

# Sales Prediction Analysis Using Machine Learning Techniques

Noorjahan P N  
PG Scholar

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

Sr Elsin Chakkalackal S H  
Assistant Professor,

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

*Abstract*—The commercial enterprise agencies relies upon on information base and sales traits prediction. Here we are going to in brief examine the idea of sales information and sales forecast. In these studies, we've got performed an in depth have a look at and evaluation of models to enhance the future sales. Machine Learning models such as Random Forest Classifier and Gradient Boosting Classifier are used here to examine the idea of sales information and sales forecast.

*Keywords*—Forecasting, Random Forest Classifier, Gradient Boosting Classifier

## I. INTRODUCTION

The business must make an educated guess about how consumers will behave and how the market will react to the product. This enables the sectors to introduce cutting-edge products that may be a great success and help the businesses avoid losing money on investments. Many different methods are frequently employed to predict future market trends. In general, forecasting is based on historical statistics available from institutes related to their fields.

An accurate sales forecast enables businesses to manage their cash flow and allocate resources for future growth. Sales projections enable executives to set early course corrections and serve as benchmarks for future improvements. Sales quotas and sales expectations can be aligned, and leaders in revenue can optimize for further wins. conveys assurance to the board and the management team that your business is supported by a reliable forecasting tool in order to scale well in the future. Sales forecasts aid in the creation of strategic strategies and inform you of your capacity for enacting and implementing them as quickly as possible.

In our article, we suggested applying machine learning algorithms to data gathered from brands' prior sales. Here we are going to in brief examine the idea of sales information and sales forecast. In these studies, we've got performed an in depth have a look at and evaluation of models to enhance the future sales. Here, machine learning models like Random Forest Classifier and Gradient Boosting Classifier are employed.

In order to gain a comprehensive understanding of the information, analysis and exploration of the acquired data have also been accomplished. The corporate agencies could

use analysis to make a probabilistic decision at each important stage of their marketing plan.

## II. RELATED WORKS

There are many machine learning techniques are used to examine the idea of sales information and sales forecast.

A. Gokilavani et. al. [1], A two-level strategy is used for sales forecasting, which generates superior predictive standard overall performance than any of the popular single model predictive masters, in order to estimate achievable income sales and to create a better approach to earnings of a product from an outlet. The department store is the target of the strategy.

Naveen Kumar et. al. [2], Business enterprises rely on knowledge bases and forecasts of sales trends.. For enhancing accuracy and efficiency, Data mining techniques proved effective in extraction of hidden knowledge. Going to analyse the idea of sales data and sales forecast in a brief manner. Analysing models to boost upcoming sales.

Purvika Bajaj et. al. [3], Machine Learning is reworking each stroll of lifestyles and has end up a primary contributor in actual global scenarios. Sales prediction is carried out the usage of Machine Learning models. Items such as item weight, item fat content, item visibility, item kind, item MRP, outlet established order year, outlet length, and outlet region type are among the information parameters used in the forecast.

Sunitha Cheriyan et.al. [4], The accuracy of sales projections has a significant impact on business. On the basis of this research, understandable predictive models are developed to improve future sales projections. With the vast records and precision of sales forecasting, forecast frameworks are challenging to manage. This shows the highest degree of forecasting and sales prediction accuracy.

Varshini S et. al. [5], Machine Learning performs a critical position in the area of sales and marketing. In this research, we provide an efficient and accurate sales forecasting model primarily based totally on numerous Machine Learning techniques which can deal with massive volumes of data, along with the Big Mart dataset, which includes an excessive ML model.

### III. METHODOLOGY

In order to create models that are complete and trustworthy, the major goal of this research is to assess and examine the usage of machine learning techniques for sales forecasting.

#### 1. Importing the Libraries

Import the libraries required and then import and print the dataset.

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import os
import scipy as sp
import warnings
import datetime
warnings.filterwarnings("ignore")
%matplotlib inline
```

Fig 1

#### 2. Loading the Dataset

A dataset is an established collection of information typically related to a completely unique frame of work. The dataset contains ten variables such as Product Name, Product ID, Listing Price, Sales Price, Discount, Brand etc.

