

# E-Commerce Website Sales Prediction using Machine Learning Algorithm

Mr. Ankush Y. Jondhale<sup>1</sup>, Mr. Mohid T. Khan<sup>2</sup>, Mr. Avinash B. Wagh<sup>3</sup>,  
Mr. Samarth<sup>4</sup>, Prof. A. S. Kolhe<sup>5</sup>

Department of Information Technology<sup>1,2,3,4,5</sup>

Matoshri Aasarabai Polytechnic, Eklahare, Nashik, Maharashtra, India

**Abstract:** *The business model used in the e-commerce web site, for which the model is implemented, includes many sellers that sell the same product at the same time at different prices where the company operates a market place model. The demand prediction for such a model should consider the price of the same product sold by competing sellers along the features of these sellers. In this study we first applied different regression algorithms for specific set of products of one department of a company that is one of the most popular online e-commerce companies. Supply and demand are two fundamental concepts of sellers and customers. Predicting demand accurately is critical for organizations in order to be able to make plans. In this paper, we propose a new approach for demand prediction on an e-commerce web site. The proposed model differs from earlier models in several ways.*

**Keywords:** game sales, k-Neighbor Regression, E-Commerce, Machine learning, Accuracy, Modal development, dataset, data Modification

## I. INTRODUCTION

Predicting right demand of a product is an important phenomenon in terms of space, time and money for the sellers. Sellers may have limited time or need to sell their products as soon as possible due to the storage and money restrictions. Therefore demand of a product depends on many factors such as price, popularity, time, space etc. Forecasting demand is being hard when the number of factors increases. Demand prediction is also closely related with seller revenue. If sellers store much more product than the demand then this may lead to surplus. On the other hand storing less product in order to save inventory costs when the product has high demand will cause less revenue. Because of these and many more reasons, demand forecasting has become an interesting and important topic for researchers in many areas such as water demand prediction

## II. PURPOSE

We have developed the modal to provide an prediction to most of sales company to focus on a accurate high demand product to make high profit over a amount of time which will leads to find an good product over an many product moving in the market. In current time its difficult to find which product is about to be in fame in next few months basically its simply possible by this modal. where the sales department of the company will give an good report to the company for farther work in next few week. it will easily decres the loss and incesse the profit

## III. LITRATURE SURVEY

Accuracy. According to the research conducted by [1], they have used the convolution neural network (CNN) algorithm to do sales forecasting in e-commerce. This research was being done to solve the identified limitation which was method require case-by-case manual feature engineering for specific scenarios which is difficult, time-consuming and requires a lot of expert knowledge. However, the goal for this research was to identify if this approach can automatically extract the effective features and provide the sales forecasting based on extracted features was mention by [1].

Based on the researches which was conducted by [2] and [3], they have both chosen neural network algorithm. But both of this neural network algorithm have their own approach where Nonlinear Autoregressive Neural Network (NARNN)

is used by the 2018 research and the 2019 research have conducted the research using Recurrent Neural Networks (RNN) and Long Short-Term Memory Networks (LSTM) algorithm which is a special neural network. These researches have used this approached algorithm to find sales prediction and demand of e-commerce. The problem that the researches have stated is similar which is difficulty in identifying the different cross-product demand/sales pattern and the correlations which are available.

Based on the other researcher [4], have conducted a research on forecasting of Walmart sales using machine learning algorithms. The key for this research was done by implementing several different classification algorithms in the sales data from all different Walmart locations all over the united states. The problem which was highlighted in this research was creating a competitive comparative analysis to find the best algorithm.

#### **IV. METHODOLOGY**

##### **Data Selection**

The dataset for this research will be a transactional data set. This is because we will be doing a sales prediction which require all the past transactional data to predict the future sales. The dataset transactional will be from one of the e-commerce which is open source and can be used without any restriction. The dataset which have been obtained is from Kaggle.com which have listed a Brazilian ECommerce Public Dataset by Olist Store (E-Commerce Site). There are about 100,000 transactional order history data provided..

##### **Data Preprocessing**

Data preprocessing is a process which explain how the selected data will be cleaned from all the noise or outliers. This means that cleaning up the data which have a huge amount of additional meaning which is meaning less and not required. For example, in this dataset there is product review and there is no need of product review in sales prediction, therefore the noise is the product review and it needs to be removed. Not only that, if the dataset has missing values for the sales and price values, we will need to handle it appropriately by replacing the missing value with the average value or use the mean or median imputation to keep the data consistent.

##### **Data Transformation**

Data transformation is a process of converting data from one format to another different format to satisfy the needs. This process also can be referred to the ETL process which means Extract, Transform and Load. Transformation have become a really important task as the data volume have increasing tremendously. Therefore, robust data transformation will allow user to focus on data which satisfy the business needs.

##### **Modelling**

Modelling is a process where you identify the algorithms which you are going to be using for the project research purpose. For this research, we will be using two different algorithms which are Gradient Boosting and Random Forest. The algorithms are being selected is because there are commonly related in prediction analysis. Gradient boosting is a machine learning technique which involves classification and regression to product based on weak prediction models such as decision tree. Random Forest contain a huge amount of decision tree that works in a group and each of the individual tree will provide a class prediction.

#### **V. CONCLUSION**

By doing this project of using machine learning for forecasting the ecommerce sales, it was noticed that the in this project, there are many different method of forecasting the sales of the ecommerce platform but the researcher was only able to focus on only four algorithms which are commonly being used when forecasting the sales of the future. The researcher was able to build and test all of the selected machine learning models which have been selected. The model which have the best prediction range, where the predicted value and the actual value are almost similar is chosen as the best algorithm. This best algorithm will then be integrated into a web application which will also built by the researcher.

**REFERENCES**

- [1] Smt. Rajeshwari and M. Shettar, "Emerging Trends of E-Commerce in India: An Empirical Study", *International Journal of Business and Management Invention ISSN (Online)*, vol. 5, no. 9, pp. 2319-8028, September. 2016..
- [2] E.W.T. Ngai and A. Gunasekaran, "A Review for Mobile Commerce Research and Applications", *Decision Support Systems*, pp. 43-15, 2007.
- [3] S. H. Sastry, P. Babu and M. S. Prasada, "Analysis & Prediction of Sales Data in SAP-ERP System using Clustering Algorithms", *International Journal of Computational Science and Information Technology (IJCSITY)*, vol. 1, no. 4, November 2013.
- [4] T Chaithralaxmi and N Shruthi, "E-Commerce in India - Opportunities and Challenges", *International Journal of Latest Trends in Engineering and Technology Special Issue SACAIM 2016*, pp. 505-510, 2015.
- [5] S. K. and M. L. Saini, "The Analytical Study of an E-Business Model for Establishing the Concept of Customer Retention (A case study of Myntra.com)", *(August 29 2019) International Conference on Advancements in Computing & Management (ICACM)*, 2019.
- [6] Jen-Hao, Chao-Wen Huang and Chia-Chun Shih, "The Exploration of Machine Learning for Abnormal Prediction Model of Telecom Business Support System", *19th Asia-Pacific Network Operations and Management Symposium (APNOMS) IEEE (2017)*, November 2017.
- [7] S. Kulshrestha and M. L. Saini, "Establishing customer loyalty by analyzing selling rate of automobile spare parts through E-commerce", *Global Journal of Engineering Science and Research Management*, July 2018, ISSN 2349-4506.