File Ref.No.24326/GA - IV - J2/2013/CU

UNIVERSITY OF CALICUT

Abstract

General and Academic - B.Voc Programme in Data Science under modified B.Voc Regulations 2014 - Scheme and Syllabus - Approved - Implemented w.e.f 2018 Admissions - Orders issued.

G & A - IV - J

U.O.No. 13256/2018/Admn Dated, Calicut University.P.O, 14.11.2018

Read:-1. U.O.No. 7404/2018/Admn dated 19.06.2018
2. Item No.1 in the minutes of the B.Voc Regulations Committee held on 13.09.2018
3. Request from Convenor, B.Voc Regulations Committee Dated 06.10.2018
4. Item No. 1 in the minutes of the meeting of Board of Studies in Statistics UG held on 01.11.2018
5. Remarks of Dean, Faculty of Science dated 05.11.2018

ORDER

Vide paper read as (1), the modified BVoc Regulations has been implemented and vide paper read as (2), B.Voc Regulation Committee decided to place the syllabi of new B.Voc Programmes which are sanctioned by UGC, in various colleges under University of Calicut, before various Boards of studies for approval.

The Convenor, B.Voc Regulations committee vide paper read as (3), pointed out that UGC has directed to start the newly sanctioned programmes in colleges by 15th October 2018 and hence requested to initiate urgent steps to approve the syllabi of newly sanctioned B.Voc Programmes at various colleges with a suggestion to submit the syllabi to the Chairmen of Boards of Studies concerned with a request to approve the syllabi in circulation with other Board members (as provided under CUFS 1976), which was approved by Vice Chancellor and the syllabus of B.Voc Programme in Data Science has been forwarded to the Chairman Board of Studies in Statistics UG accordingly.

Vide paper read as (4), Board of Studies in Statistics UG approved the syllabus for BVoc Programme in Data Science and the Dean, Faculty of Science vide paper read as (5), approved the resolution of the Board of Studies.

The Vice Chancellor, exercising the powers of Academic Council, approved the resolution of the Board of Studies, subject to ratification by Academic Council.

Sanction has, therefore, been accorded for the implementation of the Scheme and Syllabus of BVoc Programme in Data Science under modified BVoc Regulations 2014, in the University, w.e.f 2018 Admissions.

Orders are issued accordingly.

(Syllabus is herewith attached)

Biju George K

Assistant Registrar

To

All Affiliated Colleges offering the B.Voc Programme in Data Science

Copy to : PA to CE/JCE VII/EG I/JCE I/University Librarian/GA I F/SF/DF/FC

Forwarded / By Order

Section Officer

BACHELOR OF VOCATION (DATA SCIENCE)
PROGRAMME STRUCTURE

Sem No	Course No	Code	Paper	Credits/ hrs/ week	Marks		
					Internal	External	Total
1	1.1	GEC1EG01	Transactions Essential English Language Skills A01	4	20	80	100
	1.2	GEC1ML01 GEC1HD01	Malayalam Bhashayum Sahithyavum I MAL1A01 (2) Prose and One Act Plays A07 (3)	4	20	80	100
	1.3	GEC1MT01	Mathematics for Data Science I	4	20	80	100
	1.4	SDC1CS01	Programming in C	4	20	80	100
	1.5	SDC1ST01	Fundamentals of Statistical data analysis	5	20	80	100
	1.6	SDC1CS02(P)	Programming in C – Lab	4	20	80	100
	1.7	SDC1ST02(P)	Statistical Data Analysis Using MS Excel – Lab	5	20	80	100
2	2.1	GEC2EG02	Ways with Words A02	4	20	80	100
	2.2	GEC2ML02 GEC2HD02	Malayalam Bhashayum Sahithyavum II MAL2A02 (2) Poetry and Short Stories A09 (3)	4	20	80	100
	2.3	GEC2MT02	Mathematics for Data Science II	4	20	80	100
	2.4	SDC2CS03	Introduction to RDBMS and SQL	4	20	80	100
	2.5	SDC2ST03	Probability & Random variables	5	20	80	100
	2.6	SDC2CS04(P)	SQL – Lab	5	20	80	100
	2.7	SDC2ST04(Pr)	Mini Project	4	0	100	100
3	3.1	GEC3EG03	Writing for Academic and Professional Success A03	4	20	80	100

	3.2	GEC3MT03	Mathematics for Data Science III	4	20	80	100
	3.3	GEC3ST01	Optimization techniques	4	20	80	100
	3.4	SDC3ST06	Probability Distributions	4	20	80	100
	3.5	SDC3CS05	Python Programming	5	20	80	100
	3.6	SDC3CS06 (P)	Python – Lab	4	20	80	100
	3.7	SDC3ST07 (P)	Statistical Data Analysis Using SPSS – Lab	5	20	80	100
4	4.1	GEC4EG04	Zeitgeist: Readings on Society and Cultures A04	4	20	80	100
	4.2	GEC4ST02	Statistical simulation techniques	4	20	80	100
	4.3	SDC4CS07	Artificial Intelligence	4	20	80	100
	4.4	SDC4ST08	Statistical inference	4	20	80	100
	4.5	SDC4ST09	Applied Multivariate Techniques	5	20	80	100
	4.6	SDC4ST10 (P)	R Programming – Lab	5	20	80	100
	4.7	SDC4CS08 (Pr)	Mini Project	4	0	100	100
5	5.1	GEC5CS01	Data Science - Industrial perspective	4	20	80	100
	5.2	SDC5CS09	Data mining Techniques	4	20	80	100
	5.3	SDC5ST11	Design of experiments	4	20	80	100
	5.4	SDC5ST12	Categorical data Analysis	4	20	80	100
	5.5	SDC5ST13	Stochastic Modeling	5	20	80	100
	5.6	SDC5CS10	Machine Learning and Data Analytics	4	20	80	100
	5.7	SDC5ST14 (P)	Statistical Data Analysis Using SAS – Lab	5	20	80	100
6	6.1	SDC5CS11 (Pr	Internship and Project (900hrs)	30	0	100	100

NSQF Level – 7 **Qualification Packs:** SSC/Q0401 – Data Scientists
SSC/Q2601 – Analyst - Research

CONDITIONS FOR ADMISSIONS

ELIGIBILITY

- The admission to B Voc programme will be as per the rules and regulations of the University for UG admissions.
- Candidates who have passed (Eligible for Higher Studies) the HSE of the Kerala State Board of Higher Secondary Examination or any other recognized equivalent examination with Mathematics/ Statistics/ Computer Application as one of the subjects are eligible for getting admission to B Voc Data Science.
- Grace Marks may be awarded to a student for meritorious achievements in co-curricular activities such as Sports/Arts/ NSS/NCC.
- Preferred subjects and index mark calculations will be decided by the respective Board of Studies.

DIPLOMA HOLDERS

Diploma holders (after 10+2) in the parent courses, approved by the University, who satisfies eligibility criteria can be admitted to the higher diploma(3^{rd} semester) based on the availability of the seats and is under the sole discretion of the principal of the college/ B. Voc consortium.

Syllabus of common English courses (GEC1EG01, GEC2EG02, GEC3EG03 and GEC4EG04) and additional language courses are (GEC1ML01, GEC1HD01, GEC2ML02 and GEC2HD02) are same as that of the LRP programmes of University of Calicut.

