BIG DATA ANALYTICS
[As per Choice Based Credit System (CBCS) scheme]
(Effective from the academic year 2017 -2018)
SEMESTER – VI

Subject Code	17CS662	IA Marks	40
Number of Lecture Hours/Week	4	Exam Marks	60
Total Number of Lecture Hours	40	Exam Hours	03

CREDITS – 03

Module – 1	**Teaching Hours**
Introduction to Data Analytics and Decision Making: Introduction, Overview of the Book, The Methods, The Software, Modeling and Models, Graphical Models, Algebraic Models, Spreadsheet Models, Seven-Step ModelingProcess.**Describing the Distribution of a Single Variable**:Introduction,Basic Concepts, Populations and Samples, Data Sets,Variables,and Observations, Types of Data, Descriptive Measures for Categorical Variables, Descriptive Measures for Numerical Variables, Numerical Summary Measures, Numerical Summary Measures with StatTools,Charts for Numerical Variables, Time Series Data, Outliers and Missing Values,Outliers,Missing Values, Excel Tables for Filtering,Sorting,and Summarizing. **Finding Relationships among Variables**: Introduction, Relationships among Categorical Variables, Relationships among Categorical Variables and a Numerical Variable, Stacked and Unstacked Formats, Relationships among Numerical Variables, Scatterplots, Correlation and Covariance, Pivot Tables.	**08 Hours**
Module – 2	
Probability and Probability Distributions:Introduction,Probability Essentials, Rule of Complements, Addition Rule, Conditional Probability and the Multiplication Rule, Probabilistic Independence, Equally Likely Events, Subjective Versus Objective Probabilities, Probability Distribution of a Single Random Variable, Summary Measures of a Probability Distribution, Conditional Mean and Variance, Introduction to Simulation. **Normal,Binormal,Poisson,and Exponential Distributions**:Introduction,The Normal Distribution, Continuous Distributions and Density Functions, The Normal Density,Standardizing:Z-Values,Normal Tables and Z-Values, Normal Calculations in Excel, Empirical Rules Revisited, Weighted Sums of Normal Random Variables, Applications of the Normal Random Distribution, The Binomial Distribution, Mean and Standard Deviation of the Binomial Distribution, The Binomial Distribution in the Context of Sampling, The Normal Approximation to the Binomial, Applications of the Binomial Distribution, The Poisson and Exponential Distributions, The Poisson Distribution, The Exponential Distribution.	**08 Hours**
Module – 3	
Decision Making under Uncertainty:Introduction,Elements of Decision Analysis, Payoff Tables, Possible Decision Criteria, Expected Monetary Value(EMY),Sensitivity Analysis, Decision Trees, Risk Profiles, The Precision Tree Add-In,Bayes' Rule, Multistage Decision Problems and the Value of Information, The Value of Information, Risk Aversion and Expected Utility, Utility Functions, Exponential Utility, Certainty Equivalents, Is Expected Utility	**08 Hours**

Maximization Used? **Sampling and Sampling Distributions**: Introduction, Sampling Terminology, Methods for Selecting Random Samples, Simple Random Sampling, Systematic Sampling, Stratified Sampling, Cluster Sampling, Multistage Sampling Schemes, Introduction to Estimation, Sources of Estimation Error, Key Terms in Sampling, Sampling Distribution of the Sample Mean, The Central Limit Theorem, Sample Size Selection, Summary of Key Ideas for Simple Random Sampling.	
Module – 4	
Confidence Interval Estimation: Introduction, Sampling Distributions, The t Distribution, Other Sampling Distributions, Confidence Interval for a Mean, Confidence Interval for a Total, Confidence Interval for a Proportion, Confidence Interval for a Standard Deviation, Confidence Interval for the Difference between Means, Independent Samples, Paired Samples, Confidence Interval for the Difference between Proportions, Sample Size Selection, Sample Size Selection for Estimation of the Mean, Sample Size Selection for Estimation of Other Parameters. **Hypothesis Testing**:Introduction,Concepts in Hypothesis Testing, Null and Alternative Hypothesis, One-Tailed Versus Two-Tailed Tests, Types of Errors, Significance Level and Rejection Region, Significance from p-values, Type II Errors and Power, Hypothesis Tests and Confidence Intervals, Practical versus Statistical Significance, Hypothesis Tests for a Population Mean, Hypothesis Tests for Other Parameters, Hypothesis Tests for a Population Proportion, Hypothesis Tests for Differences between Population Means, Hypothesis Test for Equal Population Variances, Hypothesis Tests for Difference between Population Proportions, Tests for Normality, Chi-Square Test for Independence.	**08 Hours**
Module – 5	
Regression Analysis: Estimating Relationships: Introduction, Scatterplots : Graphing Relationships, Linear versus Nonlinear Relationships,Outliers,Unequal Variance, No Relationship,Correlations:Indications of Linear Relationships, Simple Linear Regression, Least Squares Estimation, Standard Error of Estimate, The Percentage of Variation Explained:R-Square,Multiple Regression, Interpretation of Regression Coefficients, Interpretation of Standard Error of Estimate and R-Square, Modeling Possibilities, Dummy Variables, Interaction Variables, Nonlinear Transformations, Validation of the Fit. **Regression Analysis**: Statistical Inference:Introduction,The Statistical Model, Inferences About the Regression Coefficients, Sampling Distribution of the Regression Coefficients, Hypothesis Tests for the Regression Coefficients and p-Values, A Test for the Overall Fit: The ANOVA Table,Multicollinearity,Include/Exclude Decisions, Stepwise Regression,Outliers,Violations of Regression Assumptions,Nonconstant Error Variance,Nonnormality of Residuals,Autocorrelated Residuals ,Prediction.	**08 Hours**
Course outcomes: The students should be able to:	
• Explain the importance of data and data analysis • Interpret the probabilistic models for data • Illustrate hypothesis, uncertainty principle • Demonstrate the regression analysis	
Question paper pattern: The question paper will have ten questions. There will be 2 questions from each module.	

Each question will have questions covering all the topics under a module. The students will have to answer 5 full questions, selecting one full question from each module.
Text Books:
1. S C Albright and W L Winston, Business analytics: data analysis and decision making, 5/e Cenage Learning
Reference Books: