THE NATIONAL COLLEGE, JAYANAGAR, BANGALORE

AUTONOMOUS

Bachelor of Computer Application (Data Science)

Syllabus

THE NATIONAL COLLEGE JAYANAGAR, BANGALORE-70

BACHELOR OF COMPUTER APPLICATION (Data Science)

COURSE MATRIX

I SEMESTER							
Part	**Paper**		**Hours/week**	**Marks**			**Credit**
	Code	Title		IA	Exam	Total	
Part 1	Language1	English	4	30	70	100	2
	Language2	Kan/San/Hin/Japanese	4	30	70	100	2
Part 2	B(DS)1.1	Mathematics-I	4	30	70	100	4
	B(DS)1.2	Statistics-I	4	30	70	100	4
	B(DS)1.3	Computer Organization & Architecture	4	30	70	100	4
	B(DS)1.4	Programming in C	4	30	70	100	4
	L1.1	Programming in C Lab	3	15	35	50	1
	L1.2	Mathematics – I and Statistics - I Lab	3	15	35	50	1
Part 3		Mandatory Paper	1	15	35	50	1
Total Marks and credits			**31**	**225**	**525**	**750**	**23**

II SEMESTER							
Part	**Paper**		**Hours/week**	**Marks**			**Credit**
	Code	**Title**		**IA**	**Exam**	**Total**	
Part 1	Language1	English	4	30	70	100	2
	Language2	Kan/San/Hin/Japanese	4	30	70	100	2
Part 2	B(DS)2.1	Mathematics-II	4	30	70	100	4
	B(DS)2.2	Statistics-II	4	30	70	100	4
	B(DS)2.3	Data Structures	4	30	70	100	4
	B(DS)2.4	RDBMS- MySQL	4	30	70	100	5
	L2.1	Data Structures Lab	3	15	35	50	1
	L2.2	Mathematics – II and Statistics - II Lab	3	15	35	50	1
Part 3		Mandatory Paper	1	15	35	50	1
Total Marks and credits			**31**	**225**	**525**	**750**	**24**

III SEMESTER							
Part	**Paper**		**Hours/week**	**Marks**			**Credit**
	Code	**Title**		**IA**	**Exam**	**Total**	
Part 1	Language1	English	4	30	70	100	2
	Language2	Kan/San/Hin/Japanese	4	30	70	100	2
Part 2	B(DS)3.1	Statistical Inference	4	30	70	100	4
	B(DS)3.2	Analysis and Design of Algorithms	4	30	70	100	4
	B(DS)3.3	Python	4	30	70	100	4
	L3.1	Statistics for Data Science (SAS/SPSS) Lab	3	15	35	50	1
	L3.2	Python Lab	3	15	35	50	1
	L3.3	Analysis and Design of Algorithms LAB	3	15	35	50	1
Part 3		Open Elective	2	15	35	50	1
Total Marks and credits			**31**	**210**	**490**	**700**	**20**

IV SEMESTER							
Part	**Paper**		**Hours/week**	**Marks**			**Credit**
	Code	**Title**		**IA**	**Exam**	**Total**	
Part 1	Language1	English	4	30	70	100	2
	Language2	Kan/San/Hin/Japanese	4	30	70	100	2
Part 2	B(DS)4.1	Machine Learning - I	4	30	70	100	4
	B(DS)4.2	Data Mining	4	30	70	100	4
	B(DS)4.3	Web Technologies	4	30	70	100	4
	L4.1	Tableaux (Data Visualization)	3	15	35	50	1
	L4.2	Machine Learning Lab	3	15	35	50	1
	L4.3	Web Technologies Lab	3	15	35	50	1
Part 3		Open ELECTIVE	2	15	35	50	1
Total Marks and credits			**31**	**210**	**490**	**700**	**20**

V SEMESTER							
Part	**Paper**		**Hours/week**	**Marks**			**Credit**
	Code	**Title**		**IA**	**Exam**	**Total**	
Part 2	B(DS)5.1	Machine Learning - II	4	30	70	100	4
	B(DS)5.2	Natural Language Processing	4	30	70	100	5
	B(DS)5.3	Cloud Computing	4	30	70	100	5
	B(DS)5.4	Big Data Analytics	4	30	70	100	4
	B(DS)5.5	Applications of Data Science	4	30	70	100	5
	L5.1	Machine LearningLab	3	15	35	50	1
	L5.2	Big Data Analytics Lab	3	15	35	50	1
	L5.3	Mini Project	6	30	70	100	2
Total Marks and credits			**32**	**210**	**490**	**700**	**27**

VI SEMESTER							
Part	**Paper**		**Hours/week**	**Marks**			**Credit**
	Code	**Title**		**IA**	**Exam**	**Total**	
Project/Internship			32	210	490	700	24
Total Marks and credits			**32**	**210**	**490**	**700**	**24**

All Six Semester Matrix

Semester	Hours/week	Marks			Credit
		IA	**Exam**	**Total**	
First	**31**	**225**	**525**	**750**	**24**
Second	**31**	**225**	**525**	**750**	**24**
Third	**31**	**210**	**490**	**700**	**20**
Fourth	**31**	**210**	**490**	**700**	**20**
Fifth	**32**	**210**	**490**	**700**	**27**
Sixth	**32**	**210**	**490**	**700**	**24**
Total Marks and Credits				**4300**	**139**

SEMESTER I

I SEMESTER							
Part	**Paper**		**Hours/week**	**Marks**			**Credit**
	Code	Title		IA	Exam	Total	
Part 1	Language1	English	4	30	70	100	2
	Language2	Kan/San/Hin/Japanese	4	30	70	100	2
Part 2	B(DS)1.1	Mathematics-I	4	30	70	100	4
	B(DS)1.2	Statistics-I	4	30	70	100	4
	B(DS)1.3	Computer Organization & Architecture	4	30	70	100	4
	B(DS)1.4	Programming in C	4	30	70	100	4
	L1.1	Programming in C Lab	3	15	35	50	1
	L1.2	Mathematics – I and Statistics - I Lab	3	15	35	50	1
Part 3		Mandatory Paper	1	15	35	70	1
Total Marks and credits			**31**	**225**	**525**	**750**	**24**

TITLE: MATHEMATICS-I

PAPER CODE: B(DS)1.1	CREDITS : 4	TOTAL NO OF HRS: 52

Objectives:

On completion of the course, the student will be able to

- ✓ Analyze and understand big and small numbers and their different forms of representation that relate to business. Comprehend algebraic solutions to simple mathematical and business problems.
- ✓ Solve linear and quadratic equations using multiple methods.
- ✓ Understand information organized in row and column format (matrix), and use algebraic methods to interpret them. Elementary processed in differentiation and appreciate the need for continuous and discrete functions as needed in Business and Management.

MODULE 1	**Number Systems** Introduction to numbers, Integers, Rational numbers, Irrational numbers, Real numbers, Imaginary numbers, Complex numbers, Prime numbers, Algorithm to test if a number is prime. LCM, HCF, Divisibility criteria, Expression of a number as a product of its prime factors, Perfect squares and perfect cubes – Surds, Conjugate surds, Rationalization of surds. Number systems – Binary, Octal, Hexadecimal representation, Change of base, Conversion from one base to the other, Computer representation, Scientific notation.	**06 hrs**
MODULE 2	**Vectors** Vectors, Adding and subtracting of vectors, scalar and dot product of vectors, gradient of a vector,distance between two vectors,sum of the squares and magnitude of vectors.	**06 hrs**
MODULE 3	**Linear Algebra** Algebra of powers – Indices, Logarithms, Factorials, Law of indices. Polynomials, Roots of polynomials, Descartes rule of sign, Quadratic equations, Tracing quadratics. Ratio and proportions, Binomial theorem, Use of nCr, nPr, Maximum value of nCr, Symmetric nature of binomial coefficients.	**10 hrs**
MODULE 4	**Matrices** Matrices – Nomenclature, Matrix operations – Addition, Subtraction, Multiplication, Inversion. Types of matrices, Characteristics equation of a square matrix, Cayley – Hamilton theorem. Determinants – Evaluation of a determinant, Identical rows and columns, Properties of determinants.	**10 hrs**

MODULE 5	**Solution to Systems of Linear Equations** System of linear equations and criteria for unique solutions, Solution of linear equations using Cramer's rule, Elementary row operations, Gauss elimination method, Row echelon form, Iteration solutions to linear equations, Matrix method of solutions.	**10 hrs**
MODULE 6	**Differential Calculus** Limits, Continuity, Derivative, Derivatives of standard functions (results only), Derivatives of a constant, Derivative of exponential and logarithmic functions, Derivatives of sum, product and quotient of two functions, Differentiation of composite functions – Chain rule, Differentiation of parametric functions.	**10 hrs**

Text Books:

1. P. Kandasamy, K. Thilagavathy, Mathematics for B.Sc. Vol-I, II, III & IV, S Chand & Company Ltd., New Delhi-55.
2. G.K. Ranganath, A Text Book of B.Sc. Mathematics, S Chand & Company Ltd., New Delhi- 55.
3. G.K. Ranganath, A Text Book of BCA Mathematics, Himalaya Publishing House.
4. A.P. Verma, Business Mathematics and Statistics, Asian Books Private Limited.
5. Dr. S.R. Arora, Dr. Kavitha Gupta, Taxmann's Business Mathematics, University Edition.
6. Dr. J.H. Thukral, Business Mathematics and Statistics.
7. Dr. P.R. Vittal, Business Mathematics and Statistics.
8. D.C. Sancheti, V.K. Kapoor, Business Mathematics, Sultan Chand and Sons.
9. Digambar Patri, D.N. Patri, Business Mathematics, Kalyani Publishers
10. B.G. Umarani, Dr. P.G. Umarani, Mathematics for II Year Pre-University Course, Quality Publishers.
11. Dr. B.S. Grewal, Higher Engineering Mathematics, Khanna Publishers

TITLE: Statistics-I

PAPER CODE:	**CREDITS :**	**TOTAL NO OF HRS: 52**

Objectives of the course are:

- ✓ This paper will help students to have a thorough knowledge of descriptive statistics.
- ✓ To understand measures of central tendency and use them to analyze data.
- ✓ Students will be able to find out how spread out data values are on number line.