Semester I: Mathematics for Data Science– I

Module 1: Sets and elements – Operations on sets – least upper bounds – Sequence of real numbers –Definition of sequence and sub sequence – Limit of a sequence - Convergent sequence – Bounded sequence –Monotone sequence – Operations on convergent sequence.

Module 2: Series of real numbers – Convergence and divergence – Series with non negative terms – Alternating series – Conditional convergence and absolute convergence – Tests for absolute convergence.

Module 3: Functions continuous at a point on the real line – The Derivative – Rolle's theorem - Mean value theorem – Taylor's theorem – Maclaurin theorem – simple problems.

Module 4: Riemann Integrability – Upperp and Lower sums – Upper and Lower integral – The Riemann integral – Riemann criterion for integrability – Fundamental theorem of calculus – Improper integral – simple problems. Laplace transform – Laplace transform of- Inverse Laplace transform to the above standard functions – Applications to ordinary differential equation.

Books for Study:

1. Bartle, R.G., &Shebert, Introduction to Real Analysis, Wiley Eastern & Sons,
2. Gold berg, R.R., Methods of Real Analysis, Oxford and IBH,1970.

Books for Reference:

1. Apostol, T.M., Mathematical Analysis, Narosa Publications, 1985.
2. Singaravelu, A., Allied Mathematics, A.R.S. Publications, 2014.
3. Vittal, P.R, Allied Mathematics, Margham Publications, 2015.

Semester 1: Programming in C

Module 1 : Introduction to C Programming, overview and importance of C, C Program Structure and Simple programs, Creation and Compilation of C Programs under Linux and Windows Platforms. Elements of C Language and Program constructs. Character Set,C Tokens, Keywords and Identifier, Constants, Variables, Data types, Variable declaration and assignment of values, Symbolic constant definition. C-Operators, Arithmetic operators, relational operators, and logical operators, assignment operators, increment and decrement operators, conditional operators, special operators, arithmetic expressions, evaluation of expressions, precedence of arithmetic operators, Type conversion in expressions, operator precedence and associativity, Mathematical Functions, I/O operations.

Module 2: Decision making, Branching and Looping. Decision making with IF statement, Simple IF statement, If. . .else statement, Nesting of If. . .else and else. . .if Ladder, Switch statement, Conditional operator, Go-to statement. Looping: While loop, Do-While, and For Loops, Nesting of loops, jumps in loop, skipping of loops.

Module 3: Array & Strings - One dimensional array, two dimensional array and multi-dimensional array, strings and string manipulation functions. The Concept of modularization and User defined functions-Multi-function Program, calling functions, various categories of functions, Nesting of functions and recursion, functions and arrays, scope and life-time of variables in functions, multi-file programs. Structures & Union structure definition , giving values to members, structure initialization, comparison of structure variables, arrays of structures, arrays within structures, structures within arrays, structures and functions, Unions, bit-fields.

Module 4: Pointers and Files: Understanding pointers, accessing the address of a variable, declaring and initializing pointers, accessing a variable through its pointer, pointer expressions, pointer and arrays, pointer and character string, pointers and functions, pointers and structures, pointer to pointer - dynamic memory allocation. Files : Defining, Opening and closing files - i/o operations on files - error handling on files random access of files command line operations. Preprocessor directives: Macro substitution directives - simple macros - macros with arguments - nesting of macros, Compiler control directives.

Text Books

1. Programming in ANSI C by E. Balaguruswami.

References

1. The C Programming Language by Brian W. Kernighan & Dennis M. Ritchie.
2. Let us C by Yashavant P. Kenetkar.
3. Byran Gotfried, Schaums Outline series- ‘Programming with C’.

Semester 1: Fundamentals of Statistical Data Analysis

Module 1: Data and statistics – Data, elements, variables, observations. Scales of measurement, Qualitative and quantitative data, Cross-sectional and time series data. Data sources, primary data, secondary data. Univariate, bivariate, multivariate data.

Module 2: Descriptive Statistics: Tabular and Graphical Methods - Frequency distribution, Relative frequency distribution, Percent frequency distribution, Cumulative frequency distribution, Cross tabulations. Bar graphs, Pie graphs, Dot plots, Histogram, Stem and Leaf plot, Box plot, Ogive, Scatter diagram.

Module 3: Descriptive Statistics: Numerical Methods - Measures of central tendency, mean, median, mode, percentiles, quartiles. Measures of variability, Range, Inter quartile range, Quartile deviation, Mean deviation, Variance, Standard deviation, Coefficient of variation.

Module 4: Measures of association between two variables: Covariance, Correlation, Rank correlation, Curve fitting - Principle of least squares - linear, nonlinear, exponential and growth curves, Simple linear Regression and Properties.

Books for Study:

1. Anderson, Sweeney and Williams (2012). Statistics for Business and Economics, South-Western, Cengage Learning, USA.
2. Gupta. S.C. &Kapoor,V.K. (2002) . Fundamentals of Mathematical Statistics , Sultan Chand & Sons Pvt. Ltd. New Delhi.

Books for Reference:

1. Goon A.M. Gupta. A.K. & Das Gupta, B (1987). Fundamentals of Statistics, Vol.2, World Press Pvt. Ltd., Calcutta.
2. Kapoor, J.N. & Saxena, H.C. (1976). Mathematical Statistics, Sultan Chand and Sons Pvt. Ltd, New Delhi.

Semester I: Programming in C (Lab)

Course Outline

1. Display the currency denomination of a given amount.
2. Find the sum of the digits and reverse of a number.
3. Find the factorial of a number.
4. Print N Fibonacci numbers
5. Print Armstrong numbers up to N.
6. Display multiplication table from 5 to 10.
7. Sort a list of numbers in ascending order
8. Matrix addition
9. Find the transpose of a Matrix
10. Display the count of positives, negatives and zeros in a set of N numbers.
11. To read and print the details of N students using structure.
12. Add two complex numbers using structure.
13. To check whether given string is palindrome or not.
14. Reverse a string using recursion.
15. To store and read from a text file.
16. Swapping of two numbers using function.
17. Find the sum and average of N elements in an array.
18. Display the prime numbers in a given range
19. Find the sum of the digits and reverse of a number using functions.
20. Find the factorial of number using recursion.
21. Find the number of vowels in a string
22. Read the marks of m students in n subjects and calculate the subject-wise and studentwise totals using structures
23. Swapping of two numbers using pointers.
24. Copy the contents of one file into another
25. Create a file with a set of numbers and write odd and even numbers into separate files

Semester I: Statistical Data Analysis Using MS Excel – Lab

Module 1: Introduction to MS Excel - MS Excel Options – Ribbon - Sheets - Saving Excel File as PDF, CSV and Older versions - Using Excel Shortcuts - Copy, Cut, Paste, Hide, Unhide, and Link the Data in Rows, Columns and Sheet - Using Paste Special Options - Formatting Cells, Rows, Columns and Sheets - Protecting & Unprotecting Cells, Rows, Columns and Sheets with or without Password - Page Layout and Printer Properties

Module 2: Functions: - Logical Functions - Date and Time Functions -Information Functions -Math and Trigonometry Functions - Statistical Functions - Text Functions - Charts:- Simple Bar Chart – Multiple Bar Chart – Subdivided Bar Chart – Pie Chart – Donut Chart - Line Chart – Histogram – Scatter Plot - Radar Chart – Bubble Chart – Bi-Axis chart – Plotting Density Function and Distribution Function.

