



# CUSTOMER SALES PREDICTION USING MACHINE LEARNING

Aniket Shukla And P. Selvaraj

Department of Computing Technologies, College of Engineering and Technology,  
Faculty of Engineering and Technology,  
SRM Institute of Science and Technology,  
Kanankulathur, 603203

**Abstract-** Customer sales prediction is the process of estimating a future value based on previous values and a variety of other criteria. I used the Keras functional API to develop a sales forecasting model in this model. It involves calculating current or prospective sales utilising data such as previous sales, seasonality, holidays, economic situations, and so on. As a result, after being given a set of inputs, this model will forecast future sales of a specific product/item.

Forecasting sales is a crucial part of modern market intelligence. A solid revenue projection can help a business save money on superfluous products, better plan for the future, and enhance profits. Grocery sales forecasting is linked to anticipating the potential sales of establishments including supermarkets, retail outlets, and bakeries. It allows firms to effectively divide cash, estimate realistic sales revenue, and plan a stronger strategy for the store's future growth.

## I. INTRODUCTION

Sales prediction is playing a major role in creating a noteworthy employment in various fields. Customer sales forecasting is vital for a variety of businesses, particularly car arrangements and other usual businesses. The aims are usually achieved by employing verifiable methods, such as the autoregressive–moving–typical (ARMA), which take into account recorded data. These approaches are only useful for explicit data. We can detect prospective models and headway consistency from most data using data mining. We can perfectly forecast it by using ARMA, MA, and AR models.

We will be able to set clear goals and expectations for our company by forecasting future sales. Our goals will be based on data rather than numbers conjured up out of thin air. We'll be able to look back on your previous achievements and expand on them.

A forecast of future sales might assist you in staffing and other anticipated requirements. When you know what's coming, you'll be able to prepare the employees you'll need as well as the merchandise you'll need to fulfil the increased demand. You'll be able to plan your resources and anticipate market shifts.

## II. LITERATURE SURVEY

1. One of the main goals of this research is to find a reliable mechanism for predicting sales trends, which can be implemented using data mining techniques to achieve the best possible revenue. The company now manages several data warehouses. Data volumes are expected to continue to grow exponentially. Steps are needed to adjust the speed of the transaction process and improve the expected growth in data volume and customer behavior. The e-commerce industry urgently needs new data mining methods and smart models that predict marketing trends with the highest possible level of accuracy and reliability. The sales forecast provides an overview of how the company manages employees, cash flows and resources. This is a basic requirement for business planning and decision making. It allows companies to effectively plan their business strategies.
2. The article discusses various machine learning algorithms for use in various industries, such as retail, logistics marketing, and so on. The paper discusses useful knowledge of machine learning techniques. This



leads to the conclusion that Rule Induction (RI) is the most widespread machine learning technique in enterprise data mining applications compared to the other four machine learning paradigms. Work performed describing the sales forecast of a pharmaceutical distribution company.

3. This article investigates if Facebook data can be used to make accurate Nike sales estimates and how Nike-related events affect the activity of Nike's Facebook pages. The paper applies the marketing framework AIDA (Awareness, Interest, Desire, and Action) to the sales model Big Social Data, as well as the method of social set analysis from the computational field of social sciences. The dataset comprises of (a) a selection of Nike Facebook pages with the number of likes, comments, posts, and other activities recorded for each page per day, and (b) business data in the form of Nike's quarterly global sales statistics revealed in financial reports. The Social Set Visualizer was also used to conduct an event study (SoSeVi). According to the studies, Facebook data contains a lot of information. The prediction accuracy of some of the simplest modification models is quite good. Due to data set features such as absolute multicollinearity, many variables have poorer prediction accuracy and create analytical hurdles. The event investigation discovered aberrant activity around a number of specific Nike events, but detailed analysis is required to determine whether the surges in activity are purely event-related or coincidental. Our expertise can assist in determining the value of Big Social Data data for a company's marketing strategy, marketing operations, and supply chain..
4. In retail, revenue forecasting for real sales is critical for any retail operation, such as a Big Mart or a Mall, to be profitable. Traditional predictive models, such as the statistical model, are frequently used to forecast future sales, but they need a significant amount of time to estimate sales and are unable to handle nonlinear data. As a result, nonlinear and linear data are managed using machine learning approaches. Vast volumes of data, such as the Big Mart dataset, can benefit from machine learning approaches due to the large amount of consumer data and the unique nature of the data item. The store is looking for a model that can accurately estimate sales so it can track future client demand and change inventory ahead of time. In this paper, we offer a Grid Search Optimization (GSO) technique for optimising parameters and selecting the optimal tuning hyperparameter, as well as a set of Xgboost strategies for predicting future sales of a retail company like Big Mart, and we discovered that our model performs better.
5. In this study, a two-tier statistical strategy is defined as a set of prediction strategies for layering data mining. Because the stacking is done in two layers, the lower layer contains one or more learning algorithms and the upper layer contains one learning algorithm, this is called two levels. Stacked Generalization is another name for stacking. It entails teaching the higher layer learning algorithm to combine the lower layer algorithms' predictions. All learning algorithms are trained using a big mart data set in the first stage, and all predictions made by the lower layer algorithms are loaded into the combiner algorithm in the second step to obtain a clear forecast. Because composing contains more information for prediction, it is superior to any model.
6. One of the main goals of this research is to find a reliable mechanism for predicting sales trends, which can be implemented using data mining techniques to achieve the best possible revenue. Today, the company manages several data warehouses. Data volumes are expected to continue to grow exponentially. Steps are needed to adjust the speed of the transaction process and improve the expected growth in data volume and customer behavior. E+ The commercial industry urgently needs new data mining methods and smart models to predict marketing trends with the highest levels of accuracy and reliability. The sales forecast provides an overview of how the company manages employees, cash flows and resources. This is a basic requirement for business planning and decision making. It allows companies to effectively plan their business strategies.



