

Ajish Rohit

→ Incomplete data means the missing values which can be due to unfilled forms in survey etc. These missing values are treated as NULL and this cannot be processed on machine learning algorithms. We need to clear this missing data. Following are the ways:-

- 1.) Deleting rows with missing values
 - 2.) Predicting missing values.
- This is done using other features which are non null.

Regression or classification can be used to predict missing values.

- 3.) Using algorithms like KNN that are robust to missing values

- 4.) Imputation i.e, replacing missing values with most frequent values

- 5.) Imputating using mean.

- 6.) Expectation maximization

→ These methods can be applied in following manner:-

- S1. Ensure that data is coded correctly
- S2. Identify missing values
- S3. Look for patterns. if found replace by method. (2 + 6)
- S4. If no remove rows