Module 3: Vlookup, Hlookup, Index, Address, Match, Offset, Transpose - Conditional Formatting - Data Sorting and Filtering - Pivot Tables - Chart Templates – Adding Add-Ins in Excel - Solver – Goal Seek.

Module 4: Descriptive Statistics: Tabular and Graphical Methods, Measures of central tendency, mean, median, mode, percentiles, quartiles. Measures of variability, Range, Inter quartile range, Quartile deviation, Mean deviation, Variance, Standard deviation, Coefficient of variation. Covariance, Correlation, Rank correlation, Curve fitting - linear, nonlinear, exponential and growth curves.

Books for Study:

1. Microsoft Excel 2016 Step by Step by Curtis Frye
2. Microsoft Excel Functions & Formulas by Bernd Held

Books for Reference:

1. Excel Functions and Formulas Paperback by Bernd Held
2. Microsoft Excel 2010 Data Analysis and Business Modeling Paperback by Winston

Semester 2: Mathematics for data science II

Module I : Solution of Algebraic and Transcendental Equation- Introduction, Bisection Method, Method of false position , Iteration method , Newton-Raphson Method , Ramanujan's method , The Secant Method. Finite Differences, Forward differences, Backward differences, Central differences, Symbolic relations and separation of symbols, Differences of a polynomial.

Module II : Interpolation - Newton's formulae for intrapolation, Central difference interpolation formulae, Gauss' Central Difference Formulae, Interpolation with unevenly spaced points, Langrange's interpolation formula, Divided differences and their properties, Newton's General interpolation formula, Inverse interpolation Numerical Differentiation and Integration, Numerical differentiation (using Newton's forward and backward formulae), Numerical Integration, Trapizaoidal Rule, Simpson's 1/3-Rule, Simpson's 3/8-Rule

Module III : Matrices and Linear Systems of equations, Solution of Linear Systems – Direct Methods, Gauss elimination, Gauss-Jordan Method, Modification of Gauss method to compute the inverse, LU Decomposition, LU Decomposition from Gauss elimination, Solution of Linear Systems – Iterative methods, The eigen value problem, Eigen values of Symmetric Tridiazonal matrix

Module IV : Numerical Solutions of Ordinary Differential Equations, Solution by Taylor's series, Picard's method of successive approximations, Euler's method, Modified Euler's Method, Runge-Kutta method, Predictor-Corrector Methods, Adams-Moulton Method, Milne's method

References

1. S. Sankara Rao : Numerical Methods of Scientists and Engineer, 3rd ed., PHI.

2. F.B. Hidebrand : Introduction to Numerical Analysis, TMH.

3. J.B. Scarborough : Numerical Mathematical Analysis, Oxford and IBH.

Semester 2: Introduction to RDBMS and SQL

Module 1: Introduction to database systems, File Systems Versus a DBMS, View of data – Data abstraction, View levels, Data models, Instances and Schemas, Data Independence, Database languages, Database architecture, Database users , Database administrator, Role of DBA . The Entity – Relationship (ER) model - Entity sets, Relationship sets, Attributes, Constraints, Mapping Cardinalities, Keys, ER diagrams, Weak entity sets, Strong entity sets.

Module 2: Data Definition in SQL - Data types, Creation, Insertion, Viewing, Updation, Deletion of tables, modifying the structure of the tables, Renaming, Dropping of tables. Data Constraints – I/O constraints, Primary key, foreign key, unique key constraints, ALTER TABLE command.

Module 3: Database Manipulation in SQL - Computations done on table data - Select command, Logical operators, Range searching, Pattern matching, Grouping data from tables in S QL, GROUP BY, HAVING clauses, Joins – Joining multiple tables, Joining a table to itself. Views - Creation, Renaming the column of a view, destroys view, Granting and revoking permissions: Granting privileges, Object privileges, Revoking privileges

Module 4: Program with SQL - Data types: Using set and select commands, procedural flow, if, if /else, while, goto, global variables, Security - Locks, types of locks, levels of locks. Cursors - Working with cursors, Error Handling, Developing stored procedures, create, alter and drop, passing and returning data to stored procedures, using stored procedures within queries, building user defined functions, creating and calling a scalar function, implementing triggers, creating triggers , multiple trigger interaction.

References:

1. Fundamentals of Database Systems, Elmasri & Navathe, Pearson Education.
2. Database System Concepts Abraham Silberschatz, Henry F Korth,S.
3. Introduction to Database Systems, CJ Date, Addison Wesley

Semester 2: Probability & Random variables

Module 1: Introductory Notions of Probability- Random Experiments – Sample Space and Events. Axiomatic Approach to Probability – Addition Law – Problems in Axiomatic Approach.

Module 2: Combinatorics and Classical Probability Elements of Combinatorics. Classical definition of Probability. Problems in Classical approach.

Module 3: Conditional Probability – Occupancy Problems. Stochastic Independence and related concepts - Independence of events – Pair wise and Mutual Independence. Multiplication Law, Law of Total Probability, Baye's Theorem. Bernoulli Trials – Problems.

Module 4: Random Variables - Discrete and Continuous Random Variables; Probability mass function and Probability density function- Properties and examples. Cumulative distribution function – Properties and examples; Change of variable (Univariate case).

Books for Study:

1. Gupta, S.C. and Kapoor, V.K. (2002). Fundamentals of Mathematical Statistics. Sultan chand and Sons. New Delhi
2. Parzen, E.(1960). Modern Probability Theory. John Wiley & Sons, New York

Books for Reference:

1. Hogg, R.V. and Craig, A.T. (2002). Introduction to Mathematical Statistics. Pearson Education India

Semester 2: SQL Lab

Lab 1:

Create a database for an Employee management system of an ABC organization. The details about different tables are given below

Department:

Department name - Not NULL unique
Department number - Primary Key
ManagerId - Refers to employee-id of employee table.
Manager
date of joining - Not NULL.

Employee:

First name - Not NULL
Middle initials
Last name - Not NULL
Employee id - Primary Key
Date of Birth - Not NULL
Address
Gender - M or F
Salary - Range of 5000 to 25000
Date of Joining
Department number - Refers to Department Number of Department table.
Department location:
Department number - Refers to Department number of department table.
Department location - Not NULL.
Department number & Department location are combined Primary Key

Project:

Project name-Not NULL.
Project number-Primary Key.
Project location-Not NULL.
Department number-Refers to department number of Department table.
Works-on:
Employee-id - Not NULL refers to employee-id of employee table
Project number- Not NULL refers to Project number of Project table.
Hours - Not NULL.
Employee-id & Project numbers are combined primary key.

Now enter a few sets of meaningful data and answer the following queries.

1. List the department wise details of all the employees.
2. Find the list of projects.
3. Find out the list of employees working on a project.

Lab 2:

Create a database for an Employee management system of an ABC organization. The details about different tables are given below

Department:

Department name - Not NULL unique

Department number - Primary Key
ManagerId - Refers to employee-id of employee table.
Manager
date of joining - Not NULL.