### III. LIMITATIONS OF EXISTING SYSTEM

- **Time-consuming completion** - Although there are several sales forecasting methods, two broad approaches include manual processes and data-driven processes. In any case, it takes a lot of time to make predictions.
- **Expensive technological tools** - Manual processes are not like technology, but computer tools such as spreadsheets are often used.

### IV ALGORITHM USED

- **Machine Learning:** Machine learning is used to forecast sales, which is a popular and important application (ML). Sales predictions can be used to set benchmarks, assess the incremental effects of new efforts, allocate resources in response to anticipated demand, and forecast future budgets.

- **Auto Regressive (AR) Model :** Uses Past values to make prediction

$$Y_n = \alpha_0 + \alpha_1 y_{n-1} + \alpha_2 y_{n-2} + \dots + \alpha_k y_{n-k}$$

- **Moving Average (MA) Model:** Uses past errors to make prediction

$$Y_n = \alpha_0 + \alpha_1 \epsilon_{n-1} + \alpha_2 \epsilon_{n-2} + \dots + \alpha_k \epsilon_{n-k}$$

- **ARMA Model:** Combination of both MA and AR.

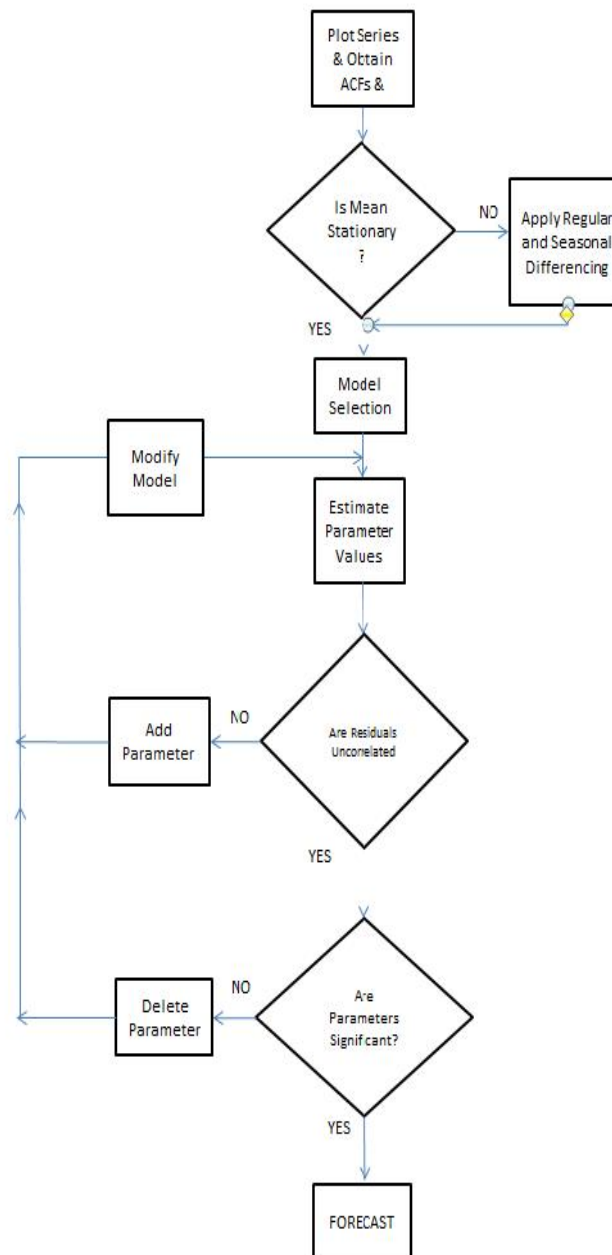
$$Y_n = \beta_0 + \beta_1 y_{n-1} + \alpha_1 \epsilon_{n-1} + \beta_2 y_{n-2} + \alpha_2 \epsilon_{n-2} + \beta_3 y_{n-3} + \alpha_3 \epsilon_{n-3} + \dots + \beta_k y_{n-k} + \alpha_k \epsilon_{n-k}$$

### V. OBJECTIVE

#### Objectives:

The goal of this paper is to forecast sales and give data that will help you make informed business decisions. If your prediction shows a 30% growth in product or service sales, for example, you can begin exploring for a wider business area and hire more personnel to satisfy demand.