```
data = pd.read_csv('Adidas Vs Nike.csv')
```

Fig 2

#### 3. Training and Testing Data

The training data is the largest subset of the original dataset, that is used to train or fit the machine learning model. The test dataset is any other subset of original data, that is independent of the training dataset.

```
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(x,y,random_state=0,test_size=0.2)
```

```
print(len(x_train))
print(len(x_test))
print(len(y_train))
print(len(y_test))
```

2614  
654  
2614  
654

Fig 3

#### 4. Fit Gradient Boosting Classifier to the Dataset

```
from sklearn.ensemble import GradientBoostingClassifier
gbc=GradientBoostingClassifier()
gbc.fit(x_train,y_train)
```

GradientBoostingClassifier()

```
y_pred_gbc=gbc.predict(x_test)
from sklearn.metrics import accuracy_score,classification_report,confusion_matrix
from sklearn.metrics import r2_score
from sklearn.metrics import mean_squared_error
print("Classification Report is:\n",classification_report(y_test,y_pred_gbc))
print("Confusion Matrix:\n",confusion_matrix(y_test,y_pred_gbc))
print("Training Score:\n",gbc.score(x_train,y_train)*100)
acc=accuracy_score(y_test, y_pred_gbc)
print("Accuracy:", (acc*100))
```

Classification Report is:	precision	recall	f1-score	support
1	0.97	0.99	0.98	221
2	0.98	0.98	0.98	185
3	0.95	0.90	0.93	114
4	1.00	1.00	1.00	134
accuracy			0.98	654
macro avg	0.98	0.97	0.97	654
weighted avg	0.98	0.98	0.98	654

```
Confusion Matrix:
[[219 0 2 0]
 [ 0 182 3 0]
 [ 7 4 103 0]
 [ 0 0 0 134]]
```

Training Score:  
99.6557000765111  
Accuracy: 97.55351681957187

Fig 4

#### 5. Fit Random Forest Classifier to the Dataset

```
from sklearn.ensemble import RandomForestClassifier
rfc=RandomForestClassifier()
rfc.fit(x_train,y_train)
```

RandomForestClassifier()

```
y_pred_rfc=rfc.predict(x_test)
from sklearn.metrics import accuracy_score,classification_report,confusion_matrix
from sklearn.metrics import r2_score
from sklearn.metrics import mean_squared_error
print("Classification Report is:\n",classification_report(y_test,y_pred_rfc))
print("Confusion Matrix:\n",confusion_matrix(y_test,y_pred_rfc))
print("Training Score:\n",rfc.score(x_train,y_train)*100)
acc=accuracy_score(y_test, y_pred_rfc)
print("Accuracy:", (acc*100))
```

Classification Report is:	precision	recall	f1-score	support
1	0.96	0.98	0.97	221
2	0.96	0.97	0.97	185
3	0.94	0.89	0.91	114
4	1.00	1.00	1.00	134
accuracy			0.96	654
macro avg	0.97	0.96	0.96	654
weighted avg	0.96	0.96	0.96	654

```
Confusion Matrix:
[[216 2 3 0]
 [ 2 180 3 0]
 [ 8 5 101 0]
 [ 0 0 0 134]]
```

Training Score:  
100.0  
Accuracy: 96.48318042813455

Fig 5

#### 6. The Actual and Predicted Values from the dataset:

