Team 20 : Retail Intelligence: Enhancing Demand Forecasting and Inventory Planning for Holidays

AAI 695: Applied Machine Learning

Team 20 Final Project

Prof. Shucheng Yu

Team Members

Namratha Nagathihalli Anantha
Master's in Data Science
Dept. Mathematical Science
Hoboken, USA
nnagathi@stevens.edu

Rohit Tiwari
Master's in Data Science
Dept. Mathematical Science
Hoboken, USA
rtiwari3@stevens.edu

Suvarsha Sai Punyala
Master's in Data Science
Dept. Mathematical Science
Hoboken, USA
spunyala@stevens.edu

Problem Statement

- Demand forecasting plays a pivotal role in the do main of Supply Chain Management (SCM) within organizations.
- It is imperative to predict sales figures accurately, especially when sales patterns are influenced by various factors such as festival seasons like Thanksgiving, Diwali, Christmas ,characterized by unique seasonal promotions resulting in diverse sales trends.

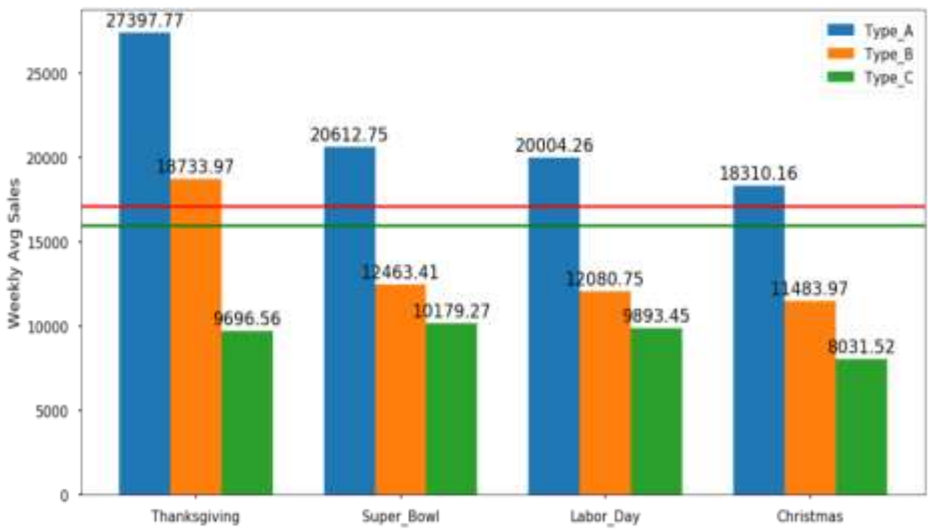
Key Issues:

- Inaccurate Demand Predictions
- Manual and Time-Consuming Processes
- Lack of Adaptability to Market Dynamics

Introduction

- In the fast-paced field of business, the ability to forecast and adapt to changing consumer demands is a crucial factor that can define the success of enterprises.
- The ability of demand planners to predict future trends is a cornerstone in achieving operational efficiency and responsiveness.
- Accurate sales predictions are very important, particularly when influenced by various factors, ranging from festive seasons such as Thanksgiving, Diwali, and Christmas to special events like Labor Day, each characterized by unique promotional dynamics that shape diverse sales trends.
- Machine learning gives forecasting models the capacity to analyze large datasets, spot trends, and gain knowledge from past performance.

- Machine learning models, as opposed to static forecasting techniques, adapt and get better with each new set of data, allowing companies to more precisely manage the challenges of demand prediction.
- Moreover, the sales trajectory of products is further complicated by continuous growth or declining demand, each exhibiting distinctive patterns. To address this complexity and enhance the precision of sales trend forecasts, we propose the utilization of advanced machine learning models such as Linear Regression, Decision Tree, Random Forest, XG Boost, and Feedforward Neural Network.
- By embracing these cutting- edge technologies, organizations can not only meet current consumer demands more accurately but also position themselves to forecast and adapt to future market trends with exceptional precision.



Related Work

1) M. A. Khan et al., ”Effective Demand Forecasting Model Using Business Intelligence Empowered with Machine Learning,” in IEEE Access, vol. 8, pp. 116013-116023, 2020,doi:10.1109/ACCESS.2020.3003790.
   In this study, effectiveness of time series and rule-base forecasting has been analyzed. As per forecasting calculations, the performance of DeepAR models is highly accurate and comparative. This concludes the percentage error values are relatively small so DeepAR models give a high percentage on accuracy for forecasting.

2) M. A. I. Arif, S. I. Sany, F. I. Nahin and A. S. A. Rabby, ”Comparison Study: Product Demand Forecasting with Machine Learning for Shop,” 2019 8th International Conference System Modeling and Advancement in Research Trends (SMART), Moradabad, India, 2019, pp. 171-176, doi: 10.1109/SMART46866.2019.9117395. Out of the many existing models, K Nearest Neighbor, Gaussian Nave Bayes and Decision Tree Classifier algorithms are used for predicting the demand of a product. Customer’s behavior, seasonal weather, are calculated in the study. 58.92% accuracy was obtained on Gaussian Nave Bayes and this gave the maximum demand prediction. The drawbacks were not implementing on the real-life markets, limited dataset and to find more decision-making attributes for prediction.

3) M. E. Hoque, A. Thavaneswaran, S. S. Appadoo, R. K. Thulasiram and B. Banitalebi, ”A Novel Dynamic Demand Forecasting Model for Resilient Supply Chains using Machine Learning,” 2021 IEEE 45th Annual Computers, Software, and Applications Conference (COMPSAC), Madrid, Spain, 2021, pp. 218-227, doi: 10.1109/COMPSAC51774.2021.00040.

   In this study, MMSE forecasts of the future demand are obtained by fitting an appropriate ARMA time series model. The results show that the performance of the proposed forecasts is better than the commonly used MMSE forecasts of the future demand. However, a major drawback of the MMSE forecasting method is that it does not provide the associated risk forecasts.

4) M. Tan, S. Yuan, S. Li, Y. Su, H. Li and F. He, "Ultra-Short- Term Industrial Power Demand Forecasting Using LSTM Based Hybrid Ensemble Learning," in IEEE Transactions on Power Systems, vol. 35, no. 4, pp. 2937-2948, July 2020, doi:10.1109/TPWRS.2019.2963109.

   In this work, a novel deep ensemble learning model was proposed to forecast ultra-short-term industrial power demand. The model obtains the smallest forecasting errors which makes it accurate. But the drawback is that the model performance, especially peak demand forecasting accuracy needs to be further improved to support more advanced power demand control.

5) P. K. Bala, ”Decision tree based demand forecasts for improving inventory performance,” 2010 IEEE International Conference on Industrial Engineering and Engineering Management, Macao, China, 2010, pp. 1926- 1930, doi: 10.1109/IEEM.2010.5674628

Induction of the decision tree is the most important feature in the model suggested. In this context, the proposed model is appropriate for items for which a substantial proportion of sale is attributed strongly to a particular profile or profiles. For such cases, the suggested model can be extended further to understand the dynamics of migration and immigration of people with specific profiles of interest in a locality. But, care should be taken while identifying the attributes for classification prior to feature selection.

Data Pre-processing

- The sample data of a retail chain considered for this implementation is from Kaggle.
- It contains Walmart data withing the timeframe of 2010 to 2012 across 40 of its stores.
- The data set consists of the following features:
    1. Store - the store number
    2. Date - the week of sales
    3. Weekly Sales - sales for the given store
    4. Holiday Flag - week 1 – Holiday week; 0 – Non holiday week)
    5. Temperature - Temperature on the day of sale
    6. Fuel Price - Cost of fuel in the region
    7. CPI – Prevailing consumer price index
    8. Unemployment - Prevailing unemployment rate

Walmart Data

- Sample Data and Headers:

	store	date	weekly_sales	is_holiday	temperature	fuel_price	cpi	unemployment
0	1	05-02-2010	1643690.90	0	42.31	2.572	211.096358	8.106
1	1	12-02-2010	1641957.44	1	38.51	2.548	211.242170	8.106
2	1	19-02-2010	1611968.17	0	39.93	2.514	211.289143	8.106
3	1	26-02-2010	1409727.59	0	46.63	2.561	211.319643	8.106
4	1	05-03-2010	1554806.68	0	46.50	2.625	211.350143	8.106

Data Pre-processing

1. Check for blank and duplicate values:
   No Blank or Duplicate Values were found in the data set

2. Transform the Date Column:
   The Date values of DD/MM/YYYY was transformed to extract the year, quarter, month, weak, day of the week and season values