Employee:

First name - Not NULL
Middle initials
Last name - Not NULL
Employee id - Primary Key
Date of Birth - Not NULL
Address
Gender - M or F
Salary - Range of 5000 to 25000
Date of Joining
Department number - Refers to Department Number of Department table.
Department location:
Department number - Refers to Department number of department table.
Department location - Not NULL.
Department number & Department location are combined Primary Key

Dependent:

Employee-id - Refer to employee table employee id field
Dependent name - Gender - M or F
Date of Birth - Not NULL
Relationship - Not NULL

Now enter a few sets of meaningful data and answer the following queries.
1. List the dependents of the employee whose employee id is 001
2. Find out all those departments that are located in more than one Location

Lab 3:

Develop a prototype database of the College library management system

Book Records:

Accession Number
ISBN Number

Books:

ISBN Number
Author
Publisher
Price

Members:

Member Id, Member Name, Maximum Number of books that can be issued, Maximum Number of days for which book can be issued

Book Issue:

Member Id, Accession Number, Issue Date, Return Date

You must create constraints, including referential integrity constraints, as appropriate. Pleasenote accession number is unique for each book. A book, which has no return date, can be considered as issued book. Enter suitable data into the tables.

Now answer the following:
1. Display the structure of the tables.
2. Using SELECT statement, write the queries for performing the following
(a) Get the list of all books (No need to find number of copies).
(b) Get the Accession number of the books which are available in the library.

Lab 4:

Develop a prototype database of the College library management system

Book Records:

Accession Number, ISBN Number

Books:

ISBN Number, Author, Publisher, Price

Members:

Member Id, Member Name, Maximum Number of books that can be issued, Maximum Number of days for which book can be issued

Book Issue:

Member Id, Accession Number, Issue Date, Return Date.

You must create constraints, including referential integrity constraints, as appropriate. Please note accession number is unique for each book. A book, which has no return date, can be considered as issued book. Enter suitable data into the tables.

Now answer the following:
(a) Get the list of all members.
(b) On return of a book by a member calculate the fine on that book.
(c) List of books issued on 01-Jan-2005

Lab 5:

Develop a prototype database of the College library management system

Book Records:

Accession Number, ISBN Number

Books:

ISBN Number, Author, Publisher, Price

Members:

Member Id, Member Name, Maximum Number of books that can be issued, Maximum Number of days for which book can be issued

Book Issue:

Member Id, Accession Number, Issue Date, Return Date

You must create constraints, including referential integrity constraints, as appropriate. Please note accession number is unique for each book. A book, which has no return date, can be considered as issued book. Enter suitable data into the tables.
Now answer the following:
1. Display the structure of the tables.
2. Using SELECT statement, write the queries for performing the following
(a) Get the list of all books having price greater than Rs. 500/-
(b) Get the list of members who did not have any book issued at any time.

Lab 6:

Develop a prototype database of the College library management system

Book Records:

Accession Number, ISBN Number

Books:

ISBN Number, Author, Publisher, Price

Members:

Member Id, Member Name, Maximum Number of books that can be issued, Maximum Number of days for which book can be issued

Book Issue:

Member Id, Accession Number, Issue Date, Return Date.

You must create constraints, including referential integrity constraints, as appropriate. Please note accession number is unique for each book. A book, which has no return date, can be considered as issued book. Enter suitable data into the tables.
Now answer the following:
1. Using SELECT statement, write the queries for performing the following
(a) Get the list of members who have not returned the book.
(b) the number of copies of each book (A book accession number would be different but ISBN number would be the same)

Lab 7:

Create the following table named customer with the following
Customer id Character 10
Name Character 25
Area Character 3
Phone Numeric 7

Insert the appropriate data into table and do the following.
1. Update Phone numbers of all customers to have a prefix as your city STD Code
2. Print the entire customer table
3. List the names of those customers who have e as second letter in their names.
4. Find out the Customer belonging to area abc
5. Delete record where area is NULL.

Lab 8:

Develop a prototype database of the College library management system

Book Records:

Accession Number, ISBN Number

Books:

ISBN Number, Author, Publisher, Price

Members:

Member Id, Member Name, Maximum Number of books that can be issued, Maximum Number of days for which book can be issued

Book Issue:

Member Id, Accession Number, Issue Date, Return Date

You must create constraints, including referential integrity constraints, as appropriate. Please note accession number is unique for each book. A book, which has no return date, can be considered as issued book.

You must create a view to know member name and name of the book issued to them, use any inbuilt function and operators like IN, ANY, ALL, EXISTS.

a. List the records of members who have not been issued any book using EXISTS operator.

b. List the members who have got issued at least one book (use IN /ANY operator).

c. List the books which have maximum Price using ALL operator.

d. Display Book Name, Member Name, Issue date of Book. Create a view of this query of the currently issued books.

Lab 9:

Write simple PL/SQL blocks for displaying whole numbers from 1 to 100, odd numbers from 1 to 100 and even numbers from 1 to 100

Lab 10:

Write simple PL/SQL blocks for displaying positive whole numbers up to a given number, odd numbers from 1 to a given number and even numbers from 2 to a given number

Semester 2: Mini Project

Objectives

On completion of this course, the student should be able to:

- Convert information to data and do basic analysis.
- Provide a solution for a real life situation.
- Get a chance to utilize and implement the skill acquired.

Semester 3: Mathematics for Data Science – III

Module 1: Matrix Algebra – Some special types of matrices – Determinants –Properties, Adjoint and Inverse of a matrix – solution of linear equations – homogeneous and non-homogeneous system of equations using cramer's rule and matrix inverse method – characteristic roots and vectors – Verification of Cayley Hamilton theorem – Computation of the inverse by functions – Differentiation of one function with respect to another function.

Module 2: Successive differentiation – Leibnitz theorem (statement only) and simple problems – Meaning of the derivative – Maxima and Minima of functions of one variable (exclude rate of change, acceleration, velocity) – Concavity and Convexity,

Module 3: Points of inflexion – Partial. Differentiation – Maxima and Minima of functions of two variables.

Module 4: Integration – simple problems – rational algebraic function, irrational functions – properties of definite integrals – integration by parts (exclude inverse function) – reduction

Books for Study:

1. P.K. Mittal, Matrices, Vrinda Publications (P) Ltd., 2007.
2. Narayanan, S. and ManickavachagamPillai, T.K., Calculus Vol. I, S. Viswanathan Printers & Publishers, 1996.
3. Narayanan, S. and ManickavachagamPillai, T.K., Calculus Vol. II, S. Viswanathan Printers & Publishers, 1996.

Books for Reference

1. Graybill, F.A., Matrices with applications in statistics, second edition, Wads Worth, 1983.
2. Narayanan, S. and ManickavachagamPillai, T.K., Ancillary Mathematics Vol II, S. Viswanathan Printers & Publishers, 1996.
3. Shanthi Narayanan, A textbook of matrices, S. Chand & Co., 1959.
4. Singaravelu, A., Allied Mathematics, A.R.S. Publications, 2014.
5. Vittal, P.R, Allied Mathematics, Margham Publications, 2015.

Semester 3: Optimization techniques

Module 1: Introduction to OR - Linear programming problem - Formulation - Graphical method – Basic solution - Optimum solution - Simplex method - Various cases - Unbounded solution -Unrestricted variables, alternative optimum.

Module 2: Need for artificial variables - Two phase method - Big M method - Primal, Dual relationship – Dual simplex method.

Module 3: Transportation problem- North-west corner rule – least cost method- Vogel's Approximation Method – Modified Method, Assignment problem.

Module 4: Networks - CPM and PERT - problems. Decision under uncertainty - Laplace criterion - Minimax criterion - Savage criterion – Hurvitz theorem - Games - Two person zero sum games - Saddle point - Solving by graphical method - solving by LPP.