```
data = pd.DataFrame({'Actual': y_test, 'Predicted': y_pred})
data
```

	Actual	Predicted
2466	2	2
1515	2	2
2376	1	1
654	3	1
1998	3	3
...	...	...
3260	4	4
40	1	1
3066	4	4
1491	3	3
454	1	1

654 rows × 2 columns

Fig 6

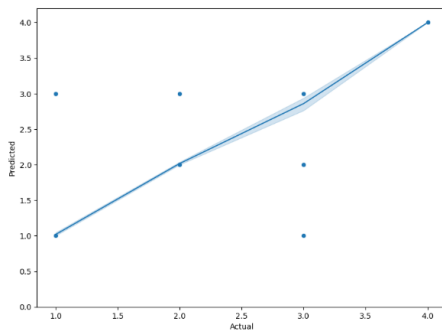


Fig 7

#### IV. PROPOSED APPROACH

The proposed work provides an opportunity for accuracy of findings by implementing data exploration and analysis algorithms. The algorithms Random Forest Classifier and Gradient Boosting Classifier are chosen as it provides a better result.

##### Gradient Boosting Classifier

A group of machine learning techniques known as gradient boosting classifiers combine numerous weak learning models to produce a powerful predictive model.

As the name implies, gradient boosting classifiers are a specific class of methods that can be applied to classification applications. Reduce the loss, or the discrepancy between the true class value of the training instance and the expected class value, using Gradient Boosting classifiers.

##### Random Forest Classifier

A supervised machine learning algorithm is random forest. A random forest is a meta estimator that uses averaging to improve predicted accuracy and control over-fitting by fitting various decision tree classifiers on various subsamples of the dataset. Regression or classification issues can be solved with the random forest classifier. A random forest classifier uses data with discrete labels or higher, often known as classes.

Compared to other algorithms, training Random Forest requires less time. Even with a large dataset, it operates effectively and predicts results with an extraordinary level of precision. Additionally, accuracy can be maintained even when a significant amount of data is absent.

#### V. RESULT

Accuracy is a key component of the machine learning algorithms' overall performance. High accuracy machine learning models are used by businesses to make wise business decisions.

Performance assessment of several machine learning methods:

Algorithms	Accuracy
Gradient Boosting Classifier	97%
Random Forest Classifier	96%

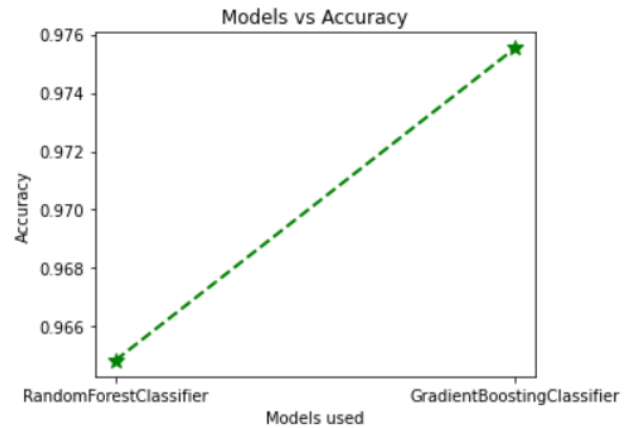


Fig 8

#### VI. CONCLUSION AND FUTURE SCOPE

The corporate agencies may use analysis to make a probabilistic decision at each important stage of their marketing plan. Here, accuracy is the primary focus of total performance. Machine learning techniques provide a strong mechanism for prediction and data mining since they solve the issue with traditional techniques. These machine learning techniques improve performance with more predictability and better results while also reducing the statistics. In our upcoming work, we can apply cutting-edge machine learning methods to improve the model's performance through sophisticated optimization.

#### REFERENCES

- [1] A. Gokilavani, T. P. Banupriya, S. Bhagavathi, A. Divya Bharathi, T. Tamilselvan, "Sales Forecasting Using Prediction Analytics Algorithm", International Journal of Research and Innovation in Applied Science (IJRIAS), 2020 Volume:V, Page: 67-70
- [2] Naveen Kumar, Jegan, Yogesh, Kavitha, "Sales Prediction Analysis", International Research Journal of Engineering and Technology (IRJET), 2021, Volume: 08, Page: 1932-1934
- [3] Purvika Bajaj, Renesa Ray, Shivani Shedge, Shravani Vidhate, Prof. Dr. Nikhilkumar Shardoor, "Sales Prediction Using Machine Learning Algorithms", International Research Journal of Engineering and Technology (IRJET), 2020, Volume: 07, Page: 3619-3625
- [4] Sunitha Cheriyan, Shaniba, Ibrahim, Saju Mohanan, Susan Treasa, "Intelligent Sales Prediction Using Machine Learning Techniques", IEEE, 2018, Page: 53- 58
- [5] Varshini S, Dr. D. Preethi, "An Analysis of Machine Learning Algorithms to Predict Sales", International Journal of Science and Research (IJSR), Volume:11, Page: 462- 466