**Mini Project Report on**



**SALES FORECASTING**



**Submitted in partial fulfilment of the requirement for the award of the degree of**

**BACHELOR OF TECHNOLOGY**

**IN**

**COMPUTER SCIENCE & ENGINEERING**

**Submitted by:**

**Name - Garvit Jain** **Univ. Roll No - 2017327**

***Under the Mentorship of***

**Dr. Jyoti Agarwal**



**Department of Computer Science and Engineering**

**Graphic Era (Deemed to be University)**

**Dehradun, Uttarakhand**

**July-2023**



**CANDIDATE’S DECLARATION**

I hereby certify that the work which is being presented in the project report entitled **“Sales Forecasting Using Machine Learning”** in partial fulfilment of the requirements for the award of the Degree of Bachelor of Technology in Computer Science and Engineeringof the Graphic Era (Deemed to be University), Dehradun shall be carried out by the under the mentorship of **Dr. Jyoti Agarwal**, Department of Computer Science and Engineering, Graphic Era (Deemed to be University), Dehradun.

Student Name: **Garvit Jain**   University Roll No. : **2017327**

**Table of Contents**

|  |  |  |
| --- | --- | --- |
| **Chapter No.** | **Description** | **Page No.** |
| Chapter 1 | Introduction | 1 |
| Chapter 2 | Literature Survey | 2 |
| Chapter 3 | Methodology | 3-9 |
| Chapter 4 | Result and Discussion | 10-11 |
| Chapter 5 | Conclusion and Future Work | 12 |
|  | References |  |

**Chapter 1**

**Introduction**

**1.1 Motivation for work**

Sales Forecasting is the procedure for estimating future sales. It is easier for big established companies to predict their sales using the past data , but in this era where entrepreneurship has gotten a boom , sales forecasting is an important technique . Sales forecasting gives insight into how a company should manage its workforce, cash flow, and resources . In addition to helping a company allocate its internal resources effectively, predictive sales data is important for businesses when looking to acquire investment capital.To help newly found business , this sales forecasting model is build .

**1.2 Problem Statement**

Develop an accurate and reliable sales forecasting model to help stores and businesses predict their sales based on historical data . With the sales record of a store of previous months , we will predict the sales for the upcoming months . The objective of this project is to predict future sales using regression model and help store and small businesses manage their resources efficiently according to sales pattern.

**1.3 Objectives**

Objectives of these project are:

1.3.1 Predicting future sales from a given dataset.

1.3.2 Find the best algorithm that will predict sales with the greatest accuracy.

1.3.3 To understand the implementation of regression model used .

**Chapter 2**

**Literature Survey**

Forecasting is supposed to be one of the oldest management activities. Nowadays it is becoming increasingly necessary for companies to make forecasts; those that do not give the prospect to their competitors a clear advantage. No forecasting is a main cause of most of today’s business failures. In the past, goods could be sold on company reputation alone and forecasting was not too important. In today’s more competitive times, sentiment does not apply, and firms that do not challenge their selves to make an accurate forecast on which to base their future production will find it increasingly difficult to survive (Lancaster G.A. & Lomas R.A., 1985).

Forecasting methods can be divided in three basic categories:

1. Quantitative or Statistical
2. Qualitative or Judgmental
3. Time Horizon

Quantitative Forecasts base on mathematical models and suppose that past data and other relevant factors can be combined into reliable predictions of the future (The Journal of Business Forecasting, fall 2000). In preparing a quantitative forecast it should begin with a number of observed values, past data, or observations (Markakis and Wheelwright, 1989).

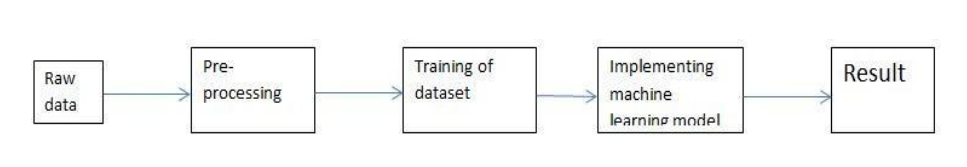
These observations may represent many things, from the actual number of units sold to the cost of producing each unit to the number of people employed Qualitative Forecasts (The Journal of Business Forecasting, fall 2000) are based on opinions, knowledge and skills rather than more formal analysis. They are used where there is no historical data. These types of forecasts are one of the simplest and widely used forecasting approaches available (Makridakis and Wheelwright, 1989).

