# BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI
# WORK INTEGRATED LEARNING PROGRAMMES

First Semester 2025-2026

Assignment 1
Total number of Questions: 4
Submission Date: 14-09-2025
Total Marks: 10

---

Q1. A hospital is analysing the probability of patients testing positive for a rare disease (D) using a diagnostic test. The disease occurs in 0.5% of the general population. The test has a 98% true positive rate and a 95% true negative rate. However, the hospital observes that the test's accuracy varies by age group: for patients under 30, the true positive rate drops to 90%, while the true negative rate increases to 97%. Further assume that the percentage of people whose age is under 30 is 40. Using the law of total probability, calculate the overall probability that a randomly selected patient tests positive for the disease.

Comment on the accuracy of the test (whether good or bad) with reason being carried out at the hospital for the disease D. [1.5 + .5 = 2M]

Q2. A region has recorded that over a 30-year period, there were 6 instances of a 140 mm/h rainfall for 10 minutes. The total number of rainfall events recorded in this period was 150. Calculate the annual probability of this specific rainfall event occurring using the formula for annual exceedance probability. Then, explain why this calculated probability differs from the result of simply dividing the number of occurrences (6) by the number of years (30). In your explanation, address how the total number of rainfall events in the 30-year period influences the annual probability and why this distinction is critical for risk assessment in hydrology. [3M]

Q3. A factory produces light bulbs with a 4% defect rate. A quality inspector randomly selects 20 bulbs. Calculate the probability that at least 3 are defective using the binomial distribution formula. Additionally, explain why the binomial distribution is an appropriate model for this scenario and discuss how this probability might influence the factory's quality control policies if the defect rate were to increase to 6%. [2m]

Q4. Implement a Naïve Bayes classifier for a categorical dataset with Laplace smoothing. Given the training data [('Sunny', 'High', 'Yes'), ('Sunny', 'Low', 'Yes'), ('Rainy', 'High', 'No'), ('Rainy', 'Low', 'No')] and test instance ('Sunny', 'High'), compute the predicted class label. Assume all features are categorical and use a smoothing parameter of 1. Discuss how the use of Laplace smoothing in this scenario affects the probability estimates for rare feature-class combinations. Output: predicted class label. [3M]

---