**Problem-1**

**Create a Naive Bayes Model and Decision Tree Model to predict whether a person has risk of heart attack or not.**

Please find Jupiter-notebook with name ‘**problem-1/heart\_data-decision\_tree.ipynb**’

View with result in ‘**problem-1/heart\_data-decision\_tree.ipynb.html**’

**Perform following activities on the given dataset:**

* **Perform exploratory data analysis and Data preprocessing (like problem 1)** 
  + **First 10 rows of the data:** Done
  + **5-point summary:** Done show ('min', '25%', '50%', '75%', 'max')
  + **Information about the column (data types):** Done show data after encoded.
  + **Number of outliers (extra points):** Done with Outliers-IQR, Z-Score and Seaborn-Boxplot
  + **Any missing value:** Find with isnull()
  + **Correlation between variables:** Done with Pandas-corr() and heatmap graph
  + **Distribution of the data:** Done with Seaborn-Histplot
* **Data Preprocessing**
* **Impute the missing values:** we remove records with blank and NA values.
* **Outlier treatment:** For treatment with drop outlier records with both IQR and Z-Score.
* **Encode categorical:** In insurance data gender value store in form of string. Encode this data as follows:
  + gender (male: 1 and female: 0)

* + **Split the dataset:** Train Test split (80-20) is perform with *sklearn.model\_selection import train\_test\_split* function.
  + **Model preparation:** We have created model with both **Naive Bayes** and **Decision Tree.**
  + **Evaluation:** For model evaluation we use accuracy\_score(), f1\_score() and ConfusionMatrixDisplay of sklearn.metrics

Submit separate Jupiter-notebooks: “**heart\_data-decision\_tree.ipynb”**

Final model pickle file: “**heart\_model.pkl”**

**Summary**

If we go with IQR treatment data

**Naive Bayes** Test Accuracy: 0.5109

**Naive Bayes** Test F1 Score: 0.4415

**Decision Tree** Test Accuracy: 0.8585

**Decision Tree** Test F1 Score: 0.9105

If we go with Z-Score treatment data

**Naive Bayes** Test Accuracy: 0.8339

**Naive Bayes** Test F1 Score: 0.8570

**Decision Tree** Test Accuracy: 0.8589

**Decision Tree** Test F1 Score: 0.9007

So overall **Decision Tree** model is performing better than **Naïve Bayes**.