Books for study

1. Don T. Philips, Ravindran, A, James J. Solberg (2007), Operations Research: Principles and Practices, John Wiley & sons.
2. Hadley (2006), Linear Programming, Addison - Wesley publishers.
3. Hamdy A. Taha (2008) Operations Research - An Introduction (fourth edition), Macmillan publishers.

Books for reference:

1. Hillier, F.S. and Lieberman, G.J. (1974), Introduction to Operations Research, Holden Day Publishing, San Francisco.
2. KantiSwarup, Gupta, P.K., Manmohan (1993), Operations Research, Sultan Chand Publishers.
3. Mittal, K.V. (1976), Optimization Methods in Operations Research, Wiley Eastern.

Semester 3: Probability Distributions

Module 1: Mathematical expectations-definition, raw and central moments (definition and relationships), moment generating function and properties, Probability generating function and characteristic function (definition and basic properties).

Module 2: Bivariate random variable, joint pmf and joint pdf, marginal and conditional probability, independence of random variables.

Module 3: Skewness and kurtosis using moments, conditional mean and variance, covariance, Karl Pearson Correlation coefficient, independence of random variables based on expectation.

Module 4: Standard distributions-Degenerate distribution, Discrete type-Bernoulli, Binomial, Poisson, Geometric, negative binomial (definition, properties and applications), Uniform, Continuous type-Uniform, exponential, gamma, Beta, Normal (definition, properties and applications), Lognormal, Pareto and Cauchy (Definition only).

References

1. V. K. Rohatgi, An Introduction to Probability Theory and Mathematical Statistics, Wiley Eastern.
2. S.C.Gupta and V. K. Kapoor Fundamentals of Mathematical Statistics, Sultan Chand and Sons.
3. A.M. Mood, F.A. Graybill and D C Bose, Introduction to Theory of Statistics, Mc Graw Hill.
4. John E Freund, Mathematical Statistics (6th edn), Pearson Edn, New Delhi.

Semester 3: Statistical Data Analysis Using SPSS - LAB

Module 1: Data handling: open SPSS data file – save – import from other data source – data entry – labeling for dummy numbers - recode in to same variable – recode in to different variable – transpose of data – insert variables and cases – merge variables and cases. Data handling: Split – select cases – compute total scores – table looks – Changing column – font style and sizes

Module 2: Diagrammatic representation: Simple Bar diagram – Multiple bar diagram – Sub-divided Bar diagram - Percentage diagram - Pie Diagram – Frequency Table – Histogram – Scatter diagram – Box plot.

Module 3: Descriptive Statistics - Mean, Median, Mode, SD- Skewness- Kurtosis. Correlation – Karl Pearson's and Spearman's Rank Correlation, Regression analysis: Simple and Multiple Regression Analysis [Enter and stepwise methods]

Module 4: Obtain probabilities and cumulative probabilities from standard discrete and continuous random variables.

Books for Study:

1. Clifford E.Lunneborg (2000). Data analysis by resampling: concepts and applications. Dusbury Thomson learning. Australia.
2. Everitt, B.S and Dunn, G (2001). Applied multivariate data analysis. Arnold London.

Books for reference:

1. Jeremy J. Foster (2001). Data analysis using SPSS for windows. New edition. Versions 8-10. Sage publications. London.
2. Michael S. Louis – Beck (1995). Data analysis an introduction, Series: quantitative applications in the social sciences. Sage. Publications. London.

3. Semester 3: Python Programming

Module 1: Introduction to Python: Python Introduction - History of Python - Python features, Python interpreter, Overview of programming in Pytho - Python built in types, Arithmetic - Program input and output, Variables and assignment. Advanced data types :Python Strings and string manipulation Assigning values in strings, String special operators, String formatting operators, Triple Quotes, Raw String, Unicode String, Build-in-String methods. Python List: Introduction - Accessing values in list, List manipulations, List Operations, Indexing, slicing & matrices - Python Dictionary -Introduction, Accessing values, Properties, Functions in Dictionary. Python Tuples: Introduction, Operation, Accessing, Function and methods in tuples and Data Type Conversion.

Module 2: Python - Basic Operators: Arithmetic Operators, Comparison Operators, Logical (or Relational) Operators, Assignment Operators, Conditional (or ternary) Operators. Conditional Statement: Branching (if, else-if, nested), Looping: while statement, for statements, Control Statements: break, continue and pass Statements. Functions: Defining a function, Calling a function, Types of functions, Function Arguments Anonymous functions - Global and local variables. Modules: Importing module, Math module Random module. Packages – Composition - Exception Handling

Module 3: OOPs concept: Class and object – Attributes – Inheritance - Overloading, Overriding - Data hiding. Regular expressions - Match function, Search function, Modifiers, Patterns. Files: reading and writing files, methods of file objects. Standard library functions - dates and times. GUI Introductions: Introduction to GUI Programming, Tkinter programming, Tkinter widgets. Database: Python database application programmers interface (DB-API), connection and cursor objects. Type objects and constructors - python database adapters - Visualization: Bar chart, Polar plot, Pie Charts, Histograms, Contour Plot, Heat Map.

Module 4: Networking: Socket, Socket Module and methods. Client and server Internet modules – Multithreading - Web Services and XML - JSON and the REST Architecture - Web Programming: Creating simple web clients - Introduction to CGI, CGI module, building CGI applications. Python web application frameworks: django.

References:

1. Core Python Programming by Wesley J. Chun, 2nd Edition ,Pearson Education
2. An Introduction to Python by Guido Van Russom, Fred L. Drake, Network Theory Limited.
3. Beginning Python: From Novice To Professional By Magnus Lie Hetland, Second Edition.
4. Programming in Python 3 by Mark Summerfield, Pearson Education
5. Online version of An Introduction To Python
6. http://www.network-theory.co.uk/docs/pytut

Semester 3: Python – Lab

Course outline

1. Programs using Loops and decisions
2. Programs for constants and String Manipulations
3. Programs for Functions, arrays , tuple, list, Dictionary
4. Programs for Sessions and request handling
5. Programs for Modules, Input-Output, Exception Handling, OOPs concept
6. Programs for Database management, Multithreading Installation of WAMP/XAMPP Server, MySQL db, and Python MySQL interface
7. Exchange of data between web page and server
8. Storage /Retrieval/Updation of form data in MySQL DB

Semester 4: Statistical Simulation Techniques

Module 1: Random numbers: Pseudorandom number generation, Using random numbers to evaluate integrals. Probability integral transformation

Module 2: Generating discrete random number variables: The inverse transformation method, generating a Poisson random variable, generating a Binomial random variable, The acceptance-rejection technique, the composition approach, the Alias method for generating discrete random variable. Generating random vectors.

Module 3: Generating continuous random variables: The inverse transformation method, generating uniform random variable, generating Normal random variable, The acceptance-rejection technique, the composition approach, the Alias method for generating random variable. Generating random vectors.

Module 4: Multivariate Normal distribution and copulas: Multivariate Normal, Generating Normal random vectors, Copulas. Generating Normal random vectors by copulas.

