

Study for the Prediction of E-Commerce Business Market Growth Using Machine Learning Algorithm

Shilpi Kulshrestha

Department of Engineering & Technology
JaganNath University
Jaipur, India
shilpi.kuls2303@gmail.com

M. L. Saini

Department of Computer Engineering
Poornima University
Jaipur, India
mlsaini@gmail.com

Abstract— Learning is a key to perform ideas adequately. Machine Learning empowers IT organizations to identify the patterns on the basis of currently available algorithms and data frames to cultivate acceptable solution concepts. Online business market and customer retention is a relation like the two sides of a coin. It is a nonlinear relationship. Prediction of Business growth is a very sensitive issue of E-Commerce market with its future existence. Online vendors of business market manage their inventories on virtual prediction bases for full filling the basic need of demand-supply chain of customers. Authorizing traditional ways and analysis methods are not ensuring the rate of reliability of the sales prediction. To produce more precise predictions and analysis, we use ML algorithm. In this paper, we utilized the selling data set of an E-commerce company and segregated it, in different quarters then calculating the sale income per quarter. After that we divided the dataset in the proportion of 70% and 30% for Training data set and Testing data set. By applying machine learning algorithm, we will be predicting income of next quarters as well as analysis the maximally sold commodities with their frequencies of purchase per quarter. Then provide analysis results and prediction of customer's purchase patterns to the business organization to make a strategy to take a competitive advantage by sustaining and accumulating for their goods management and planning for inventories.

Keywords—QoS, Machine Learning, E-commerce, Inventory, Prediction, Customers Retention

I. INTRODUCTION

Online E-Commerce market shows an imperative contribution in speedy commercial expansion in the revenue development of our country. So same as other country's progress rate in mercantile, could also base on the revenue of sales and purchase in online market of E-Commerce [1]. In the era of internet and fast speed, probably in all the countries, more than 50% of the population engaged with online shopping and ordering goods because of the dynamic and spontaneous purchasing market with in one click only. As time changes, human's starts evolving different type of market structures and several ways of selling – purchasing techniques. This online technique is very easy, convenient, less expensive compare to traditional methods, fast and no time boundation with easy return policies [2]. All these facilities and many more options are available for the online/e-shopping customers but on the other side of coin, online service and goods providing vendors, they are facing

many problems too. In current situations and scenarios, Market-competition is on the highest point, especially online business market strategies. These E-business organizations involve in strive to find out better and innovative ways to make a scheme to take a competitive advantage. For sustaining and accumulating goods and inventories with some types of prediction or managing ideas [3].

The techniques of prediction in business, especially for online E-commerce market vendors can shows a critical role in their capital investments for inventory. If the predicted results are more favorable to vendors and shop-keepers, then they make good profits out of that. Even they can study the mindset or the patterns of the purchasing habits of customers. Many online and offline tools are available which associated with Data mining techniques. Which ready to help ERP model to predict consumer's behaviors and future tendencies in dynamic business clarifications to knowledge driven decisions [2][4]. Basically with the use of these artificial Intelligent data analysis models or tools creates results and unseen analysis. which we are not able to produce by traditional methods, it's hiding information beyond expert's predictions [5]. Earlier data extraction techniques are unfamiliar. Generally, people don't know the impact of that, but it has potential and this is use for forming beneficial facts and results from data sets. In general, databases addresses are used and this process is discovered as knowledge discovery (KDD) from data sets [6].

II. RELATED WORKS

In business, random predictions are very problematic and confusing because numbers of online orders and purchasing are very uncertain, it's totally depends on customers purchasing trends and behavior. Even it is also time variant means, according to different seasons and festivals. In some last decays' numerous scholars and data scientists have used innovative and interesting ways for sales prediction methods for business trading decisions. Here, I try to explain a short review of the substantial investigators and scholars.

