

# Sale Prediction System for Industrial Revolution Using Machine Learning

**Akansha Pandey, Mehul Mahrishi**

Department of Computer Science & Engineering, Swami Keshvanand Institute of Technology,  
Management & Gramothan Jaipur-302017, (INDIA)

*Email: akansha@ppandey027@gmail.com*

Received 16.07.2019 received in revised form 27.07.2019, accepted 30.07.2019

**Abstract-** This research paper provides an algorithm which deals in sales forecasting to enable sales organizations to make business decisions and predict sales forecast for short-term and long-term performance. This research work is to plan a model to foresee the stock through Machine Learning. In this, we have old sales data of products of the stores and from that sales data, we will predict the future sale of items. This research is carried out as companies can base their forecasts on the previous sales data. The target of this research work is to improve forecasting practices. The hybrid approach sales prediction is presented to get more accurate results. This work has been implemented in Python using Machine Learning models. Different models are implemented using data set available on the Kaggle.com, for research work to be done by the researchers for better-predicted results. The results are accurate due to the hybrid approach of Auto Regression Moving Average, Rolling Forecast and Exponential Smoothing algorithm as tested on the data available for 2017. The goal of this research work aims the performance of the organization—more profit, more revenue, financial improvements. The result introduced in this research recommends that by utilizing the hybrid approach, forecasting will deliver request figures that are considerably more precise than those given by the well-known algorithm.

**Keywords** – Sales Forecasting, Machine Learning, Auto Regression Moving Average, Rolling Forecast, Exponential Smoothing

## 1. INTRODUCTION

Machine learning is the computerized reasoning of artificial intelligence that enables the creation of artificially intelligent machines and programs [1]. Because of the purchase out approach of requesting, franchisee store chiefs perform request forecasting and stock control dependent on their business execution and experience. In this examination, relapse investigation is utilized with the feed-forward neural system as a forecasting apparatus [2]. The dataset used for the simulation is accessible on kaggle.com [3]. We are furnished with every day chronicled sales information. The errand is to estimate the aggregate sum of items sold each shop for the test set. Time series forecast for financial procedures is a point of expanding interest. The techniques previously

studied by the sales predictors are Simple Moving Average method, Exponential Smoothing method, auto-regressive integrated moving average, Neural Network algorithm, Croston forecasting method [4-6]. The difficulties with traditional forecasting methods are listed below.

### 1.1 Simple Moving Average

- Forecast based on mean computed for most recent observations.
- Equal weights are assigned to each past observation.

### 1.2 Exponential Smoothing

- Recent observations have a higher influence on the forecast.
- Exponentially decreasing weights are assigned to observations.

### 1.3 ARIMA

- Makes a prediction based on previous  $p$  points (auto-regressive) and on the previous  $q$  residuals (moving-average) [4].
- Inappropriate for Intermittent data as it allows values that aren't non-negative integers [7].

## 2. PROBLEM DEFINITION

In this research work, Autoregressive and Exponential Smoothing (SES) are utilized in a hybrid approach by utilizing the ideas of Machine Learning for a transient figure of the clearance of articles in stores. These methodologies are utilized beforehand in past research works. Presently the time has come to execute the crossbreed approach.

AI is the utilization of man-made brainpower (Artificial Intelligence) that enables systems to thus take in and improve for a reality without being unequivocally tweaked[8]

- To predict future sales to use it as the basis of planning time and resources.
- To look after stock – A precise deals anticipating bits of help in evaluating the measure of stock required for future

objectives. It helps in keeping the stock up for pinnacle periods [9-10].

- It aids in distinguishing Clients and Timelines: - Notwithstanding what is selling, an exact deals figure can recognize who is purchasing the most and when they are purchasing. On the off chance that our deals are regular, we know which months are

moderate and can be utilized for prospecting. [11]

**3. LITERATURE REVIEW**

Forecasting is one of the significant parts of the organization. Many researchers are working on sales predictions for better results. Table 1 summarises the literature survey of existing research in prediction and forecasting sales.

