

1. Iris dataset file is given to you. Perform following tasks on the given dataset.

1.1. Calculate the percentage of total data points missing. [10]

1.2. Calculate the feature wise percentage of total data points missing. [10]

2. Use following techniques to handle missing data. [60]

a. mean imputation: Fill the missing value with the mean of that column. [10]

b. Class-wise mean imputation: Calculate class-wise mean of that column. Fill missing value as per the class label. [10]

c. KNN imputation [K=5]: Implement KNN imputation by calculating distances between data points (using metrics like Euclidean distance on available features), finding the k nearest neighbors for a point with missing data, and then averaging their values to fill missing data. [20]

Compare the statistics after applying each imputations. Write your conclusion. [20]

3. A list of DOBs given to you in a txt file. [20]

Convert all the DOBs in one common format:- YYYY-MM-DD

Example:-

04-09-1978 → 1978-09-04
Dec 11, 2001 → 2001-12-11
1995/10/21 → 1995-10-21

Display the output.