MODULE 1	**Introduction** Population and sample, Types of data – Qualitative, Quantitative, Univariate, Multivariate, Cross sectional, Time, Series, Discrete, Continuous, Primary, Secondary, Scales of measurement – Nominal, Ordinal, Interval, Ratio, Variables and attributes, Organization and presentation of data, Construction of frequency distributions (univariate and bivariate), Presentation of data through diagrams (bar and pie) and graphs (frequency curve, histogram, cumulative frequency curves), Stem and leaf plot.	**18 hrs**
MODULE 2	**Measures of Central Tendency** Measures of location or central tendency – Arithmetic mean, Median, Mode, Geometric mean, Harmonic mean – Properties, Positional averages or quartiles – Quartiles, Deciles and Percentiles	**7 hrs**
MODULE 3	**Measures of Dispersion** Measures of dispersion – Absolute measures – Range, Mean deviation, Quartile deviation, Standard deviation – Statement of properties, Coefficient of variation, Skewness and Kurtosis – Concept and measures. Discussions on the suitability of the different measures in practical situations	**7 hrs**
MODULE 4	**Correlation and Regression** Linear correlation – Scatter diagram, Product moment correlation coefficient – Properties, Spearman's rank correlation coefficient, Simple regression, Prediction.	**20 hrs**

Text Books:

1. Freund, Ronald, E. Walpole, Mathematical Statistics, Fourth Edition (1987), Prentice Hall of India, New Delhi.
2. B.L. Agarwal, Basic Statistics (2009), New Age Publishers.
3. J. Medhi, Statistical Methods – An Introductory Text, New Age Publishers.
4. A.M. Goon, M.K. Gupta and B. Das Gupta, Fundamentals of Statistics, Vol. 1, Sixth Edition, World Press, Calcutta.
5. Gupta and Kapoor, Fundamentals of Mathematical Statistics, Sultan Chand and Sons.
6. G.W. Snedecor, Cochran, Statistical Methods, Eighth Edition, Wiley.
7. Sheldon M. Ross, Introductory Statistics, Second Edition, Acadamic Press.
8. Pal, Sarkar, Statistics Concepts and Applications, Second Edition, PHI.
9. David Freedman, Robert Pisani, Roger Purves, Statistics, Fourth Edition, Viva.
10. Roger. E. Kirk, Statistics, An Introduction, Fourth Edition, Harcourt Brace College Publishers.
11. Walpolw, Myers, Probability and Statistics for Engineers and Scientists, Eighth Edition, Pearson Education .
12. S. Sundararajan, Monograph on Statistics and Probability. (No Publication).
13. Dr. B.S. Grewal, Higher Engineering Mathematics, 40th Edition, Khanna Publishers.
14. Harry Frank, Steven C. Althoen, Statistics Concepts and Applications, Cambridge University Press.
15. Murray R. Spiegel, Larry J. Stephens, Statistics, Third Edition, Schaum's Outlines.
16. C.M. Chikkodi, B.G. Satyaprasad, B.Com Business Statistics, Himalaya Publishing House.
17. Dr. B.N. Gupta, Statistics, (No Publication)
18. V. Sundarapandian, Probability, Statistics and Queueing Theory, PHI Learning Private Limited.
19. Vijay K. Rohatgi, A.K. Md. Ehsanes Saleh, An Introduction to Probability and Statistics, Second Edition, Wiley Series in Probability and Statistics.

TITLE: ComputerOrganizationandArchitecture

PAPER CODE:	CREDITS :	TOTAL NO OF HRS: 52

Objectives:

On completion of the course, the student will be able to

- ✓ To conceptualize the basics of organizational and architectural of a digital computer.
- ✓ Be familiarwith thehistoryanddevelopmentofmoderncomputers.Befamiliarwith Number System and Boolean algebra.
- ✓ Be familiar with Combinational and logic circuits. Be familiar with organization and design of modern computer and its architecture.
- ✓ BefamiliarwithI/OorganizationandMemoryorganization

MODULE 1	**Number System and Boolean algebra** Binary, octal, Hexadecimal Number systems, base conversions, signed binary numbers, binary arithmetic, subtraction using compliments, Binary codes, weighted-BCD-8421 code, Gray code, excess-3 code, ASCII code.	**8hrs**
MODULE 2	**Boolean algebra and logic gates:** Boolean laws, Demorgen's theorems, Minimization of Boolean expressions-using Boolean postulates and Karnaugh maps technique(sop). AND, OR, NOT gate using Transistor NAND, NOR as universal gates : X-OR,X-NOR gates	**8hrs**
MODULE 3	**Combinational and logic circuits:** Half adder, half subtractor, full adder, full subtractor, Multiplexer ,De-multiplexer, Encoder, Decoder, Flip-Flops: JK, T, D master slave JK flip flops Shift registers: SISO, SIPO, PISO, PIPO (block diagrams), and 4-bit SISO shift register using D-flip-flop. Counters: Synchronous and Asynchronous.	**9hrs**
MODULE 4	**Basic computer organization and design** Introduction,Instruction codes, Computer registers, Computer instructions, timing and control, hard wired control,micro programmed control, execution and instruction, input output interrupt. Design of computer	**9hrs**
MODULE 5	**Central Processor Organization** Processorbusorganization,arithmeticlogicunit (ALU),Instruction formats, Addressing modes, data transfer and manipulation, program control, microprocessor organization.	**9hrs**
MODULE 6	**Input-output organization and memory organization** Peripheral devices, asynchronousdatatransfer, directmemoryaccess,(DMA),priority Interrupt, input output processor,Introduction, memory hierarchy, main memory, auxiliary memory, cache memory .	**9hrs**

Text Books:

1. Digital Principles and applications by Malvino,albertpaul; Publisher Mcgraw hill, 1975.
2. Computerarchitecture byJohn.5th edition;publisherMorganKaufmann,2011.

Reference Books:

1. Digitalcomputerelectronicsby Albertp.MalvinopublisherCarrerEducation,1992.
2. BasicDigitalElectronicsbyAlvis.J.Evans;publisher masterpublishing,1996.

TITLE: ProgrammingInC

PAPER CODE:	CREDITS :	TOTAL NO OF HRS: 52

Objectives:

- ✓ To study about algorithms, flowcharts and programs. To solve problems through logical thinking.
- ✓ To clearly understand the logic of the problem. To analyze the given problem and write the algorithm, flowchart.
- ✓ To write structured C programs, this is the foundation of any programming language.

MODULE 1	**Introduction to Programming Concepts** Software, Classification of Software, Modular Programming, Structured Programming, Algorithms and Flowcharts with examples. Overview of C Language: History of C, Character set, C tokens, Identifiers, Keywords, Data types, Variables, Constants, Symbolic Constants , Operators in C, Hierarchy of Operators, Expressions, Type Conversions and Library Functions.	**8hrs**
MODULE 2	**Managing Input and Output Operation**: Formatted and Unformatted I/O Functions, Decision making, branching and looping: Decision Making Statements - if Statement, if–else statement, nesting of if-else statements, else–if ladder, switch statement,?: operator, Looping - while, do-while, for loop, Nested loop, break, continue, and goto statements.	**8hrs**
MODULE 3	**Functions**: Function Definition, prototyping, types of functions, passing arguments to functions, Nested Functions, Recursive functions.	**9hrs**
MODULE 4	**Arrays:** Declaring and Initializing, One Dimensional Arrays, Two Dimensional Arrays, Multi Dimensional Arrays - Passing arrays to functions. Strings: Declaring and Initializing strings, Operations on strings, Arrays of strings, passing strings to functions. Storage Classes - Automatic, External, Static and Register Variables.	**9hrs**
MODULE 5	**Structures** Declaring and Initializing, Nested structure, Array of Structure, Passing Structures to functions, Unions, typedef, enum, Bit fields. Pointers – Declarations, Pointer arithmetic, Pointers and functions, Call by value, Call by reference, Pointers and Arrays, Arrays of Pointers, Pointers and Structures. Meaning of static and dynamic memory allocation, Memory allocation functions.	**9hrs**

MODULE 5	**Files** File modes, File functions, and File operations, Text and Binary files, Command Line arguments. C Preprocessor directives, Macros – Definition, types of Macros, Creating and implementing user defined header files.	**9hrs**

Text Books:

1. E. Balaguruswamy, "Programming In ANSI C", 4th edition, TMH Publications, 2007
2. Ashok N. Kamthane, "Programming with ANSI and Turbo C", Pearson Education, 2006

Reference Books:

1. . Ashok N. Kamthaneet. al., "Computer Programming and IT", Pearson Education, 2011 2.Mahapatra, " Thinking In C ", PHI Publications, 1998.

