

BVRIT HYDERABAD College Of Engineering For Women

BOOK SALES PREDICTION

Team - 6

N. Srilekha 20WH1A6627

K. Spandana 20WH1A6628

B. Priyanvitha 20WH1A6629

V. Sai Sohitha 20WH1A6630

N. Sahithi Sri 20WH1A6631

Saturday 1st April, 2023

Problem Statement

- Predicting a full year worth of sales for 4 items from two competing stores located in six different countries during the year 2021.

Python Packages Used

- numpy
- pandas
- matplotlib
- seaborn
- sklearn

Algorithm

- Linear Regression
- Decision Tree Regressor
- Random Forest Regressor
- XGBoost Regressor

Linear Regression

- Linear regression analysis is used to predict the value of a variable based on the value of another variable.
- SMAPE Score: 32.24
- RMSE Score : 76.76921795839812
- Mean Absolute Percentage Error: 37.93543839647395
- Mean Absolute Error : 57.1075020403918
- r^2 - Score : 0.6264958346796725

Decision Tree Regressor

- Decision tree builds regression or classification models in the form of a tree structure.
- SMAPE Score: 8.91
- RMSE Score : 28.915451904340802
- Mean Absolute Percentage Error: 8.909509946772861
- Mean Absolute Error : 17.74743584685675
- r^2 - Score : 0.947011553824499

Random Forest Regressor

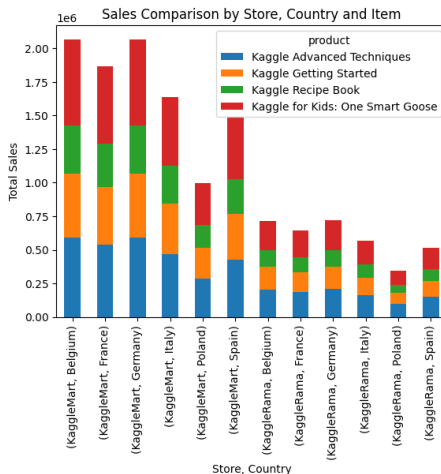
- A random forest is a meta estimator that fits a number of classifying decision trees on various sub-samples of the dataset and uses averaging to improve the predictive accuracy and control over-fitting.
- SMAPE Score: 9.83
- RMSE Score : 30.497562680880684
- Mean Absolute Percentage Error: 9.967484700618884
- Mean Absolute Error : 19.003732354279393
- r^2 - Score : 0.9410543879366279

XGBoost Regressor

- Image result for xgboost regressor XGBoost is an efficient implementation of gradient boosting that can be used for regression predictive modeling
- SMAPE Score:8.88
- RMSE Score : 28.83960898898401
- Mean Absolute Percentage Error:8.877592003143803
- Mean Absolute Error :17.68779175920322
- r2 - Score : 0.947289158183472

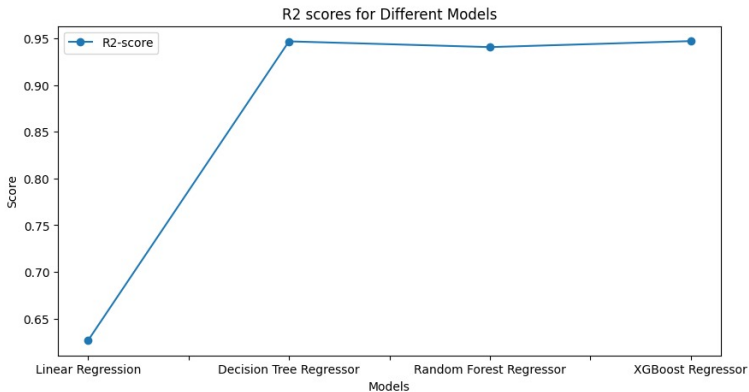
Sales graph

- Sales Comparison between Store, Country and Item



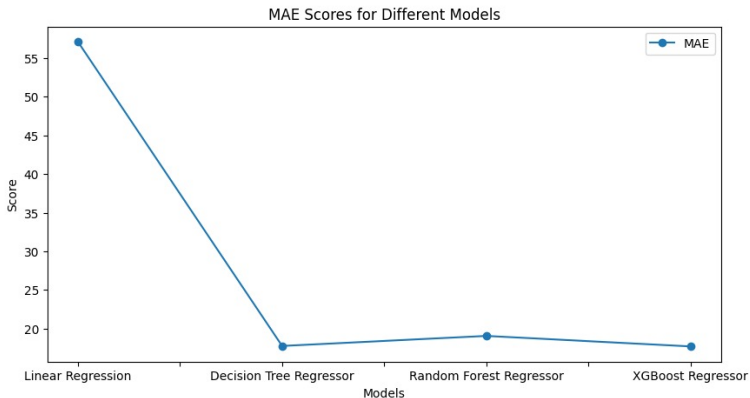
R2 Scores

- R2 scores for different models



MAE Scores

- MAE scores for different models



Comparison Table

- Comparison table for different metrics.

	SMAPE	MAPE	MAE	rmse	R2 - score
Linear Regression	32.241089	37.935438	57.107502	76.769218	0.626496
Decision Tree Regressor	8.913633	8.909510	17.747436	28.915452	0.947012
Random Forest Regressor	8.914060	8.910410	17.747021	28.916296	0.947008
XGBoost Regressor	8.865123	8.861414	17.657916	28.804622	0.947417

- Book sales prediction for the year 2021.

	row_id	num_sold
0	70128	431.901031
1	70129	311.026489
2	70130	277.744507
3	70131	437.036865
4	70132	146.048309
5	70133	108.909271
6	70134	93.818504
7	70135	156.334167
8	70136	434.636353
9	70137	313.101135

[https://colab.research.google.com/notebooks/ML*project.ipynb*](https://colab.research.google.com/notebooks/MLproject.ipynb)

THANK YOU!