Department of Information Technology (In-house)

Syllabus of B.Sc. in Information Technology (Data Science)

(Effective from academic session 2019-20)

Semester-IV

Name of the Course: B.Sc. in Information Technology (Data Science) **Subject:** Machine Learning for Data Science in R & Machine Learning for Data Science in R Lab	
Course Code: BITDS401 & BITDS491	**Semester: IV**
Duration: 36 Hrs.	**Maximum Marks: 100+100**
Teaching Scheme	**Examination Scheme**
Theory: 3 hrs./week	End Semester Exam: 70
Tutorial: 0	Attendance : 5
Practical:4 hrs./week	Continuous Assessment:25
Credit: 3+2	Practical Sessional internal continuous evaluation:40
	Practical Sessional external examination:60

Aim:

Sl. No.	
1.	To learn R
2.	To introduce the basic concepts and techniques of Machine Learning
3.	To develop the skills in using recent machine learning software for solving practical problems

Objective:

Sl. No.	
1.	To expose to basic terms and terminologies of Machine Learning.
2.	To study the various algorithms related to supervised and unsupervised learning.
3.	To understand the different types of Machine Learning models and how to use them.

Department of Information Technology (In-house)
Syllabus of B.Sc. in Information Technology (Data Science)
(Effective from academic session 2019-20)

Pre-Requisite:			
Sl. No.			
1.	Strong programming skills (Knowledge of C)		
2.	Data computational skill		
Contents		**Hrs./week**	
Chapter	**Name of the Topic**	**Hours**	**Marks**
01	**Introduction To R** Introduction to mechanism for statistics, data analysis, and machine learning; Introduction of R Programming, How to install and run R, Use of R help files, R Sessions, R Objects Vectors, Attributes, Matrices, Array, Class, List, Data Frames etc. Operators in R. R Programming Structures, Control Statements, Loops, Repeat and Break, R-Function, RVector Function, Recursive Function in R. R Packages (Install and Use), Input/Output Features in R, Reading or Writing in File. Data Manipulation in R. Rearranging data, Random Number and Simulation, Statistical methods like min, max, median, mean, length R Programming Structures, Control Statements, Loops, Repeat and Break, R-Function, RVector Function, Recursive Function in R.	**3**	**5**
02	**Supervised Learning (Regression/Classification)** Basic methods: Distance-based methods, Nearest-Neighbours, Decision Trees, Naive Bayes. Linear models: Linear Regression, Logistic Regression, Generalized Linear Models Support Vector Machines, Nonlinearity and Kernel Methods Beyond Binary Classification: Multi-class/Structured Outputs,	**8**	**15**

	Ranking		
03	**Unsupervised Learning** Clustering: K-means/Kernel K-means Dimensionality Reduction: PCA and kernel PCA Matrix Factorization and Matrix Completion Generative Models (mixture models and latent factor models)	**4**	**10**
04	Evaluating Machine Learning algorithms and Model Selection, Introduction to Statistical Learning Theory, Ensemble Methods (Boosting, Bagging, Random Forests)	**4**	**10**
05	Sparse Modeling and Estimation, Modeling Sequence/Time-Series Data, Deep Learning and Feature Representation Learning	**8**	**10**
06	Scalable Machine Learning (Online and Distributed Learning) A selection from some other advanced topics, e.g., Semi-supervised Learning, Active Learning, Reinforcement Learning, Inference in Graphical Models, Introduction to Bayesian Learning and Inference	**6**	**10**
07	Recent trends in various learning techniques of machine learning and classification methods	**3**	**10**
	Sub Total:	**36**	**70**
	Internal Assessment Examination & Preparation of Semester Examination	**4**	**30**
	Total:	**40**	**100**

Practical:

Skills to be developed:

Intellectual skills:

1. Identify the purpose of the analysis.
2. To describe the relationship between factors of the analysis.

3. Information can be useful, used to create new things to achieve objective.
4. Can use a variety of techniques to extend the original idea.

List of Practical: Sl. No. 1& 2 compulsory & at least three from the rest)

1. Implement and demonstrate the FIND-S algorithm for finding the most specific hypothesis based on a given set of training data samples. Read the training data from a .CSV file.

