import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

from sklearn.preprocessing import StandardScaler

from sklearn.model\_selection import train\_test\_split, cross\_val\_score

from sklearn.linear\_model import LogisticRegression

from sklearn.tree import DecisionTreeClassifier

from sklearn.neighbors import KNeighborsClassifier

from sklearn.ensemble import RandomForestClassifier, BaggingClassifier, GradientBoostingClassifier

from sklearn.svm import SVC

from sklearn.metrics import accuracy\_score

from sklearn.metrics import confusion\_matrix, classification\_report

import warnings

warnings.filterwarnings('ignore')

data = pd.read\_csv("/kaggle/input/pima-indians-diabetes-database/diabetes.csv")

data.head()

print("Shape of data is", data.shape)

print("="\*50)

data.info()

data.isnull().sum()

data.duplicated().sum()

data.hist(figsize=(15, 10))

plt.show()

corr = data.corr()

corr

plt.figure(figsize=(8,5))

sns.heatmap(corr, cmap='YlGnBu', annot=True, vmin=-1, vmax=1)

plt.title('Relation between features and Diabetes')

plt.show()