

# Gujarat University



## School of Emerging Science and Technology

### M.Sc. (Integrated) Five Years Program

#### M.Sc. (Int) Data Science

(For Batches 2019-2024, 2020-2025, 2021-2026, 2022-2027)

**For students completing 3 years of the above program  
with an exit option**

**B.Sc. Data Science**

**For students completing 4 years of the above program  
with an exit option**

**B.Sc. (Honours) Data Science**

# Gujarat University

## School of Emerging Science and Technology

### M.Sc. (Integrated) Five Years Program

#### Data Science

#### SEMESTER - 1

Sr No.	Course No.	Subject	Credit	Exam
1	(#)EC-101	Yoga	2	Theory
2	(#)FC-101	Communication Skills	2	Theory
3	CC-101	Mathematical Basics	3	Theory
4	CC-102	Descriptive Statistics and Probability	3	Theory
5	CC-103	Introduction to Data Science	3	Theory
6	CC-104	Basics of Computer Organization & Architecture	3	Theory
7	CC-105	Information and Communication Technology (ICT)	3	Theory
8	CC-106 (P)	Practical Based on CC-101 & CC-102	3	Practical
9	CC-107 (P)	Practical Based on CC-105	3	Practical
		<b>Total</b>	<b>25</b>	

(#) Marks and credits of this subject will not be considered for class or total.

## **EC-101 Yoga**

Introduction to Yoga, Yoga and Personality Development, Sitting and standing warm-up exercise, Basic stretching, Surya Namaskar, Ardhamatsyendrasana, Vakrasana, Pranayama, Tadasana, Hastotthanasana, Padahastanasana, Pawanmuktasana, Halasana/Ardhalasana, Naukasana, Sky-diving Yoga.

Flexibility Booster, Bhujangasana, Sarpasna (Snake Pose), Shalabhasana, Dhruvasana (Tree Pose), Katiasana, Trikonasana, Marjariasana, Mandukasana (Frog Pose), Ushtrasana, Relaxation, Exercises for Obesity, Meditation, Akarna Dhanurasana, Janushirasana, Bhunamana, Sarvangasana, Shashankasana, Simhasana (Lion Posture), Breathing Modulation, Yoga Mudra, Matsyasana, Relaxation Exercise with Smiley Ball.

### **Reference Books:**

1. Yoga: A Healthy way of living (Secondary Stage); New Delhi, India, NCERT, 2015
2. The Art of Living
3. Ministry of AYUSH, Government of India

# **FC-101 Communication Skills**

## **Unit - 1 Introduction to Communication**

Objective and Role, Types of Communication, Inter personal Communication

## **Unit - 2 Vocabulary and Written Communication**

Words, Word order, Phrases, Idioms, Format, draft, structure, Precise Writing, Letter / email

## **Unit - 3 Listening**

Types – Active, Passive, Barriers, Dicto-composition Format, draft, structure, Precise Writing, Letter / email

## **Unit - 4 Oral Communication**

Interviews, Group Discussions, Debates

### **Reference Books:**

1. Effective English Communication by Mohan, Raman (Tata Mc. Graw Hill Education)
2. Managerial Communication by Khuman, Gupta (Mahahjan Publication House)
3. English Grammar & Compositions by Wren, Martin (S Chand)

# CC-101 Mathematical Basics

## Unit - 1 Set theory

Introduction to set theory, method of representation of a set, operations on set/properties (with logical and Venn- diagrammatic proof), example related to Venn-diagram, Cartesian product

## Unit - 2 Functions

Definition of functions, Types of functions, some special functions and its examples, domain and range of function, Plotting of graphs – Rational Functions

## Unit - 3 Series and sequence

Definition of sequence, series, Arithmetic progression, Arithmetic mean, Geometric progression, Harmonic progression, Geometric mean, its examples, relationship between A.M and G.M, Sums to a term of special series, Concept of Annuities, A.P., G.M., H.M. uses and its merits and demerits

## Unit - 4 Continuity and Limit

Concept of limits, examples based on definition of limits, definition of Continuity, properties of continuity, types of discontinuity and examples based on discontinuity, Convergence of limit, Indeterminate forms  $(\frac{0}{0}, \frac{\infty}{\infty}, 1^{\infty})$

## Reference Books:

1. Thomas' Calculus by Hass, Heil, Weir (Pearson 14<sup>th</sup> edition)
2. Calculus by Smith, Milton (Mc. Graw Hill 4<sup>th</sup> edition)
3. Advanced Mathematics by Dr. Ravi Gor (Nirav Prakashan 5<sup>th</sup> Revised Edition)

# **CC-102 Descriptive Statistics and Probability**

## **Unit - 1 Index Numbers**

Definition, Construction of I.N., Problems thereof for weighter and unweighted I.N. including Laspeyre's Paasche's, Edgeworth- Marshall and Fisher's chain I.N., Conversion of foxed based to chain based I.N. and vice-versa consumer price I.N

## **Unit - 2 Data types, Classification and its Representation**

Types of data: Primary, Secondary, Internal and External data and their sources. Classification of data: Qualitative. Quantitative: Discrete, Continuous, Chronological Nominal, Ordinal, Interval and Ratio data. Frequency: Grouped and ungrouped data, cumulative frequency distribution. Graphical representation of grouped data: Histogram, frequency curve, frequency polygone ogives. Diagrammatic representation of data: Bar- diagram, Simple, multiple, sub- divided percentage. Two dimensional diagram: Rectangles and pie diagram

## **Unit - 3 Measures of Central Tendency and Dispersion**

Measures of central tendency: Concept of central tendency, relationship their merits and demerits, Empirical relation between mean, median and mode. Use of appropriate measure of central tendency in different situation

Measures of Dispersion: Range, Quartile Deviation, Mean Deviation, Standard Deviation, Coefficient of Variation, Moments, Skewness and Kurtosis, Sheppard's corrections.

Measures of inequality: Gini's coefficient and Wrenz curve

## **Unit - 4 Probability**

Random experiment, Trial, Sample point, Sample space, Definition of Equally likely, mutually exclusive and Exhaustive events. Definition of probability: Classical, Relative and Axiomatic approach & its properties, Conditional Probability, Multiplicative law of Probability, Independent events, Law of Total Probability, Bayes theorem and its application

## **Reference Books:**

1. Statistics for Business and Economics by Anderson, Sweeney, Williams (Cengage Publishers 12<sup>th</sup> edition)
2. Applied Business Statistics by Black (Wiley 7<sup>th</sup> edition)
3. Statistics for Management by Levin, Rubin (Pearson 8th edition)

# **CC-103 Introduction to Data Science**

## **Unit - 1 Introduction to Data Science**

What is Data Science, what is Data Analytics, Why Data Science, Pros and Cons of Data Science, why is Data Science Important, Important skills for data science, Data Science Process, Application of Data Science

## **Unit - 2 Foundation of Data Science**

Data Ecosystem: Introduction to data, Univariate, Bivariate and Multivariate data

Importing Datasets: Identifying, Gathering, and Importing data, how to use and transform it

Data Pre-processing: Data cleaning and wrangling using pandas, Feature scaling and Normalization. Data handling and manipulation

Basics of Machine Learning and Natural Language Processing, Model selection and evaluation, Data Warehousing, Data Mining, Business Intelligence and Big Data

## **Unit - 3 Exploratory Data Analysis for Data Science**

Introduction to Data Analysis, why is it important in data science, how to use different libraries to visualize data, what are the benefits of data analysis in data science

## **Unit - 4 Data Science and Data Analysis Tools**

Basic Introduction of Python, R, SQL, Excel, Tableau, Hadoop and Weka

### **Reference Books:**

1. Data Science from Scratch by Joel Grus
2. Python Data Science Handbook by Jake VanderPlas
3. Python for Data Analysis by Wes McKinney
4. Data Science for Dummies BY Pierson (John Wiley & Sons)
5. Introducing Data Science By Davy Cielen, Arno D. B. Meysman

# CC-104 Basics of Computer Organization and Architecture

## Unit - 1 Basics of Number System and Arithmetic with Character Coding

**Number System:** Binary, octal, Hexadecimal number system, Conversion from one system to another

**Computer Arithmetic:** Binary addition, subtraction, Multiplication & Division, Complements and its use for representing negative numbers & subtraction, floating point representation.

**Codes for character representation:** Hexadecimal, BCD, Excess-3, Gray code, ASCII Error detection codes (parity)

## Unit - 2 Basics of Logic Gates and Boolean Algebra

**Logic gates:** AND, OR, NOT. Other gates like NAND, NOR, EX-OR, EX-NOR. Their truth tables, symbols and equation, Simple logic circuits using gates, building all basic gates using NAND, NOR gates, Design using NAND, NOR gates

## Unit - 3 Basics of Combinational Circuit

**ALU:** Construction of ALU, Binary Half Adder, Full Adder, Parallel Binary Adder, Binary Adder-Subtract

**Digital components:** Integrated circuits, decoders and its expansion, Encoders, Multiplexers and its expansion

## Unit - 4 Basics of Input/ Output Device, Memory Unit and Address, Data & Control Bus

**Address, Data & Control bus Input/output devices:** Input/output devices: Key Board, Mouse, Display Unit, Printer (types), Scanner, OCR, OMR, MICR, Input/output interface, Asynchronous data transfer, and modes of data transfer. Concepts of programmed I/O, DMA

**Memory:** Memory Hierarchy, primary Memory – RAM, SRAM, DRAM, ROM, Types of ROM, Secondary Memory (Magnetic Disk, Magnetic Tape), Optical Memory (CDROM), Concept of Virtual Memory, Concept of Cache and their need

## Reference Books:

1. Fundamentals of Digital circuits - By PHI/Pearson Education – A, Anand Kumar
2. Computer System Architecture - By M. Morris Mano – PHI/Pearson Education
3. Digital Computer Fundamentals - By Thomas C. Bartee – Tata McGraw Hill



# **CC-105 Information and Communication Technology (ICT)**

## **Unit - 1 Introduction to Computers**

Overview of Computers, Characteristics of Computers, Evolution of Computers, Different Generation of Computers, Types of Computers, Components of Computer System, Primary and Secondary Memory, Input and Output device, Computer Language, Application of Computers, Central Processing Unit (CPU) and Its components

**Introduction of Internet:** Evolution of Internet, Basic Internet Terms: World Wide Web, Web Page, Web Site, Home Page, Browser, Uniform Resource Location, Hypertext, Internet Service Provider, Web Server, Getting Connected to Internet, Internet Application, Electronic mail and Introduction, Search Engine, Data over Internet, Internet and Viruses

## **Unit - 2 Data Communication and Computer Network**

### **Introduction to Data Communication**

Transmission Media: Guided and Unguided, Analog and Digital Data Transmission, Modulation Technique

Multiplexing, Switching, Computer Network, Network Topology, Communication Protocol, Network Devices: NIC, Hub, Repeater, Switch, Bridge, Router, Gateway

## **Unit - 3 Introduction to Programming Language**

Introduction To Machine Level Language, Assembly language, Higher level language, Limitations and Features. Classification of Computer Language- Procedural Language and Non-Procedural Language, Basic block diagram and functions of various components of computer, Compiler and interpreter, Algorithm Development and Flow Chart

### **Getting Started with C Language**

Executing C program, Character set & C Tokens, Identifiers & Keywords, Data Types, Constants and Variables, Type Casting, Comments

