

INTELLIGENT SALES PREDICTION USING MACHINE LEARNING TECHNIQUES

Mr. Akshay Godse, Miss. Poonam Pawar, Mr.Sairaj Sawant, Miss.Shirin Mujawar, Prof. Prashant Karkalle.
*Department of Computer Engineering, SKN Sinhgad Institute of Technology and Science,
Lonavala, Savitribai Phule Pune University, Pune*

Abstract: With the rise of e-commerce business, sales forecasting plays an increasingly important role, for accurate and speedy forecasting can help e-commerce companies solve all the uncertainty associated with request and supply and reduce inventory cost. Most of the business organizations heavily depend on information base and demand prediction of sales trends. The accuracy in sales forecast provides a big impact in business. Data mining methods are very actual tools in extracting hidden knowledge from an enormous dataset to enhance accuracy and efficiency of forecasting. The E-commerce industry is badly in need of new data mining techniques and intellectual prediction model of sales trends with highest possible level of accuracy and reliability and traditional prediction systems are difficult to deal with the big data and accuracy of sales forecasting. These issues could be overcome by using various data mining techniques. In this paper, we briefly analysed the concept of sales data and sales forecast. On the basis of a presentation evaluation, a best suited predictive model is suggested for the sales trend forecast and summarized in terms of reliability and accuracy of efficient techniques taken for prediction and predicting. The studies found that the best fit model is Gradient Boost Algorithm, which shows maximum accuracy in forecasting and future sales prediction.

Keywords - Sales prediction, Data mining techniques, Machine learning algorithm, Sales forecasting, Reliability.

I. INTRODUCTION

Earlier research on sales prediction has always used a single prediction model. However, no single model can perform the best for all kinds of merchandise. Accurate prediction results for just one product are meaningless to sellers. A universal prediction for all commodities is needed. One of the major objectives of this research work is to find out the reliable sales trend prediction mechanism which is implemented by using data mining techniques to achieve the best possible revenue. Today's business handles huge repository of data. The volume of data is expected to grow further in an exponential manner. The measures are mandatory in order to accommodate process speed of transaction and to enhance the expected growth in data volume and customer behaviour. The E-commerce industry is badly in need of new data mining techniques and intelligent prediction model of sales trends with highest possible level of accuracy and reliability. Sales

forecasting gives insight into how a company should manage its workforce, cash flow and resources. It is an important requirement for enterprise planning and decision making. It allows companies to plan their business strategies effectively. In other words, sales forecasting is sales prediction that is based on the available resources from the past. An in-depth knowledge of the past resources allows to prepare for the upcoming needs of the business and increases the likelihood to succeed irrespective of external circumstances. Businesses that treat sales forecasting as the primary step tend to perform better than those who don't. Accurate predictions allow the organization to improve market growth with higher level of revenue generation. Forecasting of the future demand is central to the planning and operation of retail business at both macro and micro levels. At the organizational level, forecasts of sales are needed as the essential inputs to many decision activities in various functional areas such as marketing, sales, production/purchasing, as well as finance and accounting. Sales forecasts also provide basis for regional and national distribution and replenishment plans. The importance of accurate sales forecasts to efficient inventory management at both disaggregated and aggregate levels has long been recognized the relationship between retail stocks and sales at the aggregate level and found that successful inventory management depends to a large extent on the accurate forecasting of retail sales.

Data mining techniques are very effective in tuning huge volume of data into useful information for cost prediction and sales forecast, it is the basic of sound budgeting At the organizational level, forecasts of sales are essential inputs to many decision making activities in various functional areas such as operations, marketing, sales, production and finance. The studies proceed with a new perspective that focuses on how to choose an appropriate approach to forecast sales with high degree of precision. Initial dataset considered in this research had a large number of entries, but the final dataset used for analysis having much smaller size compared to the original due to the riddance of non-usable data, redundant entries and irrelevant sales data. Data tuning process and predictions are highlighted with visual representation of generated results. The predictive analytics and methodology on sales price also discussed. The performance evaluations of various prediction algorithms using machine learning approaches are stated. Finally, the result is analysed and

concluded by summarizing the research reactions and upcoming possibility.

