International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 05 | May 2020 www.irjet.net p-ISSN: 2395-0072

E-COMMERCE SALES FORECASTING

Sunny Khatri[1], Merin Meleet[2]

[1]*Student, Dept. of Information Science and Engineering, R.V. College of Engineering, Karnataka, India*
[2]*Assistant Professor, Dept. of Information Science and Engineering, R.V. College of Engineering, Karnataka, India*

Abstract - *Many household goods are marketed through various distribution network subsidiaries located in different areas. Supply chain inefficiencies can arise at different locations where the retailers do not measure the demand capacity. Retailers also don't find it easy to grasp the consumer situation at various geographical locations. The retail store network organization needs to consider the business environment to increase the purchase and selling of its products in order to draw more consumers in that direction. Business forecast allows retailers to see the big picture by predicting revenue. If improvements are appropriate, we get a general view of the coming years so that adjustments can be made to the target of the retail company in order to achieve success more profitably. It also enables the consumer to be satisfied when they have the goods they want in the required period. Sales forecasting allows retailers to know the demand of the company. In this paper we make an effort to understand the driving forces of the retail store business by examining Walmart store's sales data which is geographically distributed at various locations and the revenue prediction for few weeks to come is finished. The distribution networks are assisted by revenue forecasts so that the resources are handled efficiently*.

Key Words: **Hive, Hadoop, Map reduce, Big data, PCA.**

1. INTRODUCTION

We're all excited about our future! So excited about what is going to happen to us all the next time, tomorrow, retailers will also be curious about their market, their growth, their cuts. The factors that can harm or may income can be avoided by taking such actions. For this aim of forecasting of revenue, the data from various industries are collected. The data analysis does not require an efficient understanding of the observed data through specific steps, since the data are so large and the masses of organisation's data are so shaped that the significance can be accomplished by a detailed understanding of the correct measures. Consider the retail stores network, for example, Walmart, an example for giant shops, large bazaars etc .. The retail stores sell and benefit from household goods. There are numerous branches of the retail network, whose offices in different geographical locations most often do not enable retailers to understand the customers' needs so they may determine their business potential in that area on special occasions when the sales or shopping rate is more regular By tracking these purchases, the history of buying each commodity in each store and department is preserved, allowing the awareness of profit and loss within that year. Let us remember Christmas in particular during the present session. In Christmas, sales are more in the apparel section, footwear ref, and so on, then in summer sales of cotton clothing are more frequent, in winter sales for sweaters are more frequent. Consumer sales evolve as a result of this market history, and market can be expected for the future. It provides a solution to the turmoil in the retail network market. Supply chain management is a competitive advantage; the key features of supply exchange management are to increase the sales profit and control inventory turnover.

2. BACKGROUND

2.1 Fundamentals

The key aim of retail store network planning is to get the most from knowing the details and when and in which stores to invest profitably. To succeed, businesses should be able to know each customer in an allocated area, the multiple combinations in the position of the allocation options, the options will be evaluated with some mathematical model or algorithms, which will aid in determining how to set up the shop network with the perfect strategy, i.e. shop structure, ability of the store Specific storage formats may be useful for restaurants, convenience stores, luxury department stores, supermarkets, shopping malls etc.

3. CHALLENGES AND MOTIVATION

Several problems remain in the retail shop network preparation, some of which are retailers failing to determine the demand potential. Retailers disregard the randomness of the season. Inefficiencies in the supply chain when the goods are in high demand are not available. Employees are not accessible inefficiently whenever required. Retailers face issues with the inventory management system; often retailers neglect market rivalry.

Retailers create strategies for the progress and the highly focused program. The plans will help to achieve the full benefit. The new product lines should be created or purchased confidently. The system of the supply chain should be effective.

3.1 Forecasting

The overall market turnover of a particular item is based on an estimate of the maximum selling volume possibilities in a given time and under certain defined conditions. Business

demand depends on the actions of the customer, the competitors, supporting services, etc. These affect the profits of a specific store in the future. Preparation of the store is critical and the strategy process at the strategic and organizational level is systematic. Goods for purchase and sale, supply chain management and control of human capital, spatial control work in store planning. By looking at the past of shops, it helps get an understanding of the profits of the shop and make adjustments to the target so that it becomes more successful.

In the prediction of sales the fundamental information given by the current store is very important. In this paper, we looked at the walmart store results. The main data sets to be used in the retail sector for research and reporting are:

Revenue data — point of sales data, gross margins, turnover, gross margin return on investment.

Business data-market position, price of competitors, product lines of competitors, market share of competitors and customers income. Advertising and marketing data — Consumers who support sales, their reviews, overall advertising costs, are complied with as a result of their previous success.

4. SALES FORECASTING IMPLEMENTATION

4.1 The Forecasting Process

The forecasting cycle is a set of methods for estimating sales. It is started after the goal is decided. The sales sum can include dollars, the number of employees to be hired. The expected outcomes such as revenue statistics or the number of workers to be hired over the next year. The market factor involves factors such as the availability of goods in the store, the price and the demand of the commodity. The market factor is the percentage sum that is measured according to other basis contents. As the level of the market is increased, revenues in the industry are increased. The index comprises various market variables such as quality, population of the region and personal income that are available. Then in the forecasting process prevision and data analysis methods are decided. If procedures have not been used in the past, the organization might wish to review the procedures. The data is then compiled and analyzed. Some assumptions are made about the expected revenue. The revenue estimate will then be updated and the results are measured as the time progresses. The forecasting is done by PCA analysis.

Procedure:

1. Cleaning:

Cleaning train dataset

Cleaning test dataset

2. Splitting dataset into features and target:

3. Feature selection/feature elimination:

RFE- Recursive feature elimination

4. Data standardization:

Data processing workflow that converts data of complex dataset into common data format.

