data is used	*Data Pre-processing:* Map reduce Features used are Sales data, Market data and Promotional data.	Holtwinters	80% accuracy in sales prediction
[7]	3 years POS data from supermarket situated in the Kanto Region of Japan	Products are divided into 3 categories: Binarized labelling is used	Deep Learning	Predictive accuracy of 86% is achieved
[8]	RFID data of a supermarket from Japan is used.	An RFID tag is attached with each shopping kart, when the customer uses this kart then the movements of the customer in the store is identified by this RFID tag. Using RFID tag, the data such as how much time the customer spent on selecting a particular product, type of product etc are identified	SVM (Support Vector Machines)	SVM is compared with Linear Regression and Bayes Classifier. SVM gives better result than others.
[9]	POS data of a shopping mall	The model used multiple sets of data such as i) Retail sales data ii) Holiday data iii) Weather data iv) Environmental data	Empirical Mode Decomposition and Deep Learning	Accuracy is high compared to KNN, GDBT and XGBoost algorithms.
[10]	POS data of a retail shop	i. Season Ratio (S) ii. Profit Ratio (P) and iii. Sales Volume (V)	SPV model ID3 algorithm for classification, K-means clustering algorithm for clustering	Accuracy of proposed model is high compared to Markov model

