Assignment # 1

Course Name: Artificial Intelligence lab

Instructure: Muhammad Amjad Raza

Assignment Name: Preprocessing

Due Date: 14-03-2023

Total Marks: 10

Pre-processing Techniques in Python Assignment

In this assignment, you will be working with a dataset of your choice and using various pre-processing techniques in Python to prepare the data for further analysis or modelling. You can choose any dataset that interests you, such as a dataset from Kaggle or a dataset from your own domain.

Task 1: Data Exploration

1. Import the necessary libraries (e.g., pandas, NumPy, matplotlib, seaborn).

2. Load the dataset into a panda Data Frame.

3. Display the first 5 rows of the dataset to get a general idea of its structure.

4. Check for missing values and handle them appropriately (e.g., impute missing values or drop rows/columns with missing values).

5. Use descriptive statistics and visualization techniques to explore the dataset and identify any issues that need to be addressed (e.g., outliers, skewed distributions).

Task 2: Data Cleaning

1. Handle any inconsistencies or errors in the dataset (e.g., typos, incorrect data types).

2. Remove any irrelevant or redundant features (e.g., columns with constant values or high correlation with other columns).

3. Transform any categorical variables into numerical variables (e.g., one-hot encoding, label encoding).

4. Scale or normalize any numerical variables as needed (e.g., standardization, min-max scaling).

Task 3: Feature Engineering

1. Create new features from existing features (e.g., combining features, extracting information from text data).

2. Use dimensionality reduction techniques (e.g., PCA, t-SNE) to reduce the number of features and visualize the dataset in a lower-dimensional space.

Deliverables

1. A Jupyter notebook containing your code and analysis.

2. A brief report explaining your approach and any insights you gained from pre-processing the dataset.

3. The pre-processed dataset saved as a new file.

Good Luck!