Annexure F-1

FACULTY OF COMPUTER SCIENCE & IT

SYLLABUS

of

Additional / Optional paper for specialization in Data Science

for
Bachelor of Science (Information Technology)/
Bachelor of Computer Applications

(Semester I-IV)

(Under Credit Based Continuous Evaluation Grading System)

Session: 2024-25

The Heritage Institution
KANYA MAHA VIDYALAYA
JALANDHAR
(Autonomous)

Kanya Maha Vidyalaya, Jalandhar (Autonomous)

SCHEME AND CURRICULUM OF EXAMINATIONS OF THREE YEAR DEGREE PROGRAMME

Bachelor of Science (Information Technology)/ Bachelor of Computer Applications

Credit Based Continuous Evaluation Grading System

Session 2024-25

Additional / Optional paper for Specialization in Data Science

Bachelor of Science (Information Technology)/ Bachelor of Computer Applications Semester – I										
Course Code	**Course Name**	**Course Type**	**Hours per week**	**Credit**		**Marks**				**Examination Time (in Hours)**
			L-T-P	**L-T-P**	**Total**	**Total**	**Ext.**		**CA**	
							L	**P**		
BITL-1119/ BCAL-1119	***Computational Data Science**	O	3-1-0	3-1-0	4	100	80	-	20	3
	Total				4	100				

Bachelor of Science (Information Technology)/ Bachelor of Computer Applications Semester – II										
Course Code	**Course Name**	**Course Type**	**Hours per week**	**Credit**		**Marks**				**Examination Time (in Hours)**
			L-T-P	**L-T-P**	**Total**	**Total**	**Ext.**		**CA**	
							L	**P**		
BITL-2118/ BCAL-2118	***Statistical Techniques for Data Science**	O	3-1-0	3-1-0	4	100	80	-	20	3
	Total				4	100				

Bachelor of Science (Information Technology)/ Bachelor of Computer Applications Semester – III										
Course Code	**Course Name**	**Course Type**	**Hours per week**	**Credit**		**Marks**				**Examination Time (in Hours)**
			L-T-P	**L-T-P**	**Total**	**Total**	**Ext.**		**CA**	
							L	**P**		
BITM-3117/ BCAM-3117	***Data Visualization**	O	3-0-2	3-0-1	4	100	50	30	20	3
	Total				4	100				

Bachelor of Science (Information Technology)/ Bachelor of Computer Applications Semester – IV										
Course Code	**Course Name**	**Course Type**	**Hours per week**	**Credit**		**Marks**				**Examination Time (in Hours)**
			L-T-P	**L-T-P**	**Total**	**Total**	**Ext.**		**CA**	
							L	**P**		
BITM-4117/ BCAM-4117	*** Foundation of Statistical Computing**	O	3-0-2	3-0-1	4	100	50	30	20	3+3
	Total				4	100				

Note:

O - Optional

***One additional/optional paper will be studied by the candidate if she opts for Specialization in Data Science**

Bachelor of Science (Information Technology)/ Bachelor of Computer Applications
Semester I
(Session 2024-25)
COURSE CODE: BITL–1119/ BCAL-1119

COMPUTATIONAL DATA SCIENCE

Course Outcomes:

After the completion of this course, the student will be able to:

CO1: Comprehend terminology associated with data and its processing.

CO2: Comprehend various types of functions in set theory.

CO3: Apply Algorithms of polynomial algebra to solve problems.

CO4: Apply various counting principles, permutations, combinations and averages to solve basic set of problems.

Bachelor of Science (Information Technology)/ Bachelor of Computer Applications
Semester I
(Session 2024-25)
COURSE CODE: BITL–1119/ BCAL-1119

COMPUTATIONAL DATA SCIENCE

L-T-P: 3-1-0 **Max. Marks: 100**
Credit: 4 **Theory: 80**
Examination Time: 3 Hrs. **CA: 20**

Instructions for Paper Setter -

Eight questions of equal marks (16 marks each) are to set, two in each of the four sections (A-D). Questions of Sections A-D should be set from Units I-IV of the syllabus respectively. Questions may be divided into parts (not exceeding four). Candidates are required to attempt five questions, selecting at least one question from each section. The fifth question may be attempted from any section

UNIT I

Data Processing: Basic Terminology of Data, Types of Data, Information and Knowledge, Preprocessing the Data, Data cleaning, Data transformation, Data reduction.

Introduction to Data Science, Evolution of Data science, Need of Data Science, Components of Data Science, Application Areas.

UNIT II

Functions: Functions and their types, Quadratic Functions and Equations, Inverse Function, Logarithmic Functions and Equations.

UNIT III

Algebra of Polynomials: Addition, Subtraction, Multiplication and Division Algorithms
Graphs of Polynomials: X-intercepts, multiplicities, end behavior and turning points, Graphing & Polynomial Creation.