Dai Wei, Peng Geng, Liu Ying [7] combines two areas, E-commerce website search data and structure time series model (STS) specially to forecast the sales volume for women's clothing of company "Taobao clothes Suppliers and Manufacturers". With their terminology, they achieved good results of predictions. Their work is segregated in three parts. Firstly, try to setup of combination of web search data and structure time series model, and fit the trend and seasonal

factor from the volume of sales. Then next they start comparing with traditional forecasting methods, so that they can replicate the variations of external factors through the consumers' web search behavior. Dai Wei's research represent the correlation between search index and e-commerce transactions.

Working scheme in [8] at shows an analysis of the behavior pattern of purchasing items and create a strategy to hold customers in an endless loop of purchasing and discounts to give consecutive profit growth in business with the scheme of customer retention.

Economic forecaster Mr. Harry Dent, apply an approach of common sense. Analyzing these predictable trends generate accurate long-term forecasts which can be utilized by both businesses and consumers. He always emphasized the importance of demographics [9].

Jen-Hao, Chao-Wen Huang, Chia-Chun Shih were explore several tools and machine learning algorithms. They test and compared many algorithms like linear SVM, SVM with RBF kernel, random forest and fully connected neural network, to construct a prediction model for online system failure. Their model can predict 15% of the failure occurred advance by 60 minutes, which is a very good attempt and its useful for business. They used F-score to measure the model [10].

An approach given in [11] as for earning growth rate in a specific business, by make some amendment in ERP online models. she's suggest an improve procedure of serving with our customers as compare to the normal method of goods and items deliveries. By adopting this procedure, company can improve in the sale of precise products mention in paper, customer relationship, profit increments and buyers 'retention'.

A different approach in [12] used different approach such as intelligent sales prediction system to address this problem. This strategy is used for industry to manage vast capacity of records. Major decisions of business are generally based on accuracy and speed of selling – purchasing of data processing techniques. In Sunitha Cheriyan's paper, she evaluates around 80,000 to 90,000 data records. After then, she did a comparison between ML processes and algorithms to seek better results. Her team concisely evaluate with the fact and figures for forecast and data of sales. After analyzing the performance, they created a model for forecast of sales & trend. The evaluated results are precise on the majors of trustworthiness and exactness of efficient techniques (Gradient Boost Algorithm) taken for forecasting.

III. PREDICATION METHOD

Time series prediction or forecasting is an imperative zone of Machine Learning. In many fields, Time series analysis and forecasting is very crucial because we have so many prediction issues and complication that include a time component or a time series. These types of problems are more complicated to solve and problematic to handle. Basically these types of series contain data points, which are successive in nature and we can have mapped these points at progressive interval of time. it grouped the functions that attempt to suspicion a time series in terms of understanding either the underlying concept of the data points in the time series or signifying or creating predictions [5]. For

forecasting generally, we use data analysis and algorithms on the behalf of sell and purchase history and past outcomes. With the help of some substantial Time series model We can predict better future assumptions. Mostly identified time-series models, which are analyzed by me for evaluating a better prediction approach are explained as follows with forecasting equations and smoothing parameter (constants).

