**KAHENI PETER**

**P101/1740G/21**

**COM 404E MACHINE LEARNING**

**CAT 2**

**Question 1**

**a) Assumptions made by Naive Bayes algorithm**

1. Features are independent of each other (Feature Independence Assumption).
2. Each feature contributes equally to the outcome (Equal Weight Assumption).

#### ****b) Role of labeled data in supervised learning****

1. Labeled data provides the input-output pairs that enable supervised learning models to learn patterns or relationships.

2. It acts as a guide for the model to map inputs to outputs effectively.

#### ****c) Rainfall records****

##### **i) Program to display average rainfall in each month**

rainfall = [5, 2, 12, 17, 18, 12, 10, 7, 3, 12, 18, 13]

for month, rain in enumerate (rainfall, start=1):

print (f "Month {month}: {rain} mm")

##### **ii) Code snippet to calculate the yearly average rainfall**

rainfall = [5, 2, 12, 17, 18, 12, 10, 7, 3, 12, 18, 13]

average\_ rainfall = sum(rainfall) / len(rainfall)

print (f "Average Rainfall for the Year: {average\_rainfall:.2f} mm")

### ****Question 2****

#### ****a) What does the term Loss function refer to?****

A loss function quantifies the difference between the predicted output of a model and the actual target value.

It helps in updating model parameters during training.

#### ****b) Why is overfitting a concern to data scientists?****

Overfitting occurs when a model performs well on training data but fails to generalize to unseen data.

This leads to poor performance on test datasets and real-world applications.

c) Program to create and write to a file

text = ("A new technique enables an AI agent to be guided by data crowdsourced "

"asynchronously from nonexpert human users as it learns to complete a task "

"through reinforcement learning.")

with open ("notes.txt", "w") as file:

file. write(text)

d) Program to read the file and process text

with open("notes.txt", "r") as file:

content = file. read()

data\_count = content.lower().count("data")

content\_upper = content.upper()

print(f"The word 'data' appears {data\_count} times.")

print("Text in uppercase:")

print(content\_upper)

### ****Question 3****

#### ****Program to predict yellow beans yield for 2024****

##### **Using Linear Regression**

import numpy as np

from sklearn.linear\_model import LinearRegression

# Data

rainfall = [800, 1200, 1500, 1400, 1000, 1100, 1600]

yield\_per\_acre = [1300, 1850, 1550, 1650, 1900, 1950, 1450]

# Reshape data for sklearn

X = np.array(rainfall).reshape(-1, 1)

y = np.array(yield\_per\_acre)

# Train the model

model = LinearRegression()

model.fit(X, y)

# Predict for 2024 rainfall (900 mm)

predicted\_yield = model.predict([[900]])

print(f"Predicted yield for 2024: {predicted\_yield[0]:.2f} kg/acre")

### How the Model Works

* **Linear Regression** assumes a linear relationship between rainfall (X) and yield (y), expressed as:

y= m⋅ X + c

where:

* + m is the slope (rate of change in yield with rainfall).
  + c is the intercept (yield when rainfall is 0).
* The model computes m and c using the provided data and uses these parameters to predict the yield for 2024's rainfall (900 mm).