

KD Lab Final project 2024

Part 1 Decision tree regressor

- Find suitable dataset for regression models
- Perform data preprocessing including handling missing values in a suitable way, removing duplicates, removing outliers, encoding data, scaling.
- Perform data visualization and find correlation between features
- Train decision tree regressor model on the prepared dataset
- Apply suitable evaluation metrics on the results
- Predict the output for a new data point

Part 2 K means clustering algorithm

- Find suitable dataset for clustering models
- Perform data preprocessing including handling missing values in a suitable way, removing duplicates, removing outliers, encoding data, scaling.
- Find the best value for K using the elbow method
- Train k means model on the prepared dataset
- Apply suitable evaluation metrics on the results
- Predict the output for a new data point

Bonus

- Learn an unsupervised algorithm called Apriori algorithm and do a summary on the way it works and when is it used.
- Apply k means clustering algorithm on both labeled and unlabeled data and use the different evaluation methods for both.

Due date: 3/9/2024

The projects are to be completed individually.