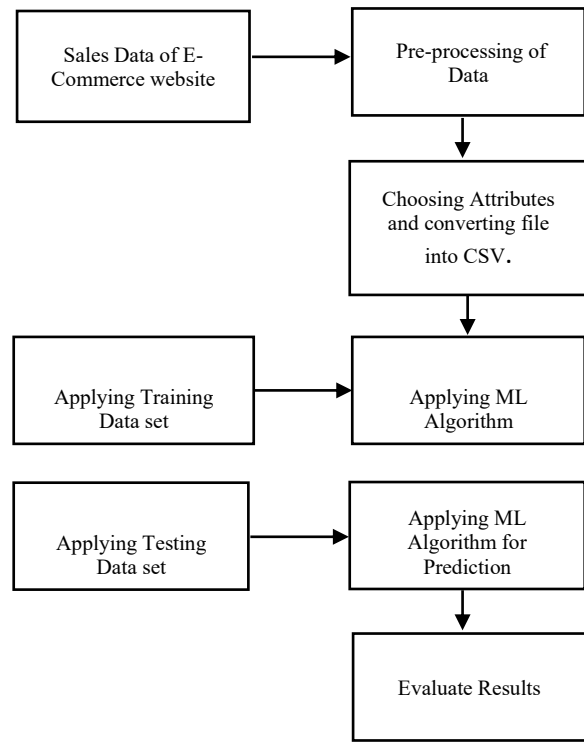


Fig. 1. Flow Diagram for Prediction Sales values

A. Naive Approach:

In this type of predicting approach, we assume that the next upcoming data point is equal to the last point of observation is known as Naive Method [13]. We use Y_t and Y_{t+1} periods for forecasting in naive approach. Hence $Y_{t+1} = Y_t$.

B. Moving Average:

It is a common approach, as named Simple Moving Average; we take an average of data points in the set of time series. In MA, it predicts the expected values by calculating the average of all the past data observations. we can determine the Moving Average by two techniques, first is Simple Moving Average (SMA), Hence $Y_{x+1} = Y_1 + Y_2 + \dots + Y_n / n$. and second type is Exponential Moving Average (EMA). $Y = 1/m (W_1 * Y_{i-1} + W_2 * Y_{i-2} + \dots + W_m * Y_{i-m})$. Probably these methods of forecasting are not best for predicting future entities but for the essential conditions, where these approaches set perfectly and show the best results.

C. Simple Exponential Smoothing:

In time series data, Exponential Smoothing allocates more and more rapidly decreasing values for data points in the sequence of latest to eldest formations. In another understanding, as the data observation become older, the

weightage of that observations starts decreasing exponentially, priorities become lesser. In comparison of old data point, new data observations perceived more reliable and relevant. According to relevancy of new data observations, are assigned more load. Hence $S_t = \alpha Y_{t-1} + (1 - \alpha) S_{t-1}$. Here we use a constant α to identifying the smoothing parameter (constant). The use of α is to determine the loads for data points. The process of smoothing become very slow, when the observations of α (smoothing parameter) is nearer to 0. The range of constant α is from 0 to 1, and t stands for time period. The technique of Exponential smoothing is more used to smooth out data or do preprocessing of data to make better predictions [14]. Generally, this technique is used, when we have the data with unclear or random data patterns for forecast. For example, it is commonly used for economics and finance for short term predictions, because this approach is not quite untrustworthy for extended time period forecasting.

D. Holt's Linear Trend Model:

After developing the technique of forecasting (SES) as denoted as "Simple Exponential Smoothing", Dr. Holt extended this approach to allow the prediction of data points with a new component "trends". In this new approach, he develops a forecasting equation and two new equations for smoothing parameter (constant), one is regrading level and other is trend. Forecasting equation $Y_{t+h|t} = l_t + hb_t$, Equation of level $l_t = \alpha y_t + (1 - \alpha) (l_{t-1} + b_{t-1})$ and equation of Trend is as $b_t = \beta (l_t - l_{t-1}) + (1 - \beta) b_{t-1}$. Here l_t define the estimated value of the level of the series for data at the time t , and b_t shows an estimate of slope or trend regarding time series at time t . Now here α represent as a smoothing parameter for the level in the range of, $0 \leq \alpha \leq 1$, and β represent as second smoothing constant for the trend in the range of, $0 \leq \beta \leq 1$ [15]. In holt predicting method, we find forecasting function with a new quality of trending. Now this function is h-step-ahead forecast and is equal to the last estimated level plus h times the last estimated trend value.