## **Unit - 4 C Language Operators and Decision Making**

Types of Operators and Expression, Precedence & Associativity, Console based I/O and related built-in I/O function, printf(),scanf(), getch(), getchar(), putchar(), Concept of HeaderFile and #include,#define, Decision Making Structure If, If-else, Nested If-else, Switch Control Structure & Array, Loop Control Structure While, Do-While, For, Nested loop, Other Statements: break, continue, goto, exit, array, one dimensional array & two dimensional array

### **Reference Books:**

1. Introduction to Computer Science Pearson Publication
2. Programming in ANCI C, Seventh edition, by Balagarusamy E, Tata McGraw-Hill Publishing Company Limited
3. C Programming language, Second edition, by Kernighan B W and Ritchie D M Prentice Hall

## **CC-106(P) Practical Based on CC-101 & CC-102**

### **Practical Based on CC-101 Mathematical Basics, CC-102 Descriptive Statistics and Probability**

#### **Unit - 1 Numerical Applications based on Set Theory and Functions and Graph**

Numerical examples on elementary operations of sets, Decision making based numerical examples on Union or Intersections of sets, Analogy questions on Venn Diagram

Sketching the graph of Rational polynomial function, Numerical examples on Exponential and Logarithmic functions, Applications of functions on Employment, Depreciation calculation, Break-Even Analysis

#### **Unit - 2 Numerical Applications based on Sequence and Series and Continuity and Limits**

Numerical based on Present value, Future value, and annuity, Problems on Arithmetic and Geometric Progression, Sum of Special series, Applications of concept of sequence and series on the problems based on Income, Biology, Recreation, Depreciation, Geometry, Health, Inventory, Real Estate, Investments, Physics

Numerical examples based on definition of limits and continuity, Problems on the sales decay in the Absence of Advertising, problems on different types of discontinuities and limit of Indeterminate forms

#### **Unit - 3**

Construction of I.N., Problems thereof for weighted and unweighted I.N.

Calculation of Laspeyre's index number Paasche's index number, Edgeworth- Marshall index number and Fisher's chain index number, Conversion of fixed based to chain based I.N.

Classification of data: Qualitative. Quantitative: Discrete, Continuous, Chronological Nominal, Ordinal, Interval and Ratio data., Frequency: Grouped and ungrouped data, cumulative frequency distribution

Graphical representation of grouped data: Histogram, frequency curve, frequency polygon ogives.

Diagrammatic representation of data: Bar- diagram, Simple, multiple, sub- divided percentage.

Two-dimensional diagram: Rectangles and pie diagram

#### **Unit 4**

Measures of central tendency: mean, Median, Mode Identify the appropriate measure for the given data.

Measures of Dispersion: Range, Quartile Deviation, Mean Deviation, Standard Deviation, Coefficient of Variation, Moments, Skewness and Kurtosis, Sheppard's corrections

Measures of inequality: Gini's coefficient and Lorenz curve

Problems related to Probability and Bayes' theorem

**Reference Books:**

1. Thomas' Calculus by Hass, Heil, Weir (Pearson 14<sup>th</sup> edition)
2. Mathematics for Management: An Introduction by M Raghavachari, Tata McGraw Hill
3. Algebra 2 by Holt, Rinehart and Winston, A Harcourt Education Company
4. Statistics for Business and Economics by Anderson, Sweeney, Williams (Cengage Publishers 12<sup>th</sup> edition)
5. Applied Business Statistics by Black (Wiley 7<sup>th</sup> edition)
6. Statistics for Management by Levin, Rubin (Pearson 8th edition)
7. Fundamental of Mathematical Statistics by Gupta and Kapoor

## **CC-107(P) Practical Based on CC-105**

### **Practical Based on CC-105 Information and Communication Technology (ICT)**

#### **Unit - 1 Word Processing Concepts**

Saving, Closing, Opening and exiting document, Selecting text, Editing text, Finding and Replacing text, Printing documents, Creating and Printing Merge documents, Character and Paragraph Formatting

#### **Unit - 2 Spreadsheet Concepts**

Navigation worksheet, entering & editing data into cells, formulas & operators, range of cells, moving – coping –transposing data, absolute and relative cell reference, spell checking, insert-delete-hide-show rows & columns, change-column width & row height, Formatting data, sorting data, Searching & Replacing data

#### **Unit - 3 Getting Started with C Language**

Make an Algorithm & Flowchart for Program Say Hello Word, Addition, subtraction, multiplication, division, if, if-else, if else if, switch case, array, Write a Program Say Hello World, use header file <stdio.h> <conio.h> using printf(), comments, identifiers, constant, variable, addition, subtraction, multiplication, division

#### **Unit - 4 Functions of C Language**

**Functions:** Scanf(), getch(), getchar(), putchar(), Concept of HeaderFile and #include,#define

**Conditional Statement:** If, If-else, Nested If-else or else-if, Switch Control Structure, use two conditional statement in one program

**Loop Control Structure:** While, Do-While, For, Nested loop

**Other Statements:** break, continue, goto, exit, array, one dimensional array & two-dimensional array

#### **References Books:**

1. Microsoft office, Professional for windows 95, By Dienes, Sheila S, Instant Reference; BPB Publication, Delhi
2. The Compact Guide to Microsoft Office By Mansfield, Ron, BPB Publication, Delhi
3. Programming in ANCI C, Seventh edition, by Balagarusamy E, Tata McGraw-Hill Publishing Company Limited

# Gujarat University

## School of Emerging Science and Technology

### M.Sc. (Integrated) Five Years Program

#### Data Science

#### SEMESTER - 2

Sr No.	Course No.	Subject	Credit	Exam
1	(#)EC-111	Environmental Studies	2	Theory
2	(#)FC-111	Commercial Communication	2	Theory
3	CC-111	Calculus and Introduction to Matrices	3	Theory
4	CC-112	Numerical and Statistical Methods	3	Theory
5	CC-113	Algorithms and Data Structures	3	Theory
6	CC-114	Object-Oriented Programming with JAVA	3	Theory
7	CC-115	Programming with PYTHON	3	Theory
8	CC-116 (P)	Practical Based on CC-114	3	Practical
9	CC-117 (P)	Practical Based on CC-115	3	Practical
		<b>Total</b>	<b>25</b>	

(#) Marks & Credits of this subject will not be considered for Class or Total

# **EC-111 Environmental Studies**

## **Unit - 1 Ecology and Natural Resources**

Types of Ecology, Earth's Climate, Biomes Distribution, Concept of Environment and Biosphere, Introduction of Population model, Renewable and Non-renewable resources and their conservation, Forest resources, Water resources, Mineral resources, Food resources, Energy resources and Land resources

## **Unit - 2 Sustainable Ecosystems**

Introduction, Type, Component and Function of Ecosystems, Energy flow, Food chain, Food web, Ecological pyramid, Biogeochemical Cycles (Nitrogen, Carbon, Water and Phosphorus nutrient cycles)

## **Unit - 3 Biodiversity**

Levels of biodiversity (Genetic, Species and Ecosystem diversity), Importance and Loss of Biodiversity, Structural magnitude of biodiversity in World and India, Types of biodiversity conservation (Ex-situ and In-situ conservation, Hotspots), National Parks and Sanctuaries of India

## **Unit - 4 Environmental Pollution and their Impact**

Types of pollution (Air, Water, Noise, Soil, Radioactive and Solid Waste Pollution), Control of pollution, Disaster and their management in India, Greenhouse gases and Climate change

## **Reference Books:**

1. Campbell Biology 11<sup>th</sup> edition (2017) by Lisa A. Urry, Michael L. Cain, Steven A. Wasserman, Peter V. Minorsky, Jane B. Reece; New York Pearson Publication
2. Ecology and Environment (2019) by R N Bhargava, V Rajaram, Keith Olson, Lynn tiede; CRC Press Publication
3. Textbook for Environmental Studies for UG courses of all branches of higher Edu. (2014) by Erach Bhaucha; University Grant Commission
4. Ecology Concepts and applications 8<sup>th</sup> edition (2019) by Manuel C. Molles Jr., Anna A. Sher; McGraw Hill Edu. Publication
5. Ecology and Environment 13<sup>th</sup> edition (2018) by P. D. Sharma; Rastogi Publication
6. Fundamentals of Ecology 5<sup>th</sup> edition (2017) by Eugene P. Odum, Gray W. Barrett; Cengage Publication

# **FC-111 Commercial Communication**

## **Unit - 1 Employee-Management relations**

Importance, Behavioral Traits, Dimensions and smoothening of inter-cultural differences

**Written Communication:** Enquiry, Reply, Offer, Sales Letters, Grievances and Replies, Memos, Reports, Business Proposal

## **Unit - 2 Presentations – Slides**

Planning and drafting, ppt presentation enhancing – style, fonts, background, content, presentation to audience

## **Unit - 3 Presentations – audio – visual**

Planning and drafting, AV presentation – Aids, Style, Content, Shooting and Recording

## **Unit - 4 Debates**

### **Reference Books:**

1. Business Communication by Urmila Rai & S. M. Rai, Himalaya Publishing House, Mumbai
2. Business Communication by Asha Kaul, Prentice-Hall, New Delhi
3. Essentials of Business Communication by Rajendra Pal & J. S. Korlahalli, Sultan Chand & Sons, New Delhi
4. Business Communication (Principles, Methods and Techniques) by Nirmal Singh, Deep & Deep Publications, New Delhi
5. Effective Business Communication by Murphy G. A., Hildebrandt W. H., Thomas J. P., Tata McGraw Hill, New Delhi

# **CC-111 Calculus and Introduction to Matrices**

## **Unit - 1 Differentiation**

Definition of derivatives, properties and formulas, formulas-based examples, chain rule for a composite function

## **Unit - 2 Integration**

Definition of indefinite integral (only statements), Some theorems on antiderivative, formulas based standard integrals and its examples, properties of integration, Definition of definite integral and its based examples

## **Unit - 3 Applications of differentiation and integration**

Rate of change, velocity and acceleration-based examples, business application examples, maxima and minima of function based on one variable, mean value theorem and its examples, Maclaurin's and Taylor's series base example, integration: Area bounded by the curve (excluding volume), business problems-based application

## **Unit - 4 Matrices and Determinants**

Definition,  $2 \times 2$ ,  $3 \times 3$  order, Minors and cofactors-based examples, Matrices: definition of matrices, different type of matrices, elementary operations on matrices, adjoint of matrices, Inverse matrices, Boolean matrix, application

## **Reference Books:**

1. Thomas' Calculus by Hass, Heil, Weir, Pearson 14<sup>th</sup> edition
2. Calculus by Smith, Milton, Mc. Graw Hill 4<sup>th</sup> edition
3. For Unit 4: Discrete Mathematical Structures, Bernard Kolman, Robert Busby, Sharon C. Ross, Pearson 6<sup>th</sup> edition
4. For Unit 4: Discrete Mathematics and Its Applications, Kenneth H. Rosen, Tata Mc. Graw Hill 4<sup>th</sup> edition



# CC-112 Numerical and Statistical Methods

## Unit - 1 Roots of Equations

Theory of errors, Intermediate value theorem, Bisection method and its examples, Newton-Raphson method and its examples, De'cartes rule of sign, Budan's theorem and its examples

## Unit - 2 Interpolation

Newton's - forward, backward interpolation formula and its examples, Lagrange formula for equal and unequal intervals, Newton's divided difference formula for unequal intervals, Inverse interpolation formula (Lagrange), Extrapolation

## Unit - 3 Correlation and Regression

Definition of Correlation, Types of Correlation, Methods of Correlation – (i) Karl Pearson Correlation Coefficient (ii) Spearman Rank Correlation Method, Definition of Regression, Types of Regression (both straight line  $x$  on  $y$  and  $y$  on  $x$ ) and examples based on data

## Unit - 4 Curve Fitting

Least Square methods, Straight Line fitting to the given data, Quadratic type, Exponential fit of curve, Logarithm fit of curve to the given data and its examples

## Reference Books:

1. Numerical Methods, E Balaguruswami, TMH
2. Numerical Analysis and Computational Procedures, S. A. Mollah, Book and Allied Ltd.
3. Fundamentals of Mathematical Statistics, S. C. Gupta & V. K. Kapoor, Sultan Chand & Sons
4. Statistical Methods, S. P. Gupta, Sultan Chand & Sons

# CC-113 Algorithms and Data Structures

## Unit - 1 Introduction to Algorithms and C - Programming Language

Algorithms, Flowcharts, Datatypes and Variables, Conditional Statements (If statements), Loops, Arrays, Pointers, Built-in and User Defined Functions, User Defined Datatypes, Dynamic Memory Allocation

## Unit - 2 Searching - Sorting Algorithms and Introduction to Data Structures

Asymptotic Notations and Complexity, Bubble sort, Selection sort, Insertion sort, Quick sort, Merge Sort, Introduction to data structures, Linear and Non-Linear data structures, Link-Lists and their types

## Unit - 3 Stacks and Queues

Introduction to stack, Applications of stacks, operations on stack using arrays and link lists, Introduction to Queues and their implementations using array and link lists, Circular queue, Dequeue

## Unit - 4 Trees and Graphs

Introduction to Tree, Traversal in tree – In-Order, Pre-Order and Post-Order traversal. Types of tree – Full binary tree, Complete Binary tree, Perfect binary tree, Introduction to AVL tree, Insertion in AVL tree, Introduction to Graph, Terminologies related to graph, Traversal in graph, Representation of Graph, Prims, Kruskal and Dijkstra algorithms

## Reference Books:

1. Introduction to Algorithms, Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein
2. Algorithms Unlocked, Thomas H. Cormen
3. The Algorithm Design Manual, Steven S. Skiena
4. Data Structures and Algorithms Made Easy: Data Structures and Algorithmic Puzzles, Narasimha Karumanchi
5. Grokking Algorithms: An illustrated guide for programmers and other curious people, Aditya Bhargava
6. Algorithms, Robert Sedgewick and Kevin Wayne
7. Advanced Data Structures, Peter Brass
8. Automate This: How Algorithms Came To Rule Our World, Christopher Steiner

## Links:

<https://www.youtube.com/watch?v=NwenwITjMys>

<https://www.hackerearth.com/practice/algorithms/graphs/minimum-spanning-tree/tutorial/>

<https://www.gatevidyalay.com/prims-algorithm-prim-algorithm-example/>

# **CC-114 Object-Oriented Programming with JAVA**

## **Unit - 1 Introduction to OOPS**

Basic concepts of OOPS, Characteristics of OOPS, Advantages of OOPS, Java Introduction, Key features of Java, JVM, JDK tools, Variables, Operators, Decision making statements, Array

## **Unit - 2 Array, Classes and Objects**

Array, Class Introduction, Access modifiers, Copying objects, Constructors, This keyword, Final keyword, String and string buffer class

## **Unit - 3 Inheritance, Packages and Interface**

Inheritance introduction, Types of inheritance, Super keyword, Abstract class, Introduction, Package types, Interface

## **Unit - 4 Exception Handling, Thread**

Introduction, Exception handling, Exception classes, Creating your own exception, Creating Thread, Thread Priority

### **Reference Books:**

1. Programming in JAVA , Sachin Malhotra, Saurabh Choudhary, Oxford
2. JAVA Programming, Hari Mohan Pandey, Pearson
3. The Complete reference JAVA, Herbert Schildt, TMH Publication

# **CC-115 Programming with PYTHON**

## **Unit - 1 Python Introduction**

Download and Installation of Python, Features of Python, Comparison between C and Python, Java and Python, Introduction to IDLE, Keyword and Identifiers, Statement, Type of Comments, Basic Syntax and Input-Output, Variable, Exception Handling

## **Unit - 2 Python Data types and Operators**

Built-in Data Types: Number, String, List, Tuple, Set, Dictionary.

Operators: Arithmetic Operators, Assignment Operators, Relational Operators, Membership Operators, Logical Operators, Boolean Operators, Identity Operators

## **Unit - 3 Conditional Statements**

Python IF, ELSE, ELIF, Nested IF, Switch Case Statement, For Loop, While Loop, Nested Loop, Enumerate, Break Statement, Continue Statement, Pass Statement, Assert and Return Statement

## **Unit - 4 Python Function**

Definition of Function, Syntax of Function, Difference between Function and Method, Types of Function, Defining a Function, Function Calling, Return Statement, Function Argument, Types of Arguments, Global and Local Variables, Recursion, Advantage and disadvantage of Recursion, Anonymous Function or Lambda Function

## **Reference Books:**

1. Core Python Programming, By Dr. R. Nageswara Rao.
2. Python Programming A Modular Approach, By Sheetal Taneja & Naveen Kumar, 2018
3. Fundamentals of Python Programming, Halterman R., Southern Adventist University
4. Introduction to Computation and Programming Using Python, Guttag J.V., Prentice Hall India
5. Core Python Programming, Chun W., Prentice Hall

## CC-116 (P) Practical Based on CC-114

### Practical Based on CC-114 Object-Oriented Programming with JAVA

1. Write a JAVA Program for only print “Hello World” or any sentences.
2. Write a JAVA Program to calculate Sum/Multi/Div/Sub for any two or three numbers manually.
3. Write a JAVA Program to calculate Sum/Multi/Div/Sub for any two or three numbers Runtime.
4. Write a JAVA Program to calculate Average of any two or three floating point numbers.
5. Write a JAVA Program to calculate Area of Square, Triangle and Circle.
6. Write a JAVA Program to find number is positive or not.
7. Write a JAVA Program to find largest of two numbers.
8. Write a JAVA Program to find smallest of three numbers.
9. Write a JAVA Program to find number is even or odd.
10. Write a JAVA Program to display total mark wise class(grade) example:
  - a. 70 above class Distinction
  - b. 60 above class First
  - c. 50 above class Second
  - d. 40 above class Pass
  - e. Fail
11. Write a JAVA Program to using switch case take number as an input.
12. Write a JAVA Program to use switch case take string(word) as an input.
13. Write a JAVA Program to print while / do.. while & for loop.
14. Write a JAVA Program to print any star(\*) series using loop.
15. Write a JAVA Program to print any number(1) series using loop.
16. Write a JAVA Program to create default constructor.
17. Write a JAVA Program to print variable using default constructor.
18. Write a JAVA Program to create Sum/Multi/Div/Sub constructor for any two objects.
19. Write a JAVA Program to create average constructor for any two or three object.
20. Write a JAVA Program to create parameterize constructor using if- else.
21. Write a JAVA Program for constructor overloading or using “this” statement.
22. Write a JAVA Program to create inheritance using variable and various method.
23. Write a JAVA Program to create single level and multi-level inheritance.
24. Write a JAVA Program to create two separate JAVA file & use multi-level inheritance.
25. Write a JAVA Program to create Hierarchical Inheritance.
26. Write a JAVA Program to create three separate JAVA file & use Hierarchical inheritance.
27. Write a JAVA Program to calculate Sum/Multi/Div/Sub using Inheritance.
28. Write a JAVA Program to calculate average of two or three numbers using inheritance.
29. Write a JAVA Program to calculate Sum & subtraction using single level inheritance.
30. Write a JAVA Program to find Greatest & Smallest of two numbers using single level inheritance.
31. Write a JAVA Program to create interface for Addition & Subtraction.
32. Write a JAVA Program to use java.lang or java.math package in your program.
33. Write a JAVA Program to use java.io package in your program.
34. Write a JAVA Program for execute only try block.
35. Write a JAVA Program for execute only catch block.
36. Write a JAVA Program for execute only finally block.

37. Write a JAVA Program for execute catch & finally block using array.
38. Write a JAVA Program to use arithmetic exception in your program.
39. Write a JAVA Program to use array index out of bounds in your program.
40. Write a JAVA Program to use number format exception using string in your program.
41. Write a JAVA Program to use null pointer exception.
42. Write a JAVA Program to use string index out of bounds exception in your program.
43. Write a JAVA Program to use file not found exception in your program.
44. Write a JAVA Program to use class not found exception in your program.
45. Write a JAVA Program to execute nested catch in your program.
46. Write a JAVA Program to print multiple catch block in your program.
47. Write a JAVA Program to use throw keyword in your program.
48. Write a JAVA Program to use throws keyword in your program.
49. Write a JAVA Program to create your exception & execute it in your program.

## CC-117 (P) Practical Based on CC-115

### Practical Based on CC-115 Programming with PYTHON

#### String

1. Write a Python program to calculate the length of a string.
2. Write a Python program to count the number of characters (character frequency) in a string.  
Sample String: Gujarat University
3. Write a Python program to get a single string from two given strings, separated by a space and swap the first two characters of each string.  
Sample String: 'abc', 'xyz'  
Expected Result: 'xyc abz'
4. Write a Python program to change a given string to a new string where the first and last chars have been exchanged.
5. Write a Python program to remove the characters which have odd index values of a given string.
6. Write a Python program to count occurrences of a substring in a string.
7. Write a Python program to reverse a string.
8. Write a Python program to reverse words in a string.
9. Write a Python program to strip a set of characters from a string.
10. Write a Python program to count repeated characters in a string.

#### List

1. Write a Python program to sum all the items in a list.
2. Write a Python program to multiply all the items in a list.
3. Write a Python program to get the largest number from a list.
4. Write a Python program to get the smallest number from a list.
5. Write a Python program to remove duplicates from a list.
6. Write a Python program to generate all permutations of a list in Python.
7. Write a Python program to get the frequency of the elements in a list.
8. Write a Python program to generate all permutations of a list in Python.
9. Write a Python program to get the difference between the two lists.
10. Write a Python program access the index of a list.

#### Dictionary

1. Write a Python script to add a key to a dictionary.  
Sample Dictionary: {0: 10, 1: 20}  
Expected Result: {0: 10, 1: 20, 2: 30}
2. Write a Python script to concatenate following dictionaries to create a new one.

Sample Dictionary:

dic1= {1:10, 2:20}

dic2= {3:30, 4:40}

dic3= {5:50,6:60}

Expected Result: {1: 10, 2: 20, 3: 30, 4: 40, 5: 50, 6: 60}

3. Write a Python script to check whether a given key already exists in a dictionary.
4. Write a Python program to iterate over dictionaries using for loops.
5. Write a Python program to remove a key from a dictionary.
6. Write a Python program to map two lists into a dictionary.
7. Write a Python program to sort a given dictionary by key.
8. Write a Python program to remove duplicates from Dictionary.
9. Write a Python program to check a dictionary is empty or not.
10. Write a Python program to convert a list into a nested dictionary of keys.

### **Tuple**

1. Write a Python program to create a tuple with different data types.
2. Write a Python program to create a tuple with numbers and print one item.
3. Write a Python program to add an item in a tuple.
4. Write a Python program to convert a tuple to a string.
5. Write a Python program to get the 4th element and 4th element from last of a tuple.
6. Write a Python program to find the repeated items of a tuple.
7. Write a Python program to check whether an element exists within a tuple.
8. Write a Python program to convert a list to a tuple.
9. Write a Python program to remove an item from a tuple.
10. Write a Python program to find the length of a tuple.

### **Set**

1. Write a Python program to add member(s) in a set.
2. Write a Python program to create an intersection of sets.
3. Write a Python program to create a union of sets.
4. Write a Python program to check if a set is a subset of another set.
5. Write a Python program to remove all elements from a given set.
6. Write a Python program to find maximum and the minimum value in a set.
7. Write a Python program to find the length of a set.
8. Write a Python program to check if a given value is present in a set or not.
9. Write a Python program to check if two given sets have no elements in common.
10. Write a Python program to find the elements in a given set that are not in another set.



## Conditional Statement and Loop

1. Write a Python program to check the validity of password input by users.

Validation:

- At least 1 letter between [a-z] and 1 letter between [A-Z].
- At least 1 number between [0-9].
- At least 1 character from [!#\$%&@].
- Minimum length 6 characters.
- Maximum length 16 characters.

2. Write a Python program to find numbers between 100 and 400 (both included) where each digit of a number is an even number. The numbers obtained should be printed in a comma-separated sequence.

3. Write a Python program to print alphabet pattern 'A'.

Expected Output:

```
***
*   *
*   *
*****
*   *
*   *
*   *
```

4. Write a Python program to print alphabet pattern 'E'.

Expected Output:

```
*****
*
*
*****
*
*
*****
```

5. Write a Python program to print alphabet pattern 'G'.

Expected Output:

```
***
*   *
*
*   **
*   *
*   *
***
```

6. Write a Python program to print alphabet pattern 'M'.

Expected Output:

```
  *           *
 *           *
*  *       *  *
 *     *   *
 *           *
 *           *
 *           *
```

7. Write a Python program to print alphabet pattern 'Z'.

Expected Output:

```
*****
 *
 *
 *
 *
 *
*****
```

8. Write a Python program to calculate the sum and average of n integer numbers (input from the user). Input 0 to finish.

9. Write a Python program to create the multiplication table (from 1 to 10) of a number.

Expected Output:

Input a number: 6

6 x 1 = 6

6 x 2 = 12

6 x 3 = 18

6 x 4 = 24

6 x 5 = 30

6 x 6 = 36

6 x 7 = 42

6 x 8 = 48

6 x 9 = 54

6 x 10 = 60

10. Write a Python program to construct the following pattern, using a nested loop number.

Expected Output:

1

22

333

4444

55555

666666

7777777

88888888

999999999

1. Write a Python function to find the Max of three numbers.
2. Write a Python function to sum all the numbers in a list.  
Sample List: (8, 2, 3, 0, 7)  
Expected Output: 20
3. Write a Python program to reverse a string.  
Sample String: "1234abcd"  
Expected Output: "dcba4321"
4. Write a Python function to calculate the factorial of a number (a non-negative integer). The function accepts the number as an argument.
5. Write a Python function that takes a number as a parameter and check the number is prime or not.  
Note: A prime number (or a prime) is a natural number greater than 1 and that has no positive divisors other than 1 and itself.
6. Write a Python program to print the even numbers from a given list.  
Sample List: [1, 2, 3, 4, 5, 6, 7, 8, 9]  
Expected Result: [2, 4, 6, 8]
7. Write a Python function that checks whether a passed string is palindrome or not.  
Note: A palindrome is a word, phrase, or sequence that reads the same backward as forward, e.g., madam or nurses run.
8. Write a Python program to create a lambda function that adds 15 to a given number passed in as an argument, also create a lambda function that multiplies argument x with argument y and print the result.
9. Write a Python program to count the even, odd numbers in a given array of integers using Lambda.  
Original arrays:  
[1, 2, 3, 5, 7, 8, 9, 10]  
Number of even numbers in the above array: 3  
Number of odd numbers in the above array: 5
10. Write a Python program that sum the length of the names of a given list of names after removing the names that starts with an lowercase letter. Use lambda function.

# Gujarat University

## School of Emerging Science and Technology

### M.Sc. (Integrated) Five Years Program

### Data Science

### SEMESTER - 3

Sr No.	Course No.	Subject	Credit	Exam
1	(#)FC-201	Soft Skills I	2	Theory
2	CC-201	Matrix Algebra and Calculus	3	Theory
3	CC-202	Random Variables and Distributions	3	Theory
4	CC-203	Discrete Mathematics	3	Theory
5	CC-204	Operating System Concepts	4	Theory
6	CC-205	Database Management Systems using SQL and PL/SQL	4	Theory
7	CC-206 (P)	Data Analysis using Excel	3	Practical
8	CC-207 (P)	Practical Based on CC-205	3	Practical
		<b>Total</b>	<b>25</b>	

(#) Marks and credits of this subject will not be considered for class or total.

# **FC-201 Soft Skills I**

## **Unit 1: Introduction to Soft Skills**

Classification, Etiquettes at Work place, Verbal Communication skills – within and outside department / office

## **Unit 2: Telephone Skills**

Phone Etiquettes, Listening and response

## **Unit 3: Meetings**

Formal and informal meetings, Grapevine and office politics, Office Meetings, Corporate Annual Meetings, Board Meetings

## **Unit 4: Teams & Team Work**

Team building, working in groups / teams, Roles and role plays, Leading and managing

## **Reference Books:**

1. Personality Development 28th Reprint by Tata McGraw Hill New Delhi (2006).
2. Organizational Behavior by Stephen P. Robbins and Timothy A. Judge(2014), Prentice Hall (16th Edition)
3. How to Succeed at Interviews, 21st (rep.) by Andrews, Sudhir, Tata McGraw-Hill New Delhi. (1988).
4. Effective leadership. Essential Manager series by Heller Robert, Dk Publishing (2002).

# **CC-201 Matrix Algebra and Calculus**

## **Unit 1: Vectors and Geometry of Space**

Definition of vectors, Approaches of vectors: geometric approach and algebraic approach, position vector, magnitude of a vector, direction cosines, types of vectors, properties of vectors, dot product and cross product of vectors, projection of a vector on a line, application of vectors on work done by the particles and examples

## **Unit 2: Matrices and Systems of Linear Equation**

Some definitions associated with matrices, some special matrices, elementary transformations on matrices, equivalence of matrices, Echelon form of matrix, system of non-linear equations, Solution of system of linear equations: Gaussian Elimination and Gauss-Jordan Elimination method, System of Homogeneous linear system, Rank of a matrix, Application of systems of linear equation

## **Unit 3: Numerical Methods for Solution of Systems of Linear Equation**

Linear algebraic equation: Direct method and Iterative method, Gauss Elimination method with partial pivoting, Gauss-Jacobi method, Gauss-Seidel iterative method, LU factorization, Doolittle's method, Cholesky method, Well-conditioned, Ill-conditioned

## **Unit 4: Functions of Several Variable and Introduction to Partial Derivatives**

Definition of function of 'n' variable, graphs, level curves of functions of two variables, function of three variable, limit and continuity in higher dimension, partial derivatives of two variables, Clairaut's Theorem, total derivative, chain rule, implicit differentiation, tangent plane and normal to a surface, extreme value and saddle points, method of Lagrangian Multipliers, Taylor's formula for two variables.

## **Reference Books:**

1. Thomas' Calculus by Hass, Heil, Weir (Pearson 14<sup>th</sup> edition)
2. Calculus by Smith, Milton (Mc. Graw Hill 4<sup>th</sup> edition)
3. Advanced Engineering Mathematics by Erwin Kreyszing (Wiley 10<sup>th</sup> edition)
4. An Introduction to Linear Algebra by V. Krishnamurthy
5. Elementary liner Algebra: Application Version, Howard Anton & Chris Rorres (Wiley 10<sup>th</sup> edition)
6. Applied Numerical Linear Algebra by James W. Demmel (SIAM)

# CC-202 Random Variables and Distributions

## Unit 1: Random Variable and Generating Function

Random variables with its type - pmf, pdf, cdf with illustration, Expectation of Random variables with properties moments, Factorial moments, Measure of location, Skewness, kurtosis, Probability, Generating Functions(pgf), mgf, cgf, fmgf(Fractional moment generating function) with their properties and uses, Central Limit Theorem  
Probability Inequalities – Boole's, Bonferroni's, Markov's, Chebyshev's, Concept of Convex and Concave function, Jensen's, Cauchy and Schwarz

## Unit 2: Joint and Conditional Distribution

Concept of Joint Distribution, Joint probability mass function and Joint pdf, Marginal and Conditional Distribution, Independence of random variables, Expectation and Conditional expectation, mean and variance of linear combinations of random variables

## Unit 3: Discrete Distributions

Bernoulli Distribution, Binomial Distribution, Negative Binomial Distribution, Geometric Distribution, Hypergeometric Distribution, Poisson and Uniform Distribution

## Unit 4: Continuous Distributions

Uniform Distribution, Gamma Distribution, Beta Distribution, Weibull Distribution, Exponential Distribution, Normal Distribution, Lognormal Distribution, chi-square, t, F, Logistic Distribution

## Reference Books:

1. Statistics for Business and Economics by Anderson, Sweeney, Williams, Cengage Publishers (12<sup>th</sup> edition)
2. Applied Business Statistics by Black, Wiley (7<sup>th</sup> edition)
3. Fundamentals of Mathematical Statistics, S. C. Gupta & V. K. Kapoor, Sultan Chand & Sons
4. Statistical Methods, S. P. Gupta, Sultan Chand & Sons
5. Complete Business Statistics by Aczel, Sounderpandian, Saravanan, Joshi, McGraw-Hill (7<sup>th</sup> edition)
6. Statistics for Management by Levin, Rubin, Pearson (8th edition)
7. Applied Statistics in Business and Economics by David P. Doane, Lori E. Seward, McGraw-Hill (3<sup>rd</sup> Addition)
8. All of Statistics by Larry Wasserman, Springer
9. Probability and Statistics for Engineers by Richard A. Johnson, PHI (6<sup>th</sup> edition)

# CC-203 Discrete Mathematics

## Unit 1: Connectives, Normal forms and Theory of Inferences

Connectives, Negations, Disjunction, Conjunction, Exclusive Disjunction, Conditional, Bi-conditional, Well-formed formula, contingency, NAND & NOR Gates and its properties, Two state devices

Disjunctive normal forms, Conjunctive normal forms, Principal Conjunctive & Disjunctive normal forms, Valid inference using truth tables, rules, implications, Equivalence, Consistency of premises and indirect method of proof

## Unit 2: Relations & Ordering, Introduction to Group Theory

Relations, Relation in a set, Binary relation in a set, Domain and range of a relation, Total number of distinct relation from a set A to B, Graph of relations, Relations and set of ordered pairs, Types of relations in a set, Properties of relation in a set, Equivalence relation, Equivalence classes or sets, Partitions, Partial Order Relations, Hasse diagram, Upper and Lower Bounds, Minimal and Maximal element, Binary Operations, Closure Operation

Introduction to Group Theory: Binary operation, Structure of a group, Subgroup, Examples, Lagrange's theorem

## Unit 3: Posets & Lattices, Boolean algebra

Posets, Lattices as Posets, Lattices as algebraic systems, Sub-lattices, Complete Lattices, Bounds of Lattices, Modular and Distributive Lattices, Complemented Lattices, Chains

Boolean algebra: Definitions and important properties, Sub-Boolean Algebra, Boolean Homomorphism, Join irreducible, meet irreducible, Atoms and Anti-atoms, Stone's representation theorem, Boolean expression and equivalence, Min & Max terms, Value of Boolean expressions and Boolean functions, k-map

## Unit 4: Graph Theory

Definitions, Types of Graphs and properties, Hand-Shaking lemma, Sub-graphs and Isomorphic graphs, Path and Reachability, Node base, Matrix representation of Graphs, Connectedness, Node base, Trees, Matrix of a graph, Definitions and basic terms of Binary Trees, Prims and Kruskal Algorithm to find shortest path

## Reference books:

1. Mathematical Foundation of Computer Science, Ravi Gor, 6<sup>th</sup> Edition, Nirav Prakashan.
2. Discrete Mathematical and its Applications, Kenneth H. Rosen, 4th Edition, Tata McGraw Hill.
3. Discrete Mathematical Structures with Applications to Computer Science, J.P.Trembley, R. Manohar, Tata McGraw Hill.
4. Discrete Mathematical Structure, B. Kolman, R.C. Busby and S.C. Ross, 4<sup>th</sup> edition, PHI.



