**INFO 7390 FINAL PROJECT REPORT**

**Project Topic:**

Forecasting sales prediction and fraud detection using machine learning and neural networks.

**Group Number – 31**

**Team Members:**

1. Rakesh Lokam

2. Mrunal Bhosale

3. Uday Kumar Reddy Somireddy

4. Krupali Patel

**Introduction:**

Machine learning is a subset of Artificial Intelligence and can be classified as algorithms and models that have an ability to learn and adapt without explicit instructions from humans to analyze and draw inferences from patterns in data. A machine learning algorithm builds training data using the information provided, which helps with predictions and decisions. As information is loaded to the machine, the data set improves, and the algorithm’s capability enhances.

In our project we explore this ability to do Sales Forecasting for supply chain management and fraud detection for a retail business. The Machine Learning model learns based on the past sales data and the current sales transactions and forecasts the sales for the next year. These predictions help retail businesses understand how to stock up inventory based on products that sell well and take measures for fraudulent activities ensuring quick and effective transactions.

**Benefits of Fraud Detection via Machine Learning**

Machines are much better than humans at processing large datasets. They can detect and recognize thousands of patterns on a user’s purchasing journey instead of the few captured by creating rules. This makes it possible to predict fraud in a large volume of transactions with increased accuracy, faster and efficient Detection.

**Goals:**

1. This project aims to compare popular machine learning classifiers and regressors type machine learning models and measure their performance against neural network models to find out which machine learning model performs better.
2. The machine learning classifiers used in this project are Logistic Regression,Linear Discriminant Analysis, Gaussian Naive Bayes, Support Vector Machines, k - Nearest Neighbors, Random Forest classification, Extreme Gradient Boosting, Decision Tree classification for fraud detection and to predict late delivery on the basis accuracy, recall score and F1 score.
3. The regression models used are Lasso, Ridge, Light Gradient boosting, Random Forest regression, Extreme Gradient Boosting regression, Decision Tree Regression, and Linear Regression to predict sales and quantity of the products required which are compared with mean absolute error (MAE) and root mean square error (RMSE).
4. Below are the steps followed in our project:

· Preparing the dataset

· Installing dependencies

· Training the model

· Hyperparameter tuning

· Visualizing data

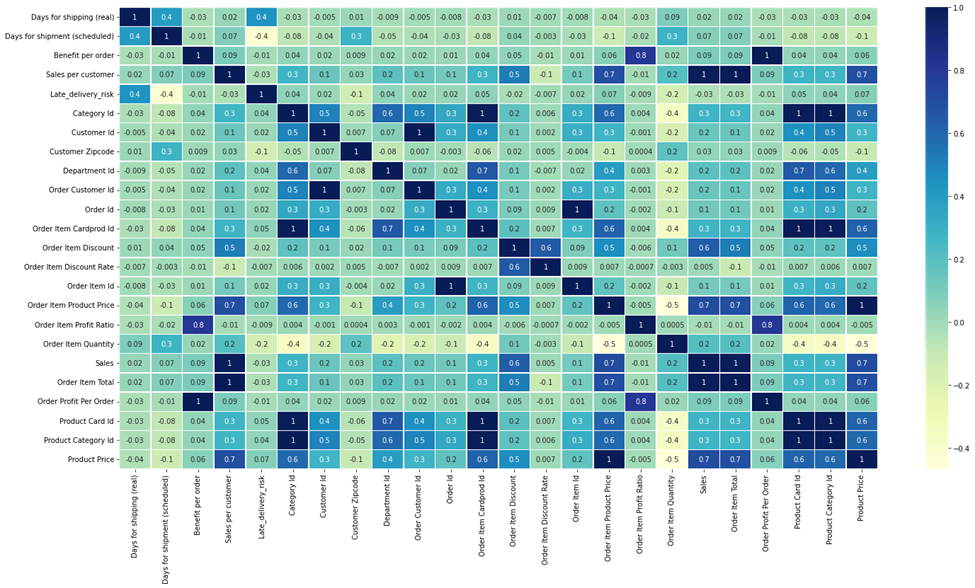
· Testing models

**Algorithms Implemented:**

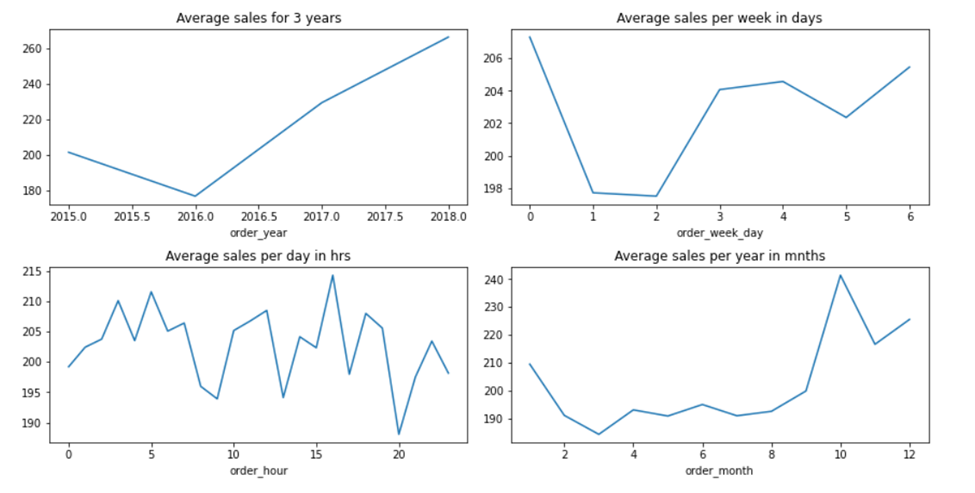
* Logistic Regression
* Gaussian naïve bayes model
* Support Vector machines
* Random Forest Classifier
* KNeighbours Classifier
* XG Boost Classifier
* Decision Tree Classifier
* Neural Network model classification

