Journal of Business and Management Studies
ISSN: 2709-0876
DOI: 10.32996/jbms
Journal Homepage: www.al-kindipublisher.com/index.php/jbms

JBMS
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

Addressing Seasonality and Trend Detection in Predictive Sales Forecasting: A Machine Learning Perspective

Md Rokibul hasan
MBA Business Analytics, Gannon University, USA
Corresponding Author: Md Rokibul hasan, **E-mail**: prorokibulhasanbi@gmail.com

| ABSTRACT
Sales prediction plays a paramount role in the decision-making process for organizations across various industries. Nonetheless, accurately predicting sales is challenging because of trends and seasonality in sales data. The prime objective of this research paper was to explore machine learning methodologies and techniques that can efficiently address seasonality and trend detection in predictive sales forecasting. The research focused on pinpointing suitable features based on correlation coefficients, which were then adopted to train the three different models: random forests, linear regression, and gradient boosting. From the performance evaluation, gradient boosting displayed relatively superior performance compared to the other two regarding R2 score and accuracy. These results highlighted the capability of sales prediction through machine learning, offering vital insights for decision-making processes. The findings of this empirical research provide an extensive guideline for executing machine learning techniques in sales forecasting and addressing seasonality and trend detection, especially when working with large datasets. Furthermore, the study shed light on possible challenges and issues encountered in the process. By resolving these issues, retailers can reinforce the reliability and accuracy of their sales predictions, thereby enhancing their decision-making capabilities in the context of sales management.

| KEYWORDS

Trend detection; Machine Learning; Gradient boosting; Random Forest; Linear regression.

| ARTICLE INFORMATION

ACCEPTED: 09 April 2024 **PUBLISHED:** 09 April 2024 **DOI:** 10.32996/jbms.2024.6.2.10

1. Introduction
According to Khan (2020), sales prediction and forecasting play an instrumental role in various sectors facilitating organizations to make informed decisions concerning resource allocation, inventory management, and strategic planning. Nevertheless, accurately predicting sales trends and patterns is a challenging task because of the presence of seasonality and trends in sales data. Bharadiya (2023), asserted that seasonality revolves around the continuous patterns that repeat at regular intervals, such as weekly, monthly, or daily. These patterns can be impacted by factors like weather, holidays, or economic conditions. On the other hand, trend denotes long-term patterns that demonstrate decline or growth over time. Detecting trends and seasonality in sales data is pivotal for accurate forecasting because failing to account for these patterns can cause inaccurate predictions. This research paper examines the adoption of machine learning techniques to combat trend and seasonality detection in predictive sales forecasting and prediction. This study presents an overview of the present methods and proposes a unique approach that utilizes machine learning algorithms to enhance the accuracy of sales forecasts.

As per Ensafi (2022), the consolidation of traditional time series frameworks coupled with modern machine learning and AI techniques for demand and sales forecasting has attracted significant interest within the forecasting spectrum. Well-established methods such as SARIMA, ARIMA, and the Exponential Smoothing State Space Model (ETS) have witnessed widespread employment because of their interpretability, straightforwardness, and potential to capture linear patterns and seasonal trends.

Their performance across different demand and sales prediction settings has solidified their position as a preferred choice among practitioners.

The consolidation of conventional time series frameworks with machine learning and Artificial Intelligence methods provides a holistic approach to forecasting. Hybrid models, which optimize the strengths of both techniques, have arisen as a promising resolution to reinforce robustness and forecast accuracy (Ensafi, 2022). Moreover, the examination of ensemble techniques, which integrate predictions from multiple frameworks, enables leveraging of diverse capabilities of distinct forecasting methods. These consolidation and ensemble techniques offer a comprehensive and powerful framework for enhancing forecasting outcomes. This research proposes a hybrid model that maximizes the capabilities of the Light-GBM algorithm, a high-performance gradient boosting framework, to produce sales predictions for individual time series.

2. Related Works Literature Review

2.1 Seasonality Detection and Modeling:

Zhou (2023) indicated that modeling and detecting seasonality are of paramount importance for accurate sales forecasting. Traditional techniques like seasonal decomposition methods (e.g., autoregressive integrated moving average (ARIMA) and seasonal decomposition of time series models have been widely utilized. Nonetheless, these techniques may struggle with complex and non-linear seasonal patterns. In that light, Machine learning algorithms like seasonal naive forecasting and seasonal-trend decomposition utilizing LOESS (STL) provide alternative techniques for capturing seasonality efficiently.