# CC-204 Operating System Concepts

## Unit - 1 Operating System Overview and Process Management

**Overview of Operating Systems:** Role of Operating Systems (OS), Operations of OS, Resource Management, Security and Protection, Virtualization, Distributed Systems, Kernel Data Structures, Computing Environments, Free and Open-Source OS, OS Services, User and OS Interface, System Calls, System Services, OS Specific applications, OS Design and Implementation, OS Structure, Building and Booting an OS, Overview of Containers and Application Containerization

**Process Management:** Process Concept, Process Scheduling, Operations on Process, Inter-process Communication, IPC in Shared-Memory Systems and Message-Passing Systems, Examples of IPC Systems, Communication in Client-Server Systems, Threads & Concurrency, Multicore Programming, Multithreading Models, Thread Libraries, Implicit Threading, Threading Issues, Concepts of CPU Scheduling, Scheduling Criteria, Scheduling Algorithms, Thread Scheduling, Multiprocessor Scheduling, Overview of Real-Time Scheduling, Algorithm Evaluation

## Unit - 2 Process Synchronization and Memory Management

**Process Synchronization:** Overview of Synchronization Tools, The Critical-Section Problem, Peterson's Solution, Hardware Support for Synchronization, Mutex Locks, Semaphores, Monitors, Liveness, Evaluation, Classical Problems of Synchronization, POSIX and Java Synchronization, Overview of Deadlocks, Deadlock in Multithreaded Applications, Deadlock Characterization, Methods for Handling Deadlocks, Deadlock Prevention, Deadlock Avoidance, Deadlock Detection, Recovery from Deadlock

**Memory Management:** Background, Contiguous Memory Allocation, Paging, Structure of the Page Table, Swapping, Virtual Memory Overview, Demand Paging, Copy-on-Write, Page Replacement, Allocation of Frames, Thrashing, Memory Compression, Allocating Kernel Memory, Considerations for Virtual Memory Management

## Unit - 3 Storage Management and File System

**Storage Management:** Overview of Mass Storage Structure, HDD Scheduling, NVM Scheduling, Error Detection and Correction, Storage Device Management, Swap-Space Management, Storage Attachment, RAID Structure, Overview of I/O Systems, Application I/O Interface, Streams, Performance of I/O

**File System:** Overview of File-System Interface, File-System Structure, File-System Operations, Directory Implementation, Allocation Methods, Free-Space Management, Efficiency and Performance, Recovery, File Systems, File-System Mounting, Partitions and Mounting, File Sharing, Virtual File Systems, Remote File Systems, NFS

## **Unit - 4 Protection in OS and Linux Shell Scripting**

**Protection in OS:** Goals of Protection, Principles of Protection, Protection Rings, Domain of Protection, Access Matrix, Implementation of Access Matrix, Revocation of Access Rights, Role-Based Access Control, Mandatory Access Control, Capability-Based Systems

**Linux Shell Scripting:** Basic commands of Linux, Process related commands of Linux, Linux APIs for process management and IPC, File related commands of Linux, Linux Shell Scripting to automate tasks

### **Reference books:**

1. Silberschatz, A., Peter B. Galvin and Greg Gagne, "Operating System Concepts", Wiley International Edition, 10<sup>th</sup> edition.
2. Venkateshmurthy, "Introduction to Unix and Shell Programming", Pearson Education (India)
3. Stallings W, "Operating Systems", 7<sup>th</sup> edition, Pearson Education.
4. Bach M J, "The Design of Unix Operating System", Prentice Hall India.
5. Flynn I. M, "Understanding Operating Systems", Cengage India Publication
6. Tanenbaum A.S., "Modern Operating Systems", 4<sup>th</sup> Edition, Pearson Education India
7. Eric Foster, John Welch, Micah Anderson, "Beginning Shell Scripting", Wrox Publication
8. Sibsankar Haldar, Alex Aravind, "Operating Systems", 1<sup>st</sup> edition Pearson Education

# **CC-205 Database Management Systems using SQL and PL/SQL**

## **Unit 1: Database Systems and Data Models**

Introducing the database and DBMS, Files and File Systems, Problems with file System and advantages of Database Management systems

The importance of Data models, Data Model Basic Building Blocks, Business Rules, The evaluation of Data Models, Degree of Data Abstraction, Relational Algebra, Current Trends in Database Systems

## **Unit 2: The Relational Database Model**

A logical view of Data, Keys, Integrity Rules, The Data Dictionary and the system catalog, Relationships within the Relational Database, Data Redundancy revisited, Indexes, Codd's relational database rules, Entity Relationship Model: The ER Model, Developing ER Diagram

## **Unit 3: Normalization of database tables**

Database Tables and Normalization, The need for Normalization, The Normal forms and High-level Normal Forms, demoralization

## **Unit 4: Transaction Management and Concurrency Control**

What is transaction, Concurrency control, and Concurrency control with locking Methods, Concurrency control with time stamping methods, concurrency control with optimistic methods, database recovery and recovery management

## **Reference books:**

1. Fundamentals of Database Systems, R. Elmasri, S. B. Navathe, Prentice Hall, New Delhi, 2007.
2. Database System Concepts, A. Silberschatz, H. F. Korth, S. Sudharshan, Tata McGraw Hill, New Delhi, 2005.
3. Introduction to Database Systems, C. J. Date, Prentice Hall, New Delhi, 2004

# **CC-206(P) Data Analysis using Excel**

## **Unit - 1 Basics of Spreadsheets**

Introduction to Excel functions and Charting – Text, Statistical, Lookup and referencing, Logical, date and time

(If, And, Or Max, Min), Count (Count, Countif, Countifs, CountA, countblank, Sum, Sum if, Sum ifs, Product, SumProduct, Average, AverageIfs), (Look up – Choose, Offset, V Lookup, H Lookup, Match, Index), (dates – year, year frac, month, day, date, days360, E date), Standard Deviation, Variance, Correlation. Charting – Column, Bar, Line, Pie Chart, Stock, XY, Surface, Doughnut, Waterfall

## **Unit - 2 Data validation and Iteration**

Concept of Data tables, Data Validation, Conditional Formatting and What if analysis

## **Unit - 3 Charts and Pivot Tables**

Charts: New Charts – Tree map & Waterfall, Sunburst, Box and whisker Charts, Combo Charts –Secondary Axis, Adding Slicers Tool in Pivot & Tables, Using Power Map and Power View, Forecast Sheet, Sparklines -Line, Column & Win/ Loss, Using 3-D Map, New Controls in Pivot Table – Field, Items and Sets, Various Time Lines in Pivot Table, Auto complete a data range and list, Quick Analysis Tool, Smart Lookup and manage Store

Pivot Tables: Creating Simple Pivot Tables, Basic and Advanced Value Field Setting, Classic Pivot table, Grouping based on numbers and Dates, Calculated Field & Calculated Items

## **Unit - 4 Probability Distributions**

Binomial Distribution, Negative Binomial Distribution, Hypergeometric Distribution, Exponential Distribution, Normal Distribution, Lognormal Distribution Gamma Distribution, Weibull Distribution, Chi-square Distribution, t Distribution, F Distribution

## **Reference Books:**

1. Microsoft® Office 2013 Bible
2. John Walkenbach - Excel 2007 Bible-Wiley (2007)
3. Excel Data Analysis For Dummies By Paul McFedries
4. Microsoft Excel 2019 Data Analysis and Business Modeling By Wayne Winston
5. MS Excel 2016 Charts & Basics of Dashboards: The Essentials of Excel 2016 Charts and Simple Steps to Create a Dashboard By Venus LearningLabs
6. Fundamentals of Mathematical Statistics, S. C. Gupta & V. K. Kapoor, Sultan Chand & Sons

## **CC-207 (P) Practical Based on CC-205**

### **Practical Based on CC-205 Database Management Systems using SQL and PL/SQL**

1. To study DDL-create and DML-insert commands.
2. Insert the data in table.
3. To Perform various data manipulation commands, aggregate functions and sorting concept on all created tables.
4. To study Single-row functions
5. Displaying data from Multiple Tables (join)
6. To apply the concept of Aggregating Data using Group functions.
7. To solve queries using the concept of sub query.
8. Manipulating Data
9. To study Transaction control commands

# Gujarat University

## School of Emerging Science and Technology

### M.Sc. (Integrated) Five Years Program

### Data Science

### SEMESTER - 4

Sr No.	Course No.	Subject	Credit	Exam
1	(#)FC-211	Soft Skills II	2	Theory
2	CC-211	Linear Algebra	3	Theory
3	CC-212	Statistical Inference Theory	3	Theory
4	CC-213	Vector Calculus	3	Theory
5	CC-214	Introduction to Natural Language Processing	4	Theory
6	CC-215	Python for Data Science	4	Theory
7	CC-216 (P)	R Programming-I	3	Practical
8	CC-217 (P)	Practical Based on CC-215	3	Practical
		<b>Total</b>	<b>25</b>	

(#) Marks and credits of this subject will not be considered for class or total.

## **FC-211 Soft Skill-II**

### **Unit 1: Soft Skills**

Etiquettes in Corporates, Communicating with internal and external customers

### **Unit 2: Listening**

Difference between hearing and listening, Active Listening and responses, Listening and salient notes, Practical – listening to article and answering questions on it

### **Unit 3: Writing Skills**

Writing short stories, Idioms, Phrases, Slangs, Use of phrases and idioms in stories

### **Unit 4: Professionalism**

Presenting yourself professionally, Ethics, Interpersonal Skills, Dressing and Grooming

### **Reference Books:**

1. Personality Development 28th Reprint by Tata McGraw Hill New Delhi (2006)
2. Organizational Behavior by Stephen P. Robbins and Timothy A. Judge (2014), Prentice Hall (16th Edition)
3. How to Succeed at Interviews, 21st (rep.) by Andrews, Sudhir, Tata McGraw-Hill New Delhi. (1988)
4. Effective leadership. Essential Manager series by Heller Robert, Dk Publishing (2002)

# CC-211 Linear Algebra

## Unit-1 Vector Spaces

Euclidean Vector Space, (Vector in  $R^n$ , Vector addition & multiplication, Properties of vectors in  $R^n$ , Inner (Dot) product in  $R^n$ , Properties of inner product in  $R^n$ , Distance in  $R^n$  & properties, Angle between vectors in  $R^n$ , Pythagoras theorem in  $R^n$ , Cauchy-Schwartz inequality in  $R^n$ ), Definition of vector spaces, Subspaces of Vector Spaces, Linear combination of vectors, Span, Linearly Dependent and Linearly Independent, Basis, Dimension of Vector Space.

