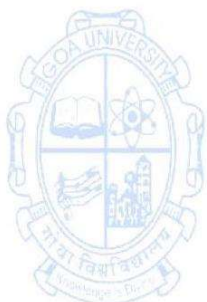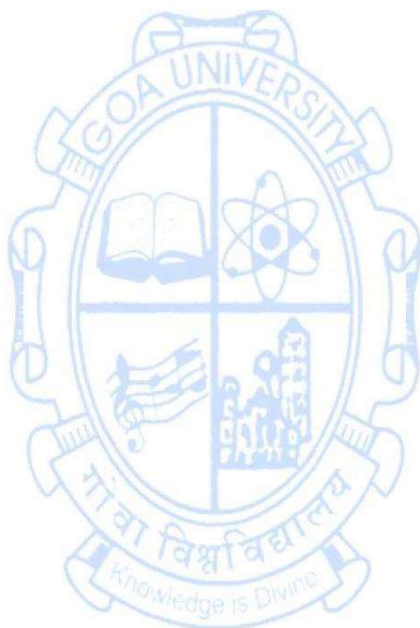**Name of the Programme** : M.Sc. in Data Science

Course Code : CSD-524

Title of the Course : Regression Analytics and Predictive Models

Number of Credits : 2 (2L-2T-0P)

Contact hours : 60 hours (30L-30T-0P)

Effective from AY : 2023-24

Pre-requisites for the Course	Probability Theory and Distributions
Course Objectives	1. Develop an understanding of regression analysis and model building. 2. Provide the ability to develop relationship between variables 3. Investigate possible diagnostics in regression techniques 4. Formulate feasible solutions using a regression model for real-life problems.

Content (Theory)	**Unit 1:** **Simple Regression Analysis** Introduction to a linear and nonlinear model. Ordinary Least Square methods. Simple linear regression model, using simple regression to describe a linear relationship. Fitting a linear trend to time series data, validating simple regression model using t, F and p test. Developing confidence interval. Precautions in interpreting regression results. **Multiple Regression Analysis** Concept of Multiple regression model to describe a linear relationship, Assessing the fit of the regression line, inferences from multiple regression analysis, problem of over fitting of a model, comparing two regression model, prediction with multiple regression equation. **Fitting Curves and Model Adequacy Checking** Introduction, fitting curvilinear relationship, residual analysis, PRESS statistics, detection and treatment of outliers, lack of fit of the regression model, test of lack of fit, Problem of autocorrelation and heteroscedasticity. Estimation of pure errors from near neighbors. **Transformation techniques** Introduction, variance stabilizing transformations, transformations to linearize the model, Box Cox methods, transformations on the repressors variables, Generalized and weighted least squares, Some practical applications.	**15 hours**

	Unit 2: **Multicollinearity** Introduction, sources of multicollinearity, effects of multicollinearity. Multicollinearity diagnostics: examination of correlation matrix, variance Inflation factors (VIF), Eigen system analysis of X1X. Methods of dealing with Multicollinearity: collecting additional data, model re- specification, and ridge regression. **Generalized Linear Models** Generalized linear model: link functions and linear predictors, parameter estimation and inference in the GLM, prediction and estimation with the GLM, Residual Analysis, and concept of over dispersion. **Model building and Nonlinear Regression** Variable selection, model building, model misspecification. Model validation techniques: Analysis of model coefficients, and predicted values, data splitting method. Nonlinear regression model, nonlinear least squares, transformation to linear model, parameter estimation in nonlinear system, statistical inference in nonlinear regression.	**15 hours**
Content for Tutorial Slots:	1. Linear Regression	**2 hours**
	2. Minimum Least Square Method	**2 hours**
	3. Calculating coefficients values	**2 hours**
	4. Ascombe's Quartet	**2 hours**
	5. Regression Equations- x on y & y on x	**2 hours**
	6. Predicting mom's height based on daughter's height	**2 hours**
	7. Regression-Solved problem-2	**2 hours**
	8. Probable Error- Calculating correlation coefficient of POPULATION	**2 hours**
	9. Predictive modelling project for credit card fraud detection	**4 hours**
	Any two Projects from below -	

	10. Predictive modeling project for customer value prediction	**5 hours**
	11. Predictive modeling project for stock market forecasting	**5 hours**
	12. Predictive modeling project for corporate bankruptcy prediction	**5 hours**

Pedagogy	Lectures/ Tutorials/Hands-on assignments/Self-study/Flipped classroom
References/ Readings	1. Draper, N. R., & Smith, H. (1998). Applied regression analysis (Vol. 326). John Wiley & Sons. 2. Johnson, R., & Wichern, D. (2007). Applied Multivariate Statistical Analysis, PHI Learning Pvt. 3. Montgomery, D. C., Peck, E. A., & Vining, G. G. (2021). Introduction to linear regression analysis. John Wiley & Sons. 4. Pardoe, I. (2020). Applied regression modeling. John Wiley & Sons.
Course Outcomes	1. Develop in-depth understanding of the linear and nonlinear regression model. 2. Demonstrate the knowledge of regression modelling and model selection techniques. 3. Examine the relationships between dependent and independent variables. 4. Estimate the parameters and fit a model.