**Experiment-8**

**AIM:**

To apply the following imputation methods for horse-colic dataset.

a.Statistical Imputation

b.KNN Imputation

c.Iterative Imputation

**SOFTWARE**

Python IDE

**THEORY**

Data imputation is a method for retaining the majority of the dataset's data and information by substituting missing data with a different value. These methods are employed because it would be impractical to remove data from a dataset each time.

**CODE:**

# statistical imputation transform for the horse colic dataset

from numpy import isnan

from pandas import read\_csv

from sklearn.impute import SimpleImputer

# load dataset

url = 'https://raw.githubusercontent.com/jbrownlee/Datasets/master/horse-colic.csv'

dataframe = read\_csv(url, header=None, na\_values='?')

data = dataframe.values

for i in range(dataframe.shape[1]):

 # count number of rows with missing values

 n\_miss = dataframe[[i]].isnull().sum()

 perc = n\_miss / dataframe.shape[0] \* 100

 print('> %d, Missing: %d (%.1f%%)' % (i, n\_miss, perc))

imputer = SimpleImputer(strategy='mean')

# fit on the dataset

imputer.fit(data)

# transform the dataset

Xtrans = imputer.transform(data)

# print total missing

print('Missing: %d' % sum(isnan(Xtrans).flatten()))

# iterative imputation transform for the horse colic dataset

from numpy import isnan

from pandas import read\_csv

from sklearn.experimental import enable\_iterative\_imputer

from sklearn.impute import IterativeImputer

# load dataset

url = 'https://raw.githubusercontent.com/jbrownlee/Datasets/master/horse-colic.csv'

dataframe = read\_csv(url, header=None, na\_values='?')

data = dataframe.values

for i in range(dataframe.shape[1]):

 # count number of rows with missing values

 n\_miss = dataframe[[i]].isnull().sum()

 perc = n\_miss / dataframe.shape[0] \* 100

 print('> %d, Missing: %d (%.1f%%)' % (i, n\_miss, perc))

# define imputer

imputer = IterativeImputer()

# fit on the dataset

imputer.fit(data)

# transform the dataset

Xtrans = imputer.transform(data)

# print total missing

print('Missing: %d' % sum(isnan(Xtrans).flatten()))

# evaluate knn imputation for the horse colic dataset

from numpy import mean

from numpy import std

from pandas import read\_csv

from sklearn.impute import KNNImputer

# load dataset

url = 'https://raw.githubusercontent.com/jbrownlee/Datasets/master/horse-colic.csv'

dataframe = read\_csv(url, header=None, na\_values='?')

data = dataframe.values

for i in range(dataframe.shape[1]):

 # count number of rows with missing values

 n\_miss = dataframe[[i]].isnull().sum()

 perc = n\_miss / dataframe.shape[0] \* 100

 print('> %d, Missing: %d (%.1f%%)' % (i, n\_miss, perc))

imputer = KNNImputer()

# fit on the dataset

imputer.fit(data)

# transform the dataset

Xtrans = imputer.transform(data)

# print total missing

print('Missing: %d' % sum(isnan(Xtrans).flatten()))

**RESULT :**

Statistical, KNN and iterative imputation methods has been implemented successfully for horse-colic dataset.