2 .YashwantKanetkar, "Let Us C", 13th Edition, PHP, 2013.

TITLE: C Programming Lab

PAPER CODE: L1.1	CREDITS : 1	NO OF HRS: 3hrs/week

Section : A

1. Printing the reverse of an integer
2. Generate first N prime numbers
3. Get a string and convert the lowercase to uppercase and vice-versa without using library functions.
4. Find the occurrence of a particular character in a string
5. Input a string and find the number of each of the vowels which appear in the string.
6. Accept N words and make it as a sentence by inserting blank spaces and a full stop at the end.
7. Print the reverse of a string.
8. Find the first N terms of Fibonacci series using arrays
9. Declare 3 pointers variables to store a character, a character string and an integer respectively. Input values into these variables. Display the address and the contents of variables.
10. Program to demonstrate structure and union.
11. Recursive program to find the factorial of an integer.
12. Find the maximum of 4 numbers by defining a macro for the maximum of two numbers.

Section : B

1. ArrangingNnumbersinascendingandindescendingorderusingbubblesort.
2. Checking whether the given matrix is an identity matrix or not.
3. Addition and subtraction of two matrices.
4. Multiplication of two matrices.
5. Convert a hexadecimal number into its binary equivalent.
6. Check whether the given string is a palindrome or not.
7. Demonstration of bitwise operations.
8. Applying linear search to a set of N numbers by using a function.
9. Createasequentialfilewiththreefields:empno,empname,empbasic.Printall the details in a neat format by adding 500 to their basic salary.
10. Arrange N names in alphabetical order

TITLE: Mathematics–I & Statistics–I Lab

PAPER CODE:	CREDITS :	NO OF HRS: 3hrs/week

Section: A

1. To find Average, Maximum, Minimum, Round and Truncation.
2. Plotting and analyzing the graphs for the given data
3. Plotting and analyzing the graphs for the given data
4. Addition, Subtraction, Scalar multiplication, Transpose, multiplication and inverse of matrices.
5. Solving determinants
6. To compute Powers, Logarithms, Factorial, ${}^{n}C_{r}$, ${}^{n}P_{r}$.
7. To solve linear equations.
8. **Staff can add some more programs**

SEMESTER II

II SEMESTER							
Part	**Paper**		**Hours/week**	**Marks**			**Credit**
	Code	**Title**		**IA**	**Exam**	**Total**	
Part 1	Language1	English	4	30	70	100	2
	Language2	Kan/San/Hin/Japanese	4	30	70	100	2
Part 2	B(DS)2.1	Mathematics-II	4	30	70	100	4
	B(DS)2.2	Statistics-II	4	30	70	100	4
	B(DS)2.3	Data Structures	4	30	70	100	4
	B(DS)2.4	RDBMS- MySQL	4	30	70	100	5
	L2.1	Data Structures Lab	3	15	35	50	1
	L2.2	Mathematics – II and Statistics - II Lab	3	15	35	50	1
Part 3		Mandatory Paper	1	15	35	50	1
Total Marks and credits			**31**	**225**	**525**	**750**	**24**

TITLE: MATHEMATICS - II

PAPER CODE:B(DS)2.1	CREDITS : 4	TOTAL NO OF HRS: 52

Objectives:

On completion of the course, the student will be able to

- ✓ Understand basics of integration and its application to business.
- ✓ Appreciate multi-variable functions, see the effect of change when a single variable is changed holding the other variables constant, applications in business.
- ✓ Student will be able to maximize beneficial values like profit,efficiency, etc. and to minimize values like expenses, effort, etc.
- ✓ Learn the basics of optimizing a business objective subject to linear resource constraints.

MODULE 1	**Integration** Definite integrals, Indefinite integrals, Integrand, Constant of integration, Integration variable. Standard formulae for integration, Methods of integration – Integration by parts, Integration of substitution, Definite integrals, Properties of limits.	**08hrs**
MODULE 2	**Partial Differentiation** Partial differentiation, Representation in suffix and differential form, Mixed derivatives, Partial derivatives of higher order. Homogeneous functions, Euler's theorem. Functions of two variables, Parametric representation, Chain rule for partial differentiation.	**08 hrs**
MODULE 3	**Maxima, Minima** Concept and rules of maxima and minima, Critical points, Inflexion points. Optimizing single variable functions, Conditions for local minimum and maximum.	**09 hrs**
MODULE 4	**Basics of Linear Programming** Basics of linear programming, Objective function, Decision variable, Constraint equations, Non- negativity constraints, Formulation of LPP, Representation of inequalities, Graphical representation, Solution space, Feasibility region of LPP, Bounded and unbounded region, Solution sets. Graphical solution to LPP.	**15 hrs**
MODULE 5	**Curve Fitting** Fitting functions to data points, Algebraic fit versus geometric fit for curves, Fitting plane curves to data points.	**12hrs**

Text Books:

1. A.P. Verma, Business Mathematics and Statistics, Asian Books Private Limited, New Delhi.
2. Stephen Ross, Randolph W Westerfield & Bradford Jordan, Fundamentals of Corporate Finance, Tata McGraw-Hill Publishing Company Limited, New Delhi.
3. P.L. Mehta, Managerial Economics, Sultan Chand & Sons, New Delhi.
4. B.G. Umarani, Dr. P.G. Umarani, Mathematics for II year pre-University Course, Quality Publishers.
5. G.B. Gururajachar, Text Book of Mathematics (BSc – I, II, III, IV Semester), Academic Excellent Series Publication.

TITLE: Statistics-II

PAPER CODE: B(DS)2.2	CREDITS : 4	TOTAL NO OF HRS: 52

Objectives:

- ✓ This paper will help students to have a thorough knowledge of descriptive basic probability and samplings.
- ✓ This course will help students to develop a deeper
- ✓ Understanding of the basis underlying probability distributions and enable them to apply the knowledge and skills to real world tasks.

MODULE 1	**Probability** Introduction to probability, Sample space and events, Axiomatic approach to probability, Addition theorem, Conditional probability, Multiplication theorem, Independent events and Baye's theorem, Rule of total probability.	**10hrs**
MODULE 2	**Random Variables** Concept of a random variable, Discrete and continuous random variable and their probability functions, Distribution function and its properties, Expectation of a random variable – Mean Variance, Bivariate probability distribution, Marginal and conditional distributions, Covariance, Independence, Conditional expectation and variance, Mean and variance of linear combination of random variables.	**12 hrs**
MODULE 3	**Probability Distributions** Bernoulli, Binomial, Poisson, Uniform, Exponential, Normal distributions – Definition through probability function, Statement of properties and applications.	**10 hrs**
MODULE 4	**Sampling and Sampling distributions** Types of sampling – Purposive, Random and mixed samples, Sampling Methods – Simple, Random, Stratified, Cluster, Relative merits and limitations of the different methods. Concepts of populations, Parameter, Random sample, Statistic, Sampling distribution and standard error, Distribution of sample mean and variance. Chi-square, t and F distributions – Definition through their probability functions, Statement of their properties, Applications, Central limit theorem and its applications.	**20 hrs**

Text Books:

1. Hogg & Tanis, Probability & Statistical Inference – Sixth Edition, Pearson Education.
2. S.M. Ross, Introduction to Probability and Statistics, John Wiley and Sons.
3. K.C. Bhuyan, Probability, Distribution theory and statistical inference – NCBA.
4. V.K. Rohatgi, A.K.MD. Ehsanes Saleh (2002), An Introdution to Probability Theory and Mathematical Statistics, John Wiley (WSE).
5. Probability and Statistics, Schaum Series.
6. Walpolw, Myers, Probability and Statistics for Engineers and Scientists, Eighth Edition, Pearson Education .
7. S. Sundararajan, Monograph on Statistics and Probability. (No Publication).
8. Dr. B.S. Grewal, Higher Engineering Mathematics, 40th Edition, Khanna Publishers.
9. Dr. Alka Chaudhary, Dr. Arun Kumar, Probability Theory, Krishna Prakashan Media (P) Ltd.
10. Vijay K. Rohatgi, A.K. Md. Ehsanes Saleh, An Introduction to Probability and Statistics, Second Edition, Wiley Series in Probability and Statistics.
11. S. Sundararajan, Monograph on Statistics and Probability. (No Publication).
12. Harry Frank, Steven C. Althoen, Statistics Concepts and Applications, Cambridge University Press.
13. Murray R. Spiegel, Larry J. Stephens, Statistics, Third Edition, Schaum's Outlines.
14. C.M. Chikkodi, B.G. Satyaprasad, B.Com Business Statistics, Himalaya Publishing House.
15. Dr. B.N. Gupta, Statistics, (No Publication)
16. V. Sundarapandian, Probability, Statistics and Queueing Theory, PHI Learning Private Limited.
17. Vijay K. Rohatgi, A.K. Md. Ehsanes Saleh, An Introduction to Probability and Statistics, Second Edition, Wiley Series in Probability and Statistics.

