

Task Title: *Exploring Dataset Sources and Feature Selection for Predictive Modeling*

Objective:

Explore reliable data sources, download a relevant dataset, and apply feature selection techniques to identify key variables for a predictive modeling task.

Steps:

1. Dataset Search and Selection:

- Explore platforms like [Kaggle](#), [UCI Machine Learning Repository](#), [Data.gov](#), or Google Dataset Search.
- Select a dataset suitable for predictive modeling (e.g., predicting house prices, customer churn, disease diagnosis). [Choose as per your interest – Evaluation will be on this]

2. Define an Objective:

- Formulate a clear objective for analysis (e.g., predicting customer churn, forecasting sales, or detecting fraudulent transactions).

3. Data Understanding and Preprocessing:

- Perform Exploratory Data Analysis (EDA) to understand the data.
- Handle missing values, detect outliers, and encode categorical variables if necessary.

4. Feature Selection:

- Apply at least three different feature selection methods:
 - **Filter Methods:** Correlation matrix, Chi-Square test, or Mutual Information.

- **Wrapper Methods:** Recursive Feature Elimination (RFE) or Forward/Backward Selection.
- **Embedded Methods:** Lasso, Ridge Regression, or Decision Tree Feature Importance.

5. **Evaluation:**

- Justify why the selected features are relevant to the objective.
- Visualize the selected features using plots like bar charts or heatmaps. [R/ Python/ Excel/or any Data Analytics tool]

6. **Documentation and Submission:**

- Provide a report explaining the dataset selection process, the chosen objective, preprocessing steps, feature selection methods, and findings.
- Submit code in a Jupyter notebook or Google Colab or dashboard of tool.