2.2 Trend Detection and Modeling

Modeling and identifying trends in sales data are instrumental for ascertaining long-term trends and making informed strategic decisions. Conventional methods such as linear regression and exponential smoothing are predominantly adopted for trend detection. Nevertheless, these methods may overlook non-linear patterns and susceptible overfitting issues (Zhou, 2023). As such, Machine learning techniques, such as support vector regression (SVR), polynomial regression, and neural networks, provided more flexibility in pinpointing complex trends and patterns.

2.2 Machine Learning for Predictive Sales Forecasting

As per Leung et al (2020), machine learning is a domain of science that revolves around developing computer algorithms capable of conducting tasks independently, without direct human supervision. Contrary to traditional programming, machine learning highly depends on the capability of computers to learn and decipher new knowledge from data, instead of depending on pre-defined rules. Machine learning algorithms have experienced widespread success across various domains. These algorithms can be broadly classified into two main categories, most notably, supervised and unsupervised machine learning, where each serves distinct purposes in the learning process.

According to Fourkiotis & Tsadiras (2024), machine learning methods provide a wide range of methodologies and algorithms for predictive sales forecasting. A significant volume of studies has examined the employment of machine learning for sales prediction and how it effectively addresses the challenges of seasonality and trend detection. For instance, one study by Zhao et al. (2020) adopted a recurrent neural network (RNN) to predict daily sales for a retail organization. The study established that the RNN model outperformed traditional models in terms of predicting sales accurately, particularly in capturing seasonality and trends. Recurrent Neural Networks, specifically Long Short-Term Memory (LSTM) networks, have captured substantial popularity in time-series predicting tasks (Javatpoint, 2024). LSTMs can efficiently capture nonlinear patterns and long-term dependencies in sales data, making them ideal for resolving trend detection and seasonality.

Besides, another research by Khan et al. (2021) recommended a deep learning method known as the seasonal trend decomposition using the Loess (STL) model for sales prediction. The STL framework decomposes sales data into seasonal, trend, and residual elements to better detect the underlying patterns. The research demonstrated that the STL framework enhanced forecasting accuracy by efficiently capturing seasonality and trends. Furthermore, supervised learning algorithms like decision trees, random forests, linear regression, and support vector machines, have been widely adopted for sales forecasting. These methods have been proven to capture complex associations between input features and sales data, facilitating the detection of both seasonality and trends.

3. Methodology

While it is difficult to find public datasets with adequate data points for predicting seasonal time series, the researcher discovered the Superstore sales dataset (Community.tableau.com, 2020). This dataset displayed a clear seasonality in its sales pattern and had minimal missing values. The dataset adopted in this research presented a comprehensive description of retail store sales from 2015 to 2019. The dataset comprised 9800 data points, each capturing details associated with client orders, sales transactions, and products. The dataset entailed 18 columns, encompassing order details, product attributes, customer information, and sales figures. Notable columns encompassed Order Date, Customer Segment, Product Category, Ship Mode, and Sales Amount.

3.1 Feature Engineering Selection

Through the data explorations, it became apparent that the sales data displayed a hierarchical structure, including various degrees of granularity such as products, individuals, stores, departments, and states. This hierarchical aspect of the data implies that adopting a hierarchical time series model could be an efficient dimension for modeling and predicting the sales data. In a hierarchical time-series model, the data is analyzed and modeled at multiple levels of aggregation. This comprises utilizing distinct models for every level, facilitating the investigator to capture the different trends and patterns of variation available at every level. At the same moment, this model also considers the relationships and inter-dependencies that exist between the different levels.

3.2 Data pre-processing

As per Maulud & Abdulazeez (2020), data pre-processing plays an instrumental role in terms of enhancing the success rate of any project. During the data pre-processing stage, four key phases were performed, most notably: handling missing values, train-test split, integrating redundant categories within every variable, and transforming categorical variables into indicator variables. The initial phase included randomly sub-dividing the presented training set into a new training set, which included 80% of the 8523 data points, and a test set entailing the remaining 20%. Successively, independent data transformation and cleaning procedures were employed for each of these sets. In the setting of this study, time-series aggregation was adopted to minimize computational resources and elevate forecasting accuracy. Subsequently, statistical methods such as Z-scores were then employed to either adjust or remove these outliers from the dataset. Bearing in mind the significant fluctuation in daily product sales before prediction, the study utilized the resampling of the daily data into a monthly frequency. Subsequently, the average daily sales values were then adopted for analysis. The key features pivotal for conducting univariate predictions are the sales amount and order date related to each data point.

3.3 Models and Metrics

Three models were chosen to fit the data: Random Forest, Linear Regression, and Gradient Boosting. Even though other techniques like Stochastic Gradient Descent were also assessed, their respective outcomes were not practical, likely because of the limited size of the training set. Therefore, they were eliminated from further consideration. All models utilized in this research were obtained from the sci-kit-learn library in Python and implemented using Python 3.9. The performance of the models was assessed using the Root Mean Square Error (RMSE):

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(P_i - O_i)^2}{n}}$$

Where,

- RMSE represents Root Mean Squared Error
- Σ stands for the sum of the squared errors for every data point.
- (Pi - Oi) denotes the distinction between the predicted value (P_i) and the noted value (Oi) for each data point i.
- n is the overall number of data points.

