

Furniture Sales Forecasting Using Machine Learning Algorithm

Arya Biju
PG Scholar,

Department of Computer Applications
Amal Jyothi College of Engineering,
Kanjirappally, Kerala
aryabiju@mca.ajce.in

Shelly Shiju George
Assistant Professor,

Department of Computer Applications
Amal Jyothi College of Engineering,
Kanjirappally, Kerala
shellyshijugeorge@amaljyothi.ac.in

Abstract — Machine learning is revolutionizing all facets of life and is becoming a significant factor in situations that occur in the real world. Machine learning has revolutionary applications in every industry, including education, healthcare, engineering, sales, entertainment, and transportation, to name just a few on an endless list. The traditional method of achieving sales and marketing objectives doesn't benefit businesses anymore because it doesn't take into account how people actually buy things. Advances in machine learning have led to significant changes in the field of sales and marketing. This paper describes how supervised machine learning methods can be applied to sales forecasting. a free platform for gathering data called Kaggle was employed to gather the project's data. Through the aid of Linear Regression is used to train in Jupyter Notebook. The model to provide precise forecasts.

Keywords — Machine learning, Simple Linear Regression, Sales Prediction

I. INTRODUCTION

The importance of focusing on sales forecasting has never diminished. For all vendors to continue supporting the efficacy of marketing organisations, forecasting must be done in an effective and optimal manner. Manually performing this work would take too much time, which is not ideal in today's fast-paced environment and could result in serious mistakes that would result in bad management of the firm. The business sectors, who are actually expected to generate the right number of products to satisfy the demands of everyone, are a significant component of the global economy.

The primary objective of business sectors is market audience targeting. It is crucial that the business has been successful in achieving this goal by utilising a forecasting system. Forecasting requires examining data from a variety of sources, including market trends, customer behaviour, and other elements. The companies would benefit from this analysis by having better financial resource management. The forecasting method can be used for a variety of things, such as estimating future demand for the product or service or estimating how much of the product will be sold in a specific time frame.

In the modern era of massive shopping malls and markets, keeping track of sales data for items or products with their numerous dependent or independent features is a crucial step. Using a variety of dependent and independent criteria,

a dataset of a collection of three months sales. Data is then cleaned up to produce more accurate predictions of exciting sales. Then, simple linear regression model issued.

II. RELATED WORKS

There are many machine learning techniques used to predict the sales. Machine learning algorithms like XG Boost Regressor, Random Forest Regressor, ANN, and SVR etc. are used.

M. Singh et. al. [1] This study analyses data gathered from a retail establishment and makes predictions about future management strategies for the establishment. The impact of numerous sequences of events, such as the weather, holidays, etc., can genuinely alter the state of several departments; therefore, it also analyses these effects and considers their impact on sales.

Mansi et. al. [2] By using approaches like Clustering Models and measurements for sales forecasts, the goal of this article is to obtain accurate findings for predicting future sales or demands of a corporation. The potential of the algorithmic methods is appraised and utilized appropriately in subsequent studies.

Michael [3] This paper describes a product recommendation system and a sales forecast system that were employed to the advantage of a set of retail outlets. Consumer demographic information has been utilized to precisely design each person's sales.

Thomas et.al [4] In this study, we tackle a problem involving a large number of identical prediction tasks. Every newspaper publisher faces the challenge of printing a big quantity of copies and delivering them to retail traders while attempting to maintain the lowest return quotation possible.

Cheriyen et.al [5] This study examines how decisions should be made based on experimental data and understandings gained from data visualization. Techniques for data mining have been used. It has been demonstrated that the Gradient Boost algorithm displays the highest degree of accuracy when imagining future transactions.

III. METHOD OF PREDICTION

The objective of this project is to create a machine learning model that correctly forecasts sales using the supplied dataset. It aids in the creation of more accurate predictions by the merchants as to which month, they should focus on more and which one has the biggest sales. The dataset is used to build the training and testing data sets. More data should be used during training so that the model can be trained accurately. The results of the model can be used to create a sales forecast model. The Simple Linear Regression technique will be used to forecast sales.