TITLE: DATA STRUCTURE

PAPER CODE: B(DS)2.3	CREDITS : 4	TOTAL NO OF HRS: 52

Objectives:

- ✓ To be able to practically implement the data structures like stack, queue, array etc. To understand and implement different searching and sorting techniques.

MODULE 1	Introduction and Overview: Definition, Elementary data organization, Data Structures, data structures operations, Abstract data types, algorithms complexity, time-space tradeoff. Preliminaries: Mathematical notations and functions, Algorithmic notations, control structures, Complexity of algorithms, asymptotic notations for complexity of algorithms.	**08hrs**
MODULE 2	Arrays: Definition, Linear arrays, arrays as ADT, Representation of Linear Arrays in Memory, Traversing Linear arrays, Inserting and deleting String Processing: Definition, Storing Stings, String as ADT, String operations, word/text processing, Pattern Matching algorithms.	**08hrs**
MODULE 3	Linked list: Definition, Representation of Singly linked list in memory, Traversing a Singly linked list, Searching a Singly linked list, Memory allocation, Garbage collection, Insertion into a singly linked list, Deletion from a singly liked list; Doubly liked list, Header liked list, Circular linked list.	**09 hrs**
MODULE 4	Stacks – Definition, Array representation of stacks, Linked representation of stacks, Stack as ADT, Arithmetic Expressions: Polish Notation, Application of Stacks, Recursion, Towers of Hanoi, Implementation of recursive procedures by stack. Queues – Definition, Array representation of queue, Linked list representation of queues Types of queue: Simple queue, Circular queue, Double ended queue, Priority queue, Operations on Queues, Applications of queues.	**09 hrs**
MODULE 5	Sorting: Bubble sort, Insertion sort, Selection sort, Searching: Linear Search, Binary search, Multidimensional arrays, Matrices and Sparse matrices.	**09hrs**
MODULE 6	Tree – Definitions, Binary trees, Representing binary trees in memory, Traversing Binary Trees, Binary Search Trees, Searching, Inserting and Deleting in a Binary Search Tree, Heap Tree.	**09 hrs**

Text Books:

1. Seymour Lipschutz, "Data Structures with C", Schaum'soutLines, Tata McGraw-Hill, 2011.

Reference Books:

1. Mark Allen Weiss, “Data Structures and Algorithm Analysis in C”, Second Edition, Pearson Education, 2013.
2. Robert Kruse, C.L.Tondo, Bruce Leung, ShashiMogalla, “Data Structures and Program Design using C”, Pearson Education, 2009.
3. Forouzan, “A Structured Programming Approach using C”, 2nd Edition, Cengage Learning India, 2008.

TITLE: Database Management Systems

PAPER CODE:B(DS)	CREDITS : 5	TOTAL NO OF HRS: 52

Objectives:

This course enables students to understand:

- ✓ Advanced topics in database management and programming including client server application development are introduced.
- ✓ Expands knowledge of data modeling concepts and introduces object-oriented data modeling techniques

MODULE 1	Introduction, Database Systems: Introduction, Database Systems Characteristics of DB Approach, Advantages of DBMS, Database Users, DB Languages, Applications of Database.	**08hrs**
MODULE 2	Data Model Concepts: Data Model Concepts, Database System Architecture-Centralized, Client/Server: Two- tier, Three-tier, Three-Schema Architecture-Physical Data Independence and Logical Data Independence, Different types of data models, Database Interfaces.	**08hrs**
MODULE 3	E-R Model concepts E-R Model concepts- Entities, Attributes, Relationship, E-R model constraints, E-R diagrams, Relational model concepts, Characteristics of relations, constraints on relations, Relational Algebra-Unary and Binary operations.	**09 hrs**
MODULE 4	**SQL** SQL:DDL - Create table/views, Drop, Alter commands, DML - Insert, Delete, Update, Select, queries ,sub-queries, nested queries, Joins – equijoin, non-equijoin, Built-in functions of SQL & grouping. Concept of Functional dependency, Normalization – 1NF,2NF,3NF.	**09 hrs**

MODULE 5	Secondary Storage device: Secondary Storage devices, Buffering of Blocks, Files on disk, Operations on files, File organization: Ordered files, Hashed files, Indexed files, Heap files, RAID organization.	**09hrs**
MODULE 6	Concurrency Control Techniques Concurrency Control Techniques, Recovery Techniques on databases, Transaction processing concepts, Database security and authorization. Introduction to Distributed databases, Data fragmentation, Replication and Allocation in distributed database, Query Processing in Databases.	**09hrs**

Text Books:

1. RemezElmasri and Shamkant B. Navathe, "Fundamentals of Database Systems", 5 th Edition, Pearson Education, 2007.

Reference Books:

1. Abrahamsi.Silberschatz,Henry.F.Korth,S.Sudarshan,"DatabaseSystemConcepts" 6th Edition, McGraw Hill, 2012.
2. C.J.Date,"Introductiontodatabasesystems",EightEdition,AddisonWesley

TITLE: Data Structures - LAB

PAPER CODE: L2.1	CREDITS : 1	NO OF HRS: 3hrs/week

SECTION-A

1. Use a recursive function to find the Fibonacci series.
2. Use pointers to find the length of a string and to concatenate two strings.
3. Use pointers to copy a string and to extract a substring from a given a string.
4. Use a recursive function for the towers of Hanoi with three discs.
5. Insert an integer into a given position in an array.
6. Deleting an integer from an array.
7. Write a program to create a linked list and to display it.
8. Write a program to sort N numbers using insertion sort.
9. Write a program to sort N numbers using selection sort.
10. Use a recursive function to find the Fibonacci series.
11. Use pointers to find the length of a string and to concatenate two strings.

SECTION-B

1. Inserting a node into a singly linked list.
2. Deleting a node from a singly linked list.
3. Inserting a node into a doubly linked list.
4. Deleting a node into a doubly linked list.
5. Pointer implementation of stacks.
6. Pointer implementation of queues.
7. Creating a binary search tree and traversing it using in order, preorder and post order.
8. Sort N numbers using merge sort.
9. Inserting a node into a singly linked list.
10. Deleting a node from a singly linked list.

TITLE: Mathematics–II & Statistics-II LAB

PAPER CODE:L2.2	CREDITS : 1	NO OF HRS: 3hrs/week

SECTION: A

1. Basic commands (Introduction).
2. Solve definite and indefinite integrals.
3. Obtain partial derivative for some standard functions.
4. Verify Euler's theorem.
5. Find extreme value of the function.
6. Find feasible region to linear programming problems.

SECTION: B

1. Probability distributions (Univariate and Bivariate probability distributions, Generation of observations from different distributions, evaluation of probabilities, etc..)
2. Construction of sampling distribution of sample mean and sample variance, Applications of Central Limit Theorem.
3. Identification of different hypotheses types and evaluation of probability of type I and type II errors and powers of tests (Discrete and Continuous distributions)
4. Tests concerning population mean and equality of two population means.
5. Tests concerning population proportion and equality of two population proportions.
6. Tests concerning population variance and equality of two population variances.
7. Chi-Square test for goodness of fit and independence of attributes.
8. Analysis of variance for a one way classified data.
9. Estimation of parameters by the methods of maximum likelihood and method of moments. Interval estimation.

SEMESTER III

III SEMESTER							
Part	**Paper**		**Hours/week**	**Marks**			**Credit**
	Code	**Title**		**IA**	**Exam**	**Total**	
Part 1	Language1	English	4	30	70	100	2
	Language2	Kan/San/Hin/Japanese	4	30	70	100	2
Part 2	B(DS)3.1	Statistical Inference	4	30	70	100	4
	B(DS)3.2	Analysis and Design of Algorithms	4	30	70	100	4
	B(DS)3.3	Python	4	30	70	100	4
	L3.1	Statistics for Data Science (SAS/SPSS) Lab	3	15	35	50	1
	L3.2	Python Lab	3	15	35	50	1
	L3.3	Analysis and Design of Algorithms LAB	3	15	35	50	1
Part 3		Open Elective	2	15	35	50	1
Total Marks and credits			**31**	**210**	**490**	**700**	**20**

TITLE: Statistics Inference

PAPER CODE: B(DS)3.1	CREDITS : 4	TOTAL NO OF HRS: 52

Objectives:

- ✓ This paper will help students to have a thorough knowledge of descriptive basic statistics.
- ✓ This course will help students to develop a deeper understanding of the basis underlying probability distributions and modern statistical inference and equip them with a statistical tool kit which will enable them to apply the knowledge and skills to real world tasks.
- ✓ Students will be able to analyze the difference among group means in a sample.

