Course code: Course Title	Course Structure			Pre-Requisite
SE331: Predictive Analytics	L	T	P	NIL
	3	1	0	

Course Objective: To make one understand the correct framework of predictive modelling process which involves data preparation, model development, hypothesis testing and model evaluation.

S. NO	Course Outcomes (CO)
CO1	Understand the concepts of classification, prediction, and regression models.
CO2	Analyzing the data and apply attribute reduction, attribute extraction, and statistical tests.
CO3	Create the predictive model and evaluate using hypothesis testing and performance evaluation measures.
CO4	Understand and apply linear regression and logistic regression technique.
CO5	Analyze and methods to resolve the problem of overfitting, class imbalance problems, and model hyperparameter tuning.

S. No	Content	Contact Hours
UNIT 1	**Introduction:** Classification & prediction, Key ingredients of predictive models, Goals of a regression analysis. Regression models, Data in a regression analysis.	6
UNIT 2	**Data Preparation & Statistical Tests:** Analyzing the metric data: Measures of central tendency, measures of dispersion, data distribution, histogram analysis, outlier analysis, correlation analysis. Attribute Reduction Methods: Univariate Analysis, Correlation-based Feature Selection, Attribute Extraction: Principal Component Analysis. Overview of statistical tests: Categories, one-tail and two-tail, Type I and Type II errors, interpreting significance results.	8
UNIT 3	**Model Development:** Model Development: Data partition, Attribute reduction, model construction, model validation, hypothesis testing, results interpretation, cross-validation.	6
UNIT 4	**Hypothesis Testing & Model Evaluation:** Steps in Hypothesis Testing, Statistical testing, model-comparison tests. Performance measures for categorical and continuous dependent variables, ROC analysis.	6
UNIT 5	**Linear and Logistic Regression Model Estimation:** Simple Linear Regression: Ordinary Least Squares Estimation, Least Squares Method, Estimating σ, Properties of Least Squares Estimates, Estimated Variances, Comparing Models: The Analysis of Variance, The Coefficient of Determination, R^2, DW Test, Confidence Intervals and Tests, The Residuals, Multiple Regression: Adding a Term to a Simple Linear Regression Model, Explaining Variability, The Multiple Linear Regression Model, Terms and Predictors, Ordinary Least Squares, The Analysis of Variance, Predictions and Fitted Values. Logistic Regression: Binomial Regression, Fitting Logistic Regression, Binomial Random Variables.	8
UNIT 6	**Overfitting, Model Tuning & Class Imbalance:** Concerns in model prediction, The Problem of Over-Fitting; Model Tuning; Data Splitting; Resampling Techniques; Choosing Final Tuning Parameters; Data Splitting Recommendations; Choosing Between Models; Computing. Remedies for Severe Class Imbalance: The Effect of Class Imbalance; Model Tuning; Alternate Cutoffs; Adjusting Prior Probabilities; Unequal Case Weights; Sampling Methods; Cost-Sensitive Training.	8
	TOTAL	**42**

REFERENCES		
S.No.	**Name of Books/Authors/Publishers**	**Year of Publication / Reprint**
1.	Max Kuhn and Kjell Johnson, "Applied Predictive Modelling", Springer Verlag.	**2013**
2.	Sanford Weisberg, "Applied Linear Regression", Wiley, Fourth Edition.	**2014**
3.	Ruchika Malhotra, "Empirical Research in Software Engineering: Concepts, Analysis & Applications", CRC press.	**2016**
4.	Samprit Chatterjee, Ali S. Hadi, "Regression Analysis by Example", John Wiley, Fifth Edition.	**2012**
5	Edward W. Frees, Richard A. Derrig, Glenn Meyers, "Predictive Modeling Techniques in Actuarial Science", Vol. I: Predictive Modeling Techniques. Cambridge University Press.	**2014**
6.	Kattamuri S. Sarma, "Predictive Modeling with SAS Enterprise Miner: Practical Solutions for Business Applications", SAS Institute, Second Edition.	**2013**
7.	Jeffrey Strickland, "Predictive Modeling and Analytics", Lulu.com.	**2012**