Springboard

Capstone Project 2

Initial Project Ideas

- ## Project 1

Analyzing Supply Chain data to identify potential late deliveries.

Problem Statement:
To identify shipments that have a high probability of late delivery based on supervised learning from available late delivery data for a supply chain firm.

Who would care and why:
Ecommerce businesses: Customer satisfaction is a huge part of the success of the ever-competitve ecommerce business.

Collection of dataset:
Multiple real-world datasets are available for supply chain businesses online. Kaggle also hosts a number of supply chain datasets.

Initial thoughts on Problem Solving:
This problem is essentially that of binary classification. The goal would be to perform machine learning using a variety of industry standard classification algorithms such as logistic regression, LDA, SVM, Decision Trees etc.
All algorithms would require hyper parameter tuning and eventually the best performing model will be selected. Feature importance scores would reveal the critical features in predicting the

- ## Project 2

Flix recommendation system

Problem Statement: To create a recommendation system that would suggest the next movie to watch for users based on a feature vector consisting of movie ratings and user tags applied to these movies

Who would care:
This project will be targeted towards India movie hosting websites like Zee5, Voot, AltBalaji etc.

Collection of dataset:
The data set to build a movie recommendation system is easily available on the internet. Kaggle itself hosts multiple movie datasets.

Initial thoughts on Problem Solving:
This problem will be that of multi classification. We can train the dataset to either have each movie considered as a separate class or we can group movies by genre or other similarities and have our classification algorithm choose based on these classes.
Industry standard classfiers such as SVM, decision trees and ensemble methods like Random Forests and extreme Gradient Boosting will be compared for building the model and the most efficient method will be selected.

- ## Project 3

Anomaly and Fault detection in the Manufacturing Industry for Predictive maintenance

Problem Statement: To detect faults/ anomalies in manufacturing machinery to predict machinery failures before they happen. Part of Predictive Maintenance

Who would care: Any manufacturing firm that relies on expensive machinery and processes to produce goods. Where down time of a single machine could mean losses more that 100x the cost of the machine.

Collection of dataset: The UC berkeley milling dataset is one of the manufacturing datasets that can be used. Other such datasets from manufacturing firms are available.

Initial thoughts of Problem Solving:
Based on the research conducted thus far, looks like the solution to this problem lies in Neural Networks. Using a technique called Variational Autoencoders, we can approach the problem of anomaly detection.