**Assignment I**

**Question1 : Steps needed to perform imputation on a dataset.**

There are 3 types of ways, normally imputation is used

1. Imputation using Statistical methods(Mean, Median and Model)
2. Imputation using Machine Learning Machine Learning Algorithms
3. Imputation using Datawig Package
4. **Imputation using Statistical methods**

Imputation is one of the most common practices usually followed for treating the missing values in the dataset. For using statistical method for imputation distribution of the data need to be checked before replacing the missing values. Data could be sometimes skewed to left or right, or it could be symmetrically distributed. Also, for numerical and categorical variables we use different approaches. Mode is usually used for categorical variables, mean and median are considered for treating missing values in numerical variables. If the data is normally distributed mean is used replace the missing values. If the data is asymmetrical using median makes more sense.

1. **Imputation using Machine Learning Machine Learning Algorithms**

Machine Learning algorithms are strong tools for solving any complex computational tasks, likewise it can also be used in treating outliers in data analysis. The first step of imputation is to create subset of the original dataset. After that choose a Model generating algorithm like Simple Imputer for example. Train a machine learning model then using trained model we will be predicting missing values of the original dataset.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Sl No** | **Student Name** | **English** | **Irish** | **Maths** |
| 1 | Robert | 96 | 66 | NA |
| 2 | Vanda | 68 | 70 | 0 |
| 3 | Ken | NA | 45 | 56 |
| 4 | Robin | 90 | NA | 48 |
| 5 | Mrunali | 56 | 50 | 51 |
| 6 | Kate | 98 | 74 | 29 |
| 7 | Lidija | NA | NA | NA |

1. **Imputation using Datawig Package**

First step of the Datawig , data need to be loaded. After that imputer need to be initialized. Secondly model need to be trained.