During the analysis, the research utilized a linear regression model with the ordinary least squares (OLS) technique. This technique presumes a straight-line association between the independent variable (x) and the dependent variable (y). The model computes a set of coefficients (βi's) that reduces the distinction between the dataset's observed values and the model's predicted values. To enhance the fit, the researcher also included an intercept term. The equation can be expressed as follows:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots$$

Random Forest was also employed in this study, random forest is an ensemble learning method that develops multiple decision trees during training and uses averaging to improve prediction accuracy and reduce overfitting. In the present experiment, the parameter "n estimator," which ascertains the quantity of trees in the forest, was stipulated at 200. Furthermore, the mean squared error function was chosen to assess the quality of splits within the trees.

$$g\,(x) = \int f0(x) + \int 1(x) + \int 2(x) + \int 3\,(x) + \cdots$$

The parameter 'g' stands for the number related to a specific model's initialization, portrayed as 'fi'. In this scenario, each single base classifier is a simple decision tree. The technique of employing multiple models to reinforce predictive performance is predominantly referred to as model ensemble. In random forests, all base models are designed independently, employing distinct subsets of data.

3.4 Experiment

3.4.1 Python

Python is an open-source high-level programming language, which also serves as the primary programming language for both preprocessing tasks and modeling. Python's versatility and popularity in data science for various applications such as web development, is well recognized. Its extensive employment in data science is unquestionable via the abundance of libraries particularly crafted for data processing and modeling. In this study, popular data science libraries such as pandas were adopted for efficient data analysis and manipulation. The NumPy library was also used for scientific calculating purposes, while matplotlib enabled the presentation of distinct data visualizations, such as charts. Moreover, the sci-kit-learn library played a pivotal role in developing computational models.

Importing Libraries

In [1]:
```
import numpy as np
import seaborn as sns
import pandas as pd
from matplotlib import pyplot as plt
```

In [5]:
```
data = pd.read_csv("train.csv")
data
```

Output:

Out[5]:

	Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	Country	City	State	Postal Code	Region
0	1	CA-2017-152156	08/11/2017	11/11/2017	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	Kentucky	42420.0	South
1	2	CA-2017-152156	08/11/2017	11/11/2017	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	Kentucky	42420.0	South
2	3	CA-2017-138688	12/06/2017	16/06/2017	Second Class	DV-13045	Darrin Van Huff	Corporate	United States	Los Angeles	California	90036.0	West
3	4	US-2016-108966	11/10/2016	18/10/2016	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	Florida	33311.0	South
4	5	US-2016-108966	11/10/2016	18/10/2016	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	Florida	33311.0	South

3.5 Loading and Exploring the Data

Before continuing with any transformations, data was loaded into the system. After the data was loaded, it underwent a structural transformation to match the input requirements of each model. In its raw form, every row in the dataset portrayed the daily sales of one of the 10 stores. Nonetheless, since the objective is to forecast monthly sales, the researcher aggregated the data by integrating sales figures from all stores and days to get the overall monthly sales. This aggregation provided the researcher with a consolidated monthly sales figure to work within our analysis.

```python
# Convert date columns to datetime with correct format
data['Order Date'] = pd.to_datetime(data['Order Date'], format='%d/%m/%Y')
data['Ship Date'] = pd.to_datetime(data['Ship Date'], format='%d/%m/%Y')
```

```python
# Handle missing values
data['Postal Code'] = data['Postal Code'].fillna(data['Postal Code'].median())
```

```python
# Encode categorical variables
data = pd.get_dummies(data, columns=['Ship Mode', 'Segment', 'Country', 'City', 'State', 'Region', 'Product ID', 'Category',
```

```python
# Drop unnecessary columns
data.drop(['Row ID', 'Order ID', 'Customer ID', 'Customer Name'], axis=1, inplace=True)
```

```python
# Feature Engineering (Example: Extract month and year from 'Order Date')
data['Order Month'] = data['Order Date'].dt.month
data['Order Year'] = data['Order Date'].dt.year
```

```python
# Normalize numerical data (Example: Sales column)
data['Sales'] = (data['Sales'] - data['Sales'].mean()) / data['Sales'].std()

# Display the preprocessed dataset
print(data.head())
```