II. LITERATURE SURVEY

In paper [1], identifying the identifying new product issues, diagnosis of manufacturing problems causes and profiling existing customers with more accurate and tangible values. This huge collection of data values is either related or not related at all and thus definitely needs to be clustered otherwise much of the data deposited will not be useful to users. Increase the sales in season wise by updating the inventory, discount offers and store layout based on the knowledge discovered in the data.

In [2], author have analysed sales data with clustering algorithms like K-Means & EM which revealed many Interesting patterns useful for improving sales revenue and achieving higher sales volume. Our study Confirms that partition methods like K-Means & EM algorithms are better suited to analyse our sales data in comparison to Density based methods like DBSCAN & OPTICS or Hierarchical methods like COBWEB.

In [3], author presented a new data clustering method for data mining in large databases. Our simulation results show that the proposed novel clustering method performs better. In addition, in all the cases we studied, our method produces much smaller errors than both the FSOM+K-means approach and GKA.

The real amounts and seasonal factors are really important to some product ranges that are near the sales forecasts results. In [4], various campaigns and marketing approaches have been proposed for the sales of company products by evaluating the forecast results. Various campaigns and marketing approaches have been proposed for the sales of company products by evaluating the forecast results.

Paper [5] illustrates a novel trigger system that can match certain kinds of commodities with a prediction model to give better prediction results for different kinds of commodities. We find some related factors for classification. The results show that the accuracy of the trigger model is better than that of a single model. This has implications for business in that sellers can utilize the proposed system to effectively predict the sales of several commodities.

Paper [6] makes an attempt to solve this problem by taking a neural network approach, at the process of sales of footwear, and arriving at an optimum neural network model. The algorithms used for developing such model through neural network are both feedforward and recurrent Elman network. The data used in this work are the weekly sales of footwear and the information about the seasonality of sales process. While solving the problem, the focus is on forecasting of weekly retail sales as per the requirement of management.

This work would reduce the uncertainty existing in the short-term/middle term planning of sales and distribution logistics of footwear over different time horizons across the entire supply chain of footwear business.

The purpose of paper [7] is to compare the accuracy of various linear and nonlinear models for forecasting aggregate retail sales. Because of the strong seasonal fluctuations observed in the retail sales, several traditional seasonal forecasting methods such as the time series approach and the regression approach with seasonal dummy variables and trigonometric functions are employed. Regression models for predicting retail sales, the performance of dummy regression models may not be robust. Furthermore, trigonometric models are not useful in aggregate retail sales forecasting.

In [8], Ming Gao et al introduce the algorithm of ELM (extreme learning machine). In addition, we subjoin many e-commerce related indicators to increase the accuracy and reliability of prediction. In sum, the new model provides a better result both in terms of speed and accuracy. Experiments are conducted with the real sales data from an ecommerce company.

A two-stage dynamic forecasting model is proposed, which is combined with both long-term and short-term predictions. The model introduces the improved adjustment methods, the main adjustment model and error forecasting model in the adjustment system collaborated with each other. The real-time data are demonstrated by applying the model in wireless mobile environment. The experiment shows that the model provides good results for fashion retail forecasting [9].

In [10], the large database to mine information that is a Data mining process and convert it into a reasonable structure for further use. Launched and order to support the organization is decision making, business planning and Data mining techniques. Data analysis, increase profitability, innovation, efficiency in resource utilization is based on important management tool in data mining. The real amounts and seasonal factors are really important to some product ranges that are near the sales forecasts results. In this context, various campaigns and marketing approaches have been proposed for the sales of company products by evaluating the forecast results.

III. PROPOSED SYSTEM

We will analyse the time sequence sales data of a company and will predict the sales of the company for the coming quarter and for a specific product. For this kind of project of sales predict, we will apply the linear regression and logistic regression and evaluate the result based on the training, testing and validation set of the data.

