JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

TEST -3 EXAMINATION- 2025

B.Tech-V Semester (CSE/IT)

COURSE CODE (CREDITS): 25B11CI512 (3) MAX. MARKS: 35

COURSE NAME: Intelligent Techniques for predictive data analytics

COURSE INSTRUCTORS: NTS* MAX. TIME: 2 Hours

Note: *(a) All questions are compulsory.*

(b) The candidate is allowed to make Suitable numeric assumptions wherever required for solving problems

Q.No	Question	CO	Marks
Q1	Apply K-Means algorithm over the data (185, 72), (170, 56), (168, 60), (179,68), (182,72), (188,77) up to two iterations and show the clusters. Initially choose first two objects as initial centroids. Consider K = 2.		7
Q2	A bank wants to classify loan applicants into *"Approved"* or *"Rejected"* categories using a dataset with the following attributes: **Applicant** | **Income** | **Employment** | **Default Before** | **Decision** A_1 | High | Stable | No | Approved A_2 | Low | Stable | No | Approved A_3 | High | Unstable | No | Approved A_4 | Low | Unstable | Yes | Rejected A_5 | Low | Stable | Yes | Rejected Using the Naive Bayes model, compute the posterior probability that a new applicant with attributes: • Income = Low • Employment = Stable • Default Before = No should be Approved.		7
Q3	Consider a binary classification task where the training data is not linearly separable. **(a)** Explain how a Support Vector Machine (SVM) deals with such data using kernel functions. **(b)** A researcher argues that *"K-Nearest Neighbor (KNN) would perform equally well without kernels."* Evaluate this statement, comparing KNN and SVM in terms of complexity, decision boundaries, and sensitivity to data distribution with help of an example.		3+4

<table>
<tr><td>Q4</td><td>(A) Discuss the differences between agglomerative and divisive hierarchical clustering approaches.
(B) Explain the need for data standardization in clustering. Compare z-score normalization, min–max scaling, and decimal scaling in terms of their impact on distance-based algorithms such as K-means and hierarchical clustering.</td><td></td><td>3+4</td></tr>
<tr><td>Q5</td><td>Consider four objects with the following distance matrix:
<table><tr><th></th><th>A</th><th>B</th><th>C</th><th>D</th></tr><tr><td>A</td><td>0</td><td>2</td><td>6</td><td>10</td></tr><tr><td>B</td><td>2</td><td>0</td><td>5</td><td>9</td></tr><tr><td>C</td><td>6</td><td>5</td><td>0</td><td>4</td></tr><tr><td>D</td><td>10</td><td>9</td><td>4</td><td>0</td></tr></table>(A) Perform one iteration of agglomerative hierarchical clustering using single linkage.
(B) State which two clusters merge first and why.
(C) What do we understand by discovering holes in a dataset? Explain its importance in clustering, and discuss methods to identify such gaps.</td><td></td><td>3+2+2</td></tr>
<tr><td></td><td></td><td></td><td></td></tr>
</table>