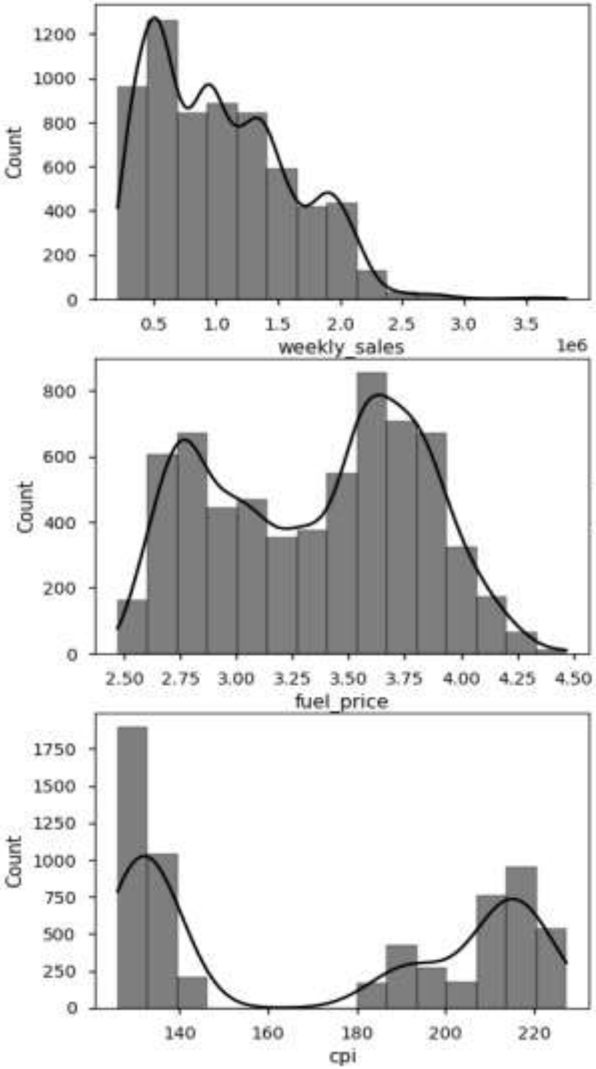
	store	date	weekly_sales	is_holiday	temperature	fuel_price	cpi	unemployment	year	quarter	season	month	month_name	week	day_of_week
0	1	2010-02-05	1643690.90	0	42.31	2.572	211.096358	8.106	2010	1	Winter	2	February	5	Friday
1	1	2010-02-12	1641957.44	1	38.51	2.548	211.242170	8.106	2010	1	Winter	2	February	6	Friday
2	1	2010-02-19	1611968.17	0	39.93	2.514	211.289143	8.106	2010	1	Winter	2	February	7	Friday
3	1	2010-02-26	1409727.59	0	46.63	2.561	211.319643	8.106	2010	1	Winter	2	February	8	Friday
4	1	2010-03-05	1554806.68	0	46.50	2.625	211.350143	8.106	2010	1	Winter	3	March	9	Friday

3. Based on the attributes the data fields were split into Numerical and Categorical features as below:

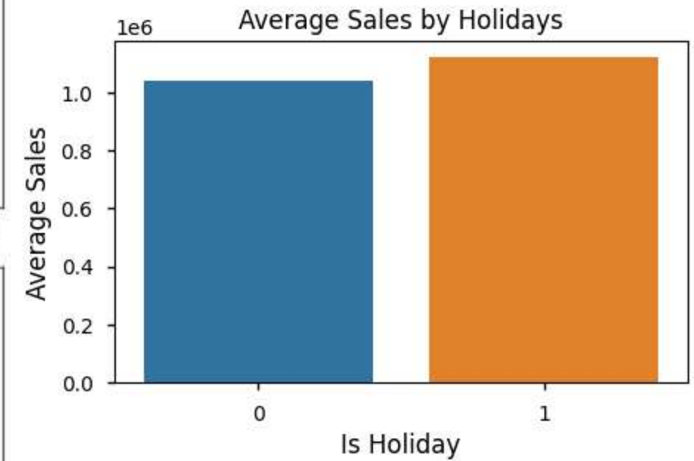
   Numerical Columns: Temperature, Fuel Price, Unemployment and CPI
   Categorical Columns: is holiday, year, season, month and day of week

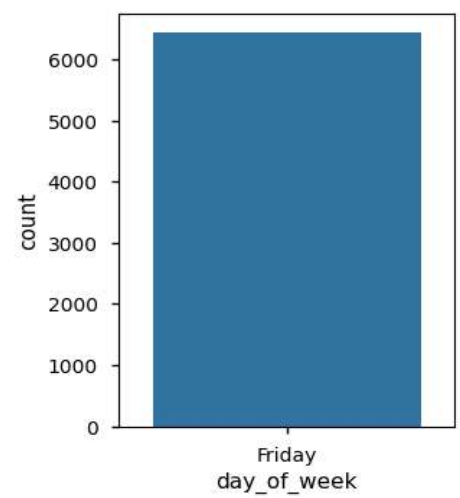
Data Pre-processing

Plotting the Numerical Columns in a histogram:



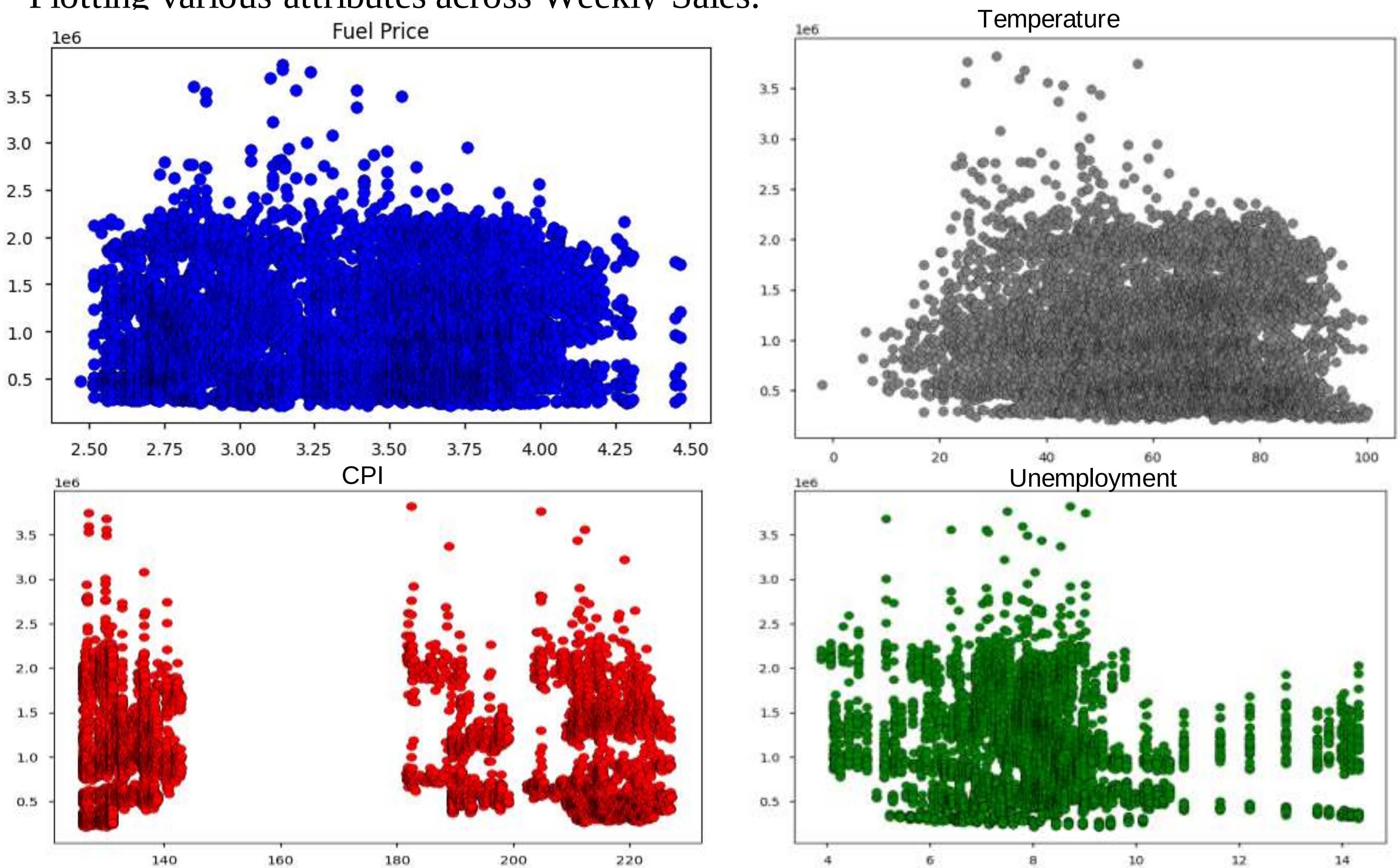
Plotting Average sales by Holiday Flag and Day of the week:





Data Pre-processing

Plotting various attributes across Weekly Sales:



Data Pre-processing

Pearson's Correlation Coefficient for Numerical Data:

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

r = correlation coefficient

x_i = values of the x-variable in a sample

$\bar{x}$ = mean of the values of the x-variable

y_i = values of the y-variable in a sample

$\bar{y}$ = mean of the values of the y-variable

Observations:

1. The Pearson's Correlation Coefficient between temperature and weekly sales is -0.0638 with a P-value of 0.0000
2. The Pearson Correlation Coefficient between fuel price and weekly sales is 0.0095 with a P-value of **0.4478**
3. The Pearson Correlation Coefficient between unemployment and weekly sales is -0.1062 with a P-value of 0.0000
4. The Pearson Correlation Coefficient between cpi and weekly sales is -0.0726 with a P-value of 0.0000

Data Pre-processing

Cramer's V for Categorical Data:

$$V = \sqrt{\frac{\varphi^2}{\min(k-1, r-1)}} = \sqrt{\frac{\chi^2/n}{\min(k-1, r-1)}}$$

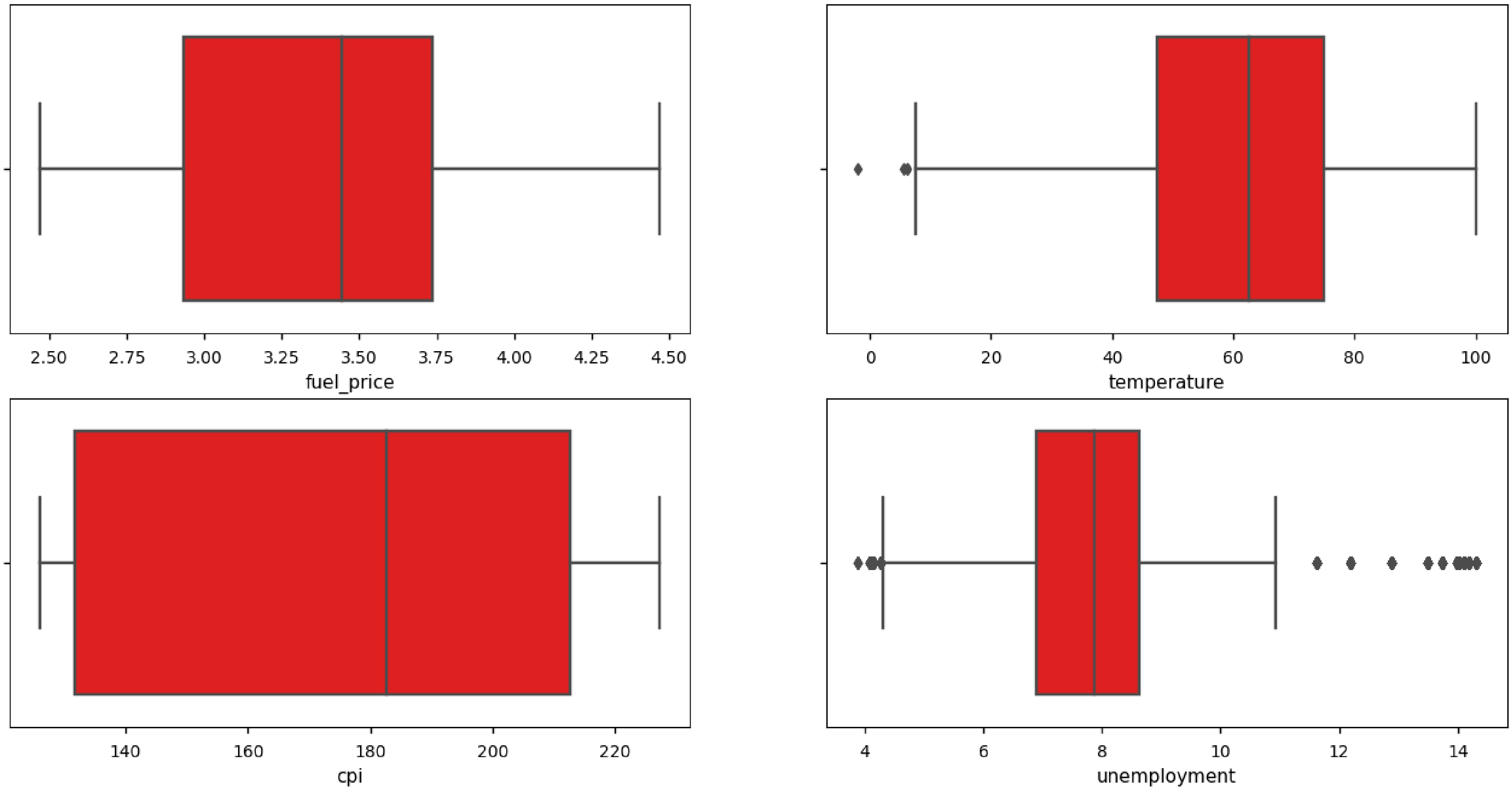
- φ is the phi coefficient.
- χ^2 is derived from Pearson's chi-squared test
- n is the grand total of observations and
- k being the number of columns.
- r being the number of rows.

Observations:

1. Cramér's V for is holiday vs. weekly sales: 1.0000
2. Cramér's V for year vs. weekly sales: 1.0000
3. Cramér's V for season vs. weekly sales: 1.0000
4. Cramér's V for month name vs. weekly sales: 1.0000
5. Cramér's V for day of week vs. weekly sales: nan

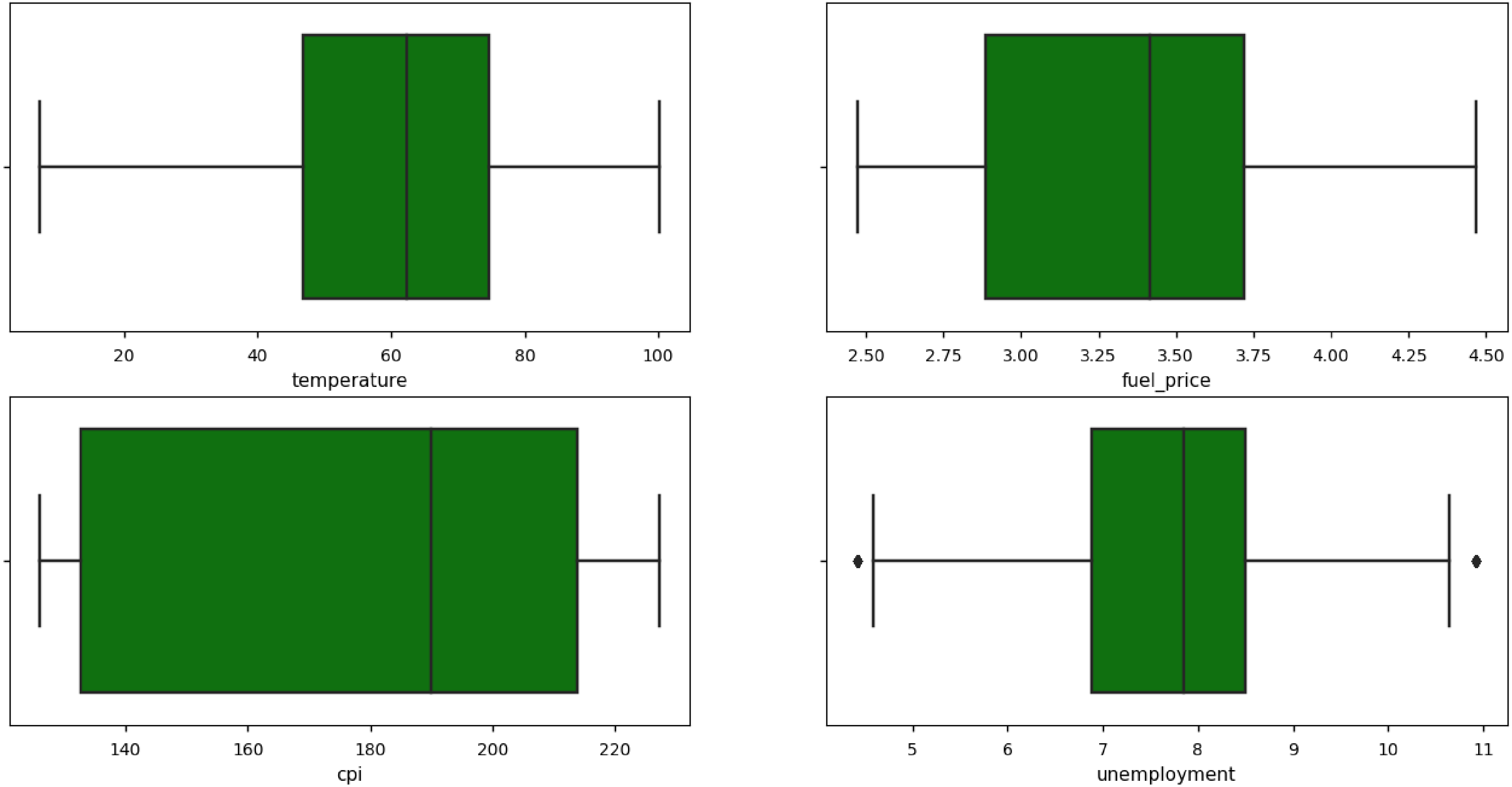
Data Pre-processing

Outlier Elimination (Identification)



Data Pre-processing

Outlier Elimination (Post-processing)



Data Pre-processing

Data Encoding:

1. For Numerical Features of 'temperature', 'fuel price', 'unemployment' and 'CPI’, we used Standard Scalar encoding.
   The standard score of a sample x is calculated as:

   $$z = (x - u) / s$$

   here: u is mean
   s is standard deviation

2. For categorical features of 'is holiday' and 'season', we used Binary Encoder
   In binary encoding, each category is assigned a unique binary code. The length of the binary code depends on the number of categories

The data is now ready for machine learning models to be implemented.