2. For a given set of training data examples stored in a .CSV file, implement and demonstrate the Candidate-Elimination algorithm to output a description of the set of all hypotheses consistent with the training examples.

3. Write a program to demonstrate the working of the decision tree based ID3 algorithm. Use an appropriate data set for building the decision tree and apply this knowledge to classify a new sample.
4. Build an Artificial Neural Network by implementing the Back propagation algorithm and test the same using appropriate data sets.

5. Write a program to implement the naïve Bayesian classifier for a sample training data set stored as a .CSV file. Compute the accuracy of the classifier, considering few test data sets.

6. Assuming a set of documents that need to be classified, use the naïve Bayesian Classifier model to perform this task. Built-in Java classes/API can be used to write the program. Calculate the accuracy, precision, and recall for your data set.

7. Write a program to construct a Bayesian network considering medical data. Use this model to demonstrate the diagnosis of heart patients using standard Heart Disease Data Set. You can use Java/Python ML library classes/API.

8. Apply EM algorithm to cluster a set of data stored in a .CSV file. Use the same data set for clustering using k-Means algorithm. Compare the results of these two algorithms and comment on the quality of clustering. You can add Java/Python ML library classes/API in the program.

9. Write a program to implement k-Nearest Neighbour algorithm to classify the iris data set. Print both correct and wrong predictions. Java/Python ML library classes can be used for this problem.

10. Implement the non-parametric Locally Weighted Regression algorithm in order to fit data points. Select appropriate data set for your experiment and draw graphs.

Assignments:

Based on the curriculum as covered by subject teacher.

Department of Information Technology (In-house)

Syllabus of B.Sc. in Information Technology (Data Science)

(Effective from academic session 2019-20)

List of Books

Text Books:

Name of Author	Title of the Book	Edition/ISSN/ISBN	Name of the Publisher
Joseph Adler	R in a Nutshell		Oreilly
Kevin Murphy	Machine Learning: A Probabilistic Perspective		MIT Press
Reference Books:			
Trevor Hastie, Robert Tibshirani, Jerome Friedman	The Elements of Statistical Learning		Springer
Christopher Bishop	Pattern Recognition and Machine Learning		Springer
Jared P. Lander	R for Everyone: Advanced Analytics and Graphics		Paperback

List of equipment/apparatus for laboratory experiments:

Sl. No.	
1.	**Computer**
2.	**R software**

End Semester Examination Scheme. Maximum Marks-70. Time allotted-3hrs.

Group	Unit	Objective Questions (MCQ only with the correct answer)		Subjective Questions			
		No of question	Total Marks	No of question	To answer	Marks per question	Total Marks

		to be set		to be set			
A	**1 to 7**	**10**					
			10				**60**
B	**1 to 7**			**5**	**3**	**5**	
C	**1 to 7**			**5**	**3**	**15**	

- Only multiple choice type question (MCQ) with one correct answer are to be set in the objective part.
- Specific instruction to the students to maintain the order in answering objective questions should be given on top of the question paper.

Examination Scheme for end semester examination:

Group	Chapter	Marks of each question	Question to be set	Question to be answered
A	**All**	**1**	**10**	**10**
B	**All**	**5**	**5**	**3**
C	**All**	**15**	**3**	**3**

Examination Scheme for Practical Sessional examination:

Practical Internal Sessional Continuous Evaluation

Internal Examination:

Continuous evaluation			**40**

External Examination: Examiner-

Signed Lab Note Book	**10**	
On Spot Experiment	**40**	

Viva voce	**10**	**60**

Name of the Course: B.Sc. in Information Technology (Data Science) **Subject:** Descriptive Analytics & Descriptive Analytics Lab	
Course Code: BITDS402A & BITDS492A	**Semester: IV**
Duration: 36 Hrs.	**Maximum Marks: 100+100**
Teaching Scheme	**Examination Scheme**
Theory: 3 hrs./week	End Semester Exam: 70
Tutorial: 0	Attendance : 5
Practical:4 hrs./week	Continuous Assessment:25
Credit: 3+2	Practical Sessional internal continuous evaluation:40
	Practical Sessional external examination:60

Aim:

Sl. No.	
1.	To interpretation of historical data to better understand.
2.	Make decision by obtain analysis of data.