E. Holt's Linear Trend Model on daily time series:

After achieving forecasting technique of Holt's Linear Trend Model, Dr. Holt and Winters added one more component with this approach to predict more accurately with the available past experiences and data observations. The new introduced component is "seasonality". This Predicting technique is also called Holt-Winters seasonal method. In this approach, they have one Forecasting c and three equations for smoothing parameter (constant), First is corresponding level and second is trend and latest parameter is seasonality with regarding the symbols of smoothing constant α , β and γ . Equations are

$$\begin{aligned} Y_{t+h|t} &= l_t + hb_t + S_{t+h-m(k+1)} \\ l_t &= \alpha (Y_t - S_{t-m}) + (1 - \alpha) (l_{t-1} + b_{t-1}) \\ b_t &= \beta (l_t - l_{t-1}) + (1 - \beta) b_{t-1} \\ S_t &= \gamma (Y_t - l_{t-1} - b_{t-1}) + (1 - \gamma) S_{t-m}, \end{aligned}$$

Here for seasonality, m depicts the frequency. In other words, frequency of seasonality describes the seasons in a year [15]. For an example, in a one year, we have 12 months so the value of m become 12 on monthly bases, but if we calculating on the bases of quarterly then it's become $m=4$. In first equation of forecasting, constant K represent the

integer part and confirms the proper estimate of the seasonal indices. Which are used for predicting values associated with the given sample of data observations for the final year. The equation of level denotes a weighted average regarding of time t in between non seasonal predictions ($l_{t-1} + b_{t-1}$) and seasonal data points ($Y_t - S_{t-m}$). The equation of trend is same as previous as in Holt's method. Point of attraction is the equation of seasonality. It describes the combination of weighted average regarding of time interval m , based on situation in between indexes of current year and previous year [15].

In our project, we are trying to find out a method or a technique to predict sales forecasting for venders, which help them to manage their stock and inventory.

IV. RESEARCH METHODOLOGY

The prospective of our research work is to appraise, improve and examine the method of Machine learning with respect of Time series analysis. This project is inclusive results for online sales prediction.

A. Data Collection and Preparation

This paper is based on the study of online business market and prediction techniques for the growth and earning profits from E-commerce portals through observing and analyzing past data observations [16]. So we used the dataset of an e-commerce based business website. In dataset, we contain Purchase dates, City, Customer ID, Customer Name, Product Category, Product Sub Category, Product, Quantity and Total Amount. we have broadly 7 product categories. These are as following:

- Kids Clothing,
- Mobile Phone Accessories,
- Personalized Gifts,
- Women's Clothing,
- Man fashion Accessories,
- Men's Clothing,
- Woman fashion Accessories,

For the consecutive years of sales data were collected for analysis and use of forecasting. Initially our fetched data set contained a good numbers of tuples entries, but after the removal of non-required data attributes, inappropriate and redundant data entries, it becomes smaller in dimension with associated the basic dataset provided by company [17].

B. Data Analysis

After the first step of data collection, we concentrate on data pre-processing, we did analysis of data. Our paper is used to evaluate the online behaviour aspect of customer purchasing through online web site of E-commerce. It will help us to understand the working of ERP model of online selling and will construct the outline scenario, technical aspect and another needs for developing a ERP (Enterprise Resource Planning) project [18]. For provide analysis results and prediction of customer's purchase patterns to the business organization to make a strategy to take a competitive advantage by sustaining and accumulating the goods and their inventories for the behalf of online venders and customer retention [19]. Table 1 shows the analysis of the collected sales amount data per quarter.

TABLE 1: SALES AMOUNT PER QUARTER

Sr. No	Quarter	Total Amount
1	Qtr-1	228153
2	Qtr-2	392276
3	Qtr-3	393327
4	Qtr-4	492057

This table describe that quarters are represented as 1,2,3,4 of three-three months respectively from December 2018 to November 2019. As regarding fourth quarter of 2019, total sales amount is increased and first quarter of 2018 appear a decrease in the total sales amount.

Table 2 describe the ratio of Unique customer verses Repeated customer per quarter respectively. Maximum repetition is shown in quarter 4 with increasing rate and minimum in quarter.

TABLE 2: UNIQUE/REPEATED USERS

Customers Types	Quarters			
	1	2	3	4
Unique	287	378	475	693
Repeat	037	035	052	084
Total	324	413	527	777

Below figure describe the generated sales amount in corresponding years and it shows the seasonality trend and high sales in month of October and November, reasonably occurrence of Indian festivals.

