

# Assignment

## Part 1: Mutual Information Classification

### Model-based Feature Selection Project

For the model-based feature selection project, you will use the "loan" dataset

Write a Python script that performs model-based feature selection using Mutual Information Classification on the "loan" dataset.

- Load the dataset using the pandas library or any other suitable method.
- Apply necessary processing on it (Hint : label\_encoding , null value handling)
- Separate the features (X) and the target variable (y : Loan\_Status) from the dataset.
- Use the SelectKBest class from the sklearn.feature\_selection module to select the top K features based on mutual information scores.
- Choose an appropriate value of K (number of features to select) for the analysis.
- Fit the feature selection model on the dataset and transform the feature matrix accordingly.
- Print the names or indices of the selected features.
- Note: Provide necessary code comments and explanations throughout your code.

Bonus Task (optional):

Implement a visualization technique to display the mutual information scores for each feature in the "loan" dataset. You can use libraries like matplotlib or seaborn to create the visualization.

## Part 2: Mutual Information Regression

### Model-based Feature Selection Project

For the model-based feature selection project, you will use the "Housing.csv" dataset

Write a Python script that performs model-based feature selection using Mutual Information Regression on the "Housing" dataset.

- Load the dataset using the pandas library or any other suitable method.
- Apply necessary processing on it (Hint : label\_encoding , null value handling)
- Separate the features (X) and the target variable (y : Price of the house) from the dataset.
- Use the SelectKBest class from the sklearn.feature\_selection module to select the top K features based on mutual information scores.
- Choose an appropriate value of K (number of features to select) for the analysis.
- Fit the feature selection model on the dataset and transform the feature matrix accordingly.
- Print the names or indices of the selected features.
- Note: Provide necessary code comments and explanations throughout your code.

Bonus Task (optional):

Implement a visualization technique to display the mutual information scores for each feature in the "Housing" dataset. You can use libraries like matplotlib or seaborn to create the visualization.

### Part 3 : Linear Regression on the Housing Dataset

In this assignment, you will perform linear regression on the Housing dataset to predict house prices. Follow the instructions below:

- Load the Housing dataset using the pandas library.
- Apply necessary preprocessing steps on the dataset, such as handling missing values and feature scaling if required.
- Separate the features (X) and the target variable (y: house prices) from the dataset.
- Split the dataset into training and testing sets using an 80:20 ratio.
- Fit a linear regression model to the training data using the LinearRegression class from the sklearn.linear\_model module.
- Predict house prices for the testing data using the trained model.
- Evaluate the performance of the model by calculating the mean squared error (MSE) and the coefficient of determination ( $R^2$ ) using appropriate functions from the sklearn.metrics module.
- Print the MSE and  $R^2$  values to assess the model's accuracy.
- Plot a scatter plot between the predicted house prices and the actual house prices for the testing data using the matplotlib library.
- Provide necessary code comments and explanations throughout your code to enhance readability.

Bonus Task (optional): Implement feature selection techniques, such as Mutual information regression, correlation etc to select the most relevant features for the linear regression model and evaluate its impact on the model's performance.

