

Capstone Project Submission

Team Member's Name, Email and Contribution:

Name – Baishnavee Mahato

Email. mbaishnavee@gmail.com

Data Understanding

- Data Analysis
- Data preprocessing and exploration

Data Visualization

- Bar plot
- Count plot
- Line plot
- Boxenpplot
- Boxplot
- Correlation Analysis

Outliers Detection and Removal

Feature Engineering

Algorithm Implementation

- Linear Regression
- Decision Tree
- Random Forest

Conclusion

Technical Documentation

GitHub Repo link.

GitHub Link: https://github.com/mbaishnav05/Retail_Sales_Prediction

Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)

Problem Statement

Rossmann operates over 3,000 drug stores in 7 European countries. Currently, Rossmann store managers are tasked with predicting their daily sales for up to six weeks in advance. Store sales are influenced by many factors, including promotions, competition, school and state holidays, seasonality, and locality. With thousands of individual managers predicting sales based on their unique circumstances, the accuracy of results can be quite varied.

I have been provided with historical sales data for 1,115 Rossmann stores. The task is to forecast the "Sales" column for the test set. Note that some stores in the dataset were temporarily closed for refurbishment.

Approach

In first step, imported the data set to carry out the descriptive analysis over the data set to understand the information of data available.

Checked for missing and repetition of values in the data set provided.

Outliers detection and removal.

Feature engineering is done by converting categorical variable to numeric and by looking at the scenario where the Stores are open and yet there is no sales on that day.

Explored all the variables of the data set

Used data visualization with different kinds of plots to explore the correlation with Sales and different variables.

Applied different ML Algorithm:
Linear Regression

Decision tree Regressor
Random Forest Regressor

Conclusion

1. Inferences from EDA

- All of the weekday has a promotion and the weekend has no promotion.
- The sales price has the highest when there is a promotion.
- All Type 'b' stores have comparatively higher sales and it mostly constant with peaks appears on weekends.
- Majority of Stores remains closed on state holidays.
- School holiday increases the sales price and customers proportionally.

2. Random Forest Regressor is the best for this Dataset Problem.