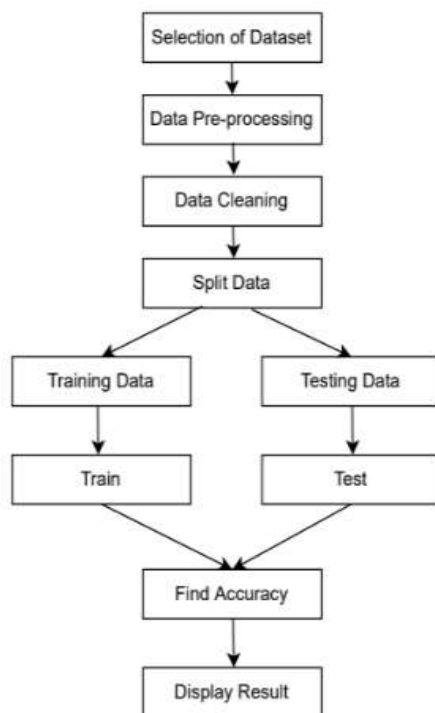


Fig.1: Sales Prediction System Methodology

The tasks that must be taken are:

- A. Data Collection
- B. Data Pre-processing
- C. Model Building
- D. Result

1.Data Collection:

For this system, we used a structured format for the dataset. Sales of each item in connection to its stores are included in the dataset being used. The dataset was divided into a training dataset and a testing dataset. We trained all of our machine learning algorithms on the training datasets that we utilized to build our models. We also provided a testing dataset so you could assess the accuracy of the trained model.

2. Data Pre-processing:

Every machine learning technique starts with this step. Data transformation, data cleaning, and data minimization are all part of this process. To increase the effectiveness of the data, some activities are undertaken. We can process the data to improve the model's accuracy and make sure that the categorization is accurate.

3. Machine Learning:

Using this, sales forecasts are made. Machine learning techniques are used to estimate sales for each store in order to assess the existing state of the stores. There are many classifiers available to find the sales. How the classifier was trained affects how accurate it is. Classifiers should be expected to have high accuracy.

IV .BUILD MODEL

The primary stage in predicting sales is model building. Algorithms are used by the user when developing the model.

1.Simple Linear Regression:

Statistics refers to a linear regression model with only one explanatory variable as simple linear regression. In other words, for two-dimensional sample points with a single independent variable and a single dependent variable, it looks for a linear function that, as exactly as possible, predicts the values of the dependent variable as a function of the independent variable. The association between the outcome variable and a single predictor is referred to as "simple." For a simple linear regression, the dependent variable must have a real or continuous value. However, it is possible to quantify the independent variable using either continuous or categorical values.

- 1) Import the libraries required and then import and print the dataset.

```

# necessary imports
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
sns.set()
%matplotlib inline
    
```

- 2) Add the dataset into jupyter.

```

In [118]: data.head()

Out[118]:
   Unnamed: 0  jan  feb  mar  april
0           1    1  200  300  400  500
1           2    2  400  300  400  500
2           3    3  100  400  500  600
3           4    4  500  400  500  600
4           5    5  600  500  600  700
    
```

- 3) splitting data into training and testing data

```

In [119]: # splitting data into training and testing data
from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size = 0.25, random_state = 0)
x_train.shape, y_train.shape, x_test.shape, y_test.shape

Out[119]: ((75, 1), (75, 1), (25, 1), (25, 1))
    
```

- 4) Then split the dataset into training and testing datasets.

```

In [122]: import statsmodels.formula.api as smf
lm = smf.ols(formula = 'sales ~ jan ', data = data).fit()
print(lm.coef)
print(lm.summary())

Out[122]:
0.4118160796115111
    
```

- 5) Creating scatterplot to find the month that has highest sales.

```
In [11]: # visualizing the relationship between the features and the label
fig, ax = plt.subplots(1, 2, figsize = (15, 7), sharey = True)
ax[0].scatterplot(x = 'Year', y = 'Sales', data = data, ax = ax[0])
ax[1].scatterplot(x = 'Year', y = 'Sales', data = data, ax = ax[1])
plt.show()
```

6) Finding the accuracy of the model.

```
In [12]: # accuracy of model
from sklearn.metrics import mean_squared_error, mean_absolute_error
from math import sqrt

print("Mean Squared Error of Model is : (mean_squared_error(y_test, y_pred))")
print("Mean Absolute Error of Model is : (mean_absolute_error(y_test, y_pred))")
print("Root Mean Squared Error of Model is : (sqrt(mean_squared_error(y_test, y_pred)))")

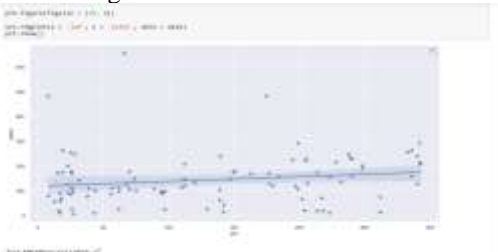
Mean Squared Error of Model is : (3431.1710862583)
Mean Absolute Error of Model is : 74.434565348908
Root Mean Squared Error of Model is : 58.56981124875
```

7) Finding the Actual and Predicted value

```
In [13]: # comparing actual and predicted values
import pandas as pd
df = pd.DataFrame({'Actual': y_test, 'Predicted': y_pred})
df.head()
```