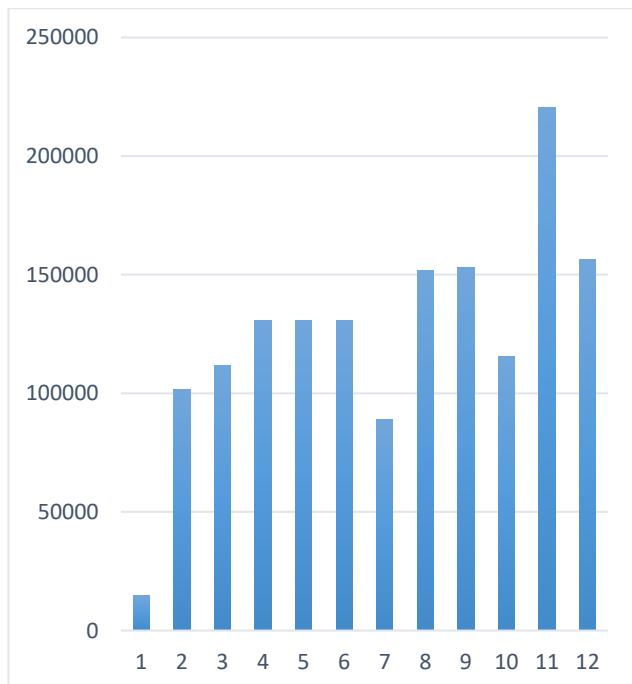


Fig. 2. Sales Amount per month

TABLE 3: PURCHASED QUANTITY BY USERS OF SALES COMMODITIES / QUARTER

Product Category	Quarters			
	1	2	3	4
Mobile Accessories	150	159	216	320
Women's Clothing	106	171	219	317
Men's Clothing	19	29	24	49
Kids Clothing	15	21	19	30
Gadgets & Accessories	15	11	20	20
Men Accessories	08	13	10	13
Women Accessories	07	05	13	17
Personalized Gifts	04	04	06	11
Total	324	413	527	777

Table 3 denotes the 7 main sales commodities and the analysis of selling history of these commodities per quarter. According the table mobile accessories and Women's Clothing are the maximum selling commodities.

C. Prediction and Trend

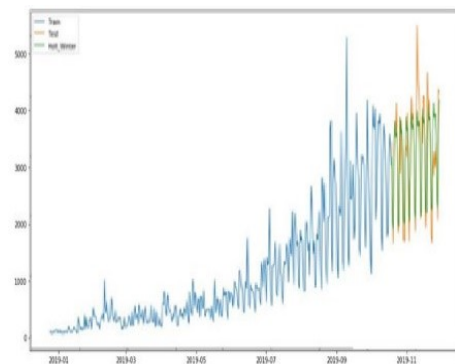


Fig. 3. Data prediction graph month wise in Python

TABLE 4: PREDICTED SALES AMOUNT VALUES FOR ONE YEAR THROUGH MACHINE LEARNING ALGORITHM

Sr. No	Months	Sale Amount
1	Dec-19	161432.6
2	Jan-20	151327.2
3	Feb-20	134270.5
4	Mar-20	131581.7
5	Apr-20	119726.0
6	May-20	137211.0
7	Jun-20	126923.4
8	Jul-20	105953.3
9	Aug-20	123025.8
10	Sep-20	108569.1
11	Oct-20	118803.8
12	Nov-20	182618.8

With the help of Machine learning algorithms, Holt's Linear Trend Model from Time series analysis, can predict values for future forecasting. These values can be helpful for business. In figure 3, it shows the increasing upward graph. So we predict values for next one year.

Above table shows the predicted sales amount values for one year through Machine Learning Algorithm. By using concepts and algorithms of ML, we can create advance and intelligent model or architecture by without any physical interference. "Machine Learning (ML) is used to optimize the performance criterion using sample data or the past experience" as stated by Ethem Alpaydin [20].