Forecasts can be classified in terms of time span they cover in the future. The basic types of time horizon forecasts are long-term, medium-term and short-term (Korpela J. et.al, 1996, p.161).

**Chapter 3**

**Methodology**

The below figure represents the process for building this predictive model . Following are the steps one needs to follow to build this model



**Fig 3.1 Process for building the model**

**3.1 Importing Libraries:**

The following libraries of python are used for this project:

* + 1. Pandas : For data analysis
    2. Numpy : For working with numerical values as it makes it easy to apply mathematical operations.
    3. Matplotlib : To plot graphs of sales and date for better visualisation.
    4. Sklearn : Offers tools for data preprocessing(MinMaxScaler ) , Regression Model(Linear Regression) , and some performance analysis functions .

**3.2 Data Collection:**

The very first step of the project is to collect data . We used a sample data for a store that contains every day sales for that store from 2013 to 2018 . The dataset contain 4 columns i.e. date , store , item and sales and 931000 rows. The data set looks like :

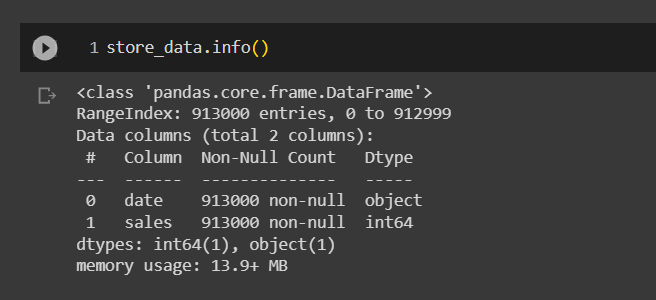


**Fig 3.2 Head function representing first 6 entries in dataset**

**3.3 Data Preprocessing:**

3.3.1. Removing unwanted columns

To forecast the future sales for a store , the necessary columns are date and sales , so there is no need for the columns store and item . It is a good choice to remove these unwanted columns from the data frame . We can now check the values in data frame.