Output:

```
  Order Date  Ship Date  Postal Code  \
0 2017-11-08 2017-11-11      42420.0
1 2017-11-08 2017-11-11      42420.0
2 2017-06-12 2017-06-16      90036.0
3 2016-10-11 2016-10-18      33311.0
4 2016-10-11 2016-10-18      33311.0

                                        Product Name     Sales  \
0                  Bush Somerset Collection Bookcase  0.049774
1  Hon Deluxe Fabric Upholstered Stacking Chairs,...  0.799760
2  Self-Adhesive Address Labels for Typewriters b... -0.344927
3      Bretford CR4500 Series Slim Rectangular Table  1.159828
4                     Eldon Fold 'N Roll Cart System -0.332563

  Ship Mode_First Class  Ship Mode_Same Day  Ship Mode_Second Class  \
0                 False               False                    True
1                 False               False                    True
2                 False               False                    True
3                 False               False                   False
4                 False               False                   False

  Ship Mode_Standard Class  Segment_Consumer  ...  Sub-Category_Furnishings
0                    False              True  ...                     False
1                    False              True  ...                     False
2                    False             False  ...                     False
3                     True              True  ...                     False
4                     True              True  ...                     False

  Sub-Category_Labels  Sub-Category_Machines  Sub-Category_Paper  \
0               False                  False               False
1               False                  False               False
2                True                  False               False
3               False                  False               False
4               False                  False               False
```

As regards the following data frame, each row was updated to portray the highest sales documented for a specific year across all stores. This implies that for every year, the row displays the maximum sales value attained among all the stores. By arranging the data in this way, the researcher was able to gain insights into the peak sales performance within every year and assess the patterns and trends related to the highest sales figures across the stores.

In [17]:

```
import matplotlib.pyplot as plt
import seaborn as sns

# Top Selling Products
top_products = data.groupby('Product Name')['Sales'].sum().nlargest(10).reset_index()
plt.figure(figsize=(12, 6))
sns.barplot(x='Sales', y='Product Name', data=top_products)
plt.title('Top Selling Products')
plt.xlabel('Sales')
plt.ylabel('Product Name')
plt.show()
```

Output:

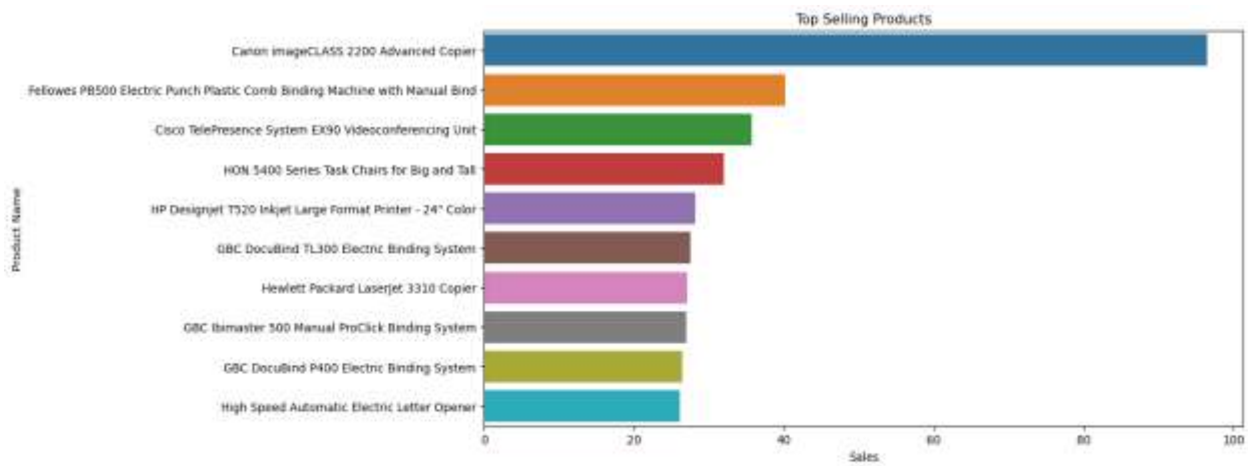


Fig 1: Displays top-selling product

From the above chart, it was evident that Canon image CLASS 2200 advanced copier was the best-selling product, followed by Fellowes PB500 punch plastic comb binding machine with manual bind, Cisco Telepresence system EX90 video conferencing unit, HON 5400 series Task chairs for Big and tall, and Hp Design-jet T520 inkjet large format printer. On the other hand, the least-selling products were high-speed automatic electric letter openers and the GBC Docu-Bind p400 Electric binding system.

The researcher also computed the distinction between the sales of each month and added it as a new column to the data frame. This conversion was conducted to make the data stationary, which assisted in modeling and assessing time series data effectively. Particularly, the sales trend () function was executed to provide information about the trend and pattern of sales. The function represented order date in terms of months, days, and years as well as their respective equivalent sales, therefore providing an extensive understanding of the period covered by the sales data.