	Actual	Predicted
0	13.0	13.737267
1	96.0	128.885008
2	89.0	121.551553
3	177.0	158.254723
4	4.7	121.179138
5	819.0	170.875058
6	125.0	128.885008
7	26.0	125.636866
8	203.0	171.478389
9	105.0	140.486764
10	128.0	150.459113
11	148.0	144.198882
12	163.0	156.125018
13	93.0	131.311221
14	113.0	134.888373
15	218.0	157.207758
16	76.0	128.871234
17	97.0	127.252827
18	172.0	122.562299
19	181.0	130.128018
20	110.0	140.886955
21	113.0	122.564295
22	188.0	136.847728
23	112.0	173.478159
24	181.0	141.363888

8) Visualizing the model



VI. RESULTS

The output is produced after fitting the data into the Simple Linear Regressor algorithm. Before that, the dataset is first loaded after which data within the dataset are trained and tested.

```
In [47]: # splitting data into training and testing data
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.25, random_state = 0)
X_train.shape, y_train.shape, X_test.shape, y_test.shape

Out[47]: ((75, 1), (75, 1), (25, 1), (25, 1))

In [48]: # fitting data to linear regression model
from sklearn.linear_model import LinearRegression
lr = LinearRegression()
lr.fit(X_train, y_train)

Out[48]: LinearRegression()
```

OLS Regression Results					
=====					
Dep. Variable:	Sales				
Model:	OLS				
Date:	Mon, 20 Oct 2023				
Time:	11:30:17				
Sample:	Obs. 100				
Residuals:	75				
Adjusted R-squared:	0.999				
Standard Error:	1.000				
Intercept:	13.0				
Year:	0.434				
=====					
coef	std err	t	Pr > t	[0.025	0.975]
Intercept	13.0	13.0	0.000	12.0	14.0
Year	0.434	0.434	0.000	0.434	0.434
=====					
variance_inflators					
tolerance	0.000	variance_inflators			
cond	1.000	cond			
maxcollinearity	1.000	maxcollinearity			
maxcondition	1.000	maxcondition			
condition	1.000	condition			

Using dataset, machine learning methods were employed, and the outcomes for each methodology were slightly varied due to the distinct working criteria of each algorithm. The results were evaluated on the basis of accuracy

Method	Accuracy	
	Training	Testing
Linear Regression	0.015	0.002

VII. CONCLUSION

The usage of machine learning approaches proves to be a crucial feature for designing business plans while taking into consideration consumer buying habits, as traditional methods are not very helpful to commercial organizations in revenue growth. Businesses can adopt effective tactics for expanding sales and stepping unafraid into the competitive world by using sales predictions based on a variety of factors, including past sales.

REFERENCES

- [1] M. Singh, B. Ghutla, R. Lilo Jnr, A. F. S. Mohammed and M. A. Rashid, "Walmart's Sales Data Analysis - A Big Data Analytics Perspective," 2017 4th Asia-Pacific World Congress on Computer Science and Engineering (APWC on CSE), 2017, pp. 114-119
- [2] Mansi Panjwani, Rahul Ramrakhiani, Hitesh Jumnani, Krishna Zanwar and Rupali Hande, "Sales Prediction System Using Machine Learning" 2020, Easy Chair
- [3] Michael Giering, "Retail Sales Prediction and Item Recommendations Using Customer Demographics at Store Level" 2018 Volume 10, Issue 2
- [4] Thomas Ragg, Wolfram Menzel, Walter Baum, Michael Wigbers, "Bayesian learning for sales rate prediction for thousands of retailers" 2022 Neurocomputing 43 (2002) 127–144
- [5] S. Cheriyan, S. Ibrahim, S. Mohanan and S. Treasa, "Intelligent Sales Prediction Using Machine Learning Techniques", 2018 International Conference on Computing, Electronics & Communications Engineering (ICCECE) 2018, pp. 53-58
- [6] Mohit Gumani, Yogesh Korke, Prachi Shah, "Forecasting of sales by using fusion of machine learning techniques", 2017, 2017 International Conference on Data Management, Analytics and Innovation
- [7] Soham Patangia, Rachana Mohite, Kevin Shah, Gaurav Kolhe, Madhura Mokashi, Prajakta Rokade, 2020, Sales Prediction of Market using Machine Learning, International Journal Of Engineering Research & Technology (IJERT) Volume 09, Issue 09 (September 2020),
- [8] Sandeep Chavan, Simsri Panchal, Tanvi Sawant, Janhavi Shinde, 2020, Predicting Online Product Sales using Machine Learning, International Journal of Engineering Research & Technology (IJERT) Volume 09, Issue 04 (April 2020),