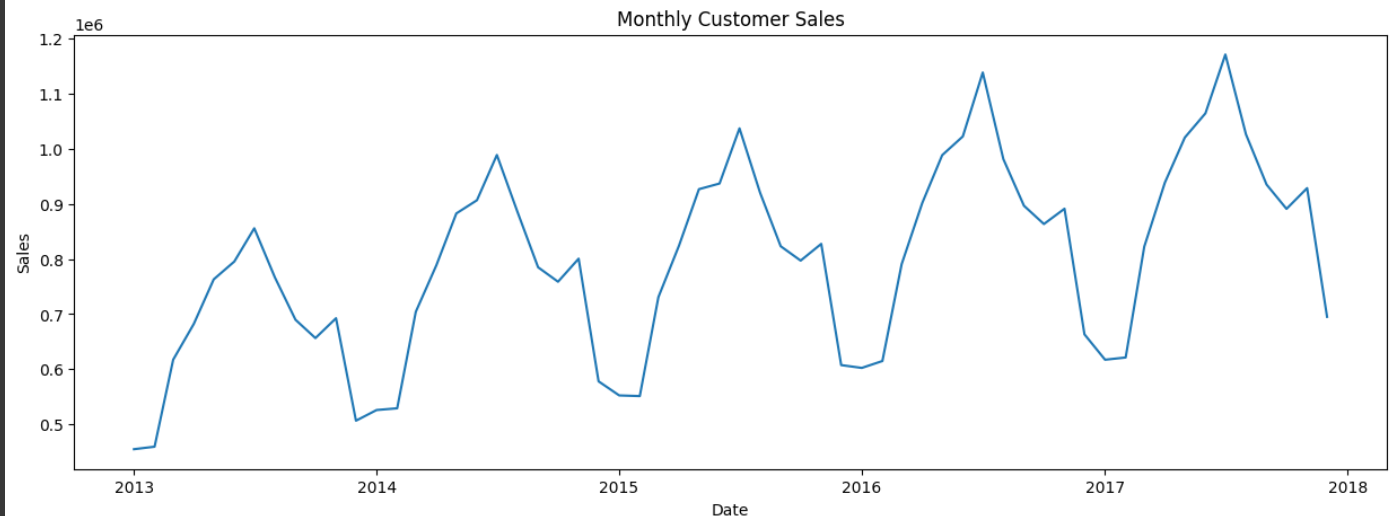
**Fig 3.3 Information about dataset**

3.3.2. Date Manipulation

We can see in the above image that the date column is of object datatype. For further manipulation we first convert it to date datatype . Next step is to convert the dates into the month period and calculating the sum of the sales in a month . This is helpful in prediction month wise . Next we convert the date datatype to timestamp.

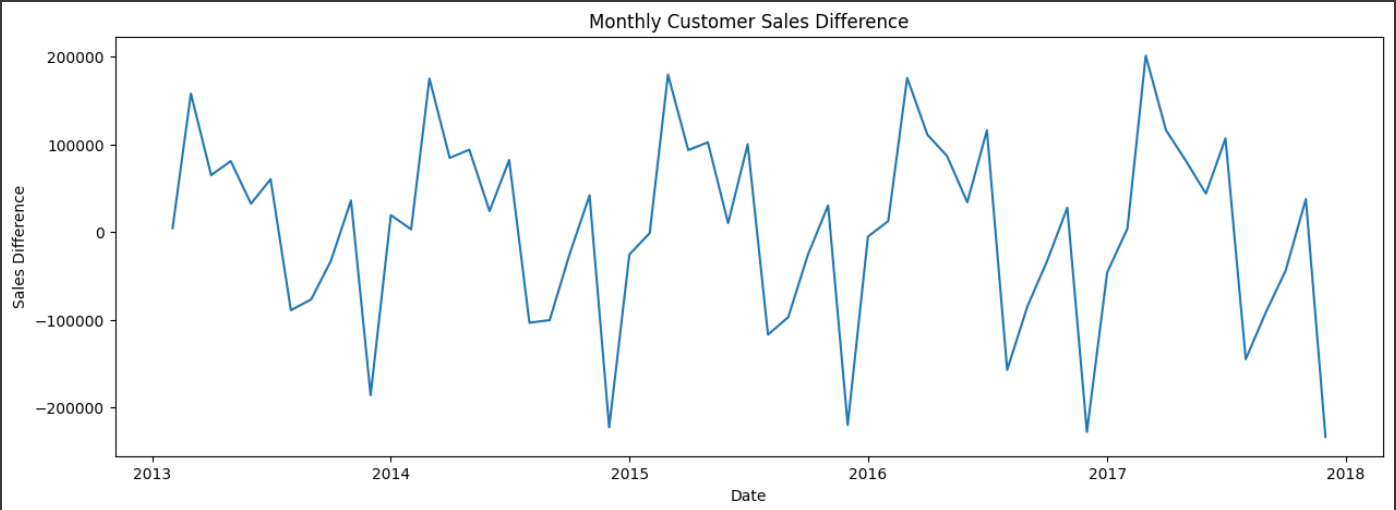
* + 1. Visualisation

Pictorial and graph representations are said to be easy to understand . Therefore we plotted a graph of sales vs. date to check themonthly customer sales and check the pattern .



**Fig 3.4 Graph of date against sales**

We need to make the data stationary to improve training phase of learning model . So for that we make a new column sales\_diff which is the difference of sales in the current month and previous month and then plot a graph of sales\_diff vs. date to check the growth pattern.