Objective:

Sl. No.	
1.	To understand the four measurement scales
2.	To interpret the utilization of mean values to describe group results.
3.	To identify the areas of strength and weakness in an organization.

Department of Information Technology (In-house)
Syllabus of B.Sc. in Information Technology (Data Science)
(Effective from academic session 2019-20)

Pre-Requisite:			
Sl. No.			
1.	Programming skills (Knowledge of R)		
2.	Elementary knowledge of data structures and algorithms		
Contents		**Hrs./week**	
Chapter	**Name of the Topic**	**Hours**	**Marks**
01	**Introduction to R Software** Basics and R as a Calculator. Calculation with Data Vectors. Built-in Commands and Missing Data Handling, Operation with Matrices. Introduction to descriptive statistics, Absolute Frequency, Relative Frequency, Frequency Distribution and Cumulative Distribution Function.	**8**	**15**
02	**Graphics and Plots, Bar Diagram** Subdivided Bar, Pie Diagrams, Histogram, Kernel Density and Stem - Leaf Plots. Central tendency of Data, Arithmetic Mean, Median, Quantiles, Mode, Geometric Mean and Harmonic Mean, Range, Interquartile Range and Quartile Deviation.	**10**	**20**
03	**Variation in Data** Absolute Deviation and Absolute Mean Deviation, Mean Squre Error, Variance and Standard Deviation, Coeffivient of Variation and Boxplots. Moments, Association of Variables, Raw and Central Moments. Sheppard's Correction, Absolute Moments and computation of moments, Skewness and Kurtosis.	**8**	**20**
04	**Association of Variables** Univariate and Bivariate Scatter Plots, Smooth Scatter Plots, Quantile	**10**	**15**

	and Three Dimensional Plots, Correlation Coefficient, Rank Correlation Coefficent, Measures of Association of Discrete and counting Variables, Least Squre Method		
	Sub Total:	**36**	**70**
	Internal Assessment Examination & Preparation of Semester Examination	**4**	**30**
	Total:	**40**	**100**

Practical:

Skills to be developed:

Intellectual skills:

1. Can provide the basis for the analysis.
2. Can determine the cause of the problem.
3. Can improve the solution to the problem.

List of Practical:

1. Data exploration (histograms, bar chart, box plot, line graph, scatter plot)
2. Qualitative and Quantitative Data
3. Measure of Central Tendency (Mean, Median and Mode),
4. Measure of Positions (Quartiles, Deciles, Percentiles and Quantiles),
5. Measure of Dispersion (Range, Median, Absolute deviation about median, Variance and Standard deviation), Anscombe's quartet
6. Other Measures: Quartile and Percentile, Interquartile Range

Assignments:

Based on the curriculum as covered by subject teacher.

List of Books

Text Books:

Department of Information Technology (In-house)
Syllabus of B.Sc. in Information Technology (Data Science)
(Effective from academic session 2019-20)

Name of Author	Title of the Book	Edition/ISSN/ISBN	Name of the Publisher
John Fox	An R Companion to Applied Regression	Second Edition	Sage Publications
Reference Books:			
Phil Spector	Data Manipulation with R		Springer
John Fox	Applied Regression Analysis and Generalized Linear Models		Sage Publications
Robert A. Muenchen, Joseph Hilbe	R for Stata Users		Springer
List of equipment/apparatus for laboratory experiments:			
Sl. No.			
1.	Computer		
2.	R software		

End Semester Examination Scheme. **Maximum Marks-70.** **Time allotted-3hrs.**

Group	Unit	Objective Questions (MCQ only with the correct answer)		Subjective Questions			
		No of question to be set	Total Marks	No of question to be set	To answer	Marks per question	Total Marks
A	**1 to 3**	**10**	**10**				**60**

B **C**	**1 to 3** **1 to 3**			**5** **5**	**3** **3**	**5** **15**	

- Only multiple choice type question (MCQ) with one correct answer are to be set in the objective part.
- Specific instruction to the students to maintain the order in answering objective questions should be given on top of the question paper.