## Unit-2 Basis and Dimension of a Vector Space on system of linear equations

Basis and Dimension for Solution Space of the Homogeneous system, Reduction & Extension to Basis, Coordinate vector relative to a basis, Change of basis, Row space, Column space and Null space, Rank and Nullity, Inner product space, Projections, Least square approximation, Orthogonal Bases & Gram-Schmidt process, Orthogonal & Orthonormal set, Orthogonal & Orthonormal basis.

## Unit-3 Linear Transformation

Euclidean Linear transformation, Some standard linear transformation, Properties of linear transformation, Linear Operators (Reflection, Projection, Rotation, Dilation, Contraction, Shear), Linear transformation Images of Basis vectors, Composition of Linear transformation, Dimensional Theorem for Linear transformation, One-to-One transformation, Onto transformation, Isomorphism, Inverse linear transformation, The matrix of a linear transformation, Matrices of linear operations, Effect of change of bases on linear operations.

## Unit-4 Eigen values and Eigen vectors

Definition and Geometric meaning of Eigen values and Eigen vectors, Nature of Eigen values of special types of matrices, Relation between Eigen values and Eigen vectors, Cayley-Hamilton theorem, Diagonalization, Similarity of matrices, Orthogonally similar matrices, Power of matrix, Quadratic form Value class or Nature of Quadratic form, Methods to Reduce Quadratic form to Canonical form: (1) Orthogonal transformation (2) Congruent transformation.

## Reference Books:

1. Linear Algebra and Optimization for Machine Learning by Charu C. Aggarwal (Springer).
2. Calculus by Smith Milton (MC. Graw Hill 4<sup>th</sup> Edition).
3. Thomas' Calculus by Hass, Heil, Weir (Pearson 14<sup>th</sup> Edition).
4. An Introduction to Linear Algebra by V. Krishnamurthy (Affiliated East- West Press).
5. Advanced Engineering Mathematics by Erwin Kreyszing (Wiley 10<sup>th</sup> Edition).
6. Elementary Linear Algebra: Application Version, Howard Anton and Chris Rorres (Wiley 10<sup>th</sup> Edition)



# CC-212 Statistical Inference Theory

## Unit – 1: Sampling and Sampling Distribution, Confidence Interval

Random Sample Techniques, Non-random Sampling, Sampling and Non-sampling Errors, Sampling Distribution of Means, Sampling Distribution of Population, Properties of Point Estimators, Interval Estimation: Population Mean ( $\sigma$ : known & unknown), Determining Sample Size, Population Proportion

## Unit – 2: Hypothesis Test: One Sample

Null hypothesis, Alternative hypothesis, simple and composite hypothesis, Type- I & type-II Errors,

Test of significance of population Mean ( $\sigma$ : known & unknown), Test of significance of Population Proportion, Calculating the Probability of Type II errors, Determination of Sample size.

## Unit – 3: Hypothesis Test: Two Sample

Sampling Distribution of Difference of Two Sample Means, Test of significance of difference between two Population Mean ( $\sigma_1$  &  $\sigma_2$ : known & unknown), Difference between two Population Mean Dependent Samples, Test of significance of Difference between two Population Proportion, Inference about Population Variance and Inference about two Population Variances, F-distribution

## Unit – 4: Non-Parametric tests

Chi-square test: Population Proportion for three or more Proportion, Test of Independence of Attributes, Goodness of fit (Binomial and Normal Probability Distribution).

Runs Test, Sign Test, Wilcoxon Signed-Rank Test, Mann-Whitney- Wilcoxon Test, Kruskal-Wallis Test, Friedman Test, Spearman's Rank Correlation

## Reference Books:

1. Statistics for Business and Economics by Anderson, Sweeney, Williams (Cengage Publishers 12<sup>th</sup> edition)
2. Applied Business Statistics by Black (Wiley 7<sup>th</sup> edition)
3. Complete Business Statistics by Aczel, Sounderpandian, Saravanan, Joshi, McGraw-Hill (7<sup>th</sup> edition)
4. Statistics for Management by Levin, Rubin, Pearson (8th edition)
5. Applied Statistics in Business and Economics by David P. Doane, Lori E. Seward, McGraw-Hill (3<sup>rd</sup> Addition)
6. All of Statistics by Larry Wasserman, Springer
7. Probability and Statistics for Engineers by Richard A. Johnson, PHI (6<sup>th</sup> edition)

# CC-213 Vector Calculus

## Unit - 1 Vector Valued Functions and Motion in Space

Vector function of a single scalar variable, curves in space and their tangents, derivatives and motion, vector functions of constant length, tangent, normal and Bi-normal vectors, arc length in space, curvature and torsion

## Unit - 2 Scalar and Vector Point Function

Scalar point function, vector point function, vector differential operator  $\text{del}(\Delta)$ , gradient, directional derivative, maximum directional derivative, divergence, physical interpretation of divergence, solenoidal function

## Unit - 3 Curl, Integral and Vector Field

Curl, physical interpretation of curl, properties of curl, line integral of vector function, scalar potential function, vector fields, circulation and flow integral, flux, green's theorem in the plane

## Unit - 4 Counting, Mathematical Reasoning and Induction, Recursion

**Counting:** Permutation and Combinations, Pigeonhole principle, Inclusion-Exclusion principle

**Mathematical Reasoning and Induction:** Rules of inference, direct proof, Proof by Contradiction, Proof by Contrapositive, Mathematical Induction and Second law of Mathematical induction

**Recursion:** Definition, Recursion algorithm, Recurrence relations, solving recurrence relations

## Reference Books:

1. Thomas' Calculus by Hass, Heil, Weir, Pearson (14<sup>th</sup> edition)
2. Discrete Mathematical Structure, B. Kolman, R.C. Busby and S.C. Ross, PHI (4<sup>th</sup> edition)
3. Calculus by Smith, Milton, Mc. Graw Hill (4<sup>th</sup> edition)
4. Advanced Engineering Mathematics by Erwin Kreyszing, Wiley (10<sup>th</sup> edition)
5. An Introduction to Linear Algebra by V. Krishnamurthy
6. Vector Algebra by Shanti Narayan & P.K. Mittal, S. Chand

# **CC-214 Introduction to Natural Language Processing**

## **Unit-1 Introduction to Natural Language Processing**

Origin of NLP, The challenges of NLP, The Ambiguity and Uncertainty in Language, NLP application, Some Successful Early NLP Systems, NLP phases or Steps, Language Modelling, Corpus, WordNet, Sampling, TreeBank Corpus, Types of TreeBank Corpus, Applications of TreeBank Corpus, WordNet

## **Unit-2 Word, Syntax and Semantic Analysis**

Regular expression (RE), Finite State automata (FSA), Types of Finite state automation, morphemes, Stems, Word Order, Lexicon, Parsing, Top-down Parsing, Bottom-up Parsing, Derivation, Concept of Parse Tree, Context free grammar (CFG), Difference between Lexical and Semantic Analysis, Elements of semantic Analysis

## **Unit-3 Word sense disambiguation**

Evaluation of Word sense disambiguation (WSD), Methods to WSD, Major difficulties in Word sense disambiguation, Discourse Processing, Concept of Coherence, Tagging, Part-of-Speech (PoS) tagging, rule-based POS tagging, Properties of rule-based POS tagging, Stochastic POS Tagging, Properties of Stochastic POS Tagging

## **Unit-4 Natural Language Grammar and Information Retrieval**

Components of Language - Phonology, phonemes, Morphology, lexeme, Syntax, Semantic, Pragmatics, Grammatical categories – Tense, Aspect, Mood, agreement, Process of information retrieval (IR), Goal of information retrieval (IR), Types of Information Retrieval, Features of IR systems, Stemming

## **Reference Books:**

1. Natural Language Processing with Python by Steven Bird, Ewan Klein and Edward Loper  
O'Reilly Media, 2009 (1<sup>st</sup> edition)
2. Speech and Language Processing by Daniel Jurafsky and James H. Martin, Prentice Hall,  
2008 (2<sup>nd</sup> edition)
3. Foundations of Statistical Natural Language Processing by Christopher D. Manning and  
Hinrich Schuetze, MIT Press, 1999

# CC-215 Python for Data Science

**Prerequisites:** Basic Knowledge of Python Programming

## **Unit-1: OOPs Concepts in Python:**

OOPS Concepts, Creating Classes and objects, Pass Statement, Self-Parameter, Modify and Delete Object Properties, Creating Methods, Constructors, Inheritance, Super Function, Data Hiding, Data Encapsulation, Polymorphism, Type Identification, Python Operator Overloading, Special Functions, Iterators, Generators.

## **Unit-2: GUI (Graphical User Interface)**

Introduction of Graphical User Interface, Introduction to Tkinter, How to Import Tkinter module, Create the container/main window, Working with Containers, Introduction to Tkinter Widgets and their standard attributes, Tkinter Geometry Management Methods

Tkinter Attributes: Dimensions, Colors, Fonts, Anchors, Relief Styles, Bitmaps, Cursors

Tkinter Widgets: Button, Canvas, Checkbutton, Entry, Frame, Label, Listbox, Menubutton, Menu, Message, Radiobutton, Scale, Scrollbar, Text, Toplevel, Spinbox, Panedwindow, Labelframe, Messagebox

Applications based on GUI.

## **Unit-3: Important Libraries in Python**

**NumPy:** Introduction to NumPy, Installing and Importing NumPy, NumPy Standard Data Types, Basics of NumPy Array, NumPy Array Attributes, Array Indexing and Slicing, Array shape and Reshape, Array Join, Split, Search, Sort and Filter, Array Iteration, Array Transpose, NumPy Matrix, NumPy Linear Algebra, Trigonometric Functions, Statistical Functions, String Functions, Bitwise Operators, Arithmetic Operators, Broadcasting, Copies and Views.

**Pandas:** Introduction to Pandas, Installing and Importing Pandas, Features of Pandas, Benefit of Pandas, Pandas Data Structure, Difference between Series and DataFrame, How to Create a Pandas DataFrame, Viewing Data, How to Select an Index or Column from a DataFrame, How to add Row and Column to a DataFrame, How to Delete Rows and Columns from a DataFrame, How to Rename the Columns or Indices of a DataFrame, How to Iterate over a DataFrame, How to write a DataFrame to a File, Manipulating Data, Pivot Table, Working with Date and Time, Working with CSV and Excel files, Handling Missing Data.

**SciPy:** Introducing SciPy, Sub Packages in SciPy, Special Functions in SciPy, Interpolation Functions, Fourier Transform Functions, Signal Processing Functions, Image Manipulation Functions, Linear Algebra Operations, Statistic Functions, File Input/Output, SciPy Constant.

## **Unit-4: Libraries for Machine Learning and Deep Learning:**

**SciKit Learn:** Introducing SciKit Learn, Loading Data Set, Cleaning Data Set, Visualizing Data Set, Normalizing Data Set for Training, Splitting Data Set into Training and Testing, Model

Selection, Training and Testing Model, Different performance metrics for evaluating model performance.

**Tensorflow Keras API:** Introduction to Tensorflow, What is Tensor, How to create different types of Tensors, Shape, Rank, Size and Dimension of Tensor, Indexing of Tensor, Tensorflow Variable and Constant, Manipulation of Shapes, Broadcasting, Ragged Tensor, String Tensor, Sparse Tensor, Various Tensor Operations, Casting Tensor to NumPy, Gradient Tape, How to create Layers, Training and Evaluating Models, Saving and Loading Models.