**Fig 3.5 Graph of date against sales**

3.3.4. Creating supervised data

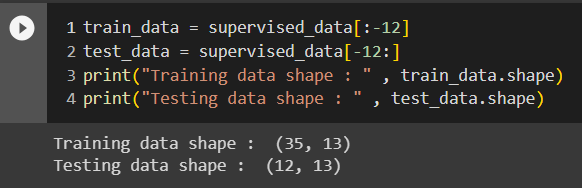
We create supervised data to feed into the machine learning model. We use previous 12 month sales to predict the next 12 month sales . In the supervised data we have only the sales\_diff . Here the previous 12 months will act as input features and next 12 months sales for output features for the machine learning model.



**Fig 3.6 Supervised Data**

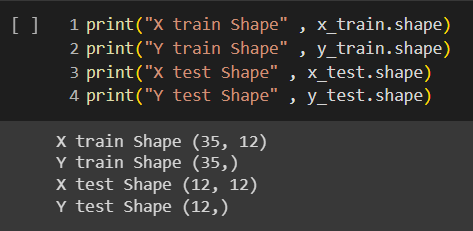
**3.4Splitting data into Training and Testing data:**

We split the data into training and testing data such that the data for sales for year 2013 to 2017 will put used as training data and sales for year 2018 will be used as testing data to check the accuracy of the model .



**Fig 3.7 Size of training and testing data**

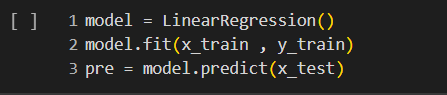
We now normalise the data by restricting the value between -1 to 1 by using MinMaxScaler. By this the relationship between original values is maintained. X\_train and x\_test contain the values of sales in 12 months and y\_train and y\_test contain the sales\_diff.



**Fig 3.8 Division of input and output**

**3.5Building the Machine Leaning Model:**

Now the dataset is ready to fit a model after performing Data PreprocessingandFeature Transformation. The training set is fed into the algorithm in order to learn how to predict values. Testing data is given as input after Model Building a target variable to predict. The regression model used is Linear Regression .



**Fig 3.9 Applying Linear Regression**

Lastly we create a pre\_test\_set which contains the predicted sales\_diff for test data and the actual sales of test data.

**Chapter 4**

**Result and Discussion**

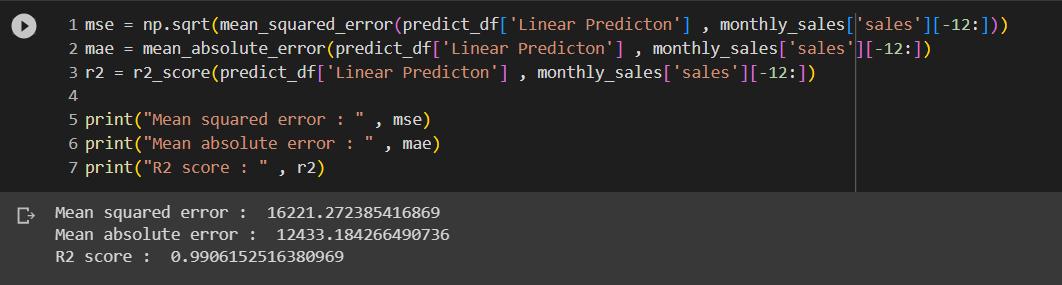
We create a predict\_df data frame which contains the predicted sales by Linear Prediction Model . The essential step in any machine learning model is to evaluate the accuracy of the model. The Mean Squared Error, Mean absolute error, Root Mean Squared Error, and R-Squared or Coefficient of determination metrics are used to evaluate the performance of the model in regression analysis.

* The Mean absolute error represents the average of the absolute difference between the

actual and predicted values in the dataset. It measures the average of the residuals in the dataset.

* Mean Squared Error represents the average of the squared difference between the original and predicted values in the data set. It measures the variance of the residuals.
* The **r2 score**varies between 0 and 100%. It is closely related to the **MSE** (see below), but not the same. It is “(total variance explained by model) / total variance.”