Examination Scheme for end semester examination:

Group	Chapter	Marks of each question	Question to be set	Question to be answered
A	**All**	**1**	**10**	**10**
B	**All**	**5**	**5**	**3**
C	**All**	**15**	**3**	**3**

Examination Scheme for Practical Sessional examination:

Practical Internal Sessional Continuous Evaluation

Internal Examination:

Continuous evaluation			**40**

External Examination: Examiner-

Signed Lab Note Book	**10**	
On Spot Experiment	**40**	
Viva voce	**10**	**60**

Department of Information Technology (In-house)

Syllabus of B.Sc. in Information Technology (Data Science)

(Effective from academic session 2019-20)

Name of the Course: B.Sc. in Information Technology (Data Science) **Subject:** Deep Learning & Deep Learning Lab	
Course Code: BITDS402B & BITDS492B	**Semester: IV**
Duration: 36 Hrs.	**Maximum Marks: 100+100**
Teaching Scheme	**Examination Scheme**
Theory: 3 hrs./week	End Semester Exam: 70
Tutorial: 0	Attendance : 5
Practical:4 hrs./week	Continuous Assessment:25
Credit: 3+2	Practical Sessional internal continuous evaluation:40
	Practical Sessional external examination:60

Aim:

Sl. No.	
1.	To improve the performance of a Deep Learning model
2.	to the reduce the optimization function which could be divided based on the classification and the regression problems

Objective:

Sl. No.	
1.	To acquire knowledge on the basics of neural networks.
2.	To implement neural networks using computational tools for variety of problems.
3.	To explore various deep learning algorithms.

Pre-Requisite:

Department of Information Technology (In-house)
Syllabus of B.Sc. in Information Technology (Data Science)
(Effective from academic session 2019-20)

Sl. No.			
1.	Calculus, Linear Algebra		
2.	Probability & Statistics		
3.	Ability to code in R/Python		
Contents		**Hrs./week**	
Chapter	**Name of the Topic**	**Hours**	**Marks**
01	**Introduction** Various paradigms of earning problems, Perspectives and Issues in deep learning framework, review of fundamental learning techniques.	**3**	**5**
02	**Feed forward neural network** Artificial Neural Network, activation function, multi-layer neural network, cardinality, operations, and properties of fuzzy relations.	**6**	**10**
03	**Training Neural Network** Risk minimization, loss function, backpropagation, regularization, model selection, and optimization.	**6**	**15**
04	**Conditional Random Fields** Linear chain, partition function, Markov network, Belief propagation, Training CRFs, Hidden Markov Model, Entropy.	**9**	**15**
05	**Deep Learning** Deep Feed Forward network, regularizations, training deep models, dropouts, Convolutional Neural Network, Recurrent Neural Network, Deep Belief Network.	**6**	**15**
06	**Deep Learning research** Object recognition, sparse coding, computer vision, natural	**6**	**10**

	language		
	Sub Total:	**36**	**70**
	Internal Assessment Examination & Preparation of Semester Examination	**4**	**30**
	Total:	**40**	**100**

Practical:

Skills to be developed:

Intellectual skills:

1. Can be able to analyze relevant data.
2. Can be able to identify a solution for the problem.
3. Can be able to provide the basis for the analysis.

List of Practical: Sl. No. 1& 2 compulsory & at least three from the rest)

Practical based on theory paper Deep Learning

Assignments:

Based on the curriculum as covered by subject teacher.

List of Books

Text Books:

Name of Author	Title of the Book	Edition/ISSN/ISBN	Name of the Publisher
Goodfellow, I.,Bengio,Y., and Courville A.,	Deep Learning		MIT Press
Satish Kumar	Neural Networks: A Classroom Approach		Tata McGraw-Hill

Reference Books:

Bishop, C. ,M.	Pattern Recognition and Machine Learning		Springer
Yegnanarayana, B.	Artificial Neural Networks		PHI Learning Pvt. Ltd
Golub, G.,H., and Van Loan,C.,F.	Matrix Computations		JHU Press

List of equipment/apparatus for laboratory experiments:

Sl. No.	
1.	**Computer**

End Semester Examination Scheme. Maximum Marks-70. Time allotted-3hrs.