**Table 1:** Literature Survey

AUTHOR	PAPER TITLE	PAPER CONCLUSION
Michael Giering(2008)[12]	Retail salesprediction and item recommendation using customer demographics at store level	Results from test store over one year time period showed that the framework realized generally extended arrangements and improved efficiencies. Through the forecast, it exhibits that business Increase by 2-3% every year.
Xiao Fang Du, Stephen C.H. Leung, Jin Long Zhang(2011) [13]	Demand forecasting of perishable farm products using support vector machine	Examination of trial results demonstrated that it is worthwhile to apply SVMs forecasting system in perishable farm products demand forecasting.
Taieb, S.B.; Bontempi, G.; Atiya, A.F.; Sorjamaa, A.(2012) [14]	A review and comparison of strategies for multi-step ahead time series forecasting based on the NN5 forecasting competition	Gave various methodologies for multi-venture ahead time arrangement anticipating are considered and looked at. Distinctive anticipating strategies joining have been examined
SamanehBeheshti-Kashi(2014) [15]	A survey on retail sales forecasting and prediction in fashion markets	The outcomes demonstrated that the best neural system structure for demand. Forecasting is the Multilayer Perception which is coincidental, the most generally utilized model.
IlhamSlimani; Ilhame El Farissi (2015) [16]	Artificial neural networks for demand forecasting: Application using Moroccan supermarket data.	Found the exactness of offers forecasts in-store network to be considered as a significant key. In the information, for example, exchanges' information of gatherings, sales and delay purchases of the perceived general store was taken. From the accessible interest information, they removed the day by day request amounts got of the pertinent item for a half year. In this study, the Artificial Neural Network was considered to foresee the interest of a client's item.
Jamal Fattah, LatifaEzzine, Zineb Aman(2018) [4]	Forecasting of demand using the ARIMA model	The outcomes got demonstrated that the model could be used to show and gauge the future excitement for this sustenance conveying. These outcomes were given to executives of this amassing reliable standard in choosing.
Maobin Li, Shouwen Ji and Gang Liu(2018) [5]	" Forecasting of Chinese E-Commerce Sales: An Empirical Comparison of ARIMA", Nonlinear Autoregressive Neural Network, and a Combined ARIMA-NARNN Model"	Recommended that with the quick improvement of online business (EC) and shopping on the web, definite and compelling forecasting of electronic business sales (ECS) is huge for making procedures for purchasing and supply of EC attempts.
Papacharalampous, TG. Tyralis, TH. Koutsoyianni(2018) [17]	D. Univariate time series forecasting of temperature and precipitation with a focus on machine learning algorithms: A multiple-case study from Greece.	Researched the consistency of time arrangement, and concentrate the presentation of various time arrangement estimating techniques.
Bohdan M. Pavlyshenko (2019) [6]		Gave the use of AI models for sales prescient examination. The essential target of their examination was to think about standard approaches and relevant investigations of using AI for sales forecasting

**4. STATISTICAL TECHNIQUES USED**

In this investigation, the relapse examination is utilized with the feed-forward neural system as a determining instrument [5]. The dataset containing prepared and test information are accessible on kaggle.com. We are given every day authentic deals information. The errand is to conjecture the

aggregate sum of items sold in each shop for the test set.

**4.1 Auto Regression Analysis**

Regression analysis is a factual strategy utilized for deciding the connection between a solitary ward (measure) variable and at least one autonomous (indicator) factors [5]. In the proposed investigation

the needy variable is predicted sale and independent variable is month and year.

For the proposed work Python will be used as the programming language in perspective on its versatility. There are two basic ways to deal with perform straight backslide in Python—with Stats-models and Scikit-learn.

**4.2 Exponential smoothing**

Exponentialsmoothing techniques areutilized in this exploration work. This strategy gives bigger loads to perceptions, and the loads decline exponentially as the perceptions become progressively removed [18]. These techniques are best when the parameters portraying the time arrangement are changing gradually after some time.