UNIT IV

Basic Principles of Counting and Factorial Concepts: Addition rule of counting, Multiplication rule of counting, Factorials.
Permutation and Combination.
Measures of Central Tendency: Mean, Median and Mode

References/Textbooks:

1. Patricia Pulliam Phillips, Cathy A. Stawarski, “Data Collection: Planning for and

Collecting All Types of Data", Wiley Publisher, First Edition, 2008.
2. Roger Sapsford, Victor Jupp, "Data Collection-and Analysis", Second Edition, Sage Publishing, 2006.
3. Kenneth Rosen, "Discrete Mathematics and Its Applications", Tata McGraw Hill, 7th Edition
4. Anshuman Sharma, Fundamentals of Numerical Methods and Statistical techniques, Lakhanpal Publications (2016)

Bachelor of Science (Information Technology) / Bachelor of Computer Applications
Semester II
(Session 2024-25)
COURSE CODE: BITL–2118/ BCAL-2118
STATISTICAL TECHNIQUES FOR DATA SCIENCE

Course Outcomes:

After the completion of this course, the student will be able to:

CO1: Comprehend the key terminology of descriptive statistics and frequency distribution

CO2: Comprehend the basic Probability terms and their usage.

CO3: Formulate hypothesis for basic problems and perform testing.

CO4: Implement statistical techniques like Chi Square test and Analysis of variance.

Bachelor of Science (Information Technology) / Bachelor of Computer Applications
Semester II
(Session 2024-25)
COURSE CODE: BITL–2118/ BCAL-2118
STATISTICAL TECHNIQUES FOR DATA SCIENCE

L-T-P: 3-1-0 **Max. Marks: 100**
Credit: 4 **Theory: 80**
Examination Time: 3 Hrs **CA: 20**

Instructions for Paper Setter -

Eight questions of equal marks (16 mark each) are to set, two in each of the four sections (A-D). Questions of Sections A-D should be set from Units I-IV of the syllabus respectively. Questions may be divided into parts (not exceeding four). Candidates are required to attempt five questions, selecting at least one question from each section. The fifth question may be attempted from any section

UNIT - I

Introduction to Statistics: Basic terminology, variables: discrete and continuous.
Introduction to descriptive Statistics: Types of data, levels of measurement, categorical variables and numerical variables. Introduction to Frequency distribution.
Measures of Dispersion: Quartile Deviation, Mean Deviation, Standard Deviation and Coefficient of Variance.

UNIT – II

Introduction to Asymmetry: Moments, Kurtosis and Skewness.
Probability: Meaning, Basic concepts, Events, Properties of Probability, Conditional Probability.

UNIT - III

Probability: Addition Theorem, Multiplication Theorem and Bayes' Theorem.
Introduction to Inferential statistics: Concept of a sample and a population, need of sampling.
Hypothesis Testing: Null and Alternate Hypothesis, Type 1 and Type 2 errors, Confidence intervals. Chi square test.

UNIT - IV

ANOVA - one way and two way.
Data Analysis Tools in Spreadsheets: Regression Analysis, Correlation Analysis, Covariance Analysis, ANOVA Analysis.

References/Textbooks:

1. S.P Gupta, Statistical Methods, Sultan Chand & Sons (2012)
2. B. L. Agarwal, Statistics For Professional Courses, CBS Professional (2011)
3. Anshuman Sharma, Fundamentals of Numerical Methods and Statistical techniques, Lakhanpal Publications (2016)
4. Stephen L. Nelson, Excel Data Analysis for Dummies, Wiley Publications (2013)

Bachelor of Science (Information Technology) / Bachelor of Computer Applications
Semester- III
(Session 2024-25)
COURSE CODE: BITM–3117/ BCAL-3117
DATA VISUALIZATION

Course Outcomes:

After the completion of this course, the student will be able to:

CO1: Comprehend Importance and applications of data visualization.

CO2: Acquaint with categories of data visualization tools.

CO3: Apply basic and specialized visualization tools for representation of data.

CO4: Apply advanced data visualization tools.

Bachelor of Science (Information Technology) / Bachelor of Computer Applications
Semester- III
(Session 2024-25)
COURSE CODE: BITM–3117/ BCAL-3117
DATA VISUALIZATION

L-T-P: 3-0-1 **Max. Marks: 100**
Credit: 4 **Theory: 50**
Examination Time: 3 Hrs. **Practical: 30**
CA: 20

Instructions for Paper Setter -

Eight questions of equal marks (10 marks each) are to set, two in each of the four sections (A-D). Questions of Sections A-D should be set from Units I-IV of the syllabus respectively. Questions may be divided into parts (not exceeding four). Candidates are required to attempt five questions, selecting at least one question from each section. The fifth question may be attempted from any section

UNIT-I

Introduction: Introduction of Data Visualization, Meaning of Data Visualization, Importance of Data Visualization, Data Visualization applications, User psychology of Visualization, UX in Data Visualization.