Reference:

1. Simulation, Fifth Edition 5th Edition Sheldon M. Ross
2. Simulation Modelling and Analysis (Mcgraw-hill Series in Industrial Engineering and Management) by Averill M Law

Reference:

1. Introduction to Probability Models, Eleventh Edition by Sheldon M. Ross
2. Simulation Modelling and Analysis (Mcgraw-hill Series in Industrial by Averill M Law
3. Top 20 MS Excel VBA Simulations!: VBA to Model Risk, Investments, Growth, Gambling, and Monte Carlo Analysis (Save Your Time With MS Excel! Book 6 by Andrei Besedin)

Semester 4: Artificial intelligence

Module 1: AI: Introduction, Brief history, Agents and rationality, task environments, agent architecture types. Search and Knowledge representation. - Search spaces, Uninformed and informed search, Hill climbing, simulated annealing, genetic algorithms,

Module 2: Logic based representations (PL, FoL) and inference, Prolog,Rule based representations, forward and backward chaining, matching algorithms,

Module 3: Probabilistic reasoning and uncertainty, Bayes nets and reasoning with them, Uncertainty and methods to handle it.

Module 4: Learning - Forms of learning- Statistical methods: naive-Bayes, nearest neighbor, kernel, neural network models, noise and over fitting, Decision trees, inductive learning. Clustering - basic agglomerative, divisive algorithms based on similarity/dissimilarity measures. Expert Systems, Applications to NLP.

Books and References:

1. Stuart Russell, Peter Norvig, Artificial Intelligence: A Modern Approach, 3rd Ed., Prentice Hall, 2009. Can also use 2nd Ed., Pearson Education International, 2003.
2. Nils Nilsson, Artificial Intelligence: A New Synthesis, Morgan Kaufmann, 1998.
3. David Poole, Alan Mackworth, Artificial Intelligence: Foundations for Computational Agents, Cambridge Univ. Press, 2010.

Other References:

1. Ronald Brachman, Knowledge Representation and Reasoning, Morgan Kaufmann, 2004.
2. Frank van Harmelen, Vladimir Lifschitz, Bruce Porter (Eds), Handbook of Knowledge Representation, Elsevier, 2008.
3. Ivan Bratko, Prolog Programming for Artificial Intelligence, 4th Ed., Addison-Wesley, 2011.
4. Stephen Marsland, Machine Learning: An Algorithmic Perspective, Chapman and Hall, 2009.
5. Christopher Bishop, Pattern Recognition and Machine Learning, Springer, 2007.

Semester 4: Statistical Inference

Module 1: Sampling distributions: Parameter, Statistic, standard error, Sampling from normal distribution: distribution of sample mean, sample variance, chi-square, students t distribution, and F distribution (definition, property and relationships only). 20 hours

Module 2: Estimation of Parameter: Point Estimation. Desirable properties of a good estimator, unbiasedness, consistency, sufficiency. Methods of Estimation; method of maximum likelihood, method of moments, method of least squares, Concept of Bayesian estimation. Interval Estimation; Large sample confidence interval for mean, equality of means, equality of proportions.

Module 3: Testing of Hypotheses; concept of testing hypotheses, simple and composite hypotheses, null and alternative hypotheses, type I and type II errors, critical region, level of significance, power of test. Most powerful tests Uniformly most powerful test, Neyman Pearson Lemma. Likelihood ratio tests, large sample properties, asymptotic distribution of LRT statistic for simple null hypothesis.

Module 4: Large sample tests concerning mean, equality of means, proportions, equality of proportions. Small sample tests based on t distribution for mean, equality of means and paired t test. Tests based on F distribution. Teats based on chi square distribution for variance, goodness of fit and for independence of attributes .Test for correlation coefficients.

References

1. V. K. Rohatgi, An Introduction to Probability Theory and Mathematical Statistics, Wiley Eastern.
2. S.C.Gupta and V. K. Kapoor. Fundamentals of Mathematical Statistics, Sultan Chand and Sons
3. A.M. Mood, F.A. Graybill and D C Bose, Introduction to Theory of Statistics, McGraw Hill
4. John E Freund, Mathematical Statistics (6th edn), Pearson Edn, NewDelhi

Semester 4: Applied Multivariate Techniques

Module 1: Basics of Matrix and Vector Algebra –Positive Definite Matrices – Square Root Matrix – Random vectors and Matrices – Mean vectors and Covariance Matrices – Matrix Inequalities and Maximization. Multivariate Normal distribution – Properties, Distributions of linear combinations, independence, marginal distributions, conditional distributions, Partial and Multiple correlation coefficient. Moment generating function.

Module 2: Maximum likelihood estimation of Mean vector and Variance - Covariance matrix of multivariate Normal distribution. Sampling distribution of mean vector and dispersion matrix. Mahalanobis D^2 and Hotelling's T^2 Statistics. Testing the equality of mean vectors. Testing the independence of subvectors.

Module 3: One – way MANOVA. Principal Components - Population Principal Components – Summarizing Sample variation by Principal Components - Scree Plot. Factor Analysis – Orthogonal Factor model, Factor rotation, Factor scores. Canonical Correlation Analysis.

Module 4: Discrimination & Classification – Fisher's method. Optimality of classification rules. Discrimination & classification for several populations. Cluster Analysis – Similarity measures, Hierarchical & Non-Hierarchical methods.

Books for study:

1. Anderson, T. W. (2003): An Introduction to Multivariate Statistical Analysis – 3rdedn. John Wiley & Sons.
2. Johnson, R. A., &Wichern, D. W. (2007): Applied Multivariate Statistical Analysis – 2ndedn. Prentice Hall International

Books for Reference:

1. Everitt, B.S & Dunn, G (2001): Applied multivariate Data analysis, second edition, Arnold publishers, London.
2. Morrison, D.F (1990): A multivariate statistical methods, Third edition, Mc graw hall, New delhi.

Semester 4: R Programming - Lab

Module 1: R as a calculator, statistical software and a programming language, R preliminaries, getting help, data inputting methods(direct and importing from other spread sheet applications like Excel), data accessing, and indexing, Graphics in R, built in functions, saving, storing and retrieving work. Looping and Decision making – *for* loop, *while* loop, *if* command, *if else* command.

Module 2: Diagrammatic representation of univariate and bivariate data - box plots, stem and leaf diagrams, bar plots, pie diagram, scatter plots, p – p plot, q – q plot. Descriptive statistics - measures of central tendency (mean, median and mode), partition values, measures of dispersion (range, standard deviation, mean deviation and inter quartile range), summaries of a numerical data, skewness and kurtosis, Correlation.

Module 3: Random sampling with and without replacement. Probability Distributions: Random number generation. Statistical Inference: One- and two-sample tests, z test, t-test, F-test, chi-square test of independence and goodness of fit, interval estimation for mean, difference of mean and variance. ANOVA (one- way and two-way). Non – parametric tests: Shapiro-Wilks test - One sample KS test- Mann-Whitney U test – Wilcoxon Signed Rank test - Kruskal Wallis test – Friedman test.

Module 4: One – way MANOVA, Principal Component Analysis, Factor Analysis, Canonical Correlation, Discriminant analysis, Cluster Analysis.

References:

1. Alain F. Zuur, Elena N. Ieno, Erik H.W.G. Meesters, "A Beginner's Guide to R" Springer, 2009
2. Roger D. Peng, "R Programming for Data Science" Lean Publishing, 2014
3. Phil Spector, Data Manipulation with R, Springer.
(2008http://www.springer.com/statistics/computanional+statistics/book/978-0-387-74730-9

Semester 4: Mini Project

The main aim of this project is to implement the theoretical knowledge gained from various areas to develop effective scientific solutions to various real life problems. The course Project is one that involves practical work for understanding and solving problems in the field of data science. Students will select individually Commercial/Technical/Research Project based on Application Development Technologies learnt in previous semesters. Each student will have to prepare proper documentation consisting of Software Requirements Specification (SRS), Modelling Techniques, Development Strategies, Implementation and Testing Strategies. The project work will be presented by students using Power Point Presentation Tool to the panel of Examiners, along with a live demonstration of the project.

