CSE445 Project

You must form a group of three to four people and choose to work on any one of the following projects. You will have to show your work during TA office hours. The datasets provided below are intended for practice purposes only and are not the final versions. The final datasets will be provided after you've formed your group and chosen your project.

Project Submission Dates: 19/05/2023

Project showcase dates: 20/05/23, 27/05/23, 03/05/23

Project Ideas:

1. **Credit Card Fraud Detection**
   Task: build a classifier that can detect fraudulent transactions
   Website: https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud

2. **Bankruptcy Prediction**
   Task: Predict the bankruptcy of a company.
   Website: https://www.kaggle.com/datasets/fedesoriano/company-bankruptcy-prediction

3. **Risk Factor prediction of Chronic Kidney Disease Data Set**
   Task: Build a model that can be used to predict chronic kidney disease
   Website:
   https://archive.ics.uci.edu/ml/datasets/Risk+Factor+prediction+of+Chronic+Kidney+Disease

4. **Housing dataset**
   Task: Predict the housing prices of a new house and analyze the most important features that determines the price of a house.
   Website: https://www.kaggle.com/datasets/deepcontractor/miami-housing-dataset/code

5. **Productivity Prediction of Garment Employees**
   Task: Predict the productivity range of garment employees.
   Website: https://archive-beta.ics.uci.edu/dataset/597/productivity+prediction+of+garment+employees

6. **E-commerce Dataset**
   Task: Analyze the data of the customer rides and find insights that can help improve the business.
   Website: https://www.kaggle.com/datasets/fivethirtyeight/uber-pickups-in-new-york-city

7. **Absenteeism at work Dataset**
   Task: Analyze the data and find insights that can help improve the business.
   Website: https://archive-beta.ics.uci.edu/dataset/445/absenteeism+at+work

8. **Online Shoppers Purchasing Intention Dataset**
   Task: Analyze the data and find insights that can help improve the business.
   Website: https://archive-beta.ics.uci.edu/dataset/468/online+shoppers+purchasing+intention+dataset

For this project you must:

- Split dataset into train, validation, and test set (Example ratio: 70: 15: 15) (if required)
- Describe the data with appropriate Exploratory Data Analysis (EDA) and Visualization.
- Deal with Missing data (Null Value) and Categorical data (if required)
- Find correlation matrix for target variable/ground truth (if required)
- Perform proper feature scaling (if required)
- Use at least 2 different algorithms
- Use relevant performance metrics.

You must also explain why you chose specific options like model parameters, data preprocessing, performance metrics, and model selection.