

Lab Assignment 2: Implementation of Data Preprocessing Techniques

Date: 19-02-2025

Duration: 2 hours

Objective: Implement the necessary preprocessing steps and prepare the dataset for further analysis.

Dataset: [UCI Credit Card](#)

Task Overview:

1. Task 1: Data Understanding and Importing

- Load the dataset into your preferred data analytics tool (e.g., Python, R).
- Print the first 10 rows of the dataset.
- Describe the structure of the dataset (number of rows, columns, types of variables).

2. Task 2: Handling Missing Data

- Identify any missing values in the dataset.
- Explain the method you will use to handle missing data and justify your choice.
- Implement the method and show the result of the dataset after handling missing values.

3. Task 3: Data Transformation

- Identify and explain any variables that need transformation or scaling.
- Apply appropriate transformation techniques (e.g., normalization, standardization).
- Visualize the effect of the transformation on one continuous variable.

4. Task 4: Encoding categorical variables

- Identify categorical variables in the dataset.
- Apply encoding techniques (e.g., one-hot encoding or label encoding) to the categorical variables.
- Explain why you chose this encoding method and how it affects the model's performance.

5. Task 5: Feature Selection and Engineering

- Remove any redundant or irrelevant features that will not contribute to model building.
- b. Justify the removal of these features.
- Apply any feature engineering techniques (e.g., creating interaction terms, new variables).

6. Task 6: Data Splitting

- Split the preprocessed dataset into training and testing sets with an 80:20 ratio.

- Explain why splitting the dataset is essential in the context of model training and evaluation.

7. Task 7: Summary of Data Preprocessing

- Write a brief summary of the data preprocessing steps you implemented.
- Discuss how these steps improve the quality of data for further analysis and modeling.

Note: Comment your code to explain the logic behind each step.