Group	**Unit**	**Objective Questions** (MCQ only with the correct answer)		**Subjective Questions**			
		No of question to be set	Total Marks	No of question to be set	To answer	Marks per question	Total Marks
A	**1 to 6**	**10**	**10**				**60**
B	**1 to 6**			**5**	**3**	**5**	
C	**1 to 6**			**5**	**3**	**15**	

- Only multiple choice type question (MCQ) with one correct answer are to be set in the objective part.
- Specific instruction to the students to maintain the order in answering objective questions should be given on top of the question paper.

Examination Scheme for end semester examination:

Department of Information Technology (In-house)

Syllabus of B.Sc. in Information Technology (Data Science)

(Effective from academic session 2019-20)

Group	Chapter	Marks of each question	Question to be set	Question to be answered
A	**All**	**1**	**10**	**10**
B	**All**	**5**	**5**	**3**
C	**All**	**15**	**3**	**3**
Examination Scheme for Practical Sessional examination:				
Practical Internal Sessional Continuous Evaluation				
Internal Examination:				
Continuous evaluation				**40**
External Examination: Examiner-				
Signed Lab Note Book			**10**	
On Spot Experiment			**40**	
Viva voce			**10**	**60**

Department of Information Technology (In-house)

Syllabus of B.Sc. in Information Technology (Data Science)

(Effective from academic session 2019-20)

Name of the Course: B.Sc. in Information Technology (Data Science) **Subject:** Real Time Analytics & Real Time Analytics Lab	
Course Code: BITDS402C & BITDS492C	**Semester: IV**
Duration: 36 Hrs	**Maximum Marks:100+100**
Teaching Scheme	**Examination Scheme**
Theory: 3 hrs./week	End Semester Exam:70
Tutorial: 0	Attendance: 5
Practical: 4 hrs./week	Continuous Assessment: 25
Credit: 3+2	Practical Sessional internal continuous evaluation: 40
	Practical Sessional external examination: 60

Aim:

Sl. No.	
1.	To be processed and analyzed as they arrive in **real time**
2.	Learn business case studies for big data analytics.
3.	It is important in situations where **real-time** processing and **analysis** can deliver important insights and yield business value

Objective:

Sl. No.	
1.	Understand the fundamentals of real time streaming data.
2.	Understand how to process real time data and store them.
3.	To visualize real time data

Department of Information Technology (In-house)

Syllabus of B.Sc. in Information Technology (Data Science)

(Effective from academic session 2019-20)

Pre-Requisite:			
Sl. No.			
1.	Database Management Systems.		
2.	Object Oriented Programming Through Java		
Contents		**3 Hrs./week**	
Chapter	**Name of the Topic**	**Hours**	**Marks**
01	**Introduction to Streaming Data** Source of streaming data, why streaming data is different, infrastructure and algorithms	**6**	**10**
02	**Designing Real-Time Streaming Architectures** Real time architecture components, features of a real time architecture, language of real time programming, real time architecture checklist, Maintaining distributed states, apache zookeeper	**10**	**20**
03	**Data Flow Management, processing and storing in Streaming Analysis** Distributed data flows, apache kafka, apache flume Distributed Processing Streaming Data, Strome, Samza, Consitent hashing, NoSQL and other technologies	**12**	**20**
04	**Analysis and Visualization** Delivering Streaming Metrics,Exact Aggregation and Delivery,Statistical Approximation of Streaming Data Approximating Streaming Data with Sketching Beyond Aggregation	**8**	**20**

	Sub Total:	**36**	**70**
	Internal Assessment Examination & Preparation of Semester Examination	**4**	**30**
	Total:	**40**	**100**

Practical:

Skills to be developed:

Intellectual skills:

1. Ability to implement algorithms to perform various operations on strome, smaza
2. Ability to process real time streaming data

List of Practical:

Hand on experiments based on theory paper

Assignments:

Based on the curriculum as covered by subject teacher.