With the help ML algorithms, we can predict values for more than one years, according to our need and convenience. By comparing the prediction results of a year and the sales data of last year, we found in forecasting outcomes and assumptions, basic growth in sales values.

V. RESULTS AND ANALYSIS

Regarding this paper, According the forecasting results of the prediction algorithm, it denotes the profitable outcomes in comparison of previous sales data of E-commerce website. It increases by the rate of 6.7%. Predicted values reflects the highest sale amounts in months of November, December, January. With the help of these values, online venders of website able to manage their mill productions, can plan no of workers, purchasing inventory, stocks and as well as purchasing of raw materials [21]. As we already explain, in data analysis section, maximum and minimum selling commodities. So now, venders have much better understanding of online business requirements and customer needs because according to predicted values which are calculated by three main majors trend, level and seasonality provides approximate time durations, stock quantity and selling commodities. With this method they able to find out, when they have to increase their production as well as for which duration. Venders can also find out the which commodities gives maximum profit and which is not in trend, minimally purchased by customers [22]. So they will invest their money accordingly and wisely. In fact, with this approach ERP systems are always ready to distribute or supply products to their customers. After using this technique, if venders are in a satisfied and profitable state then they can provide more offers, discounts, membership cards and good qualities products to customers. By availing offers and discount, customers can also show more interest to purchase items from our website and shows the properties of customer retention too.

VI. CONCLUSION

The method of forecasting for predicting sales values is very beneficial towards ERP system and as well as for customers' demands. With the help of predicted assumptions, venders can evaluate production, storage, transportation allowances and can be prepare himself as per the requirements of buyers and mostly importantly the storage capacity of a warehouse. Biggest benefit with the approach that we can avoid the problem of overstocking or understocking of goods and could plan according for future investments, budget and the strength of employees and labors for mill production.

Prediction deals with events occurring in the future. Many researchers and developers suggest that an intellectual and smart sales forecasting model or a procedure is required for obtaining the best results of business growth. With the help of that system, companies and organizations able to optimize the gigantic volume of data and can understand the power of data analysing and prediction methods [19]. Regarding the ever increasing popularity and usage of E-commerce websites or mobile apps, it gets extremely critical to focus on existing customer base. It is expected that majority, if not all, in the coming 8-10 years, businesses would have an online presence to sell their goods and services to customers. This competition of online world would demand companies to regularly revise their planing and thought processes of sales management, inventory creation and customer retention strategies. In conclusion, with the help of machine learning algorithm and its predicted values, companies and venders can take benifites and mold their business with a better approach to generate growth, satisfaction and a new vision to earn profit and sustainability.

REFERENCES

- [1]. Dr. (Smt.) Rajeshwari, M. Shettar, "Emerging Trends of E-Commerce in India: An Empirical Study" International Journal of Business and Management Invention ISSN (Online): 2319 – 8028, ISSN (Print): 2319 – 801X www.ijbmi.org || Volume 5 Issue 9 September. 2016 || PP—25-31
- [2]. E.W.T. Ngai, A. Gunasekaran, "A Review for Mobile Commerce Research and Applications" www .elsevier.com /locate/dss, Decision Support Systems 43 – 15, (2007)
- [3]. Chaithralaxmi.T and Shruthi. N "E-Commerce in India – Opportunities and Challenges" International Journal of Latest Trends in Engineering and Technology Special Issue SACAIM 2016, pp. 505-510 e-ISSN: (2278-621X), 2015
- [4]. S. H. Sastry, P. Babu, & 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).
- [5]. N. Shah, M. Solanki, A. Tambe, & D. Dhangar, "Sales Prediction Using Effective Mining Techniques", International Journal of Computer Science and Information Technologies (IJSIT) (2015)
- [6]. M. H. Dunham, & S. Sridhar, "Data Mining: Introductory and Advanced Topics", Pearson Education, New Delhi, ISBN: 81-7758-785-4, 1st Edition. (2006)
- [7]. Dai Wei1, Peng Geng1, Liu Ying1, Li Shuaipeng "A Prediction Study on E-commerce Sales Based on Structure Time Series Model and Web Search Data" 26th Chinese Control and Decision Conference (CCDC), IEEE (2014)
- [8]. S. K., Dr. 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) SSRN: <https://ssrn.com/abstract=3444797>
- [9]. Grady Maguire "E-Commerce: A Statistical Market Analysis and Forecast of Emerging Trends" International Journal of Management, Technology and Engineering, Page (41-49)
- [10]. Jen-Hao, Chao-Wen Huang, 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) (Added to IEEE Xplore: 02 November 2017)