MODULE 1	**Estimation** Point estimation: Estimator, Estimate, Unbiasedness, Consistency, Sufficiency, Methods of estimation (MLE and MME). Interval estimation: Confidence interval, Confidence coefficient, Confidence limits, One-sided and two-sided confidence internals, Confidence intervals for the mean, Difference between means, Variance, Ratio of variances, Proportions and difference between proportions for normal population(s).	**15hrs**
MODULE 2	**Statistical Hypothesis** Statistical hypotheses – Null and alternative, Simple and composite hypotheses, One-sided and two-sided, Critical and acceptance regions, Type – I and Type – II errors, Level of significance, p-value.	**10hrs**
MODULE 3	**Tests of significance** Tests of significance of a population mean, Difference between means, Variance and difference between variances, Proportion and difference between proportions, Test for goodness of fit and independence of attributes, Relations between test of hypothesis and confidence interval.	**12 hrs**
MODULE 4	**ANOVA** Analysis of variance (ANOVA) – Introduction, Logic and overview, Formulation, Decisions. Foundation of ANOVA, Linear model for ANOVA, Test statistic (Mean squares) and rejection rule for ANOVA, Two-way ANOVA, Linear model for two-way ANOVA, Hypothesis and test statistics for two-way ANOVA.	**15 hrs**

Text Books:

1. Hogg & Tanis, Probability & Statistical Inference –Sixth Edition, Pearson Education.
2. Ross S.M., Introduction to Probability and Statistics, John Wiley and Sons.
3. P. Mukhopadhyay, (1996), Mathematical Statistics, Calcutta Publishing House.
4. Irwin Mille, Maryless Miller, Mathematical Statistics with Applications, Seventh Edition, Pearson Education.
5. Y.P.Aggarwal, Statistical Methods, Concepts, Applications and Computation.
6. Prem S. Mann, Introductory Statistics, FourthEdition.
7. Bhattacharya and N.L.Johnson (1986), Statistical concepts, John Wiley.
8. B.L. Agarwal, Basic Statistics (2009), New Age Publishers.
9. Gupta and Kapoor, Fundamentals of Mathematical Statistics, Sultan Chand and Sons.
10. Walpolw, Myers, Probability and Statistics for Engineers and Scientists, Eighth Edition, Pearson Education .
11. S. Sundararajan, Monograph on Statistics and Probability. (No Publication).
12. Dr. B.S. Grewal, Higher Engineering Mathematics, 40th Edition, Khanna Publishers.
13. Harry Frank, Steven C. Althoen, Statistics Concepts and Applications, Cambridge University Press.
14. Murray R. Spiegel, Larry J. Stephens, Statistics, Third Edition, Schaum's Outlines.
15. C.M. Chikkodi, B.G. Satyaprasad, B.Com Business Statistics, Himalaya Publishing House.

TITLE: Design and Analysis of Algorithms

PAPER CODE: B(DS)3.2	CREDITS : 4	TOTAL NO OF HRS: 52

Objectives:

- ✓ This course aims to introduce the classic algorithms in various domains, and techniques for designing efficient algorithms.

MODULE 1	**Introduction to Analysis and Design of Algorithms** A simple example of Design, Insertion sort, pseudo code for insertion sort, analysis of time complexity, Asymptotic notations and time complexity and writing efficient programs (by considering some small programs). Harner's method of evaluating a polynomial at a given point, finding maximum and minimum for a given set of numbers, straight max, straight min, combinations for max and min. Analysis of linear and binary search algorithms.	**09hrs**
MODULE 2	**Divide and Conquer Algorithms** Divide and conquer algorithms, Sorting, multiplication of two long integers, Stassen's matrix multiplication	**07hrs**
MODULE 3	**The Greedy Method.** Greedy approach, optimum scheduling, fractional Knapsack problem, minimum spanning trees, single source shortest path problem.	**09 hrs**
MODULE 4	**Dynamic Programming** Dynamic programming, Design and analysis, Travelling salesman problem, optimal parameterization for product of a sequence of matrices.	**09 hrs**
MODULE 5	**Back Tracking and Branch and Bound** Back tracking and Branch and bound methods, least cost method, 4-queens problem using back tracking, travelling salesman problem using branch and bound method.	**09 hrs**
MODULE 6	**Lower Bound Theory** Introduction, Comparison tree, Order searching sorting binary insertion sort, introduction to NP-Hard and NP-hard and NP-Complete, deterministic and Non deterministic algorithms non deterministic algorithm for sorting , Halting problem, cooks' theorem nodes cover decision problem	**09 hrs**

Text Books:

1. Design & Analysis Of Algorithms by S Srikanth, Published by Skyward Publishers
2. A.M Padma Reddy, Seventh revised edition February 2014, Sri Nandi Publication

Reference Books:

1. The Design and Analysis of Computer Algorithms by Aho, Hopcroft and Ullman.
2. Fundamental Algorithms: The art of Computer programming (Vol I) by D.E. Knuth.

TITLE: Python Programming

PAPER CODE: B(DS)3.3	CREDITS : 4	TOTAL NO OF HRS: 52

Objectives:

- ✓The course is designed to provide Basic knowledge of Python.
- ✓Python programming is intended for Software development and coding in software Industry.
- ✓Python is a language with a simple syntax, and a powerful set of libraries. It is an interpreted language, with a rich programming environment, including a robust debugger and profiler. While it is easy for beginners to learn, it is widely used in many scientific areas for data exploration.
- ✓This course is an introduction to the Python programming language for students without prior programming experience.

MODULE 1	Teaching Hours RBT Levels Introduction to Computers, Programs, and Python Elementary Programming, History of Python, Basic Features of Python ,Mathematical Functions, Strings, and Objects	**09hrs**
MODULE 2	Creating Python Programs, Selections, Loops, Functions. Programming examples	**09hrs**
MODULE 3	Functional programming, Objects and Classes, More on Strings and Special Methods, GUI Programming Using Tkinter, Programming examples	**12 hrs**
MODULE 4	Lists, Multidimensional Lists, Object Oriented Programming, Inheritance and Polymorphism, Programming examples	**08 hrs**
MODULE 5	Files: Files and Exception handling, tuples, sets and dictionaries, recursion, programming examples.	**14 hrs**

Text Books:

1. Y. Daniel Liang, "Introduction to Programming Using Python", Pearson, ISBN:978-0- 13-274718-9, 2013

2.Exploring Python, Timothy A. Budd, Indian edition, McGraw Hill education, ISBN-13: 978-0-07-132122-8

Reference Books:

1. Kenneth A. Lambert , B.L Juneja , "Fundamentals of Python Programming", Cengage Learning,ISBN:978- 81-315-2903-4, 2015

2. Charles Dierbach."Introduction to Computer Science Using Python: Computational Problem-Solving Focus", Wiley,ISBN:978-81-265-5601-4,2015

3. Allen B.Downey,"Think Python",O'Reilly,First Edition,2012,ISBN:978-93-5023- 863-9

TITLE: Design and Analysis of Algorithms LAB

PAPER CODE: L3.1	CREDITS : 1	NO OF HRS: 3hrs/week

1. Sort a given set of elements using the Quicksort method and determine the time required to sort the elements. Repeat the experiment for different values of n, the number of elements in the list to be sorted and plot a graph of the time taken versus n. The elements can be read from a file or can be generated using the random number generator.
2. C program to evaluate a given polynomial by reading its coefficients in an array.
3. Write a program to Compute the transitive closure of a given directed graph using Warshall's algorithm.
4. Write a program to solve knapsack problem using greedy method.
5. From a given vertex in a weighted connected graph, find shortest paths to other vertices using Dijkstra's algorithm.
6. Find Minimum Cost Spanning Tree of a given undirected graph using Kruskal's algorithm.
7. Print all the nodes reachable from a given starting node in a digraph using BFS method.
8. Check whether a given graph is connected or not using DFS method
9. Write a program to solve sum of sub set using backtracking.
10. Find Minimum Cost Spanning Tree of a given undirected graph using Prim's algorithm.
11. Implement N Queen's problem using Back Tracking.
12. Implement graph coloring using Back Tracking.

TITLE: Python Programming Lab

PAPER CODE: L3.2	CREDITS : 1	NO OF HRS: 3hrs/week

1. Program to demonstrate mathematical functions.
2. Program to calculate Body mass Index by accepting height and weight.
3. Program to demonstrate Bank transactions using class and objects.
4. Program to generate prime numbers and calculate CPU time using time module.
5. Program to generate different permutations of a given String using functions.
6. Program to demonstrate format specifiers of python by calculating interest and Principle amount for 'n' number of years.
7. Program to sort given numbers using selection Sort.
8. Program to convert temperature to Fahrenheit and vice versa using functions.
9. Program to find different areas of shapes using functions.
10. Program to find the occurrence of Character in a given file.
11. Program to generate Login Page UI using Tkinter.
12. Program to accept data from a Excel Sheet of temperature database and calculate the maximum and minimum temperature recorded using pandas.
13. Program to demonstrate list methods.
14. Program to demonstrate String methods in python.