List of Books

Text Books:

Name of Author	**Title of the Book**	**Edition/ISSN/ISBN**	**Name of the Publisher**
Wily	Real time Analytics		Byron Ellis
Reference Books:			
Anand Rajaraman and Jeffrey David Ullman	Mining of Massive Datasets		CUP
Tom White	Hadoop: The Definitive	Third Edition	O'reilly Media

	Guide		
Chris Eaton, Dirk DeRoos, Tom Deutsch, George Lapis, Paul Zikopoulos	Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data		McGrawHill Publishing
Pete Warden	Big Data Glossary		O'Reilly

List of equipment/apparatus for laboratory experiments:

Sl. No.	
1.	Computer with moderate configuration
2.	Linux os or VM
3.	Hadoop 2.x or higher and other software as required.

End Semester Examination Scheme. Maximum Marks-70. Time allotted-3hrs.

Group	**Unit**	**Objective Questions** (MCQ only with the correct answer)		**Subjective Questions**			
		No of question to be set	Total Marks	No of question to be set	To answer	Marks per question	Total Marks
A	**1 to 4**	**10**	**10**				
B	**1 to 4**			**5**	**3**	**5**	**60**
C	**1 to 4**			**5**	**3**	**15**	

- Only multiple choice type question (MCQ) with one correct answer are to be set in the

objective part.

- Specific instruction to the students to maintain the order in answering objective questions should be given on top of the question paper.

Examination Scheme for end semester examination:

Group	Chapter	Marks of each question	Question to be set	Question to be answered
A	**All**	**1**	**10**	**10**
B	**All**	**5**	**5**	**3**
C	**All**	**15**	**5**	**3**

Examination Scheme for Practical Sessional examination:

Practical Internal Sessional Continuous Evaluation

Internal Examination:

Continuous evaluation			**40**

External Examination: Examiner-

Signed Lab Note Book	**10**	
On Spot Experiment	**40**	
Viva voce	**10**	**60**

Department of Information Technology (In-house)
Syllabus of B.Sc. in Information Technology (Data Science)
(Effective from academic session 2019-20)

Name of the Course: B.Sc. in Information Technology (Data Science) **Subject:** Graphs-Algorithms and Mining	
Course Code: BITDS403	**Semester: IV**
Duration: 48 Hrs.	**Maximum Marks: 200**
Teaching Scheme	**Examination Scheme**
Theory: 3hrs./week	End Semester Exam: 70
Tutorial: 1 hrs./week	Attendance : 5
Practical:0	Continuous Assessment:25
Credit: 4	Practical Sessional internal continuous evaluation:NA
	Practical Sessional external examination:NA

Aim:

Sl. No.	
1.	To provides students a hands-on introduction to scalable graph mining
2.	Data analysis on social networks
3.	Focusing on ways to handle large-scale networks efficiently

Objective:

Sl. No.	
1.	Understand the basic concepts of social networks
2.	Understand the fundamental concepts in analyzing the large-scale data that are derived from social networks
3.	Implement mining algorithms for social networks
4.	Perform mining on large social networks and illustrate the results.

Department of Information Technology (In-house)
Syllabus of B.Sc. in Information Technology (Data Science)
(Effective from academic session 2019-20)

Pre-Requisite:			
Sl. No.			
1.	The students should have a basic algorithmic and programming background		
2.	basic knowledge in the fields of graph theory		
Contents		**Hrs./week**	
Chapter	**Name of the Topic**	**Hours**	**Marks**
01	**Introduction to Social Network Mining, Graph Models and Node Metrics** Introduction to social network mining. Illustration of various social network mining tasks with real-world examples. Data characteristics unique to these settings and potential biases due to them. Social Networks as Graphs.Random graph models/ graph generators (Erd¨os-R´enyi, power law, preferential attachment, small world, stochastic block models, kronecker graphs), degree distributions. Models of evolving networks. Node based metrics, ranking algorithms (Pagerank). Gephi graph visualization and exploration software – practice.	**8**	**10**
02	**Social-Network Graph Analysis** Social network exploration/ processing: graph kernels, graph classification, clustering of social-network graphs, centrality measures, community detection and mining, degeneracy (outlier detection and centrality), partitioning of graphs. SNAP system for large networks analysis and manipulation.	**10**	**15**
03	**Social-Network Graph Analysis and Properties** Social network exploration/ processing and properties: Finding overlapping communities, similarity between graph nodes, counting triangles in graphs, neighborhood properties of graphs. Pregel paradigm and Apache Giraph graph processing system.	**10**	**15**
04	**Information Diffusion in Social Networks**	**12**	**15**