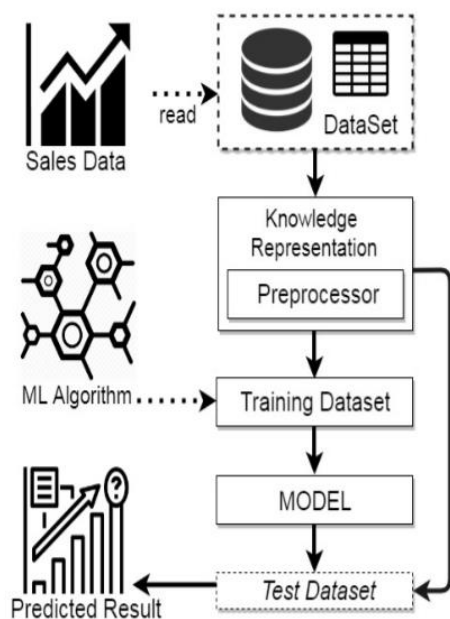


Fig system architecture

Machine learning is the study of computational and statistical methods that automate the process of knowledge acquisition from experiences. Our project has been conducted on sales forecast and some of them has been analysed. Different machine learning algorithms has been discussed that were applied to different sectors of the industry like retailing, logistic marketing etc. as per the requirement. Beneficial knowledge about machine learning techniques is mentioned in the paper.

IV. CONCLUSION

Individually and each company requirements to know the demand of the customer in any season previously to avoid the shortage of products. As time passes by, the demand of the businesses to be more accurate about the predictions will increase exponentially. So, huge study is going on in this sector to make accurate predictions of sales. Better predictions are directly comparative to the profit made by the company. Here in this paper, an effort has been made to predict sales of the product from a particular passage

accurately by using a two-level statistical model that reduces the mean entire error value. The two-level statistical model outperformed the other single model predictive techniques and donated better predictions to the big mart dataset.

V. REFERENCES

- [1] Maingi, M. N. A Survey on the Clustering Algorithms in Sales Data Mining.
- [2] Sastry, S. H., Babu, P., & Prasada, M. S. (2013). Analysis & Prediction of Sales Data in SAP-ERP System using Clustering Algorithms. ArXiv preprint arXiv: 1312.2678.
- [3] Tsai, C. F., Wu, H. C., & Tsai, C. W. (2002). A new data clustering approach for data mining in large databases. In *Parallel Architectures, Algorithms and Networks, 2002. I-SPAN'02. Proceedings. International Symposium on* (pp. 315-320). IEEE.
- [4] Shah, N., Solanki, M., Tambe, A., & Dhangar, D. Sales Prediction Using Effective Mining Techniques
- [5] Huang, W., Zhang, Q., Xu, W., Fu, H., Wang, M., & Liang, X. (2015). A Novel Trigger Model for Sales Prediction with Data Mining Techniques. *Data Science Journal*, 14.
- [6] Prasun Das and Subhasis Chaudhury, "Prediction of Retail Sales of Footwear using Feedforward and Recurrent Neural Networks", *Neural Computing and Applications*, Vol. 16(4), pp. 491-502, May, 2007.
- [7] Ching-Wu Chu and Guoqiang Peter Zhang, "A Comparative Study of Linear and Nonlinear Models for Aggregate Retail Sales Forecasting", *International Journal of Production Economics*, Vol. 86(3), pp. 217-231, December, 2003.
- [8] Ming Gao, Wei Xu and Hongjiao Fu "A Novel Forecasting Method for Large-scale Sales Prediction Using Extreme Learning Machine" 2014 Seventh International Joint Conference on Computational Sciences and Optimization.
- [9] Yanrong Ni, Feiya "A two-stage dynamic sales forecasting model for the fashion retail" Fan School of Mechanical Engineering, Shanghai Jiaotong University, Shanghai 200030, China
- [10] M. Jayasri, S. Santhoshkumar "Sales Prediction : Analysis of Time Series Data Using K-Means Based Smooth Subspace Clustering" 2018 Vol.4 Science and Technology