**MACHINE LEARNING LAB PROJECTS**

**Guidelines**

* Read through the problem and try to apply the best ML models on the dataset after proper pre-processing the data.
* Design and develop the complete ML pipeline for the above. Develop a flask application to act as Front-end and deploy it using Docker.
* You may work with simple image datasets
* You now know many ML & AML techniques. Try to incorporate the most required ones after careful study.

**Project Titles**

1. **Customer Churn Prediction**

* **Problem**: Telecom companies need to predict whether a customer is likely to churn to optimize retention strategies.
* **Dataset**: Telco Customer Churn

1. **House Price Prediction**

* **Problem**: Predicting real estate prices helps both buyers and sellers make informed decisions based on housing features.
* **Dataset**: House Prices: Advanced Regression Techniques

1. **Credit Card Fraud Detection**

* **Problem**: Detecting fraudulent credit card transactions in real-time to prevent financial losses.
* **Dataset**: Credit Card Fraud Detection

1. **Sentiment Analysis of Movie Reviews**

* **Problem**: Sentiment analysis can help movie production companies understand audience feedback from reviews.
* **Dataset**: IMDb Movie Reviews

1. **Predicting Diabetes**

* **Problem**: Early detection of diabetes in patients using health records to prevent serious complications.
* **Dataset**: Diabetes Dataset

1. **Spam Email Detection**

* **Problem**: Classifying emails as spam or not helps in filtering out irrelevant content for users.
* **Dataset**: Spam Email Dataset

1. **Stock Price Prediction**

* **Problem**: Financial institutions and individual investors seek to predict stock price movements for better investment decisions.
* **Dataset**: Stock Market Dataset

1. **Loan Approval Prediction**

* **Problem**: Banks want to predict the likelihood of a loan approval based on customer profiles.
* **Dataset**: Loan Prediction Dataset

1. **Heart Disease Prediction**

* **Problem**: Predicting the occurrence of heart disease based on patient data can aid in early diagnosis and prevention.
* **Dataset**: Heart Disease UCI

1. **Forest Fire Prediction**

* **Problem**: Forecasting forest fires based on environmental data can help in early warning and preparedness.
* **Dataset**: Forest Fires Dataset

1. **Human Activity Recognition**

* **Problem**: Wearable technology companies use this data to classify activities for health monitoring and fitness applications.
* **Dataset**: Human Activity Recognition with Smartphones

1. **Energy Consumption Forecasting**

* **Problem**: Accurate predictions of household energy consumption can assist in managing power grids and energy resources.
* **Dataset**: Household Electric Power Consumption

1. **Fake News Detection**

* **Problem**: News outlets and social media platforms need to filter out fake news articles to maintain information integrity.
* **Dataset**: Fake News Detection

1. **Air Quality Prediction**

* **Problem**: Environmental agencies can predict air quality levels to issue health warnings and control pollution.
* **Dataset**: Air Quality Dataset

1. **Plant Disease Detection**

* **Problem**: Farmers need to detect diseases in plants early to prevent large-scale agricultural losses.
* **Dataset**: Plant Village Dataset

1. **Bike Sharing Demand Prediction**

* **Problem**: Bike-sharing companies need to predict demand at different times to optimize bike distribution.
* **Dataset**: Bike Sharing Demand Dataset

1. **Supermarket Sales Forecasting**

* **Problem**: Supermarkets want to predict future sales for different product categories to manage inventory efficiently.
* **Dataset**: Supermarket Sales Dataset

18.**Predicting Solar Power Generation**

* **Problem**: Predicting solar power generation helps in balancing energy grids and managing renewable energy sources.
* **Dataset**: Solar Power Generation Dataset

19. **Flight Price Prediction**

* **Problem**: Predicting flight prices helps travelers plan their trips, and airlines optimize pricing strategies.
* **Dataset**: Flight Price Prediction Dataset

**20. Building Energy Efficiency Prediction**

* **Problem**: Property managers and governments aim to improve energy efficiency in buildings by predicting energy consumption and optimizing design.
* **Dataset**: Energy Efficiency Dataset

21. **Human Development Index Prediction**

* **Problem**: Predicting HDI helps governments and organizations monitor socio-economic progress and target areas needing improvement.
* **Dataset**: World Development Indicators Dataset

22. **Movie Recommendation System**

* **Problem**: Streaming services like Netflix use recommendation systems to suggest relevant content to users.
* **Dataset**: MovieLens Dataset

23. **Predicting Traffic Accidents**

* **Problem**: Predicting the likelihood of traffic accidents can help city planners and law enforcement improve road safety.
* **Dataset**: Traffic Accidents Dataset

24. **Taxi Fare Prediction**

* **Problem**: Predicting taxi fares in advance helps customers estimate trip costs, and taxi services optimize pricing models.
* **Dataset**: New York City Taxi Fare Prediction

25. I**nsurance Claims Prediction**

* **Problem**: Predicting the likelihood of an insurance claim can help insurance companies set premiums and manage risk.
* **Dataset**: Insurance Claims Dataset

26. **Predicting Product Demand in Retail**

* **Problem**: Retailers need to predict future product demand to manage their supply chains effectively.
* **Dataset**: Walmart Sales Forecasting

27. **Wildlife Species Classification**

* **Problem**: Wildlife researchers use image classification models to identify and classify different species from camera trap data.
* **Dataset**: Wildlife Camera Trap Dataset

28. **Sentiment Analysis on any News**

* **Problem**: Predicting market sentiment from financial news articles can help traders make informed decisions.
* **Dataset**: Financial Sentiment Dataset

29. **Early Diagnosis of Alzheimer’s Disease**

* **Problem**: Predicting Alzheimer’s disease in its early stages
* **Dataset**: Alzheimer’s Disease Dataset

**30**. Product defect identification

* **Problem**: Predict defects in any manufactured product.
* **Dataset**: Alzheimer’s Disease Dataset