Linear Regression model

Linear regression is a statistical method used to model the relationship between a dependent variable (what you want to predict) and one or more independent variables (the features or factors that may influence the prediction) by fitting a linear equation to observed data.

$Y = b_0 + b_1X^1 + b_2X^2 + . . . + b_nX^n$

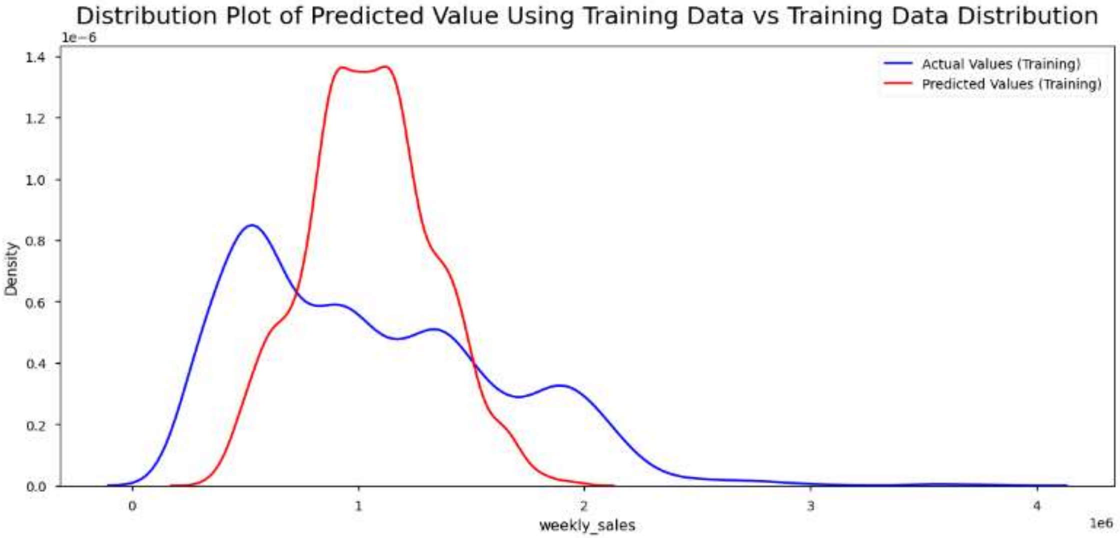
where:

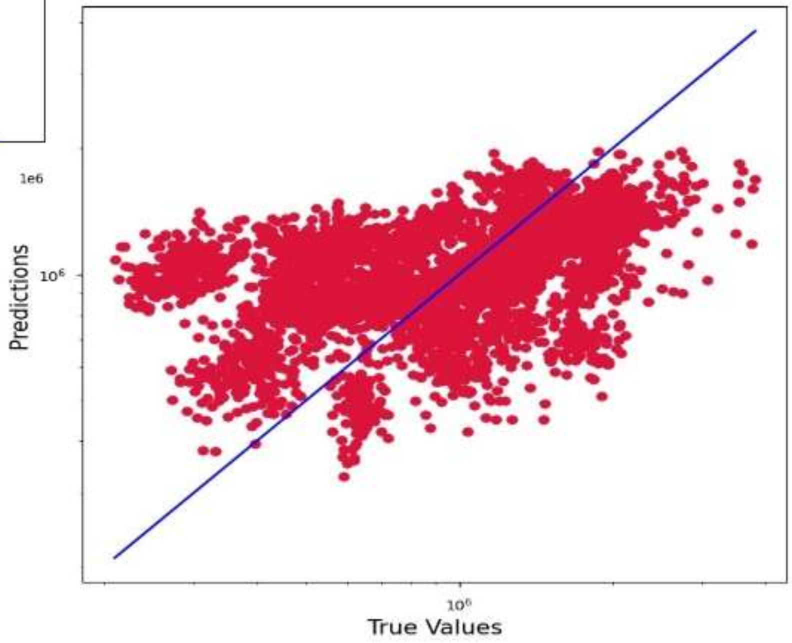
- Y is the dependent variable,
- X^1, X^2,. . . ,X^n are the independent variables,
- b_0 is the y-intercept,

Linear Regression model

Training Data vs Predicted Training Data (before hyperparameter tuning)

- **Normalized Root Mean Squared Error:** 0.1375
- **R-Square score Training:** 24.26%

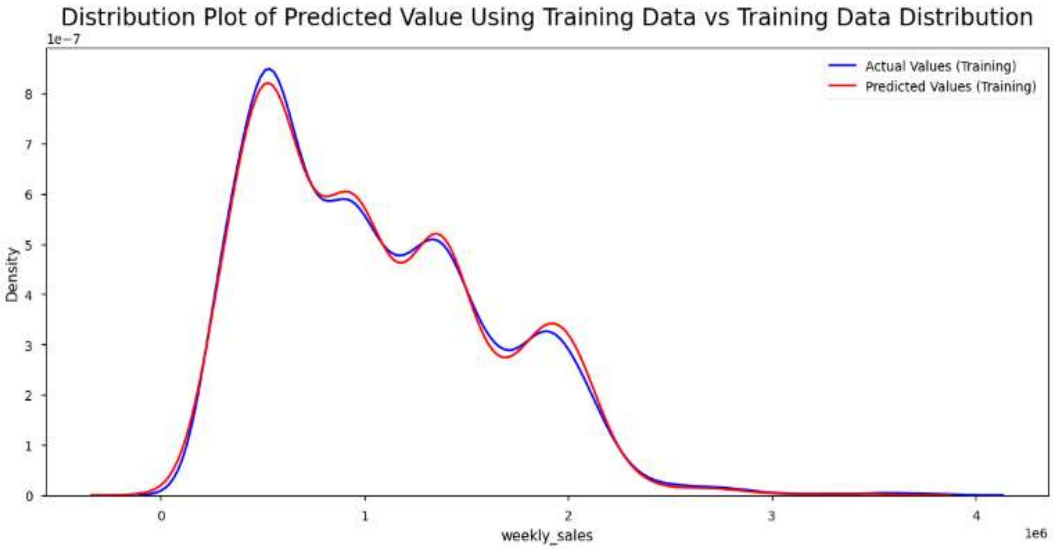


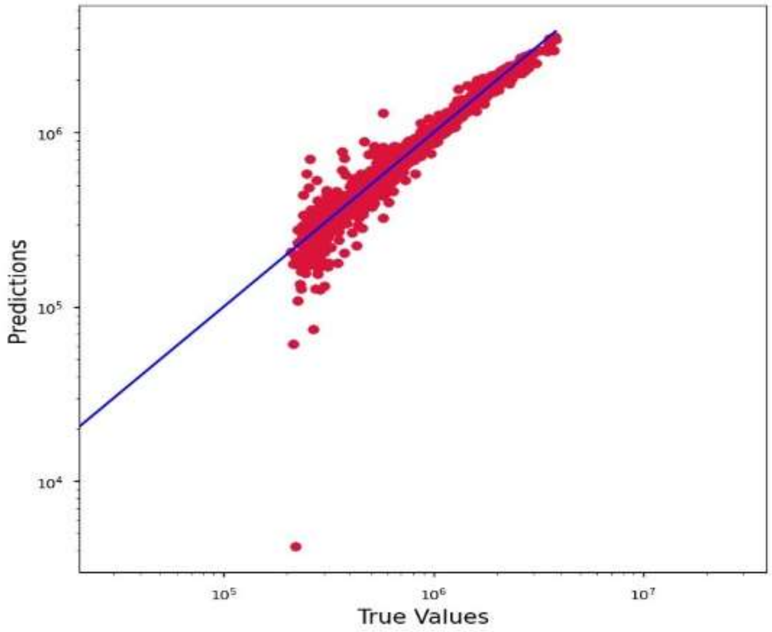


Linear Regression model

Training Data vs Predicted Training Data (after hyperparameter tuning) (poly_feat_degree)

