Name: Enrolment No:	

UPES
End Semester Examination, May 2025

Course: Application of Machine Learning in Industries — **Semester: VI**
Program: B. Tech CSE — **Time : 03 hrs.**
Course Code: CSAI3006 — **Max. Marks: 100**

Instructions: Write all the answers concisely

SECTION A
(5Qx4M=20Marks)

S. No.		Marks	CO
Q 1	Q1. A delivery company uses a binary classifier to detect fraud transactions. Assume the feature vectors follow a multivariate normal distribution for legitimate and fraud transactions respectively: $f: \mathbb{R}^d \rightarrow \{0, 1\}$ $N(\mu_0, \Sigma_0), N(\mu_1, \Sigma_1)$ (a) Derive the decision boundary under the assumption $\Sigma_0 = \Sigma_1$. (b) How would you implement a Bayes-optimal classifier in this case?	4	CO1
Q 2	In retail demand forecasting, suppose monthly sales data y(t) follows a seasonal autoregressive model: $y(t) = \alpha + \sum_{i=1}^{p} \phi_i\, y(t-i) + \sum_{j=1}^{s} \theta_j\, y(t-jS) + \varepsilon(t)$ $\varepsilon(t) \sim N(0, \sigma^2)$ (a) Explain the stationarity conditions for this model. (b) Propose how ML can improve prediction over this model and define feature transformations to be used in LSTM.	4	CO4
Q 3	In wind power forecasting, suppose $X = \{x_1, x_2, ..., x_n\}$ denotes hourly wind speeds and Y denotes the corresponding energy outputs.	4	CO2

	(a) Fit a linear regression model and derive the closed-form solution using matrix notation. (b) Apply Ridge regularization and derive the cost function. How does it promote sparsity in feature selection?		
Q 4	Consider an ML pipeline for genomics classification using Support Vector Machines (SVM). Let $\phi : \mathbb{R}^d \rightarrow \mathbb{R}^D$ be a nonlinear kernel mapping. (a) Derive the dual optimization problem of SVM using Lagrangian multipliers. (b) How does the kernel trick reduce computational complexity? Use RBF kernel as an example.	**4**	**CO1**
Q 5	A CNN-based quality control system uses a convolution kernel $K \in \mathbb{R}^{3x3}$ on an input matrix $I \in \mathbb{R}^{5x5}$. (a) Derive the output feature map using valid padding and stride 1. (b) Interpret the role of ReLU activation mathematically. (c) How would you apply cross-entropy loss if this system classifies products into Defective or Non-defective?	**4**	**CO2**
	SECTION B **(4Qx10M= 40 Marks)**		
Q 6	Design a deep learning-based fault detection system in oil refineries. Include: (a) Formulation of the classification problem. (b) CNN/RNN model suited for sensor time series data. (c) Evaluation using metrics such as F1-score, AUC. (d) Sensor-specific feature engineering strategies.	**10**	**CO1**

Q 7	In educational platforms, adaptive learning engines improve student outcomes. (a) Construct a hybrid recommender using deep matrix factorization and content filters. (b) Define the utility matrix and compute predictions using cosine similarity. (c) Suggest algorithms to detect student disengagement early.	**10**	**CO2**
Q 8	A cancer prediction pipeline uses SVM with a polynomial kernel. (a) Derive the dual formulation of the SVM optimization using Lagrangian multipliers. (b) Explain the polynomial kernel's effect on feature space and provide an example transformation.	**10**	**CO3**
Q 9	(a) In autonomous vehicles, use SVR for path prediction. Derive SVR objective and explain kernel use for nonlinear paths. OR (b) Build a system for delivery time prediction in last-mile logistics. Include: • Data cleansing and encoding • Feature engineering • XGBoost pipeline • Evaluation with MAE and R^2	**10**	**CO2**
SECTION-C (2Qx20M=40 Marks)			
Q 10	Develop an AI system for dynamic inventory management in fast-moving consumer goods (FMCG). Discuss: (a) Feature engineering for product attributes, promotions (b) Time series forecasting using GRU or Prophet (c) Evaluation with MAPE and RMSE (d) Multi-model ensemble design	**20**	**CO4**

Q 11	**(Attempt any one)** (a) Design a fraud detection framework for banking using AI. Address: • Structured and transactional data fusion • Use of graph-based anomaly detection • Precision-Recall curve evaluation • Real-time scoring and feedback loop OR (b) Build a predictive model for early detection of heart disease. Include: • Data imputation and normalization • Use of ensemble classifiers (e.g., Random Forest, LightGBM) • SHAP-based interpretability • Considerations for model updates	**20**	**CO3**