Semester 5: Data Science - Industrial Perspective

Module 1: What is Data Science?- Big Data and Data Science hype – and getting past the hype - Why now? – Datafication - Current landscape of perspectives - Skill sets needed. R.

Module 2: Exploratory Data Analysis and the Data Science Process - Basic tools (plots, graphs and summary statistics) of EDA - Philosophy of EDA - The Data Science Process - Case Study: RealDirect (online real estate firm)**.** Three Basic Machine Learning Algorithms - Linear Regression - k-Nearest Neighbors (k-NN) - k-means.

Module 3: One More Machine Learning Algorithm and Usage in Applications - Motivating application: Filtering Spam - Why Linear Regression and k-NN are poor choices for Filtering Spam - Naive Bayes and why it works for Filtering Spam - Data Wrangling: APIs and other tools for scrapping the Web.

Module 4: Feature Generation and Feature Selection (Extracting Meaning From Data) - Motivating application: user (customer) retention - Feature Generation (brainstorming, role of domain expertise, and place for imagination) - Feature Selection algorithms – Filters; Wrappers; Decision Trees; Random Forests. Data Science and Ethical Issues - Discussions on privacy, security, ethics - A look back at Data Science - Next-generation data scientists.

Text Book :

Cathy O'Neil and Rachel Schutt. Doing Data Science, Straight Talk From The Frontline.O'Reilly. 2014.

Reference Books :

1. Jure Leskovek, Anand Rajaraman and Jeffrey Ullman. Mining of Massive Datasets, Cambridge University Press. 2014. (free online)
2. Kevin P. Murphy. Machine Learning: A Probabilistic Perspective. ISBN 0262018020. 2013.
3. Foster Provost and Tom Fawcett. Data Science for Business: What You Need to Know about Data Mining and Data-analytic Thinking. ISBN 1449361323. 2013.
4. Trevor Hastie, Robert Tibshirani and Jerome Friedman. Elements of Statistical Learning, Second Edition. ISBN 0387952845. 2009. (free online)

Semester 5: Data Mining Techniques

Module 1: Data warehouse – definition – operational database systems Vs data warehouses – multidimensional model – from tables and spreadsheets to Data Cubes – schemas for multidimensional databases – measures – concept hierarchies - OLAP operations in the multidimensional data model – data warehouse architecture.

Module 2: Data mining – introduction – definition - data mining functionalities – major issues in data mining - data preprocessing – data cleaning – data integration and transformation – data reduction – data discretization and concept hierarchy generation. Association rule mining - efficient and scalable frequent item set mining methods – mining various kinds of association rules – association mining to correlation analysis – constraint-based association mining.

Module 3: Classification and prediction - issues regarding classification and prediction – classification by decision tree introduction – Bayesian classification – rule based classification – classification by back propagation – support vector machines – associative classification – lazy learners – other classification methods – prediction – accuracy and error measures – evaluating the accuracy of a classifier or predictor – ensemble methods – model section.

Module 4: Cluster analysis - types of data in cluster analysis – a categorization of major clustering methods – partitioning methods – hierarchical methods – density-based methods – grid-based methods – model-based clustering methods – clustering high dimensional data – constraint-based cluster analysis – outlier analysis. Graph mining - mining object, spatial, multimedia, text and web data - multidimensional analysis and descriptive mining of complex data objects – spatial data mining – multimedia data mining – text mining – mining the World Wide Web.

References

1. Jain Pei, Jiawei Han and Micheline Kamber, Data Mining Concepts and Techniques, 3rd Edition, Elsevier, ISBN: 9380931913.
2. Alex Berson and Stephen J. Smith, Data Warehousing, Data Mining & OLAP, Computing Mcgraw-Hill, ISBN: 0070062722.
3. K.P. Soman, Shyam Diwakar and V. Ajay, Insight into Data mining Theory and Practice, 1st Edition, Prentice Hall of India, ISBN: 8120328973.
4. G. K. Gupta, Introduction to Data Mining with Case Studies, 3 rd Edition, PHILearning Pvt. Ltd, ISBN: 8120350022.
5. Pang-Ning Tan, Michael Steinbach and Vipin Kumar, Introduction to Data Mining, 1st Edition, Pearson India, ISBN: 9332518653.

Semester 5: Design of Experiments

Module 1: Linear estimation, estimability of parametric functions and BLUE Gauss- Markov theorem-Linear Hypothesis.

Module 2: Analysis of variance, one way and two way classification (with single observation per cell), Analysis of covariance with a single observation per cell.

Module 3: Principles of design-randomization-replication-local control, completely randomized design; Randomized block design; Latin square design.
Missing plot technique; comparison of efficiency; Greco-Latin square design (Concept only).

Module 4: Basic concepts of factorial experiments, 2^2 and 2^3 factorial experiments, Duncan's multiple range test. Incomplete block design; incidence matrix; orthogonal design (Definition only), Balanced incomplete bock design (Basic concept only), partially incomplete block design (Basic concept only).

Books for references

1. S.C. Gupta and V K Kapoor, Fundamentals of applied Statistics, Sulthan Chand and Sons
2. Federer, Experimental Designs
3. M N Das and N Giri, Design of Experiments, New Age international,
4. DD Joshy, linear Estimation and Design of Experiments, Wiley Eastern
5. Montgomeri, Design of Experiments

Semester 5: Categorical Data Analysis

Module 1: Categorical Response data – Inference procedures. Contingency tables – Comparison of proportions, partial association in 2 x 2 and I x J tables. Testing independence in two-way contingency tables. Generalized Linear Model – For binary data & count data. Inference for & Fitting of GLMs.

Module 2: Logistic Regression Model – Fitting & diagnostics. Conditional associations in 2 x 2 x K tables. Multinomial logit models – Baseline logit models for nominal responses & Cumulative logit model for ordinal responses.

Module 3: Loglinear models for two-way tables; Loglinear models for Independence & Interaction in three-way tables. Loglinear - Logit model connection. Diagnostics for checking models. Ordinal Association Models. Probit Models.

Module 4: Comparison of dependent proportions. Conditional logistic regression for Binary Matched pairs. Marginal models for square contingency tables. Symmetry, Quasi-Symmetry & Quasi independence.

Book for Study:

1. Alan Agresti (2002): Categorical Data Analysis. John Wiley & Sons

Book for Reference:

1. Hosmer, D.W. & Lemeshow, S. (1989)Applied Logistic Regression (John Wiley).

Semester 5: Stochastic Modeling

Objective: To introduce the vital area of regression models applicable in a wide variety of situations. To expose the students to the wide areas of its applications.