UNIT-II

Gestalt principles of Data Visualization, Introduction to DIKW hierarchy, Goals of Data Visualization, Basic Visualization tools - Area Plots, Histograms, Bar Charts.

UNIT-III

Data Visualization tools: Introduction, characteristics, types, image and graphical visualization. Specialized Visualization tools: Pie Charts, Box Plots, Scatter Plots, Bubble Plots.

UNIT-IV

Advanced Visualization tools: Need, Application, Visualization of Maps, Storyboards in Visualization. Waffle Charts, Word Clouds, Seaborn and Regression Plots.

References / Textbooks:

1. E. Tufte, The Visual Display of Quantitative Information (2nd Edition), Graphics Press, 2001.

2. Herbert Jones, Data Science: The Ultimate Guide to Data Analytics, Data Mining, Data Warehousing, Data Visualization, Regression Analysis, Database Querying, Big Data for Business and Machine Learning for Beginners Kindle Edition, 2020.
3. E. Tufte, Envisioning Information, Graphics Press, 1990.
4. Andy Kirk, Data Visualisation: A Handbook for Data Driven Design SAGE Publications Ltd; First edition, 2016.
5. Kieran Healy, Data Visualization: A Practical Introduction Kindle Edition, Princeton University Press; First edition, 2018.

Note: The latest editions of the books should be followed.

Bachelor of Science (Information Technology) / Bachelor of Computer Applications
Semester IV
(Session 2024-25)
COURSE CODE: BITM–4117/ BCAM-4117
FOUNDATION OF STATISTICAL COMPUTING

Course Outcomes:

After passing this course the student will be able to:

CO1: Comprehend basics of Statistical Computing and managing data structures like vector, matrix, etc.

CO2: Create, operate and manage lists and data frames.

CO3: Apply control and I/O statements for generating outputs.

CO4: Simulate various descriptive and analytical algorithms using R language along with their visualization.

Bachelor of Science (Information Technology) / Bachelor of Computer Applications
Semester IV
(Session 2024-25)
COURSE CODE: BITM–4118/ BCAM-4118
FOUNDATION OF STATISTICAL COMPUTING

L-T-P: 3-0-1 **Max. Marks: 100**
Credit: 4 **Theory: 50**
Examination Time: 3 Hrs. **Practical:30**
CA: 20

Instructions for Paper Setter -

Eight questions of equal marks (10 marks each) are to set, two in each of the four sections (A-D). Questions of Sections A-D should be set from Units I-IV of the syllabus respectively. Questions may be divided into parts (not exceeding four). Candidates are required to attempt five questions, selecting at least one question from each section. The fifth question may be attempted from any section

UNIT – I

Statistical Computing: Introduction, Role of Programming and Statistical Software. Data **Statistics:** Sampling, Cumulative statistics, Statistics for Data frames, matrix objects and lists.

Introduction to R, Help functions in R, Vectors, Common Vector Operations, Using all and any function, subletting of vector. Creating matrices, Matrix operations, Applying Functions to Matrix Rows and Columns, Adding and deleting rows and columns.

UNIT - II

Lists, Creating lists, general list operations, Accessing list components and values, applying functions to lists, recursive lists

Creating Data Frames: Matrix-like operations in frames , Merging Data Frames, Applying functions to Data frames, Factors and Tables , factors and levels , Common functions used with factors , string operations

UNIT - III

Input/ Ouput:scan() , readline() Function, Printing to the Screen Reading and writing CSV and text file. Control statements: Loops, Looping Over Nonvector, Sets, if-else , writing user defined function, scope of the variable, R script file.

UNIT – IV

Graphics in R: Graph Syntax ((title, xlabel, ylabel, pch, lty, col.), Simple graphics (Bar, Multiple Bar, Histogram, Pie, Box-Plot, Scatter plot, qqplot), Low-level and High-Level plot functions. Using Analytical Algorithms (KNN, K-means,Naive Bayes) for Predictive analysis and Modelling.

References / Textbooks:

1. Andrie de Vries and JorisMeys, R Programming for Dummies, Wiley (2016), 2nd Edition.
2. Sandip Rakshit, R Programming for Beginners, McGraw Hill Education (2017), 1st Edition.
3. SandipRakshit, Statistics with R Programming, McGraw Hill Education (2018), 1st Edition.
4. Garrett Grolemund, Hands on Programming with R, O'Reilly (2014), 1st Edition
5. Mark Gardener, Beginning R: The Statistical Programming Language, Wiley (2013)
6. Tilman M. Davies, The Book of R: A first Course in Programming and Statistics, No Strach Press (2016), 1st Edition