In [23]:

```
# Sales Trend Over Time
plt.figure(figsize=(10, 6))
sns.lineplot(x='Order Date', y='Sales', data=data)
plt.title('Sales Trend Over Time')
plt.xlabel('Order Date')
plt.ylabel('Total Sales')
plt.xticks(rotation=45)
plt.show()
```

Output:

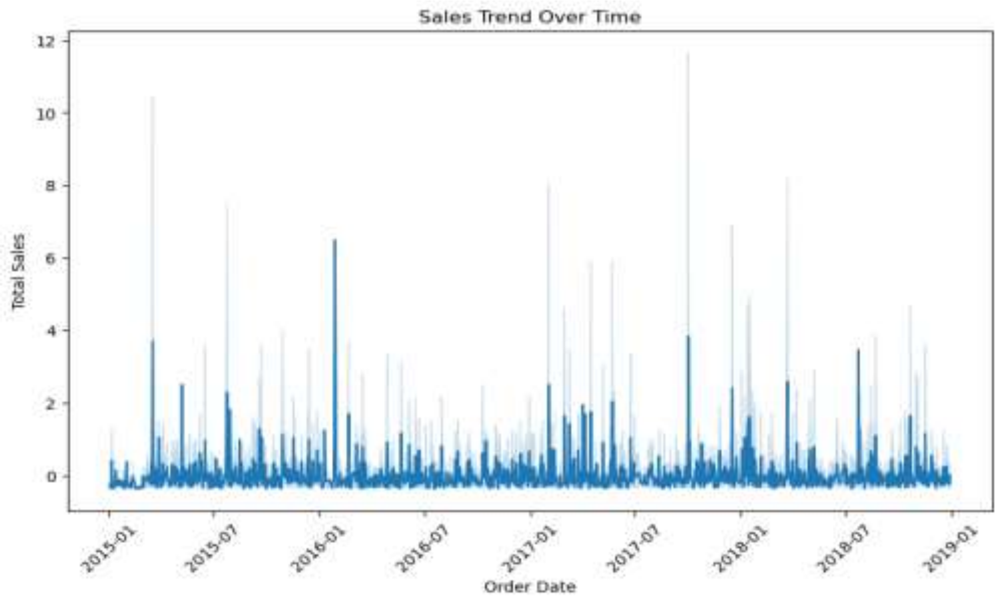


Fig 2: Showcases Sales trend over time

The chart portrayed in the chart above presents an insightful analysis regarding the average yearly sales across distinct product categories from 2015 to 2019. The chart exhibits notable trends and patterns observed within the sales data. Particularly, it is apparent that the sales display a peak during the fourth quarter and a drop in the first quarter of each year. This recurring pattern indicates the presence of seasonal impacts shaping the sales performance. The escalation in sales during the fourth quarter may be linked to various factors, such as promotions, holiday seasons, or heightened consumer spending. In contrast, the drop in sales during the first quarter may be impacted by post-holiday activities or reduced consumer activity.

In [31]:
```
# Customer Segmentation
plt.figure(figsize=(8, 6))
data['Segment_Consumer'].value_counts().plot(kind='pie', autopct='%1.1f%%')
plt.title('Customer Segmentation')
plt.ylabel('')
plt.show()
```

Output:

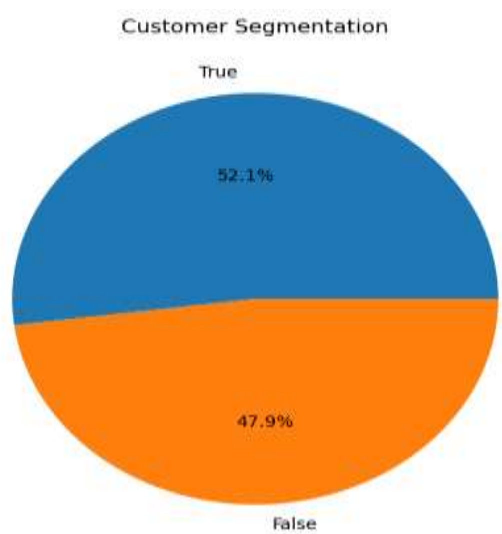


Fig 3: Exhibits customer segmentation

The chart above exhibited "Customer Segmentation," Where consumers were subdivided into two categories: true and false. "True" accounted for 52.1% of valid customers and “False” accounted for 47.9% of invalid consumers. True consumers were classified as those who made a purchase, while false consumers may be classified as those who abandoned their shopping carts or signed up for a service but did not use it.