SEMESTER IV

IV SEMESTER							
Part	**Paper**		**Hours/week**	**Marks**			**Credit**
	Code	**Title**		**IA**	**Exam**	**Total**	
Part 1	Language1	English	4	30	70	100	2
	Language2	Kan/San/Hin/Japanese	4	30	70	100	2
Part 2	B(DS)4.1	Machine Learning - I	4	30	70	100	4
	B(DS)4.2	Data Mining	4	30	70	100	4
	B(DS)4.3	Web Technologies	4	30	70	100	4
	L4.1	Tableaux (Data Visualization)	3	15	35	50	1
	L4.2	Machine Learning Lab	3	15	35	50	1
	L4.3	Web Technologies Lab	3	15	35	50	1
Part 3		Open ELECTIVE	2	15	35	50	1
Total Marks and credits			**31**	**210**	**490**	**700**	**20**

TITLE: MACHINE LEARNING I

PAPER CODE: B(DS)4.1	CREDITS : 4	TOTAL NO OF HRS: 52

Objectives:

- ✓ This course will serve as a comprehensive introduction to various topics in machine learning.
- ✓ At the end of course student be able to design and implement machine learning solutions to classification, regression and clustering problems.
- ✓ It evaluates and interpret the results of algorithms.

MODULE 1	Introduction ,What is Machine Learning?,Supervised Learning,Unsupervised Learning,Linear Regression with One Variable Model Representation,Cost Function,Gradient Descent method for linear egression. A review of Linear Algebra.	**09hrs**
MODULE 2	Linear Regression with Multiple Variables, Gradient Descent for Multiple Variables, Octave tutorial,Features and Polynomial Regression, Normal Equation	**08hrs**
MODULE 3	Logistic Regression,Classification, Hypothesis Representation, Decision Boundary,Cost Function, Simplified Cost Function and Gradient Descent,Advanced Optimization	**08hrs**
MODULE 4	Neural Networks: Representation, Non-linear Hypothesis,Neurons and the Brain,Model Representation, Examples,Multi-class Classification, Multi-class Classification and Neural Networks	**09hrs**
MODULE 5	Neural Networks: Learning, Backpropagation Algorithm, Gradient Checking, Random Initialization, Application case study, Neural Network Learning	**09hrs**
MODULE 6	Applying Machine Learning in Practice, Evaluating a Hypothesis , Model Selection and Train/Validation/Test Sets ,Bias, Variance ,Regularization and Bias/Variance ,Learning Curves	**09hrs**

Text Books:

1. Introduction to machine learning: Nils J Nilsson ,Robotics Laboratory Stanford University.
2. Pattern recognition and machine learning by Chirstopher Bishop , Springer 2006.
3. Understanding of machine learning from theory to algorithm: Shai Shalev – Shwartz, Shai Ben-David, Cambridge university.

Reference Books:

1. Fundamentals of Neural networks :architecture , algorithm and applications by Lauren Fausette, Pearson edition.

TITLE: Data Mining

PAPER CODE: B(DS)4.2	CREDITS : 4	TOTAL NO OF HRS: 52

Objectives:

- ✓ Interpret the contribution of data warehousing and data mining to the decision-support level of organizations.
- ✓ Evaluate different models used for OLAP and data preprocessing.
- ✓ Categorize and carefully differentiate between situations for applying different data-mining techniques: frequent pattern mining, association, correlation, classification, prediction, and cluster and outlier analysis.
- ✓ Design and implement systems for data mining.
- ✓ Evaluate the performance of different data-mining algorithms.
- ✓ Propose data-mining solutions for different applications.

MODULE 1	Introduction to Data Warehousing and Data Mining: Component and Processes, ETL, Data Mart, Decision Support system, Executive Information system. What is Data Mining? Motivating Challenges; The origins of data mining, Data Mining Tasks.	**08hrs**
MODULE 2	Data: Types of Data; Data Quality; Data Preprocessing; Measures of Similarity and Dissimilarity. Exploring Data: OLAP, Multidimensional Data Analysis, Data cube model,Visualization.	**08hrs**
MODULE 3	Classification: Preliminaries; General approach to solving a classification problem, Decision tree induction,ID3,CD4,CART Algorithms, Rule-based classifier; Nearest- neighbor classifier.	**09hrs**
MODULE 4	Association Analysis: Problem Definition, Frequent Item set generation; Rule Generation , Compact representation of frequent item sets, Alternative methods for generating frequent item sets. FP-Growth algorithm, Evaluation of association patterns, Effect of skewed support distribution, Sequential patterns.	**09hrs**
MODULE 5	Cluster Analysis: Overview, K-means, Agglomerative hierarchical DBSCAN, Overview of Cluster Evaluation.	**09hrs**
MODULE 6	Multidimensional analysis and descriptive mining of complex data objects; Spatial data mining, Multimedia data mining; Text mining. Applications: Data mining applications, Additional themes on Data mining; Social impact of Data mining; Trends in Data mining.	**09hrs**

Text Books:

1. Pang-Ning Tan, Michael Steinbach, Vipin Kumar: Introduction to Data Mining , Pearson Education.
2. Jiawei Han and Micheline Kamber: Data Mining – Concepts and Techniques, 3nd Edition, Morgan Kaufmann.

Reference Books:

1. K.P.Soman, Shyam Diwakar, V.Ajay: Insight into Data Mining – Theory and Practice, PHI.

TITLE: Web Technology

PAPER CODE:B(DS)4.3	CREDITS : 4	TOTAL NO OF HRS: 52

Objectives:

- ✓ Students should learn to develop object-oriented programs using C#.
- ✓ Be able to develop window forms, web forms and GUI based programs.
- ✓ Students will gain the skills and project based experience needed for entry into web application and windows applications.

MODULE 1	**Introduction to .Net Framework and C#:** The .Net Programming Frame work, .Net Languages, Common Language Run Time, The .Net Class Library Necessity of C#, Evolution of C#, Characteristics of C#, Applications, Structure of C# program, Name spaces, providing interactive inputs, multiple main methods, C# tokens, literals, variables, data types, value types, reference types, Boxing and Unboxing, for-each statement, Methods in C#, Handling Arrays.	**08hrs**
MODULE 2	**Classes and Objects:** Defining a class, Adding Variables, Adding Methods, member access modifiers, creating objects, accessing class members, static members and static constructors, constant members and read-only members, properties, indexers, Delegates and Events.	**08hrs**
MODULE 3	**Data Access with .NET.** ADO.NET overview, Using database connections, commands, The data reader, the dataset class, populating dataset class with a data adapter. The DataGridView Control, DataGridView Class Hierarchy, Data binding.	**09hrs**
MODULE 4	**Developing ASP.NET Application and Web Controls** ASP.NET Application, Code behind model, The Global. Asax application File, Understanding ASP.NET Classes, Web form Fundamentals. Basic Web control classes, Auto Post back and Web control Events, Assessing Web controls Using Visual Studio .NET.	**09hrs**
MODULE 5	**Validation and Rich Controls and State management.** Validation Controls, Validation Process, Validation Classes, Server side Validation Classes, Manual Validation, Understanding Regular Expression, Custom Validation, View State, TransferringInformation,CustomCookies,SessionState,ApplicationState.	**09hrs**

MODULE 6	**Master page , Themes, WCF:** Creating master page, simple master page, nested master page, expanding themes, creating themes, applying themes at runtime, features of wcf , routing services, default configuration, creating and using web services, creating and using wcf services.	**09hrs**

Text Books:

1. ProgramminginC#ByEBalagurusamy,FourthEdition(TataMcGrawHillPublications)
2. Comdex.NetProgrammingKit,VikasGupta, Fourtheditiondreamtechpuplication, Asp.netCompleteReference,MacDonald,TataMcGrawHillPublications

Reference Books:

1. ProfessionalC#2005byChristianNagelandOthers(WroxPublications).
2. ASP.NET3.5Unleashed,byStephenWaltherSAMSPublishing.
3. Microsoft ASP.NET and AJAX: Architecting Web Applications, by Dino Esposito Microsoft Press.

TITLE: Web Technology Lab

PAPER CODE: L4.3	CREDITS : 1	NO OF HRS: 3hrs/week

PART A: C#

1. Write a C# program toaccept students register number,nameand 3 subjects marks and perform the following.

a) Displayallstudentdetailswithtotalmarks. b)Display student details whoscored highest marks c)Displayallstudent namesinascendingorder.

Design a system using class called book with a suitable members.

2. A bookshop maintains the inventory of books that is being sold. The List includes book title, author name,price and stack position. The shop keeper Performs following activities

a) Add new books to inventory
b) Add stock to existing stock
c) Search a particular book
d)Displaystockdetails.

Design a system using class called inventory with a suitable members.

3. Write a program to create a class studentwith datamembers registernumber,name and three subject's marks.Set the values of the datamembers by using **indexers**.Calculate total marks ,averge and declare the class.Display all the information of the student with classs.

4. Write a Program to find sum and difference of two matrices using multicast delegates.

5. Write a Program to generate the first N even numbers and fobonacci numbers using events.

6. Create a database *Bank* in which create a table customer with fields *Account Number*, *Name*, *Account type* and *Total Balance*. Write a program to perform the following.

a) Display all the records of the customer table.
b) Display Account number and name of the customers whose account type is "SB"
c) Update the total balance by adding bonus amount Rs 500whose balance is greater than or equal to 10,000.

7. Create a database *Emp* in which create a table customer with fields *Employee Id*, *Name*, Designation and *Basic Salary*. Write a program to perform the following.

a) Display all the records of the Emp table.
b) Display number of records present in the table

Display the details of the employee who has highest basic salary.