- **Normalized Root Mean Squared Error:** 0.0183
- **R-Square score Training:** 98.66%

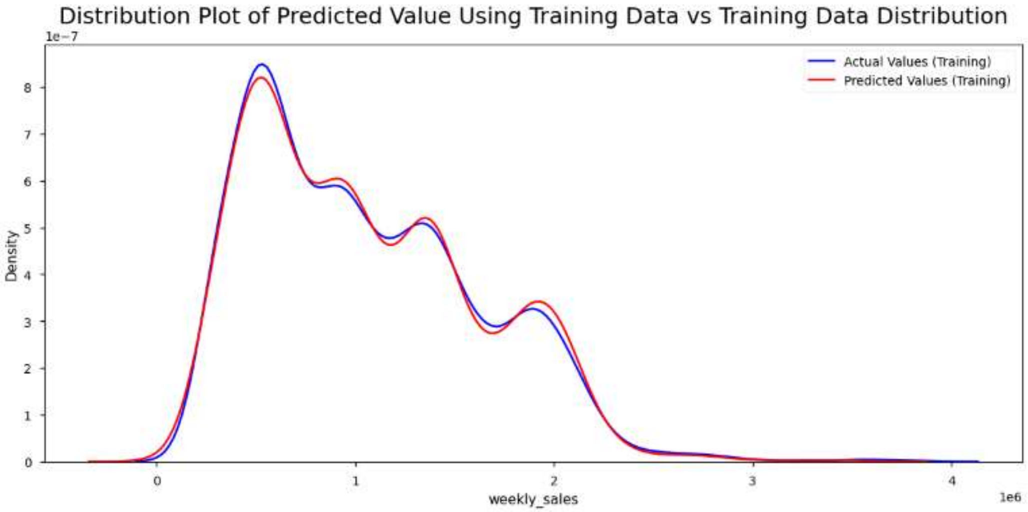


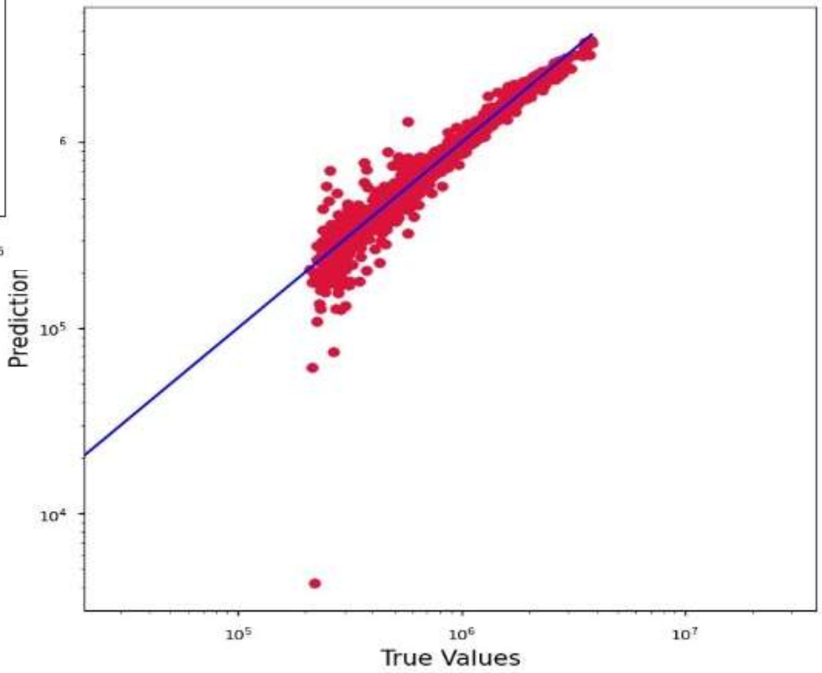


Linear Regression model

Testing Data vs Predicted Testing Data

- **Normalized Root Mean Squared Error:** 0.027
- **R-Square score Testing:** 97.13%





Decision Tree

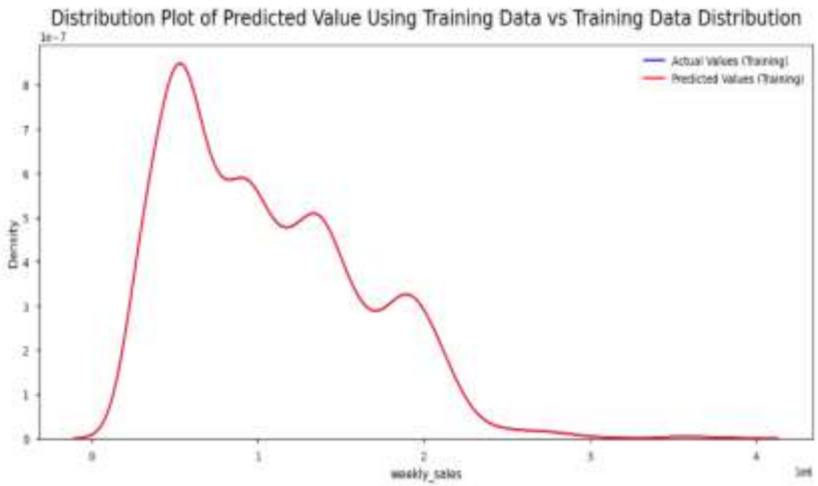
Decision trees are chosen for their interpretability, ease of implementation, and ability to handle both classification and regression tasks. They mimic human decision-making processes, breaking down complex decisions into a series of simpler ones.

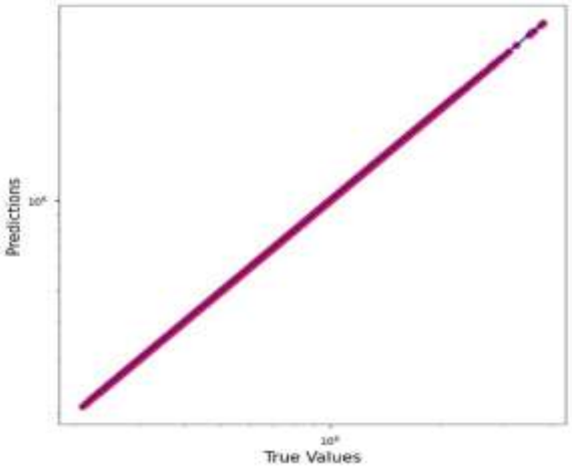
Initial parameters:

- ’max_depth’ :13
- ’min_samples_split’:30

Before hyperparameter Tuning:

- **Normalized Root Mean Squared Error:** 0.0
- **R-Square score Training:** 100.0%

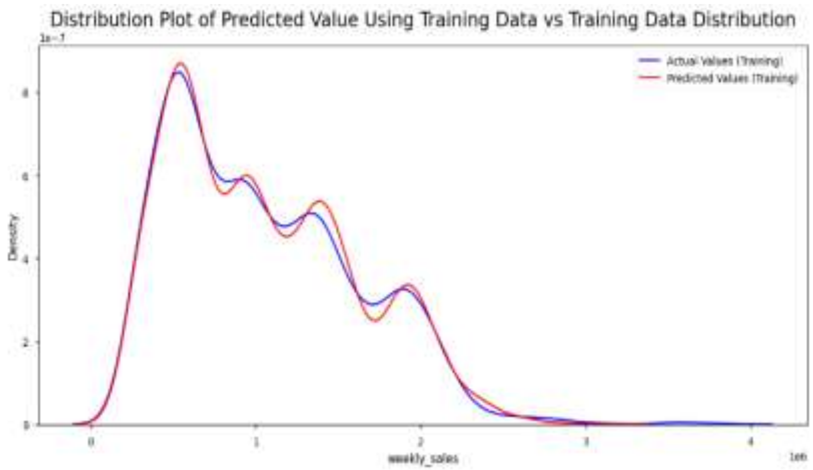


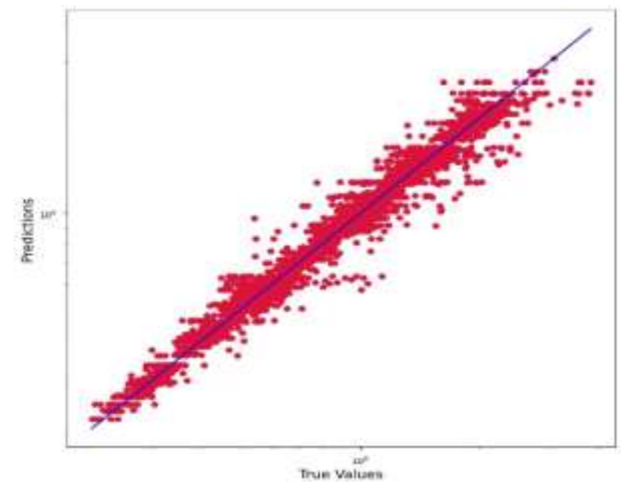


Decision Tree

After hyperparameter Tuning:

- **Normalized Root Mean Squared Error:** 0.0334
- **R-Square score Training:** 95.52%



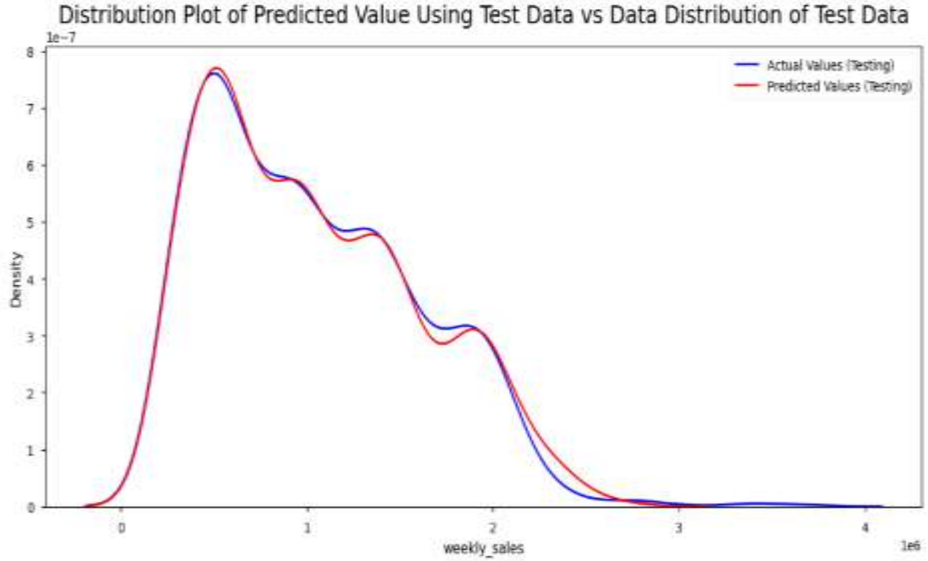


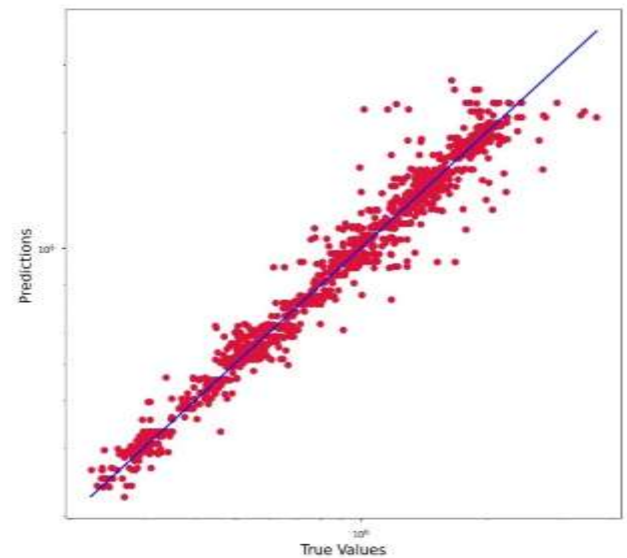
- **Cross Validation Scores:** [0.91626155 0.93674064 0.9278914 0.92195092 0.88405018 0.91331679 0.92466785 0.90900891 0.89630786 0.94763668]
- **Mean of Scores:** 91.78%
- **Standard Deviation of Scores:** 0.01765373204085632

Decision Tree

Testing the model :

- **Normalized Root Mean Squared Error:** 0.0489
- **R-Square score Testing:** 91.22%





The achieved accuracy using the decision tree model is 91.22%.

Random Forest

Random Forest is often chosen for its robustness, versatility, and high predictive accuracy. It is an ensemble learning method that constructs multiple decision trees during training and outputs the mode of the classes (classification) or the mean prediction (regression) of the individual trees. Random Forest is less prone to overfitting compared to individual decision trees, and it can handle large datasets with high dimensionality.

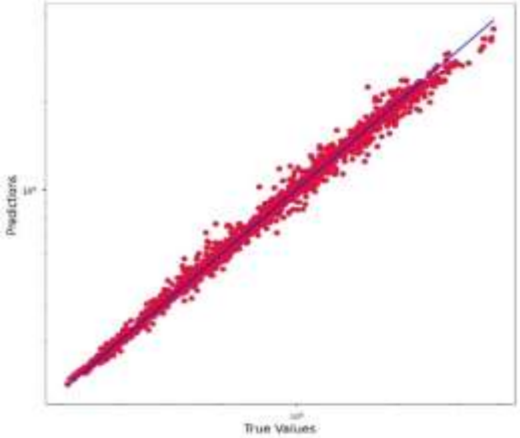
Initial parameters:

- ’n_estimators’
- ’max_features’
- ’max_depth’
- ’min_samples_split’
- ’min_samples_leaf’.

Before Hyperparameter Tuning:

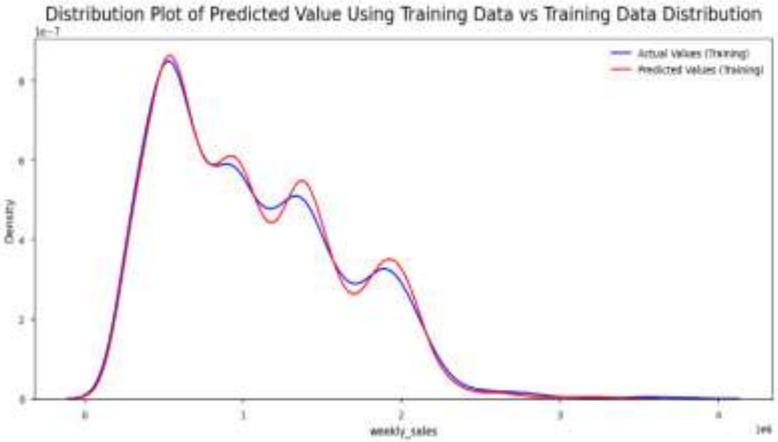
- **Normalized Root Mean Squared Error:** 0.014323497973037494
- **R-Square score Training:** 99.18%

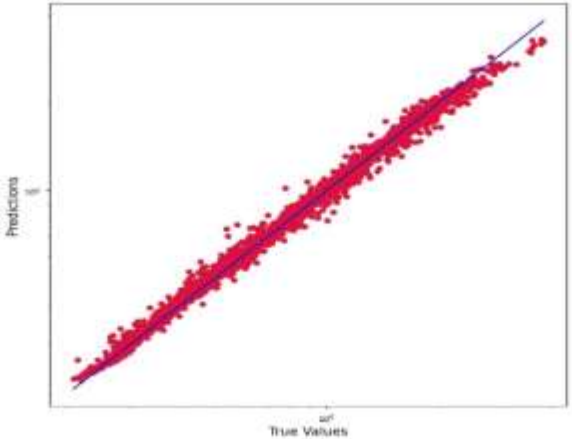
Random Forest



After hyperparameter Tuning:

- **Normalized Root Mean Squared Error:** 0.015296152891836235
- **R-Square score Training:** 99.06%



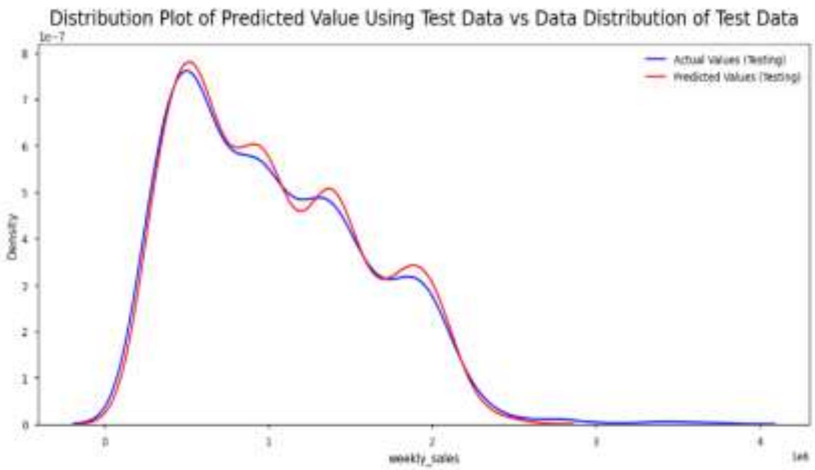


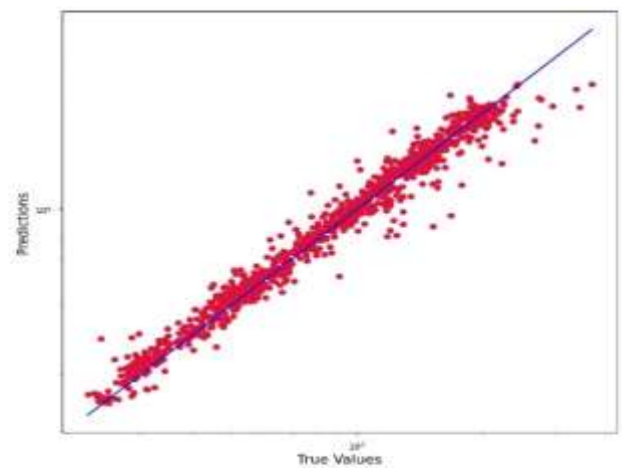
Random Forest

- **Cross Validation Scores**: [0.94372224 0.95224094 0.94774706 0.94307089 0.92997839 0.95047328 0.95215728 0.93792037 0.93657662 0.96320551]
- **Mean of Scores:** 94.57%
- **Standard Deviation of Scores:** 0.009053443430509272

Testing the model:

- **Normalized Root Mean Squared Error:** 0.03767470372904538
- **R-Square score Testing:** 94.8%





The achieved accuracy using the Random Forest model is 94.8%.