3.6 Models Performance

3.6.1 Linear Regression

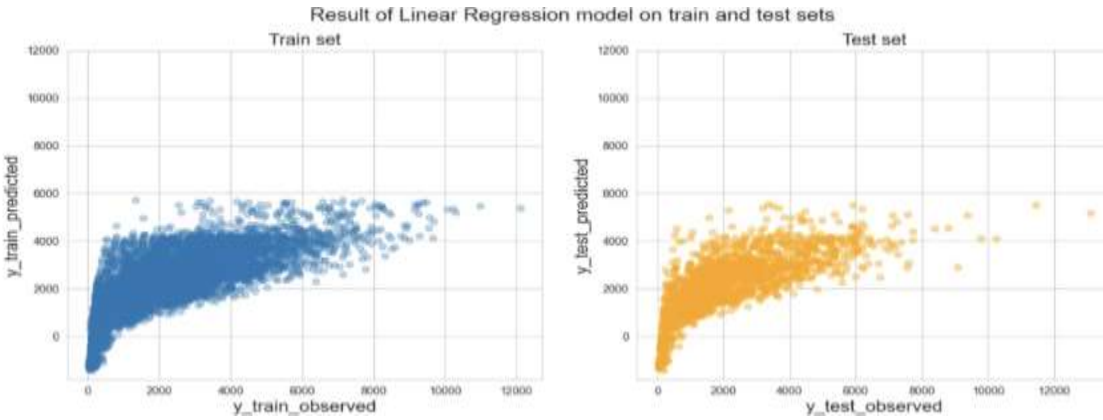
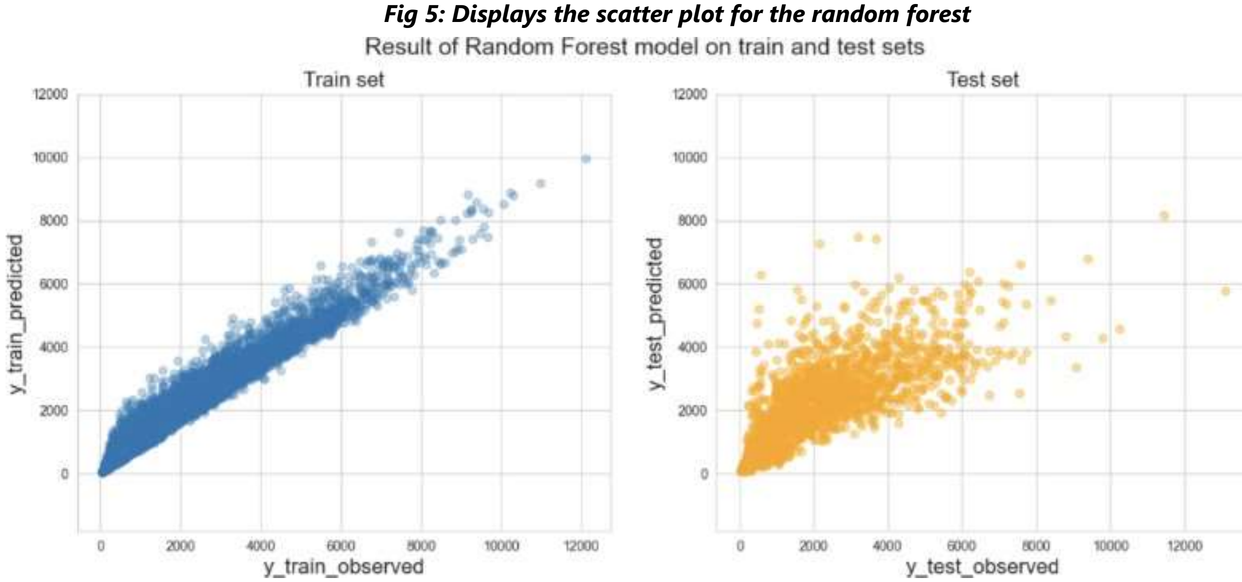


Figure 4: Portrays the scatter plot for the linear regression model

The results of the linear regression model, as portrayed in Figure 4, deviated from linearity. Instead, it displayed a curved trend with a positive but shrinking slope, reaching an upper limit of approximately 7000 for the predicted values. Particularly, for a proportion of the data where the observed values were under 1000, the model's projections fell below 0, which was relatively inaccurate since sales values cannot be negative. The unsatisfactory outcomes were attributed to the nature of the features adopted in the training process. The majority of the features employed were categorical variables, which were not appropriate for linear regression when the correlation was low. By nature, categorical variables lack a linear association with the target variable, causing sub-optimal model performance in scenarios where linear regression depends on solid correlations for accurate predictions.

