



Start-Tech Academy

Missing Value Imputation

Real-world data often has missing values. Data can have missing values for a number of reasons such as observations that were not recorded and data corruption.

Missing Value Imputation

Impact

- Handling missing data is important as many machine learning algorithms do not support data with missing values.

Solution

- Remove rows with missing data from your dataset.
- Impute missing values with mean/median values in your dataset.

Note

- Use business knowledge to take separate approach for each variable
- It is advisable to impute instead of remove in case of small sample size or large proportion of observations with missing values



Missing Value Imputation

Methods

1. Impute with ZERO

- Impute missing values with zero

2. Impute with Median/Mean/Mode

- For numerical variables, impute missing values with Mean or Median
- For categorical variables, impute missing values with Mode

3. Segment based imputation

- Identify relevant segments
- Calculate mean/median/mode of segments
- Impute the missing value according to the segments
- For example, we can say rainfall hardly varies for cities in a particular State
- In this case, we can impute missing rainfall value of a city with the average of that state