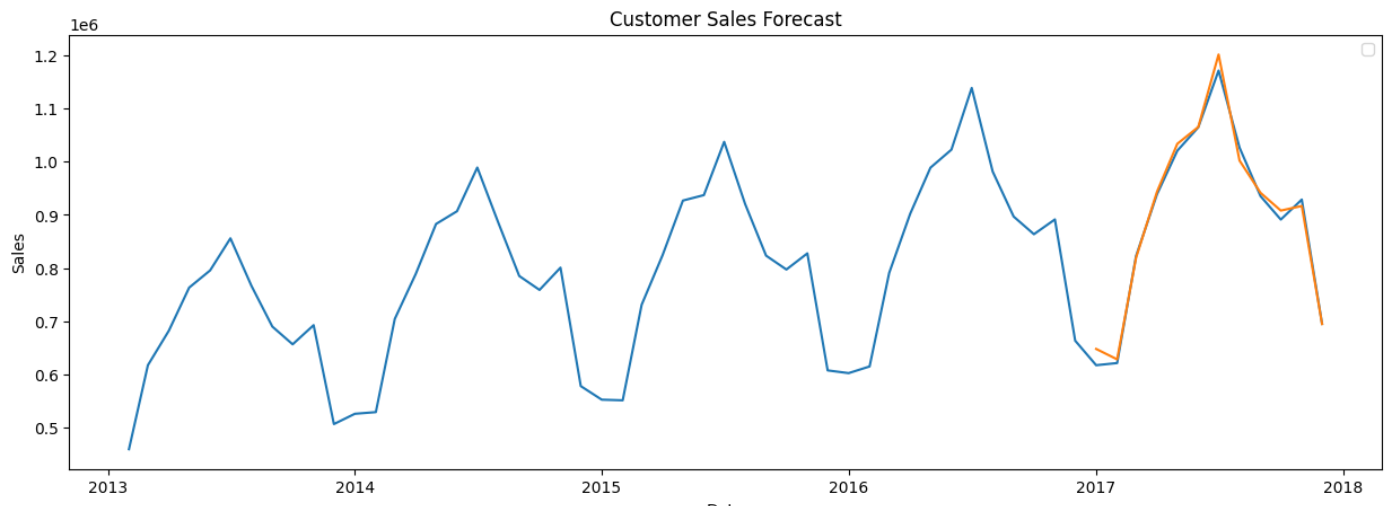
The below image shows the mean absolute error , mean squared error and r2 score of the trained model.



**Fig 4.1 Metrics of the model**

As we can see r2 score for the model is 0.9906152516380969 which depicts the predicted values and actual values are closely corelated . It means that our model is very efficient and accurate .

**Visualisation of the predicted sales against the actual sales**

****

**Fig 4.2 Graph of date against sales for actual and predicted sales**

Blue line : predicted sales

Orange line : actual sales

**Chapter 5**

**Conclusion and Future Work**

**5.1 Conclusion:**

So from this project we conclude that a smart sales forecasting program is required tomanage vast volumes of knowledge for business organizations. We also observed that the Machine algorithm i.e. Linear Regression gave us very accurate results and can be used to build a accurate and efficient sales forecasting program.

The output from this program can help businesses optimize their sales strategy and manage the cash flow .

* 1. **FutureScope :**

5.2.1 The future scope of this project is that this project can further collaborate with any other devices which are supported with an in-built intelligence by virtue of the Internet of Things(I0T) which makes it more feasible to use.

5.2.2 Multiple instances parameters and various factors are also make this sales prediction project more innovative and successful.

* + 1. The most important term for any prediction-based system that is accuracy, is often significantly increased because of the increase in the number of parameters.

**References**

[1] <https://medium.com/analytics-vidhya/mae-mse-rmse-coefficient-of-determination-adjusted-r-squared-which-metric-is-better-cd0326a5697e>

[2] <https://medium.com/mlearning-ai/sales-prediction-using-a-linear-regression-model-ffeec84eede1>

[3]<http://www.ir.juit.ac.in:8080/jspui/bitstream/123456789/3600/1/Big%20Mart%20Sales%20Prediction%20Using%20Machine%20Learning.pdf>

[4] <https://blog.hubspot.com/sales/regression-analysis-to-forecast-sales>