### VI. ARCHITECTURE DIAGRAM



## VII. MODULE DESCRIPTION

### Data Collection

At this stage, data is collected for analysis of the best proposed method. The KAGGLE data set was used in our study to identify a prediction model for this data set. This dataset contains 22,717 items and their sales.

### Data Preprocessing

- Preprocessing the dataset and formatting it into a form that is relevant to us.



- Next is plotting our time series and coming the conclusion that we have a clear decreasing trend component and a seasonal component with an interval of approximately 12 months
- Then gaining insights from our dataset by aggregating it to monthly level to form our time series and plotting several graphs like rolling mean and rolling standard deviation and other data visualization visualize our dataset.
- Next we decompose time series into its constituent component which are trend , seasonality and residue.
- We decompose our time series using two models ,one is additive and other is multiplication.

## Classifier Training

- Then we perform certain steps to stationarize our time series.
- Stationarizing our time series is very important in order to fit the time forecasting models on it.
- Then we perform AUGMENTED DICKEY FULLER TEST on our time series to check whether it is stationary or not , and we come to the conclusion that our time series is not stationary
- Then by using the difference() function we de trend our time series and using the same function but just updating the time interval to 12 months we deseasonalise our time series
- And then we perform adf test on the obtained time series and find out that the time series is stationary
- So we can now proceed further to fitting different models
- We then tried to fit several time series forecasting models like AR,MA and ARMA with different time lag values
- Then we determine the best accurate value for all the fitted models and conclude that ARMA(2,2) was the best fit..

## VIII. IMPLEMENTATION

- **Dataset**

We have used the Kaggle dataset. It consists of 22,171 instances of different items and lot of data like sales of item, item id, shop id and date of sales.

- **Training and testing**

We tried to fit several time series forecasting models like AR,MA and ARMA with different time lag values. Then we determine the best accurate value for all the fitted models and conclude that ARMA(2,2) was the best fit. And then we predict the sales properly.

## IX. RESULT DISCUSSION

- Based on that dataset we can get the result using Decision tree algorithm to classify result.
- By applying different model of testing we are able to predict the perfect customer sales in future.
- It is helpful for all kind of retailers all around the world .
- The sales forecast provide an overview of how the company manages employees, cash flows and resources. This is a basic requirement for business planning.
- We then tried to fit several time series forecasting models like AR,MA and ARMA with different time lag values.





- The best accurate value for all the fitted models and conclude that ARMA(2,2) was the best fit. And then we predict the sales properly.

## X. REFERENCES

- Sunitha Cheriyan IT Department Muscat College, Sultanate of Oman  
sunitha.cheriyam@hct.edu.om Shaniba Ibrahim IT Department Muscat College, Sultanate Oman  
shaniba.ibrahim@hct.edu.om Saju Mohan IT Department College of Technology Muscat ,  
Sultanate of Oman saju.mohanan@hct.edu.om Susan Treasa IT Department Muscat College of  
Technology, Sultanate of Oman
- Kumari Punam<sup>1</sup> M .Tech Student, Informatics and Engineering, IIT (ISM), Dhanbad, Jharkhand  
kumaripunam23@gmail.com Rajendra Pamula<sup>2</sup> Assistant Professor, Computer Science and  
Engineering, IIT (ISM), Dhanbad, Jharkhand praphulajn1@gmail.com Praphula Kumar Jain  
Junior Research Fellow, Computer Science and Engineering, IIT (ISM), Dhanbad, Jharkhand  
rajendrapamula@gmail.com
- Linda Camilla Boldt, Vinothan Vinayagamoorthy, Florian Winder, Melanie Schnittger, Mats  
Ekran, Raghava Rao Mukkamala, Niels Buus Lassen, Benjamin Flesch, Abid Hussain<sup>1</sup> and Ravi  
Vatrapu<sup>1,2</sup> <sup>1</sup>Centre for Business Data Analytics, Copenhagen Oslo Business School, Denmark  
Westerdal, Denmark Comm & Tech, Norway
- Gopal Behera<sup>1</sup> and Neeta Nain<sup>2</sup> <sup>1,2</sup>Department Of Computer Science and Engineering  
<sup>1,2</sup>Malaviya National Institute of Technology Jaipur, India 12019rcp9002@mmit.ac.in
- Kumari Punam<sup>1</sup> M .Tech Student, Informatics and Engineering, IIT (ISM), Dhanbad, Jharkhand  
kumaripunam23@gmail.com Rajendra Pamula<sup>2</sup> Assistant Professor, Computer Science and  
Engineering, IIT (ISM), Dhanbad, Jharkhand praphulajn1@gmail.com Praphula Kumar Jain  
Junior Research Fellow, Computer Science and Engineering, IIT (ISM), Dhanbad, Jharkhand  
rajendrapamula@gmail.com
- Sunitha Cheriyan IT Department Muscat College, Sultanate of Oman  
sunitha.cheriyam@hct.edu.om, Shaniba Ibrahim IT Department Muscat College, Sultanate Oman  
shaniba.ibrahim@hct.edu.om, Saju Mohanan IT Department Muscat School, Sultanate of Oman  
saju.mohanan@hct.edu.om, Susan Treasa IT Department Muscat College of Technology