XGBoost

Introduction to XGBoost:

- XGBoost is an efficient and scalable machine learning algorithm designed for tree boosting.
- It stands for eXtreme Gradient Boosting and is widely used for regression and classification tasks.
- XGBoost is an implementation of Gradient Boosted decision trees.
- In this algorithm, decision trees are created in sequential form. Weights play an important role in XGBoost. Weights are assigned to all the independent variables which are then fed into the decision tree which predicts results.

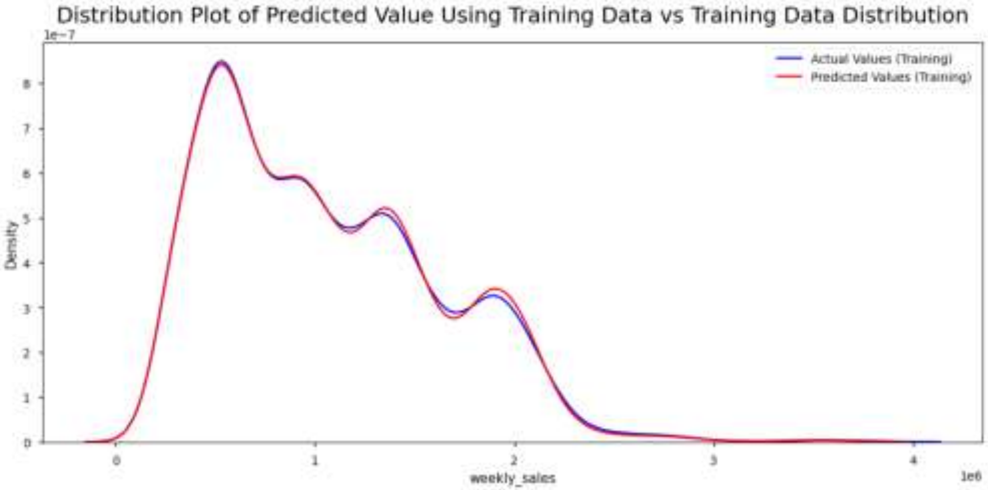
Initial parameters:

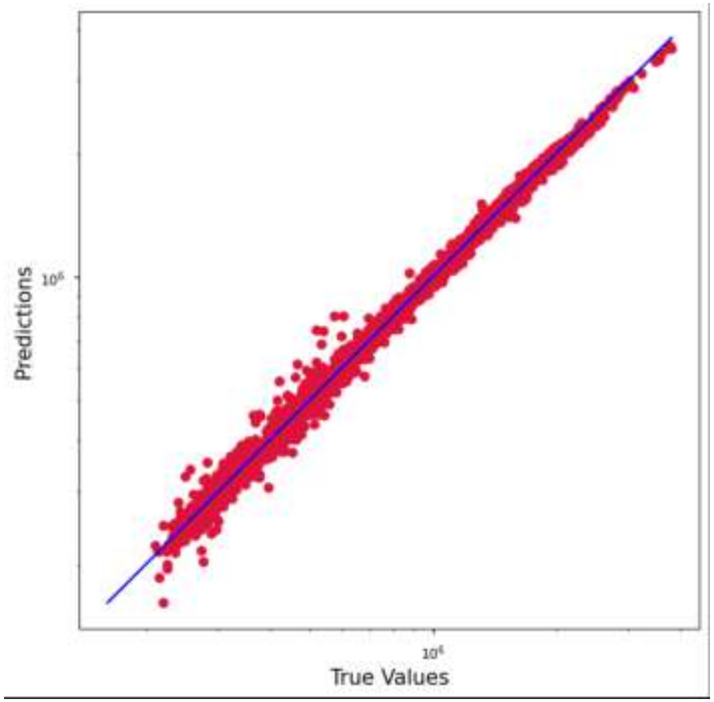
- ’n estimators’
- ‘’max depth ’

XGBoost

Before hyperparameter Tuning:

- Normalized Root Mean Squared Error: 0.008682702431196695
- R-Square score Training: {99.7} %



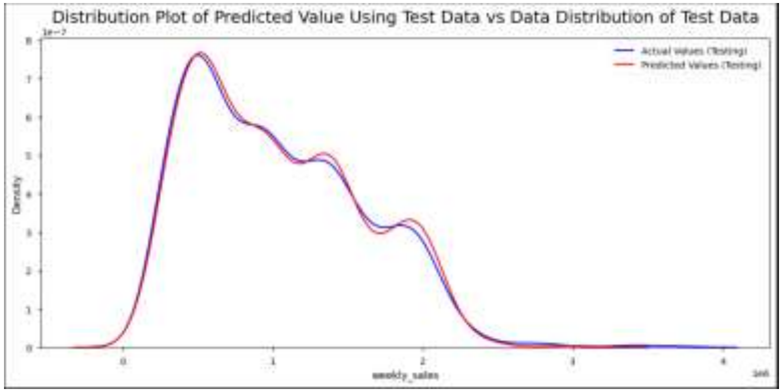


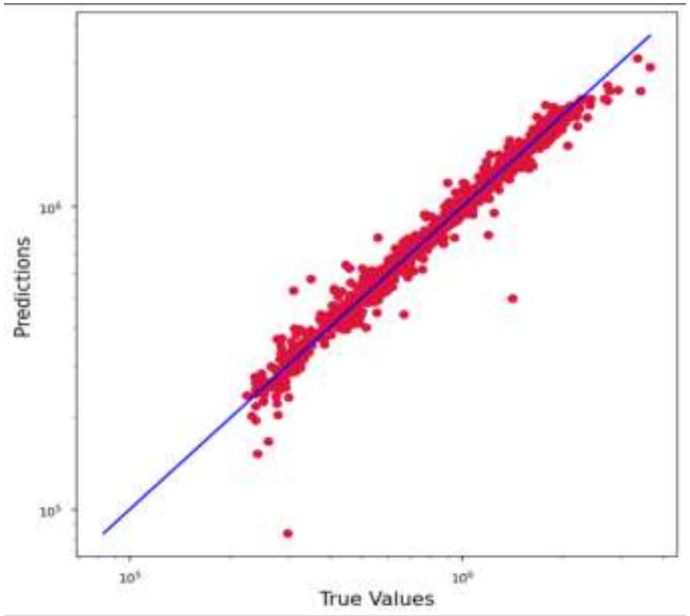
XGBoost

Outputs

Testing Accuracy

- Normalized Root Mean Squared Error: 0.025837143513686517
- R-Square score Testing: 0.975(97.55 %)





XGBoost

Performance Metrics of XGBOOST

Mean Squared Error (MSE)

- Mean Squared Error measures the average of the squared differences between predicted and actual values.

