Semester: 3	
Programme: Data Science	
Course: Predictive Analytics	
Paper code: M2C4DS26033CT / M2C4DS26033CP	**Credits: 6**
Hours/week: Theory: 4 classes per week; Practical: 3 classes per week	
Category: Core/MDC/SEC/VAC: Core	
Theory / Practical / Composite: Composite	
No of Module: 1	

Course Outcome:

Remember the fundamental concepts of model selection, regularization, classification, decision trees and generalized additive models.

Understand subset selection, shrinkage methods, classification techniques, ensemble methods and generalized additive models.

Apply model selection, regularization, classification, decision tree, bagging, random forest and boosting techniques to predictive analytics problems.

Analyze predictive models using evaluation metrics, hyperparameter tuning, ensemble learning and generalized additive modelling techniques.

Evaluate predictive models using appropriate model selection criteria, classification metrics, handling of imbalanced data and scoring techniques.

Create data-driven predictive solutions using suitable model selection, classification, ensemble and generalized additive modelling approaches.

SYLLABUS

UNIT	CONTENT	HOURS or NUMBER OF CLASSES	CO Mapping	COGNITIVE LEVEL
1	**Model selection and Regularization** Shrinkage methods - Ridge regression, LASSO and Elastic NET. Parameters versus hyperparameters. Hyperparameter tuning.	10	CO1, CO2, CO3, CO5	K1, K2, K3, K5
2	**Classification Problems** Logistic regression, multiple logistic regression, multi-category logistic regression (model, parameter estimation and prediction). Multiclass discriminant analysis. K nearest neighbour classification. Evaluation metrics. Handling imbalanced data.	20	CO2, CO3, CO4, CO5	K2, K3, K4, K5
3	**Decision trees**	12	CO2, CO3,	K2, K3, K4, K5

	Classification & regression tree (CART). Bagging. Random forest. Boosting.		CO4, CO5	
4	**Uncertainty, Calibration and Causal Inference** Uncertainty quantification, Model calibration, Basic concepts of causal inference	10	CO2, CO3, CO4, CO5	K2, K3, K4, K5
List of Practical				
Based on Theory topics				
Text Books				
1. James, Witten, Hastie and Tibshirani: An Introduction to Statistical Learning. Second edition, Springer.				
2. Hastie, Tibshirani, Friedman: The Elements of Statistical Learning, Data Mining, Inference and Prediction. Second Edition, Springer Series in Statistics.				
3. McCullagh, P & Nelder, J.A. (1995), Generalized Linear Models. Chapman and Hall.				
Evaluation				
Marks	CIA: Class Performance: 2 Attendance: 3 Mid-term Exam: 10 Total: 15 End Semester Exam: 45	Continuous Assessment: 40		
Paper Structure for Theory Semester Exam	Short Questions: 5 marks each	Long Questions: 10 marks each		
	3 out of 5	3 out of 5		

Course outcomes (COs) and Cognitive Level Mapping

COs	CO Description	Cognitive levels
CO1	**Remember** the fundamental concepts of model selection, regularization, classification, decision trees and generalized additive models.	K1
CO2	**Understand** subset selection, shrinkage methods, classification techniques, ensemble methods and generalized additive models.	K2
CO3	**Apply** model selection, regularization, classification, decision tree, bagging, random forest and boosting techniques to predictive analytics problems.	K3
CO4	**Analyze** predictive models using evaluation metrics, hyperparameter tuning, ensemble learning and generalized additive modelling techniques.	K4
CO5	**Evaluate** predictive models using appropriate model selection criteria, classification metrics, handling of imbalanced data and scoring techniques.	K5
CO6	**Create** data-driven predictive solutions using suitable model selection, classification, ensemble and generalized additive modelling approaches.	K6