Q1) Explain the Concept of Naive Bayes ~~Theorem~~ Classifier.

ANS.

- The Naive Bayes Classifier Theorem is a Fundamental concept in the Field of Machine learning and statistical Inference.
- It provides a Framework for making predictions or classifying data based on the principles of Baye's Theorem, which describes the relationship between conditional and Marginal probabilities.
- The Theorem (Naive Bayes) is particularly well-suited for classifying emails as either spam or not spam.
- It states that for a given Feature or attribute, the probability of a particular class can be calculated by multiplying the prior probability of the class with the likelihood of the Feature given that class, and then dividing the result by the probability of the Feature.
- Mathematically,

The Naive Bayes Classifier Theorem can be expressed as:-

$$P(\text{class}|\text{Feature}) = (P(\text{class} * P(\text{Feature}|\text{class})) / P(\text{Feature})$$

- $P(\text{class}|\text{feature})$ is the posterior probability of the class, given the Feature
- $P(\text{class})$ is the prior probability of the class.
- $P(\text{Feature}|\text{class})$ is the likelihood of the Feature given the class.
- $P(\text{Feature})$ is the prior probability of the Feature.

Q2) Consider the Observation For the Car Theft Scenario having 3 attributes, color, type and Origin. Find the probability of car Theft having scenarios Red SUV and Domestic.

ANS.

(1) Calculate ~~on~~ the Prior Probabilities.

⇒ i) $P(\text{stolen} := \text{yes}) = \frac{\text{Count of Yes}}{\text{Total Observation}} = \frac{5}{10} = \boxed{0.5}$

ii) $P(\text{stolen} := \text{No}) = \frac{\text{Count of No}}{\text{Total Observation}} = \frac{5}{10} = \boxed{0.5}$

(2) Calculate Conditional Probabilities for P(Stolen := Yes)

- $P(\text{Color} = \text{Red} \mid \text{stolen} = \text{YES}) = \frac{(\text{Red}, \text{Yes count})}{(\text{Total stolen Yes})}$

$= \frac{3}{5} \approx \underline{0.6}$

- $P(\text{Type} = \text{SUV} \mid \text{stolen} = \text{YES}) = \frac{(\text{SUV}, \text{Yes count})}{(\text{Total stolen Yes})} = \frac{1}{5}$

$= \frac{1}{5} \approx \underline{0.2}$

- $P(\text{Origin} = \text{Domestic} \mid \text{stolen} = \text{Yes}) = \frac{(\text{Domestic, Yes count})}{(\text{Total stolen Yes})}$

$= \frac{2}{5} \approx \underline{0.4}$

(3.) Combine Probabilities :-

$P(\text{stolen} = \text{YES} \mid \text{RED}, \text{SUV}, \text{DOMESTIC}) = 0.5 \times 0.6 \times 0.2 = \underline{0.024}$

Answers) The Probability of Car Theft having scenario Red, SUV and Domestic (car) is 0.024

Q3.) Write a Python code for the Data pre-processing mentioned in step 4 and Explain every step in Detail.

ANS. Python Code :-

```
import pandas as pd
From sklearn.preprocessing import LabelEncoder, StandardScaler
from sklearn.impute import SimpleImputer
df = pd.read_csv(("https://archive/Iris.csv")
names = ["sepal-length", "sepal_width", "petal-length", "petal_width", "class"])
```

```
imputer = SimpleImputer(strategy='mean')
df.iloc[:,:-1] = Imputer.fit_transform(df.iloc[:,:-1])

encoder = LabelEncoder()
df['class'] = encoder.fit_transform(df['class'])

scaler = StandardScaler()
df.iloc[:,:-1] = scaler.fit_transform(df.iloc[:,:-1])

print("Preprocessed Data Sample:")
print(df.head())
```

EXPLAINATION:-

→ (1.) Load the Dataset.
- The Iris Dataset (iris.csv) is read from the .csv file

(2.) Handle Missing Values.
- We use SimpleImputer to replace missing values in numerical columns with their mean.

(3.) Encode the Categorical Data.
- Convert Categorical to Numerical Labels.

(4.) Feature Scaling.
- We use StandardScaler() to normalize numerical Features, ensuring that they have a mean of 0 (zero) and standard deviation of 1.

(5.) Display the Processed Data.
- The First few row of pre-processed DF are printed to verify changes.