3.6.2 Random Forest

Fig 5: Displays the scatter plot for the random forest



As showcased in Figure 5, the Random Forest model's performance was relatively logical compared to the linear regression model. The Random Forest model displayed excellent performance on the training set, which was attributed to the inherent attribute of decision trees. The high performance on the training provided much informative value.

3.6.3 Gradient Boosting

Fig 6: Showcases scatter plot for the Gradient Boosting Model

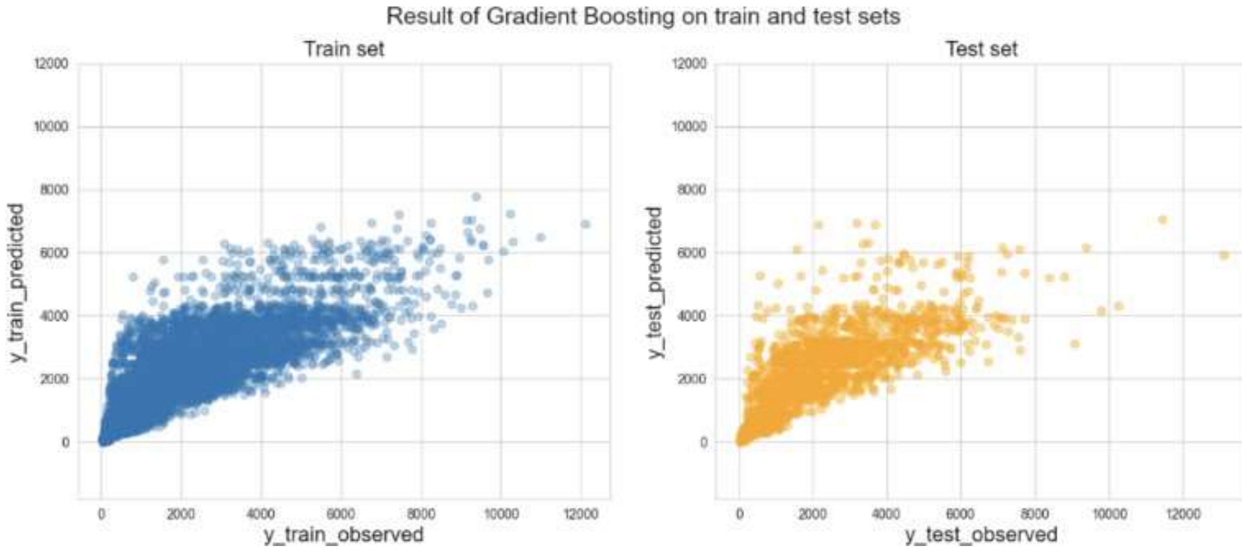


Figure 6 displays the outcomes attained from the Gradient Boosting model. Similar to Random Forest, Gradient Boosting also employs decision trees. Gradient Boosting proved to have relatively superior performance in terms of fitting the training set because of its additive nature. Besides, its performance on the test set portrayed a significant improvement, as the linear pattern of the data points became clearer. It is noteworthy that the forecasted values produced by the Gradient Boosting model were quite lower than the true values. Furthermore, the matter noticed in the Random Forest model concerning predictions for data points with true values beyond 7000 persisted in the Gradient Boosting model as well. The R2 scores and errors of different models were essentially constant with the patterns portrayed by the scatter plots, as presented in Table. 1.

	Random Forest	Linear Regression	Gradient Boost	
MAE train	293.638	834.714	727.225	
RMSE train	425.269	1124.926	1029.046	
R2_test	0.574	0.567	0.609	
RMSE_test	1133.550	1142.004	1086.291	
MAE_test	790.548	840.730	753.112	

Table 1: Exhibits the R2 scores and errors of different models

Taken together, considering the three models examined, Gradient Boosting demonstrated the best overall performance. In particular, it attained the highest R2 score and showcased the lowest errors during testing. These results matched with the general comprehension that gradient-boosted trees tended to outperform random forests. Although, random Forest, slightly less efficient than Gradient Boosting, still exhibited respectable performance. The performance rift between Random Forests and Gradient Boosting was constant with previous research that underscored the solid performance of gradient-boosted trees. Conversely, the linear regression model illustrated relatively lower accuracy as compared to the other two models. This inference was supported by both the Mean Absolute Error (MAE) and the Root Mean Squared Error (RMSE) metrics (ProAIRokibul – Repositories, 2024). Although the linear regression model's RMSE was close to that of the other models, it suffered from the greatest number of outliers in its outcomes. These inconsistencies were attributed to the models' inability to accurately forecast data with large values.