	Strategic network formation: game theoretic models for network creation/ user behavior in social networks. Information diffusion in graphs: Cascading behavior, spreading, epidemics, heterogeneous social network mining, influence maximization, outbreak detection. Opinion analysis on social networks: Contagion, opinion formation, coordination and cooperation.		
05	**Dynamic Social Networks, Applications and Research Trends** Dynamic social networks, Link prediction, Social learning on networks. Special issues in Information and Biological networks. Important applications of social network mining related to the above topics. Research trends.	**8**	**15**
	Sub Total:	**48**	**70**
	Internal Assessment Examination & Preparation of Semester Examination	**4**	**30**
	Total:	**52**	**100**

List of Books

Text Books:

Name of Author	Title of the Book	Edition/ISSN/ISBN	Name of the Publisher
David Easley and Jon Kleinberg	Networks, crowds, and markets		Cambridge University Press
Reference Books:			
Jure Leskovec, Anand Rajaraman and Jeffrey David Ullman	Mining of massive datasets		Cambridge University Press

Department of Information Technology (In-house)
Syllabus of B.Sc. in Information Technology (Data Science)
(Effective from academic session 2019-20)

List of equipment/apparatus for laboratory experiments:	
Sl. No.	
2.	**Computer**

End Semester Examination Scheme. **Maximum Marks-70.** **Time allotted-3hrs.**

Group	**Unit**	**Objective Questions** (MCQ only with the correct answer)		**Subjective Questions**			
		No of question to be set	Total Marks	No of question to be set	To answer	Marks per question	Total Marks
A	**1 to 5**	**10**	**10**				**60**
B	**1 to 5**			**5**	**3**	**5**	
C	**1 to 5**			**5**	**3**	**15**	

- Only multiple choice type question (MCQ) with one correct answer are to be set in the objective part.
- Specific instruction to the students to maintain the order in answering objective questions should be given on top of the question paper.

Examination Scheme for end semester examination:

Group	**Chapter**	**Marks of each question**	**Question to be set**	**Question to be answered**
A	**All**	**1**	**10**	**10**
B	**All**	**5**	**5**	**3**
C	**All**	**15**	**3**	**3**

Department of Information Technology (In-house)
Syllabus of B.Sc. in Information Technology (Data Science)
(Effective from academic session 2019-20)

Name of the Course: B.Sc. in Information Technology (Data Science) **Subject:** Inferential Statistics	
Course Code: BITDS404	**Semester: IV**
Duration: 48 Hrs.	**Maximum Marks: 100**
Teaching Scheme	**Examination Scheme**
Theory: 3 hrs./week	End Semester Exam: 70
Tutorial: 1 hr./week	Attendance : 5
Practical: 0	Continuous Assessment:25
Credit: 4	Practical Sessional internal continuous evaluation:NA
	Practical Sessional external examination:NA

Aim:

Sl. No.	
1.	To learn how to set up and perform hypothesis tests
2.	Use regression analysis to analyze and interpret data collected from ANOVA and ANCOVA designs.

Objective:

Sl. No.	
1.	To enable students to analyze and interpret data
2.	Understand the types of questions that the statistical method addresses
3.	To evaluate the reliability and validity of a measuring
4.	Apply the method to other examples and situations
5.	Use data to make evidence based decisions that are technically sound

Department of Information Technology (In-house)

Syllabus of B.Sc. in Information Technology (Data Science)

(Effective from academic session 2019-20)