- [11]. S. Kulshrestha, Dr. 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, ISSN 2349-4506, Impact Factor: 3.799, (5 July 2018)
- [12]. Sunitha Cheriyan ; Shaniba Ibrahim ; Saju Mohanan ; Susan Treesa “Intelligent Sales Prediction Using Machine Learning Techniques” International Conference on Computing, Electronics & Communications Engineering (iCCECE), IEEE, (2018) (Added to IEEE Xplore: 07 March 2019)
- [13]. G. Zhu, C. Zheng, Hui Hu, W. Guan. “A Kind of Demand Forecasting Model Based on Holt-Winters Model and Customer-credit Evaluation Model” International Conference on Service Systems and Service Management, 26 Feb 2007, IEEE, (2006)
- [14]. A. Sharaff, M. Choudhary “Comparative Analysis of Various Stock Prediction Technique” 2nd International Conference on Trends in Electronics and informatics (ICOEI), IEEE (2018) INSPEC Accession Number: 18290885 (Added to IEEE Xplore: 03 Dec 2018)
- [15]. Deetchiga, U. Kabila harini, M. Marimuthu, J. Radhika “Prediction of Passenger Traffic for Global Airport using Holt’s Winter Method in Time Series Analysis” International Conference on Intelligent Computing and Communication for Smart World (I2C2SW), IEEE (2018) (Added to IEEE Xplore: 13 February 2020)
- [16]. K. Asooja, G. Bordea, G. Vulcu, & P. Buitelaar, “Forecasting Emerging Trends from Scientific Literature”. In LREC. (2016) <https://www.aclweb.org/anthology/L16-1066.pdf>
- [17]. T. D. Rey, C. Wells, & J. Kaulh, “Using data mining in forecasting problems.” In SAS Global Forum 2013: Data Mining and Text Analytics (2013)
- [18]. Huei Huang Kuan, Gee-Woo Bock, Vichita Vathanophas, “Conversion and Retention at E-Commerce Websites”, Proceedings of the 38th Hawaii International Conference on System Sciences, 0-7695-2268-8/05/\$20.00 (C) 2005 IEEE
- [19]. Vicky Chrystian Sugiarto, Riyanarto Sarno , Dwi Sunaryono “Sales Forecasting Using Holt-Winters in Enterprise Resource Planning At Sales and Distribution Module” 2016 IEEE International Conference on Information & Communication Technology and Systems (ICTS) INSPEC Accession Number: 16839589 (Added to IEEE Xplore: 27 April 2017)
- [20]. Ethem Alpaydin. “Introduction to Machine Learning” (Adaptive Computation and Machine Learning), https://kkpatel7.files.wordpress.com/2015/04/alpaydin_machinelearning_2010.pdf , The MIT Press (2015).
- [21]. S. H. Liao, Y. C. Chung, Y. R. Hung, R. Widowati “The impacts of brand trust, customer satisfaction, and brand loyalty on word-of-mouth” IEEE International Conference on Industrial Engineering and Engineering Management INSPEC Accession Number: 11705957, IEEE (2010) (Added to IEEE Xplore: 23 December 2010)
- [22]. S. Tiwari, A. Bharadwaj, Sudha Gupta “Stock Price Prediction using Data Analytics” IEEE International Conference on Advance Computing, Communication and Control (ICAC3) INSPEC Accession Number: 17634101, IEEE (2017) (Added to IEEE Xplore: 19 March 2018)