**Table 2:** For item code 956014, prediction for - Jan, 2017

Year	Actual Sale [20]	Predicted sale for 2017 Simple Linear Regression	predicted sale using exponential smoothing	Linear Regression values based on combine list of monthly sale using Rolling Forecast with Exponential Smoothing Method	Final Predicted Value of 2017 using mean (predicted sale of 2017 using exponential smoothing + Linear Regression value of 2017 based on combine list of monthly sale using Rolling Forecast with Exponential Smoothing Method)/2
2013	1146	1199	1146	1193	NA
2014	1190	1181	1190	1181	NA
2015	1304	1162	1155	1169	NA
2016	1046	1144	1185	1156	NA
2017	1186	1124	1157	1144	1150.6000000000004 Or 1151

**Forecasting Formula**

*Forecasting the next point*

The forecasting formula is the basic equation  

$$St+1 = \alpha Yt + (1-\alpha)St, 0 < \alpha \leq 1, t > 0.$$

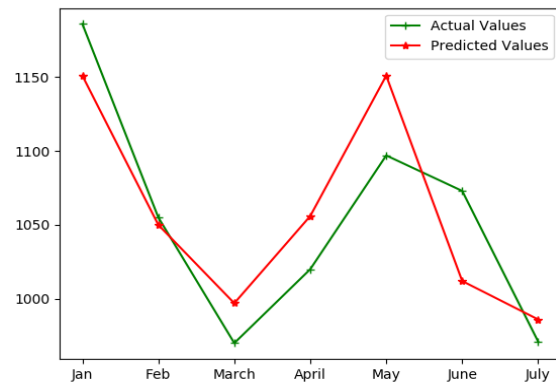
*New forecast is previous forecast plus an error adjustments*

This can be written as:  

$$St+1 = St + \alpha ct,$$

where ct is the forecast error (actual - forecast) for period t.

In other words, the new forecast is the old one plus an adjustment for the error that occurred in the last forecast.



**Figure 1:** The predicted results are compared with the data set available for 2017

**5. EXPERIMENTAL RESULTS**

Simulator for Sales Forecasting is used to get predicted sales for a particular month and year.

To accurately get insights and to find new practical approaches, it is important to find errors in previous approaches. The errors found in the previous approaches are given in the following table, we have used the hybrid approach.

**Table 3:** Results Comparisons

Model	Validation Terror	Out -of-sampleTerror
Extra Tree	14.6%	13.9%
ARIMA with Python	13.8%	11.4%
RandomForest	13.6%	11.9%
Lasso regression	13.4%	11.5%
Neutral Network	13.6%	11.3%
Stacking in ML	12.6%	10.2%

**6. CONCLUSION**

This study concluded that if we don't have information regarding our sales and producing, a business can't survive for a long duration. Exact Forecasting predictions can help in maintaining the supply chain, decrease lead time, staffing level, inventory, financial status and more. The precise

information enables firms to respond all the more rapidly to changing economic situations. There are some limitations for forecasting as we are dependent on the previous data available on the websites for research works. This data could be somewhat incorrect due to incorrect information provided by clients, information on stock prices, store and their employees due to unexpected changes in market and, unemployment insurance claims. The data of shops and products a slight extent every month. Designing a robust model that can manage such circumstances is part of the challenge. Python language is used to create a simulator to compare the results and to check the accuracy of the prediction system. Regression Metrics is more efficient to use. The Mean Squared Error (MSE) for forecasts, R Squared ( $R^2$ ) and Root Mean Square Error (RMSE) are calculated. Auto Regression Moving Average and Exponential Smoothing algorithms in a Hybrid approach are giving good results.