Module 1: Introduction to Regression – Mathematical and Statistical Equation – Meaning of Intercept and Slope – Error term – Measure for Model Fit –R^2 – MAE – MAPE – Testing Significance of Model Coefficients, Confidence interval for model coefficients. Model diagnostics - Mean predicted value, Testing normality of error term, QQ-plot, PP-plot, Anderson Darling, Kolmogrov Smirnov

Module 2: Introduction to Multiple Linear Regression Model, Partial Regression Coefficients, Testing Significance overall significance of Overall fit of the model, Testing for Individual Regression Coefficients, Estimating R^2, MAE and MAPE

Module 3: Dummy Variable trap, Study of Interaction Effects, Varying Intercept and Slope using dummy variable, Detection and Removal of Outliers. Study of Normality of Error Term using graphical and testing procedures, Testing for Multicollinearity using VIF, Testing for assumption of Homoscedasticity.

Module 4: Components of Time Series, Mathematical models of time series. Measurement of Trend Component : Graphic, Semi-Averages, Moving Averages. Least-squares – Straight Line, Second Degree Parabola, Exponential Curve, Modified Exponential Curve, Gompertz Curve, Logistic Curve.Measurement of Seasonal Variations – Simple averages, Ratio-to-trend, Ratio-to Moving average, Link Relative. Deseasonalisation of data.Measurement of Cyclic variations.

Books for Study:

1. Gujarati, D.(2004): Introduction to Econometrics. McGraw Hill, New Delhi.

Books for Reference:

1. Montgomery,D.C. ,Peck E.A, & Vining G.G.(2003). Introduction to Linear Regression Analysis. John Wiley and Sons,Inc.NY

Semester 5: Machine Learning and Data Analytics

Module1: Introduction: Definition of learning systems. Goals and applications of machine learning. Aspects of developing a learning system: training data, concept representation, function approximation. Supervised Learning techniques: Introduction to machine learning ,Supervised Learning concepts, Linear Regression, Logistic regression, K-NN classification, Naïve Bayesian classifiers, SVM - (Support Vector Machines)

Module2: Decision Tree Learning **:** Representing concepts as decision trees. Recursive induction of decision trees. Picking the best splitting attribute: entropy and information gain. Searching for simple trees and computational complexity. Occam's razor. Overfitting, noisy data, and pruning**.**

Module3: Ensemble Techniques :Introduction to Ensemble learning, Different Ensemble Learning Techniques -Bagging, Boosting, Random Forests, Stacking, Featurization, model selection, block chain and tuning: Text Analytics, Feature extraction, Model Defects & Evaluation Metrics, Model selection and tuning, comparison of machine learning models

Module 4: Recommendation Systems :Introduction to Recommendation Systems ,Types of Recommendation techniques - Collaborative Filtering, Content based Filtering, Hybrid RS, Case Study, Performance measurement, Deep Learning Concepts: Introduction to Deep Learning Concepts, Fundamentals of neural networks, Introduction to Tensor flow and Keras as Deep Learning frameworks, Tensor Flow illustrative example, Introduction to CNN, Evaluation of Deep Learning model

References:

1. Data Science and Machine Learning with Python - Hands On!/by Frank Kane
2. J. Han, Data mining: concepts and techniques/Jiawei Han Micheline Kamber, Jian Pei:Morgan Kaufmann Pub., pp. 740, 2012.

Semester 5: Statistical Data Analysis Using SAS - Lab

Module 1: Data step and Proc Step, SAS Data Libraries, Creating dataset using data lines, Importing data using INFILE statement, Importing data using Proc Import, Creating HTML Output, Sub setting observations using conditional statements, Sub setting variables using Keep/Drop, Creating variables using IF-THEN else statements, Retain statement, FIRST. , LAST. , Date functions, Character functions.

Module 2: SAS procedures, Sub setting in Procedures with the WHERE Statement, Sorting Data with PROC SORT, Printing Data with PROC PRINT, Summarizing Your Data Using PROC MEANS, Writing Summary Statistics to a SAS Data Set, Counting Data with PROC FREQ, Producing Tabular Reports with PROC TABULATE, PROC SORT, PROC SUMMARY

Module 3: Modifying a Data Set Using the SET Statement, Stacking Data Sets Using the SET Statement, Interleaving Data Sets Using the SET Statement, Combining Data Sets Using a One-to-One Match Merge, Combining Data Sets Using a One-to-Many Match Merge, Merging Summary Statistics with the Original Data, Writing Multiple Data Sets Using the OUTPUT Statement, Changing Observations to Variables Using PROC TRANSPOSE

Module 4: SAS Macro Concepts, Substituting Text with Macro Variables, Creating Modular Code with Macros, Adding Parameters to Macros, Writing Macros with Conditional Logic, Writing DataDriven Programs with CALL SYMPUT. Proc SQL, Using Proc SQL to create tables, Modifying tables, Aggregating tables, Stacking and Merging tables. PROC UNIVARIATE, PROC MEANS, PROC CORR, PROC PLOT, PROC FREQ, PROC TTEST , PROC NPAR , PROC ANOVA, PROC REG, PROC ARIMA.

Books for Reference

1. The Little SAS Book: A Primer, Fourth Edition, Lora D. Delwiche, Susan J. Slaughter
2. Learning SAS by Example: A Programmer's Guide, Ron Cody, SAS Institute

Semester 6: Internship and Project

Course Outline

The student shall undergo Industrial training and a project of four month duration. Industrial training should be carried out in an industry / company approved by the institution and under the guidance of a staff member in the concerned field. At the end of the training he / she have to submit a report on the work being carried out.

The project is designed to develop practical ability and knowledge about practical tools/techniques in order to solve real life problems related to the industry, academic institutions and data science research. The project should strictly stick to the data science principles. Students can take up any application level/system level project pertaining to a relevant domain. Projects can be chosen either from the list provided by the faculty or in the field of interest of the student. For external projects, students should obtain prior permission after submitting the details of the external guide, institution and synopsis of the work. The project guide should have a minimum qualification of ME/M.Tech/MCA/M.Sc in Statistics/Computer science/Mathematics or related fields.

At the end of each phase, presentation and demonstration of the project should be conducted, which will be evaluated by a panel of examiners. A detailed project report duly approved by the guide in the prescribed format should be submitted for end semester assessment. Marks will be awarded based on the report and their performance during presentations and demonstrations. Publishing the work in Conference Proceedings/ Journals with National/ International status with the consent of the guide will carry an additional weightage in the review process.

Qualifications Pack-Junior Data Associate

SECTOR: IT-ITeS

SUB-SECTOR: IT Services

OCCUPATION: Data Scientists

REFERENCE ID: SSC/Q0401

ALIGNED TO : NCO-2015/ 2521.0202

Junior Data Associate in the IT-ITeS Industry is also known as a Data Warehousing Engineer, Data Scientist, Business Intelligence Engineer etc.

Brief Job Description: Individuals at this job are responsible for designing and implementing processes and layouts for complex, large-scale data sets used for modeling, data mining, and research purposes. Responsibilities also include designing and implementing statistical data quality procedures around new data sources.

Qualifications Pack-Analyst – Research

SECTOR: IT-ITeS

SUB-SECTOR: Business Process Management

OCCUPATION: Knowledge Process Outsourcing - Research

REFERENCE ID: SSC/Q2601

ALIGNED TO: NCO-2015/2421.0101

Analyst-Research in the IT-ITeS Industry is also known as a Research Associate.

Brief Job Description: Primary responsibility of the individual in this job is to organise, analyse, synthesize and summarize information using appropriate analytical methodologies. He/she needs to prepare companies profiles, conducts financial analysis & valuations, benchmarking, collect data using techniques such as questionnaires, surveys, interviews and electronic data collection as part of the job.