**Text Books:**

1. Dr. R. Nageswara Rao, Core Python Programming, dreamtech, 2<sup>nd</sup> Edition, 2019
2. Martin C. Brown, Python The Complete Reference, Mc Graw Hill, 2019
3. Michael Urban and Joel Murach, Python Programming, Shroff/Murach, 2016
4. Mark Lutz, Programming Python, O`Reilly, 4th Edition, 2010
5. Python Data Science Handbook by Jake VanderPlas

**Reference Books:**

1. Programming Python: Powerful Object- Oriented Programming by Mark Lutz, O`Reilly; 4th edition
2. Python Cookbook: Recipes for Mastering Python by Brian Jones, David Beazley, O`Reilly Media; 3rd edition

**Link:**

<https://www.tensorflow.org/tutorials/quickstart/beginner>

# CC-216(P) Programming with R - I

## Unit-1 and 2: Introduction of R and RStudio

**Introduction:** Download and installation of R and RStudio, Introduction to R: Console window, shortcut keys, directory, help of RStudio, Installation of packages

**Basics of R:** Basic Math, Variables, Data Types, Vectors, Calling Function, Function Documentation, Missing data

### Data Structures:

- Vector – Character, Numeric, Integer, Complex, Logical
- Matrix & Array - Creating matrices, Matrix operations, Applying Functions to Matrix Rows and Columns, Adding and deleting rows and columns
- Data Frame - Creating Data Frames, Matrix-like operations in frames, Merging Data Frames, Applying functions to Data frames
- List - Creating lists, General list operations, accessing list components and values, applying functions to lists, recursive lists
- Factor and Table

**Flow Control:** ‘If’ statement, ‘else’ statement, ‘elseif’ statement, loop, ‘For’ loop, ‘While’ loop, ‘Repeat’ loops, function

## Unit-3 and 4: Data Handling through R

**Importing data files:** import.data function, read.table function; **Exporting data:** export.data function, cat, write, and write.table functions; Exporting Output and results - sink function, formatting output - options, and format functions, export.graph function;

**Data Visualization:** Box plot, Histogram, Scatter Plotter, Line Chart, Bar Chart, Pie Chart, Steam and Leaf Plot;

### Data Cleaning and Manipulation

**Statistical Command:** Basic statistical command, Calculating Probabilities from Probability Distributions, Descriptive statistics and graphics, one-sample and two-sample tests

### Reference Books:

1. R for Everyone by Jared P. Lander, Addison Wesley Data & Analytics Series, 2017 (2<sup>nd</sup> edition)
2. The Art of R Programming: A Tour of Statistical Software Design by Norman Matloff, no starch press, San Francisco, 2011
3. Statistics and computing: Introductory Statistics with R by Peter Dalgaard, Springer, 2002
4. Data Analysis and Graphics Using R: An example-based approach by Maindonald J. and Braum, J., Cambridge Series in Statistical and Probabilistic Mathematics, 2010 (3<sup>rd</sup> edition)
5. An R Companion to Linear Statistical Models by Hey-Jahans, C., CRC Press, 2012

## **CC-217 Practical Based on CC-215**

### **Practical Based on CC-215 Python for Data Science**

#### **GUI**

1. Write a Python GUI program to import Tkinter package and create a window and set its title.
2. Write a Python GUI program to import Tkinter package and create a window. Set its title and add a label to the window.
3. Write a Python GUI program to create a label and change the label font style (font name, bold, size) using tkinter module.
4. Write a Python GUI program to create a window and set the default window size using tkinter module.
5. Write a Python GUI program to create a window and disable to resize the window using tkinter module.
6. Write a Python GUI program to add a button in your application using tkinter module.
7. Write a Python GUI program to add a canvas in your application using tkinter module.
8. Write a Python GUI program to create two buttons exit and hello using tkinter module.
9. Write a Python GUI program to create a Combobox with three options using tkinter module.
10. Write a Python GUI program to create a Checkbutton widget using tkinter module.
11. Write a Python GUI program to create a Spinbox widget using tkinter module.
12. Write a Python GUI program to create a Text widget using tkinter module. Insert a string at the beginning then insert a string into the current text. Delete the first and last character of the text.
13. Write a Python GUI program to create three single line text-box to accept a value from the user using tkinter module.
14. Write a Python GUI program to create three radio buttons widgets using tkinter module.
15. Write a Python GUI program to create a ScrolledText widgets using tkinter module.
16. Write a Python GUI program to create a Progress bar widgets using tkinter module.
17. Write a Python GUI program to create a Listbox bar widgets using tkinter module.

#### **OOPS**

1. Write a Python program to import built-in array module and display the namespace of the said module.
2. Write a Python program to create a class and display the namespace of the said class.
3. Write a Python program to create an instance of a specified class and display the namespace of the said instance.

4. 'builtins' module provides direct access to all 'built-in' identifiers of Python. Write a python program which import the abs() function using the builtins module, display the documentation of abs() function and find the absolute value of -155.
5. Define a Python function student(). Using function attributes display the names of all arguments.
6. Write a Python function student\_data () which will print the id of a student (student\_id). If the user passes an argument student\_name or student\_class the function will print the student name and class.
7. Write a simple Python class named Student and display its type. Also, display the \_\_dict\_\_ attribute keys and the value of the \_\_module\_\_ attribute of the Student class.
8. Write a Python program to create two empty classes, Student and Marks. Now create some instances and check whether they are instances of the said classes or not. Also, check whether the said classes are subclasses of the built-in object class or not.
9. Write a Python class named Student with two attributes student\_name, marks. Modify the attribute values of the said class and print the original and modified values of the said attributes.
10. Write a Python class named Student with two attributes student\_id, student\_name. Add a new attribute student\_class and display the entire attribute and their values of the said class. Now remove the student\_name attribute and display the entire attribute with values.
11. Write a Python class named Student with two attributes student\_id, student\_name. Add a new attribute student\_class. Create a function to display the entire attribute and their values in Student class.
12. Write a Python class named Student with two instances student1, student2 and assign given values to the said instances attributes. Print all the attributes of student1, student2 instances with their values in the given format.
13. Write a Python class to convert an integer to a roman numeral.
14. Write a Python class to convert a roman numeral to an integer.
15. Write a Python class to find validity of a string of parentheses, '(', ')', '{', '}', '[' and ']'. These brackets must be close in the correct order, for example "()" and "()[]{}" are valid but "[D]", "({[D])" and "{{{" are invalid.
16. Write a Python class to get all possible unique subsets from a set of distinct integers.  
Input : [4, 5, 6]  
Output : [], [6], [5], [5, 6], [4], [4, 6], [4, 5], [4, 5, 6]
17. Write a Python class to find a pair of elements (indices of the two numbers) from a given array whose sum equals a specific target number.



Note: There will be one solution for each input and do not use the same element twice.

Input: numbers= [10,20,10,40,50,60,70], target=50

Output: 3, 4

*Difficulty: Medium. Company: Google, Facebook*

18. Write a Python class to find the three elements that sum to zero from a set of n real numbers.  
Input array : [-25, -10, -7, -3, 2, 4, 8, 10]  
Output : [[-10, 2, 8], [-7, -3, 10]]
19. Write a Python class to implement pow(x, n).
20. Write a Python class to reverse a string word by word.  
Input string : 'hello .py'  
Expected Output : '.py hello'
21. Write a Python class which has two methods get\_String and print\_String. get\_String accept a string from the user and print\_String print the string in upper case.
22. Write a Python class named Rectangle constructed by a length and width and a method which will compute the area of a rectangle.
23. Write a Python class named Circle constructed by a radius and two methods which will compute the area and the perimeter of a circle.
24. Write a Python program to get the class name of an instance in Python.

## **NumPy**

1. Write a NumPy program (using NumPy) to sum of all the multiples of 3 or 5 below 100.
2. Write a NumPy program to combine a one and a two dimensional array together and display their elements.
3. Write a NumPy program to compute the multiplication of two given matrixes.
4. Write a NumPy program to compute the cross product of two given matrixes.
5. Write a NumPy program to compute the determinant of a given square array.
6. Write a NumPy program to compute the eigenvalues and right eigenvectors of a given square array.
7. Write a NumPy program to compute the inverse of a given matrix.
8. Write a NumPy program to compute the mean, standard deviation, and variance of a given array along the second axis.
9. Write a NumPy program to compute the covariance matrix of two given arrays.
10. Create a 2d array with 1 on the border and 0 inside

## **Pandas**

1. Write a Pandas program to sort the DataFrame first by 'name' in descending order, then by 'score' in ascending order.
2. Write a Pandas program to change the name 'James' to 'Suresh' in name column of the data frame.
3. Write a Pandas program to insert a new column in existing DataFrame.
4. Write a Pandas program to change the order of a DataFrame columns.
5. Write a Pandas program to count city wise number of people from a given of data set (city, name of the person).
6. Write a Pandas program to join the two given dataframes along rows and assign all data.
7. Write a Pandas program to append rows to an existing DataFrame and display the combined data.
8. Write a Pandas program to join (left join) the two dataframes using keys from left dataframe only.
9. Write a Pandas program to split the following dataframe by school code and get mean, min, and max value of age for each school.
10. Using the following dataset find the mean, min, and max values of purchase amount (purch\_amt) group by customer id (customer\_id).

## **Scipy**

1. Implement K-means clustering using scipy. Create two clusters.
2. Find out root of  $x^3 - x - 1$  using bisection method. Consider lower and upper limit as -1 and 5.
3. Implement linear and cubic interpolation using scipy.
4. Find out  $x_1, x_2, x_3, x_4$  and  $x_5$ . using scipy.

$$\left\{ \begin{array}{l} x_1 + 9x_2 + 2x_3 + x_4 + x_5 = 170 \\ 10x_1 + x_2 + 2x_3 + x_4 + x_5 = 180 \\ x_1 + 5x_3 + x_4 + x_5 = 140 \\ 2x_1 + x_2 + x_3 + 2x_4 + 9x_5 = 180 \\ 2x_1 + x_2 + 2x_3 + 13x_4 + 2x_5 = 350 \end{array} \right.$$

5. Find out determinant of [9 , 6] , [4 , 5] using scipy.
6. Find Singular Value Decomposition (SVD) of a Matrix [4 , 0] , [3 , -5] using scipy.
7. How to read an image from a system using scipy.
8. How to crop an image using scipy.
9. How to rotate an image using scipy.
10. How to change a contrast of an image using scipy.

### **Scikit Learn**

1. Importing important libraries Numpy, Pandas, Matplotlib, Seaborn etc. and loading the dataset(.csv/.xls)
2. How to pre-process the data
3. How to check missing values in dataset
4. How to transform the data
5. Visualizing the data using Scatterplot, Heatmap etc.
6. Feature Engineering and Selecting for required Features
7. Splitting the data into training, testing and validation set
8. Fitting the Model
9. Predict on unseen data
10. Predict on testing data
11. Calculate Accuracy
12. Calculate Model Performance

## **Tensorflow**

1. How to create different types of
  - Tensors
  - Shape
  - Rank
  - Size
2. Dimension of Tensor
3. Indexing of Tensor
4. Tensorflow Variable and Constant
5. Manipulation of Shapes,
6. Various Tensor Operations
7. Casting Tensor to NumPy

# Gujarat University

## School of Emerging Science and Technology

### M.Sc. (Integrated) Five Years Program

### Data Science

### SEMESTER - 5

Sr No.	Course No.	Subject	Credit	Exam
1	(#)FC-301	Scientific Writing	2	Theory
2	CC-301	Differential Equations	3	Theory
3	CC-302	Regression Theory	3	Theory
4	CC-303	Distributed Platforms	4	Theory
5	CC-304	Machine Learning-I	4	Theory
6	CC-305 (P)	Data Visualization	3	Practical
7	CC-306 (P)	Practical Based on CC-303	3	Practical
8	CC-307 (P)	Practical Based on CC-304	3	Practical
		<b>Total</b>	<b>25</b>	

(#) Marks and credits of this subject will not be considered for class or total.

