

Branch Name:	MCA
Program Code:	CS201
Course Code	4CS2010309T
Course Title	Statistical Modeling & Predictive Analytics
Pre-requisite Course:	It's should have prior knowledge of Mathematics, Probability & Statistics, Programming (Python/R), and basic Machine Learning concepts.

Course Objective:

1. To develop a strong foundation in statistical concepts including probability, distributions, and inferential statistics for data analysis.
2. To equip students with techniques for statistical modeling such as regression analysis and hypothesis testing.
3. To enable the application of predictive analytics methods using machine learning algorithms for real-world problem-solving.
4. To train students in analyzing and interpreting data through exploratory data analysis and visualization techniques.
5. To provide hands-on experience with modern data science tools such as Python, R, and relevant libraries for building predictive models.
6. To enhance the ability to evaluate and deploy predictive models for decision-making in various domains like healthcare, finance, and business analytics.

Teaching and Examination Scheme:

Teaching Scheme (Hours per week)				Evaluation Scheme (Marks)				
Lecture	Tutorial	Practical	Credit	Theory		Practical		Total
				University Assessment	Continuous Assessment	University Assessment	Continuous Assessment	
3	-	-	3	60	40	-	-	100

Course Content

Unit No	Topics	Total Hours	Weightage (%)
1	Foundations of Statistical Modeling • Introduction to Statistical Modeling and Data Science lifecycle • Types of Data: Structured, Unstructured, Time Series • Descriptive Statistics: Mean, Median, Mode, Variance, Standard Deviation • Probability Theory Basics ◦ Conditional Probability ◦ Bayes Theorem • Probability Distributions ◦ Normal, Binomial, Poisson Distribution • Sampling Techniques and Central Limit Theorem • Exploratory Data Analysis (EDA) ◦ Data Visualization (Histograms, Boxplots, Scatterplots)	09	25

	Correlation and Covariance Analysis		
2	Regression and Statistical Inference - Concept of Statistical Inference - Population vs Sample - Estimation Techniques - Hypothesis Testing - Null & Alternative Hypothesis - t-test, z-test, Chi-square test, ANOVA - Linear Regression - Simple Linear Regression - Multiple Linear Regression - Assumptions and Model Evaluation (R^2, Adjusted R^2) - Logistic Regression - Binary Classification - Odds Ratio and Logit Function - Model Diagnostics - Residual Analysis - Multicollinearity (VIF)	10	25
3	Predictive Modeling Techniques - Introduction to Predictive Analytics - Supervised vs Unsupervised Learning - Classification Techniques - Decision Trees - Random Forest - K-Nearest Neighbors (KNN) - Naïve Bayes - Clustering Techniques - K-Means Clustering - Hierarchical Clustering - Model Evaluation Metrics - Accuracy, Precision, Recall, F1-score - Confusion Matrix - ROC Curve & AUC - Feature Engineering and Selection - Cross-Validation Techniques	09	25
4	Advanced Predictive Analytics & Applications - Time Series Analysis - Components (Trend, Seasonality, Noise) - AR, MA, ARIMA Models - Ensemble Learning Techniques - Bagging and Boosting (AdaBoost, Gradient Boosting) - Introduction to Machine Learning Pipelines - Handling Imbalanced Data - Model Deployment Basics - Case Studies and Real-world Applications - Finance Forecasting	10	25

	- o Healthcare Analytics - o Social Media Sentiment Prediction • Tools & Technologies - o Python Libraries (NumPy, Pandas, Scikit-learn, Statsmodels) - o Visualization Tools (Matplotlib, Seaborn)		
--	---	--	--

Text Books

1. Data Science and Predictive Analytics – Ivo D. Dinov, Springer
2. Predictive Data Mining Models – David L. Olson, Desheng Wu, Springer
3. Statistical Rethinking: A Bayesian Course with Examples in R and Stan – Richard McElreath

Reference Books

1. Applied Predictive Modeling – Max Kuhn, Kjell Johnson, Springer
2. The Elements of Statistical Learning – Trevor Hastie, Robert Tibshirani, Jerome Friedman
3. Causality: Models, Reasoning, and Inference – Judea Pearl
4. An Introduction to Statistical Learning – Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani

Course Learning Outcomes (CLO): On completion of this course, the students will be able to:

CLO	Description	Bloom's Taxonomy Level
CLO1	Understand statistical foundations for data analysis	2. Understanding
CLO2	Apply regression and inference techniques	3. Applying
CLO3	Build predictive models using machine learning algorithms	4. Analyzing
CLO4	Evaluate and deploy models for real-world applications	5. Evaluating

Mapping of CLOs with POs & PSOs

Course Learning Outcomes	Program Outcomes (POs)												Program Specific Outcomes (PSOs)	
	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2
CLO1	H	M	L	M	M	-	M	-	-	-	-	-	M	M
CLO2	H	H	M	H	M	-	M	-	-	-	-	-	H	M
CLO3	H	H	H	M	H	-	M	-	-	-	M	M	H	H
CLO4	H	M	H	H	H	M	H	M	M	M	M	H	H	H

H:High, M:Medium, L:Low