### **Probability Grand Retake 30marks 1hr40minutes**

### **Technology**

"Artificial Intelligence is rapidly transforming industries, with advancements in machine learning, natural language processing, and automation. From self-driving cars to AI-powered chatbots, the technology is redefining human-computer interactions, improving efficiency, and unlocking new possibilities for businesses and consumers alike."

### **Health**

"Regular physical activity and a balanced diet play a crucial role in maintaining overall well-being. Studies show that even moderate exercise can reduce the risk of chronic diseases such as diabetes and heart conditions, while proper nutrition enhances mental and physical health, leading to a longer, more active life."

### **Politics**

"In a heated debate over economic policies, lawmakers remain divided on tax reforms and public spending. While some advocate for increased government intervention to support social programs, others push for deregulation and tax cuts to boost private sector growth. The outcome of these discussions will shape the nation's financial landscape for years to come."

### **Entertainment**

"The latest blockbuster film has taken the box office by storm, breaking records and captivating audiences worldwide. With a compelling storyline, breathtaking visuals, and a star-studded cast, the movie has sparked conversations across social media, proving once again the power of storytelling in cinema."

**Question 1: Preprocessing and Vocabulary Analysis (5 marks)**

**Task**:

1. Convert all text to lowercase
2. Remove punctuation and stop words
3. Create a unique vocabulary for each domain
4. Calculate and analyze word frequency distributions

**Question 2: Probabilistic Classification of an Unknown Text (10 marks)**

**Unknown:**  "Prioritizing mental health is as important as maintaining physical well-being. Research indicates that mindfulness practices, adequate sleep, and social connections significantly reduce stress and improve cognitive function. By incorporating relaxation techniques and self-care routines, individuals can enhance their emotional resilience and overall quality of life."

**Task**:

1. Calculate the likelihood of the unknown text belonging to each domain
2. Determine the most probable domain classification
3. Explain the key features that informed the classification

**Question 3: Basic Statistical Analysis of Car Prices (5 marks)**

**Background** You have a dataset of used cars with information about their selling prices, kilometers driven, and year of manufacture.

**Tasks**

1. Descriptive Statistics (2 marks):
   * Calculate the mean, median, and standard deviation for:
     + Selling price
     + Kilometers driven
     + Year of manufacture
2. Correlation Analysis (2 marks):
   * Compute the correlation matrix between:
     + Selling price
     + Kilometers driven
     + Year of manufacture
   * Explain which variables show the strongest relationships
   * Create a scatter plot showing the relationship between these features.
3. Price Distribution Analysis (1 mark):
   * Fit a normal distribution to the selling price
   * Calculate the probability of a car being priced above 10,00,000
   * Explain what this probability represents

**Question 4: Multivariate Analysis of Car Market Characteristics (10 marks)**

**Tasks**

Here's a slightly expanded version of the preprocessing and exploration task:

1. Data Exploration and Preprocessing (3):
   * Data Cleaning:
     + Check for and handle missing values.
   * Descriptive Statistics:
     + Calculate basic statistics for numerical columns.
   * Visualization:
     + Create histograms for each numerical variable.
     + Generate a box plot to show the distribution.
2. Multivariate Normal Distribution Modeling (4 marks):
   * Construct a multivariate normal distribution using all three numerical columns.
   * Compute:
     + Mean vector for these parameters
     + Covariance matrix
     + Heatmap
3. Calculate the probability density for a hypothetical car with (3 marks):
   * + Selling price: 8,00,000
     + Kilometers driven: 50,000
     + Year of manufacture: 2016