Pre-Requisite:			
Sl. No.			
1.	Mathematics		
2.	Probability Statistics		
Contents		**Hrs./week**	
Chapter	**Name of the Topic**	**Hours**	**Marks**
01	Estimation: Concepts of estimation, unbiasedness, sufficiency, consistency and efficiency. Factorization theorem. Complete statistic, Minimum variance unbiased estimator (MVUE) and Rao-Blackwell theorem with applications. Cramer-Rao inequality and MVB estimators (statement and applications).	**9**	**10**
02	Methods of Estimation: Method of moments, method of maximum likelihood estimation.	**3**	**05**
03	Principles of test of significance: Null and alternative hypotheses (simple and composite), Type-I and Type-II errors, critical region, level of significance, size and power, best critical region, most powerful test, uniformly most powerful test,	**12**	**20**
04	Neyman Pearson Lemma (statement and applications to construct most powerful test). Likelihood ratio test and relevant problems, properties of likelihood ratio tests (without proof).	**12**	**15**
05	Interval estimation - Confidence interval for the parameters of various distributions, Confidence interval for Binomial proportion, Confidence interval for population correlation coefficient for Bivariate Normal distribution, Pivotal quantity method of constructing confidence interval, Large sample confidence intervals.	**12**	**20**
	Sub Total:	**48**	**70**
	Internal Assessment Examination & Preparation of Semester Examination	**4**	**30**

	Total:	**52**	**100**

List of Books

Text Books:

Name of Author	**Title of the Book**	**Edition/ISSN/ISBN**	**Name of the Publisher**
Goon A.M., Gupta M.K.: Das Gupta.B.	Fundamentals of Statistics		World Press
Reference Books:			
Rohatgi V. K. and Saleh, A.K. Md. E.	An Introduction to Probability and Statistics	2ndEdn	John Wiley & Sons.
Dudewicz, E. J., and Mishra, S. N.	Modern Mathematical Statistics		John Wiley & Sons.
Bhattacharjee, D. & Das, K. K.	A Treatise on Statistical Inference and Distributions		Asian Books
Hogg, R.V., Tanis, E.A. and Rao J.M	Probability and Statistical Inference	Seventh Ed	Pearson Education

End Semester Examination Scheme. **Maximum Marks-70.** **Time allotted-3hrs.**

Group	**Unit**	**Objective Questions** (MCQ only with the correct answer)		**Subjective Questions**			
		No of question to be set	Total Marks	No of question to be set	To answer	Marks per question	Total Marks
A	**1 to 5**	**10**	**10**				**60**

B **C**	**1 to 5** **1 to 5**			**5** **5**	**3** **3**	**5** **15**	

- Only multiple choice type question (MCQ) with one correct answer are to be set in the objective part.
- Specific instruction to the students to maintain the order in answering objective questions should be given on top of the question paper.

Examination Scheme for end semester examination:

Group	Chapter	Marks of each question	Question to be set	Question to be answered
A	**All**	**1**	**10**	**10**
B	**All**	**5**	**5**	**3**
C	**All**	**15**	**3**	**3**

Department of Information Technology (In-house)
Syllabus of B.Sc. in Information Technology (Data Science)
(Effective from academic session 2019-20)

Name of the Course: B.Sc. in Information Technology (Data Science) **Subject:** Technical Seminar and Communication Skill	
Course Code: BITDS481	**Semester: IV**
Duration: 36 Hrs.	**Maximum Marks: 100**
Teaching Scheme	**Examination Scheme**
Theory: 0	**End Semester Exam: 100**
Tutorial: 0	**Attendance: 0**
Practical: 2 hrs./week	**Continuous Assessment: 0**
Credit: 1	**Practical Sessional internal continuous evaluation: 40**
	Practical Sessional external examination: 60
Contents	
Students will give technical seminar and improve their communication skill.	

Name of the Course: B.Sc. in Information Technology (Data Science) **Subject: Project II**	
Course Code: BITDS482	**Semester: IV**
Duration: 36 Hrs.	**Maximum Marks: 100**
Teaching Scheme	**Examination Scheme**
Theory: 0	**End Semester Exam: 100**
Tutorial: 0	**Attendance: 0**
Practical: 4 hrs./week	**Continuous Assessment: 0**
Credit: 2	**Practical Sessional internal continuous evaluation: 40**
	Practical Sessional external examination: 60
Contents	
Students will do projects on application areas of latest technologies and current topics of societal relevance.	