### 7. FUTURE SCOPE

This work is done to prevent the problem of overstock or understock of the item. It also helps in increasing the annual sales of the store. It also helps in predicting the manufacturing of the product. This paper recognizes and portrays issues in foreseeing the precise outcomes and provides the hybrid approach to deal with actualizing the most appropriate strategies. The exact forecasting gives the company a structure for the limit of constructive goals for its business groups. Exact force helps the controller analyzes the need to extend the staff so as to enlist and prepare them. The committee can build new offices for creation and capacity when and where they will do the appreciable. The approximate idea results in valuable time for industries or the organizations so that they can harmonize the resources required for the expansion. This study can be improved by getting the latest data after 2017 to till date.

### REFERENCES

- [1] Elkan, C. (2001). Magical thinking in data mining: lessons from COIL challenge 2000. Proc. of SIGKDD01, 426-431.
- [2] NavanshuKhare, SaadYunusSait, "Credit Card Fraud Detection Using Machine Learning Models and Collating Machine Learning Models", International Journal of Pure and Applied Mathematics (2018), 118, 20
- [3] <https://www.kaggle.com/c/competitive-data-science-predict-future-sales/data>
- [4] Jamal Fattah, LatifaEzzine, Zineb Aman, "Forecasting of demand using ARIMA model", International Journal of Engineering Management, (2018)
- [5] Maobin Li, Shouwen Ji and Gang Liu, "Forecasting of Chinese E-Commerce Sales: An Empirical Comparison of ARIMA, Nonlinear Autoregressive Neural Network, and a Combined ARIMA-NARNN Model", Mathematical Problems in Engineering (2018)
- [6] Bohdan M. Pavlyshenko, "Machine-Learning Models for Sales Time Series Forecasting", MDPI(2019)
- [7] Zineb Aman, Haj El Moussami, and AbdeslamLachhab, Forecasting of demand using the ARIMA model, International Journal of Engineering Business Management Volume 10: 1-9, 2018,
- [8] Armstrong, J.S. Combining forecasts: The end of the beginning of the end? Int. J. Forecast (1989), 5, 585-588.
- [9] Ankur Pandey, ArunChaubey, Sanchit Garg, Shahid Siddiqui, SharathSrinivas."Forecasting Demand for Perishable Items", 2012
- [10] El Bahi Y, Ezzine L, Aman Z, et al., "Modeling and forecasting of fuel selling price using time series approach: case study", International conference on control decision and information technologies, Thessaloniki, Greece (2018)
- [11] SamanehBeheshti-Kashi, "A survey on retail sales forecasting and prediction in fashion markets", Systems Science & Control Engineering an Open Access Journal (2015), 3, 1, 154 161
- [12] M Giering, "Retail sale Prediction and item recommendations using customer demographics at store level", ACM SIGKDD Explorations Newsletter (2010)
- [13] Xiao Fang Du, Stephen C.H. Leung, Jin Long Zhang & K.K. Lai, "Demand forecasting of perishable farm products using support vector machine", Pages 556-567 | Received 08 Apr 2010, Accepted 06 Aug 2011, Published online: 10 Oct 2011
- [14] Yi Yangl; RongFulil; Chang Huiyou; Xiao Zhijiao "SVR mathematical model and methods for sale prediction" Journal of Systems Engineering and Electronics(2009), 18, 769-773
- [15] SamanehBeheshti-Kashi, "A survey on retail sales forecasting and prediction in fashion markets", Systems Science & Control Engineering (2014)
- [16] IlhamSlimani; Ilhame El Farissi, "Artificial neural networks for demand forecasting: Application using Moroccan supermarket data", International Conference on Intelligent Systems Design and Applications (ISDA), 2015.
- [17] V. Bhusari& S. Patil, "Application of hidden Markov model in credit card fraud detection", International Journal of Distributed and Parallel Systems (2011) 2, 6.
- [18] S.J.Stolfo, D.W. Fan, W. Lee, A.L. Prodromidis, and P.K. Chan, "Credit card fraud detection using meta-learning: issues and initial results" AAAI Workshop AI Methods in Fraud and Risk Management (2018), 83-90.
- [19] Federal Trade Commission. Consumer Sentinel Network - data book for January (2008). Retrieved Oct 20, 2016. From <https://www.ftc.gov/>