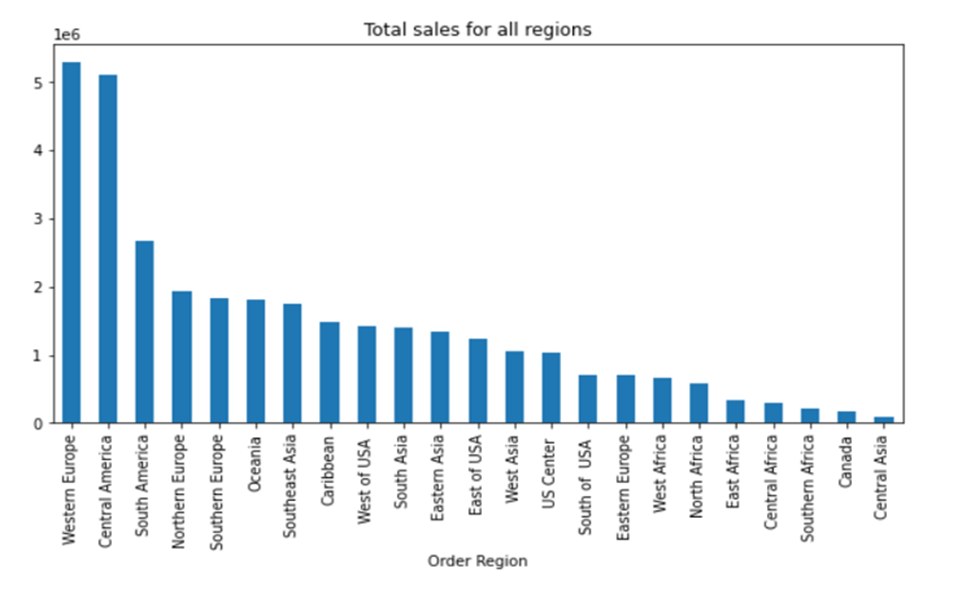
**Analysis**

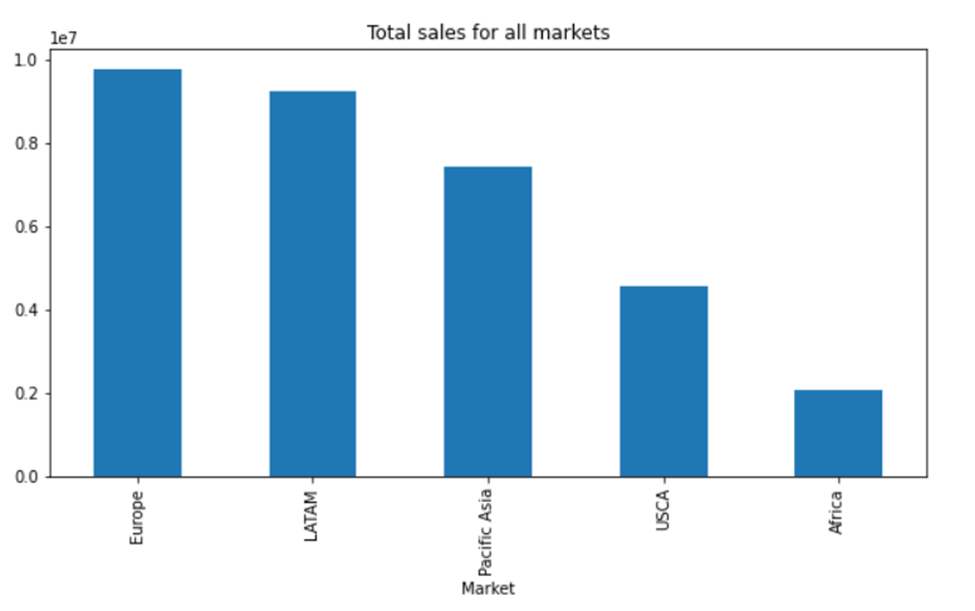
**Data correlation heat map**

****

**Average sales**

****

**Total sales across regions**

****

**Conclusion:**

Our Neural Network classifier model trained for fraud detection outperformed all machine learning classifier models with an f1 score of 0.96. ​ Decision Tree model did a good job compared with other classification models in identifying orders with later delivery and detecting fraudulent transactions with an f1 score of 0.80. ​Linear Regression model did better for predicting sales revenue, while both Random Forest and Extreme Gradient Boosting Regression outperformed and predicted the demand more accurately with MAE and RMSE scores lower than the Neural Network model.