

Classification

Credit Risk Scoring

- Dataset: [German Credit Dataset \(UCI\)](#)
- Task: Classify loan applicants as good or bad credit risks

Heart Failure Prediction

- Dataset: [Heart Failure Prediction Dataset \(Kaggle\)](#)
- Task: Predict the likelihood of heart failure based on clinical features.

Regression

Student Performance Prediction

- Dataset: [Student Performance Dataset \(UCI\)](#)
- Task: Predict final exam grades based on demographic, social, and school-related features.

Bike Sharing Demand

- Dataset: [Bike Sharing Dataset \(Kaggle\)](#)
- Task: Predict daily rental counts.

Clustering

Country data

- Dataset: [Country Data \(Kaggle\)](#)
- Task: Cluster countries based on various socio-economic indicators.

Notes

1. You can put filters on Kaggle search and sort by “Most Votes” <https://www.kaggle.com/datasets?sort=votes&tags=14101-Tabular%2C13304-Clustering>
2. For each task figure out what’s the best way to evaluate the model, and present the results.
3. Always do EDA first.
4. Use Git / Github

ToDo

1. Reply with + in Slack.
2. EOD - send team (1-3), choose project (not necessarily from above)
3. Random shuffle
4. GitHub repository