

Branch Name:	MCA
Program Code:	CS201
Course Code	4CS2010309P
Course Title	Statistical Modeling & Predictive Analytics Practical
Prerequisite 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	
-	-	4	2	-	-	25	25	50

List of Experiments:

1. **Introduction to Data Handling**
 - Install Python/R environment (Anaconda, Jupyter Notebook)
 - Load datasets using Pandas
 - Perform basic data operations (filtering, sorting, aggregation)
2. **Exploratory Data Analysis (EDA)**
 - Summary statistics (mean, median, variance)
 - Data visualization: histograms, boxplots, scatter plots
 - Correlation analysis and heatmaps
3. **Probability Distributions**
 - Implementation of Normal, Binomial, and Poisson distributions
 - Visualization of distributions
 - Sampling and Central Limit Theorem demonstration
4. **Hypothesis Testing**
 - Perform t-test, z-test, Chi-square test using datasets
 - ANOVA for comparing multiple groups
 - Interpretation of p-values
5. **Linear Regression Modeling**
 - Simple Linear Regression implementation
 - Multiple Linear Regression
 - Model evaluation using R^2 and residual analysis

6. **Logistic Regression**
 - Binary classification problem
 - Model building and prediction
 - Evaluation using confusion matrix
7. **Classification Algorithms**
 - Decision Tree implementation
 - K-Nearest Neighbors (KNN)
 - Naïve Bayes classifier
 - Compare model performances
8. **Clustering Techniques**
 - K-Means clustering
 - Hierarchical clustering
 - Cluster visualization and interpretation
9. **Model Evaluation Techniques**
 - Accuracy, Precision, Recall, F1-score
 - ROC Curve and AUC
 - Cross-validation (K-fold)
10. **Feature Engineering**
 - Handling missing values
 - Encoding categorical variables
 - Feature scaling (Normalization, Standardization)
11. **Time Series Analysis**
 - Visualization of time series data
 - Decomposition (trend, seasonality)
 - ARIMA model implementation for forecasting
12. **Mini Project (Capstone)**
 - End-to-end predictive analytics project
 - Dataset selection (real-world dataset)
 - Data preprocessing, modeling, evaluation
 - Report and presentation

Tools & Technologies

- Python (NumPy, Pandas, Scikit-learn, Statsmodels)
- R (optional)
- Jupyter Notebook / Google Colab
- Visualization: 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	L2 Understanding
CLO2	Apply regression and inference techniques	L3 Applying
CLO3	Build predictive models using machine learning algorithms	L4 Analyzing
CLO4	Evaluate and deploy models for real-world applications	L5 Evaluating

Mapping of CLOs with POs & PSOs

Course Learning Outcomes	Program Out comes(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