**Data Analytics Laboratory**

**Task 1**

**Data Pre-Processing Techniques**

**Introduction**

* Data Pre-processing refers to the steps applied to make data more suitable for data mining. The steps used for Data Pre-processing usually fall into two categories:
  + Selecting data objects and attributes for the analysis.
  + creating/changing the attributes values.
* The data can have many irrelevant and missing values. This situation arises when some data is missing in the data. It can be handled in various ways.
  + Ignore the tuples
  + Fill the Missing values
* Normalization is done in order to scale the data values in a specified range (-1.0 to 1.0 or 0.0 to 1.0)
* Attribute Selection strategy is used when new attributes are constructed from the given set of attributes to help the mining process.
* Discretization is used to replace the raw values of numeric attribute by interval levels or conceptual levels.

**Prerequisites**

Why do we need data pre-processing ?

List the any five python library & classes used for data pre-processing with description.

Why is a one-hot encoding required? Why can’t you fit a model on your data directly?

**Results**

The program is implemented in python and the output is observed.

**Faculty Signature**