8. Write a program to create a dataset company and perform the following

a) Add the table employee manually.

b) Retrieve the table Department from physical database and store in the Dataset.

c) Display the all contents of the company dataset.

PART B: ASP.NET

9. Create Student feedback form about courses and store the details in a database and display feedback details in DataGridView control.

10. Write a program containing the following controls: ListBox, Button, Image and Label.

The listbox is used to list products available in a store. When the user clicks the button. respective image will display on Image control and the cost of the selected product will be displayed on the label control**.**

11. Create a Login user page by using ***Login*** Control. If the login is successful display user name and password in another page. If the user attempts login three times block the login control.

12. Create a web page with textboxes for customer name, meter number, current reading and previous reading. Put required field validator and Compare validators. Calculate units consumed and total amount and display the same in another page.

SEMESTER V

V SEMESTER							
Part	**Paper**		**Hours/week**	**Marks**			**Credit**
	Code	**Title**		**IA**	**Exam**	**Total**	
Part 2	B(DS)5.1	Machine Learning - II	4	30	70	100	4
	B(DS)5.2	Natural Language Processing	4	30	70	100	5
	B(DS)5.3	Cloud Computing	4	30	70	100	5
	B(DS)5.4	Big Data Analytics	4	30	70	100	4
	B(DS)5.5	Applications of Data Science	4	30	70	100	5
	L5.1	Machine LearningLab	3	15	35	50	1
	L5.2	Big Data Analytics Lab	3	15	35	50	1
	L5.3	Mini Project	6	30	70	100	2
Total Marks and credits			**32**	**210**	**490**	**700**	**27**

TITLE: MACHINE LEARNING II

PAPER CODE:B(DS)5.1	CREDITS : 4	TOTAL NO OF HRS: 52

Objectives:

- ✓ To design and implement various machine learning algorithm in the range of real world applications.
- ✓ To understand neural implementations of attention mechanism and how modular components can be combined to build NLP system.
- ✓ To derive and implement and optimization techniques.

MODULE 1	Machine Learning System Design ,Performance of a machine learning system with multiple parts,Managing skewed data,Error Analysis, Error Metrics for Skewed Classes,Various Trade Offs	**8hrs**
MODULE 2	Support Vector Machines (SVM),Idea behind SVMs, Use in practice,Mathematics Behind Large Margin Classification,Kernels ,Programming SVM	**8hrs**
MODULE 3	Unsupervised Learning,Introduction,K-Means Algorithm,Optimization Objective,Random Initialization,Dimensionality Reduction,Principal Components Analysis (PCA),PCA, data compression, visualizations of complex datasets. PCA and K-Means Clustering,Casestudy	**9hrs**
MODULE 4	Anomaly Detection ,Introduction ,Gaussian Distribution,Developing and Evaluating an Anomaly Detection System,Anomaly Detection and Supervised Learning,Multivariate Gaussian Distribution,Anomaly Detection using the Multivariate Gaussian Distribution,Recommender Systems,Introduction,Content Based Recommendations,Collaborative Filtering Algorithm,Vectorization - Low Rank Matrix Factorization,Mean Normalization,Anomaly Detection and Recommender Systems	**9hrs**
MODULE 5	Large Scale Machine Learning,Introduction,Stochastic Gradient Descent,Mini-Batch Gradient Descent,Online Learning,Map Reduce and Data Parallelism	**9hrs**
MODULE 6	Application Example,Discuss a case study	**9hrs**

Text Books:

1. Machine learning A Probabilistic perspective by Kevin Murphy.
2. Python machine learning: Dr. Randal S olson.
3. Pattern classification 2nd edition Richard duda, Peter Hart David Stork

TITLE: Natural LanguageProcessing

PAPER CODE: B(DS)5.2	CREDITS : 5	TOTAL NO OF HRS: 52

Objectives:

- ✓ To understand how key concepts from NLP are used to describe and analyze the language.
- ✓ Understanding semantics and pragmatics of English language processing.
- ✓ It will focus on the computational properties of Natural languages and algorithm use to process them, as well as the match between grammar formalisms and the linguistic data that needs to be covered.

MODULE 1	Mathematics for ML and NLP:Probability review - random variables, axioms of probability, joint distribution, conditional probability, review of normal (and other) distributions, sum and product rules of probability, independent variables, **expectation maximization** (the most important concept in ML), bias variance,Linear algebra review - matrix operations, representing things as vectors, rank of a matrix	**9hrs**
MODULE 2	Naive Bayes theorem, SVMs, linear and logistic regression,Assignments in math and something simple like a spam/not spam classifier. Extra work: Intro to NLTK, scikit- learn, numpy, scipy and how to use these tools Most basic form of NLP - regular expressions and how to write them Language modeling - given a sequence of words, what is the probability of this sequence occurring in a document, n-grams, smoothing and data sparsity, Linguistics - parts of speech, lemmatization, stemming, stripping punctuation and other forms of data cleaning, tokenization (problems of how to tokenize e.g. tokenization in Chinese is different from English) State machines and sequence modeling Tf-idf, word-document frequencies.	**9hrs**
MODULE 3	Intro to perceptrons and feedforward networks. Generative and discriminative models Backpropagation algorithms Hidden Markov Models (HMM) Forward algorithm, backward algorithm, forward-backward, Viterbi algorithm.	**8hrs**
MODULE 4	Basics of ML predictions - training, testing and validation - a model can only recognize labels it's seen before, difficulties of collecting and cleaning data. Named entity recognition (NER) (also cover pitfalls and problems), examples of how understanding language is hard, even for humans Dependency parsing and understanding relationships between words ("I saw a boy with a bicycle" - did you see a boy who had a bicycle or did you see a boy and an unrelated bicycle?) Context free grammars and syntax in language (Chomsky hierarchy, CKY algorithms)	**9hrs**

MODULE 5	Maximum entropy (log-linear) classifiers Review week Application project of the above - i.e. build a maxent/HMM/other classifier for NER	**8hrs**
MODULE 6	Word embeddings and how to create them - bag-of-words, dictionary representations, tf-idf, clustering and similarity Existing word embedding datasets like word2vec	**9hrs**

Text Books:

1. Natural Language understanding: james F Allen, 2nd edition.
2. Introduction to soft computing: Neuro-fuzzy and generic algorithm by Samir roy and udit chakroborty.
3. Data mining concepts and techniques 3rd edition Jiaweihan m kamber Jian pie

TITLE: Cloud Computing

PAPER CODE: B(DS)5.3	CREDITS : 5	TOTAL NO OF HRS: 52

Objectives:

- ✓ Understand various basic concepts related to cloud computing. Technologies.
- ✓ Understand the architecture and concept of different cloud models: IaaS, PaaS, SaaS.
- ✓ Understand big data analysis tools and techniques.
- ✓ Understand the underlying principle of cloud virtualization, cloud storage, data management and data visualization.
- ✓ Understand different cloud programming platforms and tools.

MODULE 1	**Understanding Cloud** Origin and influences, A brief History, Definitions, Business Drivers, Technology Innovations , Clustering Grid Computing, Virtualization, Technology Innovations vs. Enabling Technologies, Roles and Boundaries , Cloud Consumer, Cloud Service Owner, Cloud Characteristics , On-Demand Usage , Ubiquitous Access Multitenancy (and Resource Pooling) , Elasticity , Measured Usage , Resiliency	**8hrs**
MODULE 2	**Cloud Delivery and cloud deployment models** Cloud Delivery Models, Infrastructure-as-a-Service (IaaS) , Platform-as-a-Service (PaaS), Software-as-a-Service (SaaS), Comparing Cloud Delivery Models , Combining Cloud Delivery Models , *IaaS + PaaS* , *IaaS + PaaS + SaaS* , Cloud Deployment Models . Public Clouds , Community Clouds, Private Clouds , Hybrid Clouds , Other Cloud Deployment Models	**8hrs**
MODULE 3	**Cloud Models** Introduction, Storage as a service, Amazon storage services, Compute as a service Amazon elastic compute cloud(EC2) , Cloud System matrix, Platform as Service, Windoes Azure, Google Apps Engine, Amazon Web services, Software as a Service CRM as a service, sales force.com	**9hrs**
MODULE 4	**Data Center** Introduction to Data center, Virtualization, Standardization and modularity, Automation, Remote operation and management, Data center Security and facilities, Computing hardware, storage hardware, Network hardware, LAN fabric , SAN fabric, NAS gateways.	**9hrs**
MODULE 5	**Cloud Virtualization Technologies** Server Virtualization, Hypervisor based Virtualization, Hardware support Virtualization, VMware Virtualization software, Storage Virtualization, Hardware independence, Server Consolidation, Resource replication, Virtualization Management, Hypervisor clustering architecture.	**9hrs**

MODULE 6	**Using the Mobile Cloud** Defining Mobile Market, Connecting to the cloud, Adopting mobile cloud application, Smart phones with the Cloud, Android, Apple iPhone, Black berry, Symbian, Windows mobile, Mobile web service , Mobile interoperability, Location awareness, Push Service, Using SMS, Defining WAP and other Protocol, Performing Synchronization	**9hrs**

Text Books:

1. Cloud Computing: Concepts, Technology & Architecture **By Thomas Erl, Ricardo Puttini, Zaigham Mahmood, Publication : Prentice Hall 2013(4th Edition)**

2. Moving to Cloud by Dinkar Sitaram, Geetha Manjunath, Publication: Syngress Elsevier Inc, 2014(2nd Edition)

3. Cloud Computing Second Edition by Dr Kumar Saurabh, Publication Willy INDAI (2013)

1. Cloud Computing Bible by Barrie Sosinsky, Publisher Willy INDAI (2014)

TITLE: BIG DATA ANALYTICS

PAPER CODE:B(DS)	CREDITS :4	TOTAL NO OF HRS: 52

Objectives:

- ✓ Introduce students the concept and challenge of big data (3 V's: volume, velocity, and variety). Teach students in applying skills and tools to manage and analyze the big data.