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

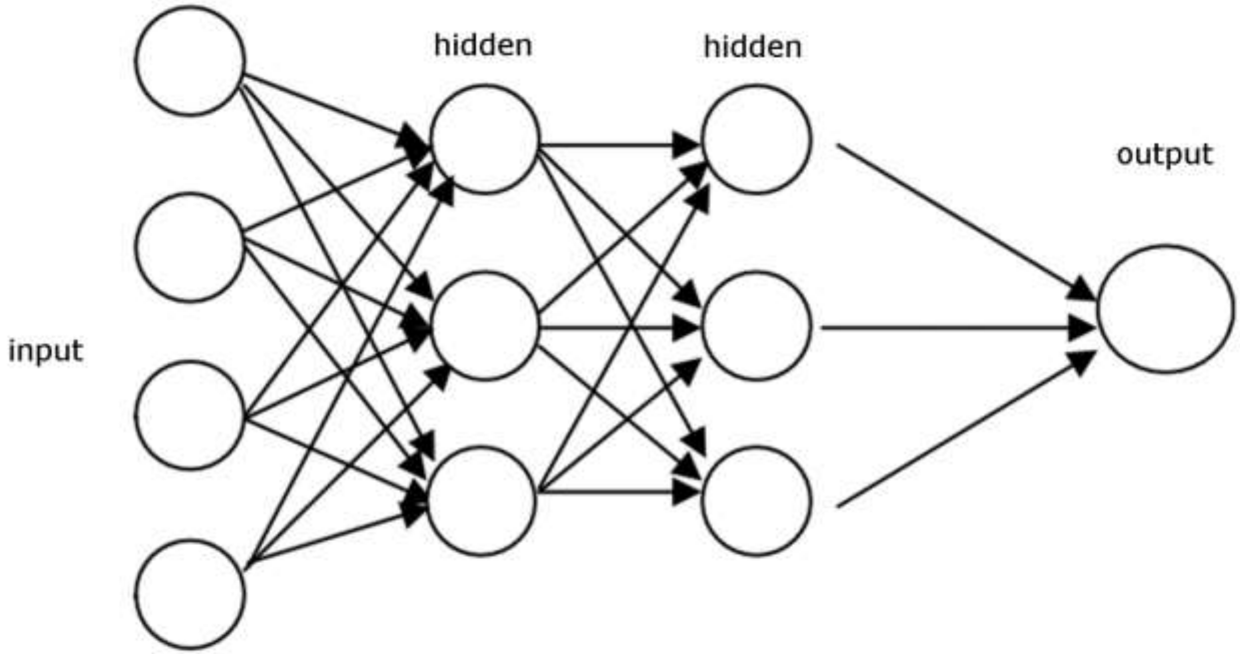
Normalized Root Mean Squared Error

- It's like understanding how well the model captures the patterns and trends in sales data

- R-squared measures the proportion of the variance in the dependent variable (sales) that is predictable from the independent variables (features)

Feed-Forward Neural Network:

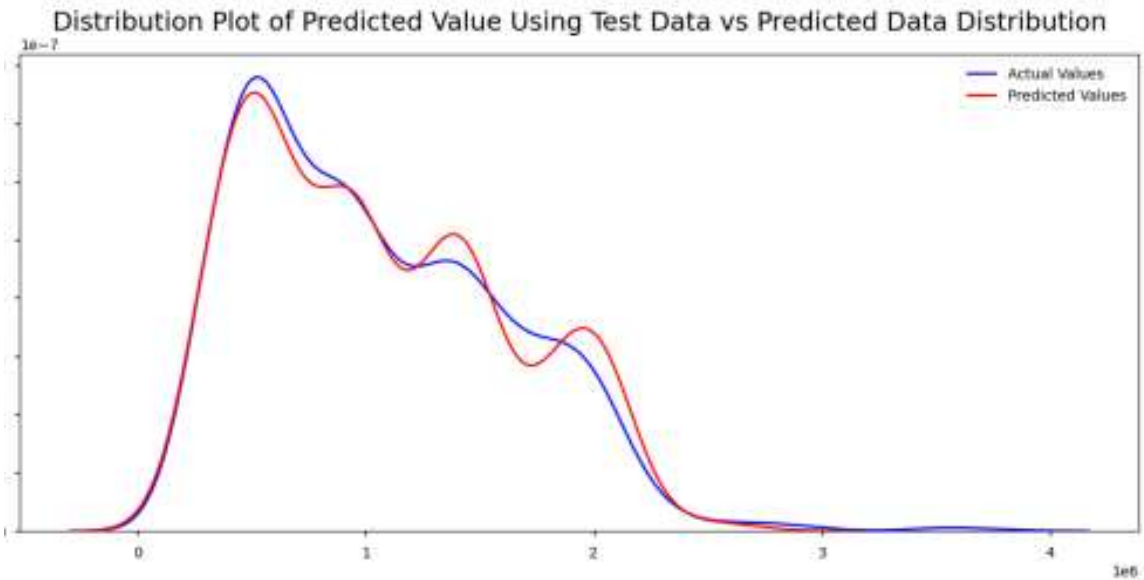
- A Feedforward Neural Network (FNN) is a fundamental type of artificial neural network where information moves in only one direction—forward—from the input layer through the hidden layers to the output layer.
- It’s called ”feedforward” because the information flows through the network in a forward direction without any feedback loops

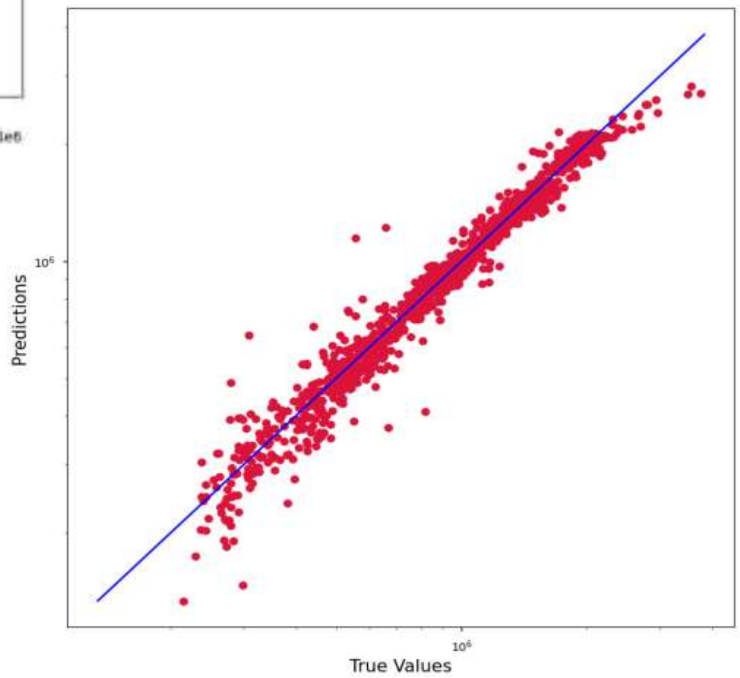


Feed-Forward Neural Network:

- For columns with numerical features, Standard Scalar for encoding
- For columns with categorical features, we use One Hot Encoder
- The neural network has one input layer for numerical features and multiple input layers for each categorical feature.
- For the given data set, we applied ReLU as activation function for the input and hidden layers and Linear activation function for the output layer.
- To compile the model, we used Stochastic Gradient Descent as an optimizer and mean square error as a metric to improve performance

Feed-Forward Neural Network:





- **Normalized Root Mean Squared Error:** 0.0187
- **R-Square score Testing:** 96.47%

Comparison

- The accuracy of the trained model on the linear regression test data was **97.13%**
- The accuracy of the trained model on the XGB test data was **97.55%**
- The accuracy of the trained model on the Random Forest test data was **94.8%**
- The accuracy of the trained model on the Decision tree test data was **91.22%**
- The accuracy of the trained model on the Feedforward Neural Network test data was **96.47%**
- We can have a comparative analysis of the results obtained by training and testing the different models across the sample dataset. This enables us to suggest the best and robust model that can be implemented for accurate demand forecasting, especially during demand surge experienced around the holiday seasons.

Metrics	Linear Regression	Decision Tree	Random Forest	XGBoost	Feedforward Neural Network
Normalized RMS	0.0279	0.0489	0.03767	0.0258	0.0187
R-square score	97.13%	91.22%	94.8%	97.55%	96.47%
Mean of Scores	96.45%	91.78%	94.57%	96.67%	-
Standard Deviation of Scores	0.00347	0.0176	0.00905	0.00419	-

Conclusion

- Sales forecasting is a crucial aspect of business planning.
- The researchers have concluded that an intelligent sales prediction system is required for business organizations to handle enormous volume of data.
- Business decisions are based on speed and accuracy of data processing techniques
- It became clear from exploring a variety of machine learning approaches from complex ensemble models that these techniques perform better than conventional ones.
- Our machine learning model of linear regression demonstrated commendable accuracy in predicting demand, as evidenced by its low mean absolute error and R-squared values.
- This indicates that our models are robust and reliable in capturing the underlying patterns in demand data.

Future Scope

In the realm of demand forecasting, there are several exciting future prospects and trends that can significantly impact the field

Artificial Intelligence (AI) and Machine Learning (ML):

- Advanced AI and ML algorithms will continue to play a pivotal role in enhancing the accuracy of demand forecasting models. Deep learning techniques and neural networks can capture complex patterns and dependencies in data

Real-time Forecasting:

- Exploring and implementing real-time demand forecasting capabilities, allowing organizations to make instantaneous adjustments based on the most recent data and market trends

Adaptive Models:

- Models that continuously learn and adapt to changing patterns in data. Continuous learning ensures that forecasting models remain relevant and effective in dynamic market environments.

Integration of Advanced Algorithms:

- Continued development and integration of advanced machine learning algorithms, beyond traditional methods, for more accurate and robust demand forecasting stevens.edu