Lab Exercise

**KNN and SVM**

You are provided with two datasets. For each dataset:

1. Perform relevant data-preprocessing techniques
2. Visualize the dataset using Matplotlib. Include at least one scatter plot, bar graph and box plot
3. Implement

1) Decision Tree

2) Naive Bayes

3) KNN

4) SVM

Perform hyperparameter tuning to improve your results.

**Dataset 1**: <https://archive.ics.uci.edu/ml/datasets/Internet+Advertisements>

**Dataset 2:**

### **Context**

According to the World Health Organization (WHO) stroke is the 2nd leading cause of death globally, responsible for approximately 11% of total deaths.  
This dataset is used to predict whether a patient is likely to get a stroke based on the input parameters like gender, age, various diseases, and smoking status. Each row in the data provides relevant information about the patient.

### **Attribute Information**

1) id: unique identifier  
2) gender: "Male", "Female" or "Other"  
3) age: age of the patient  
4) hypertension: 0 if the patient doesn't have hypertension, 1 if the patient has hypertension  
5) heart\_disease: 0 if the patient doesn't have any heart diseases, 1 if the patient has a heart disease  
6) ever\_married: "No" or "Yes"  
7) work\_type: "children", "Govt\_jov", "Never\_worked", "Private" or "Self-employed"  
8) Residence\_type: "Rural" or "Urban"  
9) avg\_glucose\_level: average glucose level in blood  
10) bmi: body mass index  
11) smoking\_status: "formerly smoked", "never smoked", "smokes" or "Unknown"\*  
12) stroke: 1 if the patient had a stroke or 0 if not  
\*Note: "Unknown" in smoking\_status means that the information is unavailable for this patient