MODULE 1	**Preparatory:** Data Science landscape, relevance and importance of data analytics, Data sources: Social data - from organizations like WHO and social sites like face book. Government data - like data.gov.in, Data from own organization, Data formats: Structured, Semi-structured, Unstructured , Excel for presentation and simple visualization of structured data. Raw and Processed Data, Components of Tidy Data, Downloading Files, Reading Local Files, Reading Excel Files, Reading XML, Reading JSON, Reading from MySQL, Reading from HDF5, Reading from The Web, Reading From APIs.	**08hrs**
MODULE 2	**Data preparation / Mugging:** Subsetting and Sorting, Summarizing Data, Handling missing values, Creating New Variables, Reshaping Data, Merging Data.	**08hrs**
MODULE 3	**Data Exploration:** Exploratory Graphs	**09hrs**
MODULE 4	**Data Modelling:** Data grouping, frequency, and aggregation, Handling missing data, Text manipulation and format conversion, Assertions and logical operations	**09hrs**
MODULE 5	**Analysis**: Mathematical functions, Sampling , Relationship between variables, Rank and percentile Time series analysis, Descriptive statistical measures, Confidence level, Analysis of variance, Correlation Covariance, Regression, Moving average	**09hrs**
MODULE 6	**Visualisation** Comparison among items, Comparison over time, Relationship - two variables and three variables, Distribution - histogram, line chart, scatter chart, 3D area chart, Composition - static and changing over time	**09hrs**

Text Books:

1. Jake VanderPlas, Python Data Science Handbook: Essential Tools for working with Data , O'Reilly, 2017
2. W Mckinney, Python for Data Analysis, O'Reilly, 2013

Reference Books:

1. Murtaza Haider, Getting Started with Data Science, IBM Press, 2015
2. Davy Cielen, Introducing Data Science: Big Data, Machine Learning, and More, Manning, 2016

TITLE: BIG DATA ANALYTICAL LAB

PAPER CODE:L5.1	CREDITS : 1	NO OF HRS: 3hrs/week

In "1.2 Two_Novels.ipynb" discussed in the class, you found that Huck's name is mentioned the least because the story is told in first person. Find counts of Huck as subject ('I') and object ('me') and add the plots to those of Jim, Tom,
and Huck.

In the two novels, count the number of occurrences of other subject and object occurrences: he, his, she, her, they, them, we, us. Plot the cumulative counts.
Look for patterns.

Draw similar plots for occurrences of names in Little Women. What pattern do you find? What do you infer from the plots?

For the two novels,

a) Count the number of sentences by chapter.
b) Lengths of chapters.
c) Average length of sentences by chapter.
d) Average length of words by chapter.

5. The following is the directory structure you now have:

/Data Analytics
    /Data
        /Notebooks

Add subdirectories to reflect the following:

/Data Analytics
    /Data
    /Notebooks
    /Pourakarmikas
        /PDF
        /TXT
        /REC

Programmatically download the PDF files in URL-PK.txt and save them at /Data Analytics/Pourakarmikas/PDF

The cost of conducting census 2011 was ₹2,200 crore. How can we benefit from this massive work product?

a) Download census data for the country. Understand the structure and data contents. Classify the data to various groups. What analyses can we do with the data?
b) Create a table with the names of 100 most populous cities of the country, and (their population, and population density) in 2001 and 2011.

Bangalore District.

16 Oct 2017

a) Read an electoral roll into a dataframe. Check its shape. Print the first 3 rows and the last two. What would be the choice index? Reindex the dataframe to what you identify as the best column to index.
b) Extract age column as a series. Find various statistics for the ages of voters.
c) Group by house and print the number of voters by the house.
d) Read KA_Age-Data.xlsx into a daraframe. You will need to cleanse the file a little to read the data conveniently.
e) Create appropriate index. Delete redundant columns. Rename columns as appropriate. Add 3 columns to show female ratio (women per 1000 men) - overall, urban, and rural.
f) Apply hierarchical index on section, house, and serial number of voters.
g) Install JupyterLab in your PC and check in which browsers it works.

25 Oct 2017

Read 'SSLC Midterm 2017-18.xlsx' into a dataframe.

a) add a column with average scores of each student.
b) find various statistics for the table.
c) what are your findings about the performance of the students and the teachers?

2. ../Data/Voters/voterREC/ has 227 voter lists for Jayanagar Constituency. The Constituency has 7 wards. From the 227 files, create 7 files - one per ward - by creating dataframes by part and concatenating by ward.

Tips:

- Read each file into a dataframe, find the ward of the part, and make a set of wards in the constituency.
- Create a dictionary with ward numbers as keys and empty dataframes as values. Thus, the dictionary will have one dataframe per ward.
- For each of the 227 voter lists, read and create a temporary DataFrame, find the ward number of voter list and refer to the DataFrame of the ward from wards dictionary, concat the temporary DataFrame to the ward DataFrame.
- Save each ward dataframe as excel file.

After an election, CEO publishes Form-20, giving the details of votes polled by booth by various candidates. This document does not have the count of voters per booth. The voters list gives the total voters. Extract necessary data from the two sources, create a dataframe by merging the two, add a column with calculated voter turnout percentage.

Tips:

- Make a dataframe with empty rows equal to the number of parts in the constituency with columns for part number and total voters.
- Voters with status '#','A', or 'O' are valid. Calculate the valid voters for

each part and assign the values to the appropriate row in the empty dataframe.

- From "../Data/Form_20/AC170_Polled.xlsx" make a dataframe with total votes polled per part.
- Merge the two dataframes.
- Create a new column with turnout%.
- Calculate quantiles for turnout%.

10 02 Nov 2017

a) 1From http://ceokarnataka.kar.nic.in/ClaimsObj.aspx download Form types 6, 7, 8, and 8A in spreadsheet format for Jayanagar constituency.
b) Reading the files, create dataframes with appropriate column names.
c) Remove unwanted rows.
d) Remove unwanted columns.
e) Create new columns where needed.
f) Attempt for the remaining constituencies of Bangalore.

11 Final data in various forms would be as follows: Form 6:

Part	Date	Name	Relative	Reln	Status	Reason
Form 7:						

Part	Serial	Name	Status	Reason
Form 8:				

Part	Serial	Name	Status
Form 8A:			

New Part	Name	Old Part	EPIC	Address	Status	Serial	Reason

Use 'apply' feature of dataframe to change contents of columns and also to create new columns

In all cases,

- Give one word status.
- Rreason would be applicable for rejections. For therest, state 'NA'
- Check contradictions like - approved for inclusion, maybe deleted

Take counts of inclusions, rejections, verification in progress for the 4 types of forms. What % of the total applications belong to each category of actions?

Compare the counts across the 28 constituencies. Do the results show any patterns? Can suggest some actions based on the results?

12 Nov 2017

claim_obj.db has 4 tables - form_6, form_7, form_8, and form_8A. The tables contain data extracted from the 'List of Claims and Objections' at CEO-KA website for the 28 constituencies of Bangalore for 01 Jan to 30 Oct 2017.

claim_obj-Jan_Oct_2017.xlsx has the set of results based on the data. Using data analytics tools, write scripts to create these tables based on the data in the tables.

Voter Enrolment Activities in Bangalore.docx discusses the findings.

16 Nov 2017

a) Form 7 for deletion has the columns: Constituency, Part, Serial, Name, Reason, Status, and Remarks
b) We want to check if the approved records are deleted.
c) To query on CEO-KA website for a record we need: district and EPIC number OR district, constituency name, name, sex, and relative's name
d) We have a copy of previous version of electoral rolls with the above columns.
e) Merge the data to get EPIC# for the deleted voters
f) The electoral roles tables doesn't have section address, but have only section numbers.
g) Section has several voters
h) The section table has sections numbers and section address.
i) for field work, we want the voters list which has a column for section address.
j) Merge the data to get the voter list with section address.
k) form 6 data does have part number, name and relative name.

from the available data, create a data frame to query CEO site for the newly added records.

VI SEMESTER							
Part	**Paper**		**Hours/week**	**Marks**			**Credit**
	Code	**Title**		**IA**	**Exam**	**Total**	
Project/Internship			32	210	490	700	24
Total Marks and credits			**32**	**210**	**490**	**700**	**24**