

Sales Prediction with the Machine Learning

- Importing the necessary libraries and loading the dataset into a pandas DataFrame.
- Data wrangling and exploratory data analysis (EDA), which involves checking for missing values, duplicates, and outliers, visualizing the distribution of each feature, and identifying any correlation between the features.
- Splitting the dataset into training and testing sets.
- Building a regression model, which can be done using various machine learning algorithms such as Linear Regression and XG Boost.
- Evaluating the model's performance using various metrics such as MSE (Mean Squared Error), MAE (Mean Absolute Error), R-Square, R-Square Adjusted and Cross-Validation accuracy score.
- Using VIF technique to improve accuracy of the model.
- Improving the model's performance by tuning hyperparameters with the help of grid search cross-validation.

Conclusion

EDA Insights:

- Most of the expenses are on TV.
 - Radio has the least effect on Sales.
 - TV advertisements have maximum effect on Sales.
 - Understanding the distribution of the data: EDA can help to identify the distribution of the data, whether it is normal, skewed, or has outliers. This information can be used to select appropriate statistical methods for analysis and modeling.
 - Identifying relationships between variables: EDA can reveal the nature and strength of relationships between variables, including correlation, causation, and interactions. This information can be used to select relevant features for modeling and to gain insights into the underlying mechanisms that drive the data.
 - Detecting anomalies and outliers: During EDA analysis we find some outliers present in the dataset so we remove them with the help of the Z-score method where I took values less than 3.
-

ML Results:

- For training the dataset we use two techniques: Linear Regression and XG Boost technique.
- For Linear Regression, Newspaper is the most important feature.
- For XGBoost Regression, TV is the most important feature.
- In Linear Regression, I found 89% accuracy of the model and In XG Boost, I found 98% accuracy of the model.