5. Regression analysis without PCA:

Without reducing dimensionality.

6. Linear regression model:

To train models and make predictions.

7. Decision tree:

Predict target variable by learning simple decision rule.

8. Random forest regression:

Supervised learning model

Hyperparameter tuning using GridSearchCV

Hyperparameter tuning using RandomizedSearchCV

9. PCA using Scikit learn:

Split data into training and testing

Standardize the data fitting PCA on training

Applying mapping to both training and testing.

5. METHODOLOGY

Map reductions: To generate large volumes of parallel data sets, distributed algorithms are used for processing on the cluster programming model called Map reduced. In the map, data is focused instead of algorithms. The large amount of data is needed for processing. MapReduce is represented by the terms repeatedly split by the generic term into the number of words, so that for each word the code or the numerical representation is allocated. Big data is used in the industry, the precise results obtained from big data help to determine the results, and in forecasting analysis Big data is helpful. Big data is used by many people such as scientists, entrepreneurs, and the government. The software foundation Apache published open source software called Hadoop which runs on commodity hardware, works on frames, distributes storage and parallels. Hadoop has many server nodes that are useful for storage and processing. In large data collections, batch processing is enabled by hadoop programming. Map Reduction is used to produce vast sets of intermediate results / key values. All intermediate values are combined in a map.

5.1 Graphical Representation

The processed data by the machine learning algorithm can be interpreted which provides efficient results, but the information received is too large and therefore complicated to reach the conclusion. We all know that the photo is worth more than a lot of words. Therefore, the interactive data visualization based on business intelligence can be understood.

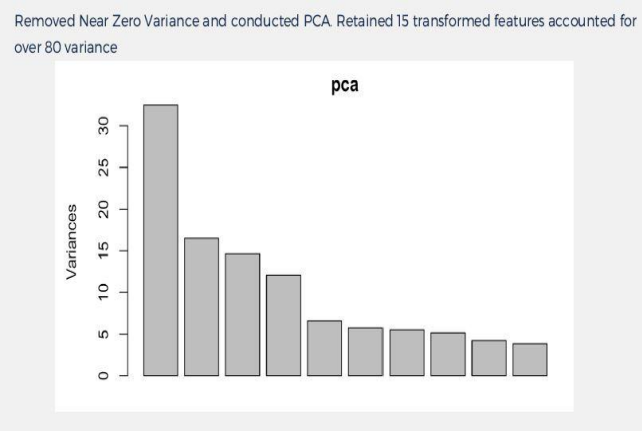

Figure 1:Results after PCA analysis

6. RESULTS AND DISCUSSION

- Typical accuracy is 86% and AUC value is 0.655.

Algorithm	Value Accuracy	Test Accuracy	AUC
Decision Tree	0.8564	0.8634	0.65
Random forest	0.8699	0.847	0.6355

- Value accuracy of decision tree is 0.8564.
- Value accuracy of random forest is 0.8699.

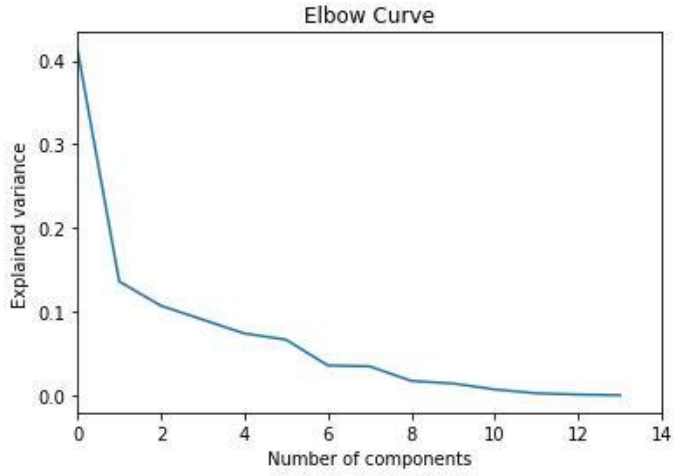

Figure 2: Elbow curve between explained variance and number of components

7. CONCLUSION

Most shopping centers aim to draw customers into the store and make full profit from them. When clients enter the shops, they can certainly purchase more from exclusive deals and receive the desired items, which are available at a good price and satisfy them. If the goods are tailored to consumers' needs, retailers can also benefit to their fullest by making improvements to their operations, store goals cause missing inventory and productive strategies can be used to obtain more profit by looking at the history of data, and current stores can clearly define the phenomenon of sales as seasonality and randomness. Forecasting has the advantage of understanding the number of workers to reach the production standard. Sales decline is bad thing because sales prediction aids research, and it can mitigate the decrease in sales in order to remain a critical factor in the market forecast.

REFERENCES

[1] Dai Wei, Peng Geng,Liu Ying ,Li Shuaipeng "A prediction study on e-commerce sales based on structure time series model and web search data".

[2] Kasun Bandara, Peibei Shi, Christoph Bergmeir, Brian Seaman "Sales Demand Forecast in E-commerce using a Long Short-Term Memory Neural Network Methodology".

[3] Chi Chun Wan "Forecasting E-commerce Key Performance Indicators".

[4] Ashwini Rekha,Banjanagari,Vijaykumar B "Retail Giant Sales Forecasting using Machine Learning"

[5] .Kui Zhao,Can Wang "Sales Forecast in E-commerce using Convolutional Neural Network".

[6] PaulaA.Hurtado, CarinaDorneles, EnzoFrazzon"Big Data application for E-commerce's Logistics: A research assessment and conceptual model".

[7] Prediction of retail sales of footwear using feed forward and recurrent neural networks. Prasun Das Æ Subhasis Chaudhury.