3.7 Business Impact of Adopting Our Model

Adopting the proposed will undoubtedly Improve sales forecasting by providing significant competitive advantage and strategic benefits, enabling companies to make well-informed choices and improve operational efficiency in the following ways:

1. ***Streamlined Inventory Management Accuracy***: In particular, by accurately forecasting sales patterns and trends, an organization can enhance its inventory levels, minimizing the risks connected to stockouts or surplus stock. This accuracy subsequently leads to an enhanced and agile supply chain (BiznestAI, 2024).
2. ***Boosted Operational Efficiency***: Predicting and forecasting demand empowers companies to streamline logistics and reduce operational expenses. Consequently, this elevated operational efficiency results in enhanced overall productivity (BiznestAI, 2024).

3. ***Strategic Financial Planning:*** Forecasting future sales trends establishes the foundation of strategic financial planning. Particularly, it facilitates companies to proactively disseminate resources, set achievable revenue targets, and pinpoint growth opportunities. This foresight is pivotal for ensuring fiscal well-being and accomplishing long-term financial goals (BiznestAI, 2024).
4. ***Data-Driven Decision Making***: Informed by predictive sales analytics, organizations can match their products more efficiently with consumer preferences. This data-driven dimension not only improves client satisfaction but also cultivates longevity and brand loyalty in a competitive market environment (BiznestAI, 2024).

4. Conclusion

In this research, the efficiency of machine learning techniques in addressing seasonality and trend detection in predictive sales forecasting sales for every product in the Super Stores was thoroughly explored. The study concentrated on pinpointing suitable features based on correlation coefficients, which were then employed to train the three different models, most notably, random forests, linear regression, and gradient boosting. From the performance evaluation, gradient boosting displayed relatively superior performance compared to the other two regarding R2 score and accuracy. These results highlighted the capability of sales prediction through machine learning, offering vital insights for decision-making processes. It is paramount to mention that while the predictive power of sales forecasting was ascertained to be moderately efficient, it still presented valuable guidance for big retailers. The findings of this research present an extensive guideline for executing machine learning techniques in sales forecasting, especially when working with large datasets. Furthermore, the study shed light on possible challenges and issues encountered in the process. By resolving these issues, retailers can reinforce the reliability and accuracy of their sales predictions, thereby enhancing their decision-making capabilities in the context of sales management.

Funding: This research received no external funding.
Conflicts of Interest: The authors declare no conflict of interest.
Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers.

References

[1] Bharadiya, J. P. (2023). Machine learning and AI in business intelligence: Trends and opportunities. *International Journal of Computer (IJC)*, *48*(1), 123-134.
[2] BIZNESTAI. (2024). Elevate your experience: AI magic in every click. Retrieved from: https://biznestai.com/
[3] Ensafi, Y., Amin, S. H., Zhang, G., & Shah, B. (2022). Time-series forecasting of seasonal item sales using machine learning – A comparative analysis. *International Journal of Information Management Data Insights*, 2(1), 100058. https://doi.org/10.1016/j.jjimei.2022.100058
[4] Fourkiotis, K. P., & Tsadiras, A. (2024). Applying machine learning and statistical forecasting methods for enhancing pharmaceutical sales predictions. Forecasting, 6(1), 170–186. https://doi.org/10.3390/forecast6010010
[5] Javatpoint. (2024.) Sales prediction using Machine Learning -. www.javatpoint.com. https://www.javatpoint.com/sales-prediction-using-machine-learning
[6] Khan, M. A., Saqib, S., Alyas, T., Rehman, A. U., Saeed, Y., Zeb, A., ... & Mohamed, E. M. (2020). Effective demand forecasting model using business intelligence empowered with machine learning. IEEE Access, 8, 116013-116023.
[7] Leung, K. H., Mo, D. Y., Ho, G. T., Wu, C. H., & Huang, G. Q. (2020). Modeling near-real-time order arrival demand in e-commerce context: a machine learning predictive methodology. Industrial Management & Data Systems, 120(6), 1149-1174.
[8] Ma, S., & Fildes, R. (2021). Retail sales forecasting with meta-learning. European Journal of Operational Research, 288(1), 111-128.
[9] Maulud, D., & Abdulazeez, A. M. (2020). A review on linear regression comprehensive in machine learning. *Journal of Applied Science and Technology Trends*, *1*(2), 140-147.
[10] ProAIRokibul – Repositories, (2024). GitHub. https://github.com/proAIrokibul?tab=repositories
[11] Webtunix Solutions (2024). RIS AI - Expert Thesis Writing Services. Ris-ai. https://www.ris-ai.com/python-sales-forecasting
[12] Zhou, T. (2023). Improved Sales Forecasting using Trend and Seasonality Decomposition with LightGBM. ResearchGate. https://www.researchgate.net/publication/371136595_Improved_Sales_Forecasting_using_Trend_and_Seasonality_Decomposition_with_LightGBM