# **CSE445 Project**

**Classification:**

1. **Credit Card Fraud Detection**Task: build a classifier that can detect fraudulent transactions  
   Website: <https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud>
2. **Parkinson’s Data Set**  
   Task: Build a model that can be used to differentiate healthy people from people having Parkinson’s disease.  
   Website: <https://archive.ics.uci.edu/ml/datasets/parkinsons>

**Regression:**

1. **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://archive.ics.uci.edu/ml/datasets/parkinsons>
2. **Risk Factor prediction of Chronic Kidney Disease Data Set**  
   Task: Build a model that can be used to differentiate healthy people from people having Parkinson’s disease.  
   Website: <https://archive.ics.uci.edu/ml/datasets/Risk+Factor+prediction+of+Chronic+Kidney+Disease>
3. **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>

**Unsupervised:**

1. **Uber Pickups in New York City**  
   Task: Analyze the data of the customer rides and find insights that can help improve business.   
   Website: <https://www.kaggle.com/datasets/fivethirtyeight/uber-pickups-in-new-york-city>

For this assignment you must:

* Split dataset into train, validation, and test set (Example ratio: 70: 15: 15)
* 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.