Unschool

Overview: Store-Sales-Prediction

Goals:

1. In this project Store Sales Prediction using machine learning and python, the task is to predict the sales of different stores based on the attributes available in the dataset. Predictive analytics can help us to study and discover the factors that determine the number of sales that a retail store will have in the future by using different machine learning techniques and trying to determine the best algorithm suited to our particular problem statement. You have to implement normal regression techniques as well as boosting techniques and relate the boosting algorithms to have better results than the regular regression algorithms.

Methodology: During data exploration for insights, while opening the data, it is found that there were many stores that were not open. Thus it is confirmed that there were no closed stores with sales, and that there were no open stores with no sales and could safely remove them from the data set. Then that sales would scale with the number of customers on that day. Thus by plotting the trend, it is able to confirm that there was a correlation between customers and sales. Continuing with this insight, it is able to separate them by the day of week, confirming that some days were correlating more strongly than others. It has been noted that there were no easter holiday data set in state holidays, thus making any test data containing Easter holiday harder to predict accurately. Additionally, having promotions had an impact on sales, but school holidays did not impact sales much.

Reference Link:

1. Pandas: https://pandas.pydata.org/docs/
2. Numpy: https://numpy.org/doc/stable/user/absolute_beginners.html
3. Matplotlib: https://matplotlib.org/stable/users/getting_started/
4. Seaborn: https://seaborn.pydata.org/tutorial/introduction.html
5. SciKit-Learn: https://scikit-learn.org/stable/getting_started.html

Popular Machine Learning Algorithms:-

https://www.kaggle.com/code/marcovasquez/top-machine-learning-algorithms-beginner

- Linear regression.
- Logistic regression.
- Decision tree.
- SVM algorithm.
- Naive Bayes algorithm.
- KNN algorithm.
- K-means.
- Random forest algorithm.

Python Docs for your reference: https://docs.python.org/3/tutorial/