## **FC-301 Scientific Writing**

What is Scientific Writing, Scientific Communication, Writing a Research Paper, How to Review Literature, How to communicate the Methodology of Research, Formulating a Research Hypothesis, Methodology, Discussion and Conclusions, How to get the research published, Writing a Research / Project Proposal

# CC-301 Differential Equations

## Unit 1: Ordinary Differential Equations and Applications

Differential equations, Ordinary Differential Equations of first order and first degree, Applications of first-order differential equations, Homogeneous Linear Differential Equations of Higher Order with constant co-efficient, Non-Homogeneous Linear Differential Equations of Higher Order with constant co-efficient

## Unit 2: Partial Differential Equations with first order

Partial Differential Equations, Formation of Partial Differential Equation, Solution of Partial Differential Equations, Linear Partial Differential Equations of first order, Non-Linear Partial Differential Equations of first order

## Unit 3: Partial Differential Equations with Higher order

Charpit's method, Homogeneous Linear Partial Differential Equations with constant co-efficients, Non-Homogeneous Linear Partial Differential Equations with constant co-efficients, Classification of Second order Linear Partial Differential Equations - Heat, Wave and Laplace equation

## Unit 4: Numerical Methods for first order and first-degree differential equations

Newton-cotes formula, Simpson's One-third Rule, Trapezoidal Rule, Euler's Method, Euler's Modified Method, Picard's Method of Successive Approximation, Runge-Kutta Methods

## Reference Books:

1. Schaum's outline of theory and problems of partial differential equations by Paul DuChateau, D. Zachmann (Schaum's Series, McGraw-Hill publication)
2. Numerical Analysis and Computational Procedures by S A. Mollah (Central publication)
3. Introductory Methods of Numerical Analysis by S.S. Sastry (Prentice-Hall of India 3<sup>rd</sup> edition)
4. Thomas' Calculus by Hass, Heil, Weir (Pearson 14<sup>th</sup> edition)
5. Advanced Engineering Mathematics by Erwin Kreyszing (Wiley 10<sup>th</sup> edition)

# **CC-302 Regression Theory**

## **Unit 1: Design of Experiment and Analysis of Variance**

Design of Experiment: Completely Randomized Design, Analysis of Variance (One Way & Two Way), Multiple Comparison Procedures, Randomized Block Design

## **Unit 2: Linear Regression-1**

Regression equation, Estimation of regression equation, Standard error of the estimate, Coefficient of Determination, assumptions of the regression model, Using Residuals to test the assumption of the regression model

## **Unit 3: Linear Regression-2**

Test for significance for the slope of the regression model and testing the overall model, Estimation: Confidence intervals to estimate the mean and Prediction intervals to estimate an individual value, using regression to develop a forecasting trend line, Case study-based examples through computer output

## **Unit 4: Time Series Analysis and Forecasting**

Components of Time Series, Forecast Accuracy, Moving Average and Exponential Smoothing, Trend Analysis and Projection, Seasonality and Trend, Time Series Decomposition

## **Reference Books:**

1. Statistics for Business and Economics by Anderson, Sweeney, Williams, Cengage Publishers (12<sup>th</sup> edition)
2. Applied Business Statistics by Black, Wiley (7<sup>th</sup> edition)
3. Complete Business Statistics by Aczel, Sounderpandian, Saravanan, Joshi, McGraw-Hill (7<sup>th</sup> edition)
4. Statistics for Management by Levin, Rubin, Pearson (8<sup>th</sup> edition)
5. Introduction to Linear Regression Analysis, Douglas C. Montgomery, Elizabeth A. Peck, G. Geoffrey Vining, Wiley (3<sup>rd</sup> edition)
6. Regression Analysis by Example, Samprit Chatterjee, Ali S. Hadi, Wiley (5<sup>th</sup> edition)
7. Applied Statistics in Business and Economics by David P. Doane, Lori E. Seward, McGraw-Hill (3<sup>rd</sup> edition)
8. All of Statistics by Larry Wasserman, Springer
9. Probability and Statistics for Engineers by Richard A. Johnson, PHI (6<sup>th</sup> edition)



# CC-303 Distributed Platforms

## Unit 1: Hadoop

**Understanding Big Data and Hadoop:** Introduction to Big Data, Introduction to Hadoop, Features of Hadoop, Hadoop Ecosystem, Why do we need Hadoop

**Hadoop Architecture:** Main Hadoop Components, Hadoop Distributed File System (HDFS)

**Hadoop Components:** Hadoop MapReduce Framework, How Pig and Hive can be used in a Hadoop environment

## Unit 2: Tableau

Introduction to Tableau, How does Tableau work, Application of Tableau

**Tableau Product Suit:** Tableau Desktop, Tableau Public, Tableau Online, Tableau Server, Tableau Reader

**Tableau Server Architecture:** Data Server, Data Connectors, Components of Tableau Server, Gateway and Clients

**Fundamentals of Tableau:** Sorting, Grouping, and Filtering of Data, Types of Joins

**Visualization with Tableau:** Visual Analytics with Tableau, Table Calculation in Tableau, Mapping, Dashboard and Storytelling with Data

## Unit 3: WEKA

Introduction to Data Mining, Introduction to Weka, What is Weka, Getting started with Weka GUI, Different File Formats in Weka, Loading Data, Preprocessing Data, Building Classifier and Regression model using Weka, Clustering Data using Weka, Attribute Selection in Weka

## Unit 4: Excel

**Excel Fundamentals:** Understanding the Excel interface, Formatting in Excel, Data Types in Excel, Applying Sort, Filter, UPPER, LOWER, LEFT, RIGHT, TRIM, LEN, CONCAT functions, Finding the Outliers and Missing values in data

**Excel for Data Visualization:** Bar, Line, Columns and Pie charts, Histogram and Scatter plot, Waterfall and Hierarchical chart, Pivot table and chart, Stock charts, Heatmaps

**Excel for Data Analytics:** Regression Model using Excel

## **Reference Books:**

### **Hadoop**

1. Hadoop: The Definitive Guide By Tom White
2. Hadoop Beginner's Guide By Garry Turkington
3. Hadoop Real-World Solutions Cookbook By Jonathan Owens, Brian Femiano, Jon Lentz
4. Hadoop MapReduce v2 Cookbook By Thilina Gunarathne
5. Hadoop with Python by Zach Radtkia & Donald Miner, O'Reilly Media 2015

### **Excel**

1. Master business modeling and analysis techniques with Microsoft Excel 2019
2. Excel Data Analysis For Dummies
3. Managing Data Using Excel

### **Tableau**

1. Beautiful Visualization, Looking at Data Through the Eyes of Experts by Julie Steele, Noah Iliinsky
2. The Accidental Analyst: Show Your Data Who's Boss by Eileen and Stephen McDaniel
3. Information Dashboard Design: Displaying Data for At-a-glance Monitoring by Stephen Few
4. Learning Tableau by Joshua N Milligan

### **WEKA**

1. Machine Learning Mastering with WEKA by Jason Brownlee

# **CC-304 Machine Learning-I**

## **Unit 1: Introduction to Machine Learning**

What is Learning? Definition of Machine Learning, Types of Machine Learning, Importance of Machine Learning, Introduction to basic machine learning concepts, tasks, and workflow, Difference between Artificial Intelligence and Machine Learning, Training, Testing and Validation Set, Applications of Machine Learning

## **Unit 2: Supervised Machine Learning -1**

Regression Algorithm: Develop various algorithms viz Simple Linear regression, Multiple Linear regression, Logistic regression, Polynomial regression, Pros and Cons of each algorithm  
Classification Algorithm: Logistic Regression, K-Nearest Neighbor(K-NN), Support Vector Machine (SVM), Naive Bayes, Pros and Cons of each algorithm

## **Unit 3: Model Evaluation, Optimization & Regularization**

Optimization: Gradient Decent, Stochastic Gradient Decent, Batch Gradient Decent, Least Square Error, Maximum Likelihood, Cost Function, Parameter tuning, Overfitting, Underfitting, Bias-variance Tradeoff, Convex and Nonconvex function, Dimensionality Reduction-PCA, Regularization

## **Unit 4: Supervised Machine Learning - 2**

Develop more advanced algorithms viz Decision Tree, Introduction to Ensemble Learning, Commonly used Ensemble Methods (Bagging and Boosting, Random forest, AdaBoost etc.), Pros and Cons of each algorithm

## **Reference Books:**

1. Machine Learning for Absolute Beginners: A Plain English Introduction by Oliver Theobald
2. Introduction to Machine Learning with Python: A Guide for Data Scientists by Andreas C. Müller (Author), Sarah Guido (Author)
3. Machine Learning by Tom M. Mitchell
4. Machine Learning using Python by U Dinesh Kumar and Manaranjan Pradhan

# CC-305 (P) Data Visualization

## Unit 1: Introduction to Data Visualization, Data Ecosystem and Importing Datasets

Introduction to Data Visualization: what is data visualization, why is it important, how to use different libraries to visualize data, what are the benefits of data visualization

Data Ecosystem: Introduction to data, Univariate, Bivariate and Multivariate data

Importing Datasets: Identifying, Gathering, and Importing data, how to use and transform it

## Unit 2: Data Visualization using Matplotlib and Plotly libraries

Matplotlib: What is Matplotlib, how to plot Line and Points, how to use Markers, Labels, Grid, how to create subplots, Scatter diagram, Pie charts, Bar charts. Box plots, Contour plot and Histograms

Plotly: What is Plotly, how to plot Line chart, Scatter charts, Bar charts, Pie charts, Bubble charts, how to create Histogram, Box plot, Violin plot, Contour plot, Heatmap, how to create subplots and insert plots

## Unit 3: Data Visualization using Seaborn and ggplot (Grammar of Graphics) libraries

Seaborn: Changing Figure Aesthetics, how to use Color Palette, how to plot Bar Plot, Point Plot, Joint Plot, Pie Chart, Violin Plot, Heatmap, Box Plot, Swarm Plot, Pair Plot, Count Plot, how to use Pair Grid, Facet Grid, how to show Linear Relationships plot

Ggplot: Introduction to Grammar of Graphics, Components of Grammar of Graphics, Making Plots with ggplot

## Unit 4: Data Visualization using Folium and geoplolib libraries

Folium: Introduction to Folium, how to create interactive map using folium, Layers and Tiles in Folium, Plotting Markers and Paths on the Map, Heatmap using folium

Geoplolib: What is geoplolib, how to plot Dot plot, Heatmap and Markers, how to create Spatial graph and Choropleth map

## Reference Books:

1. Data Visualization in python by Daniel Nelson (Text Book)
2. Mastering Python Data Visualization by Kirthi Raman

Links:

1. <https://matplotlib.org/>
2. <http://seaborn.pydata.org/>
3. <https://plotly.com/python/>
4. <https://github.com/python-visualization/folium>

# **CC-306 (P) Practical Based on CC-303**

## **Practical Based on CC-303 Distributed Platforms**

### **Unit 1: Hadoop**

How does Hadoop Ecosystem work, Why do we need Hadoop, Main Hadoop Components, How to use Hadoop Distributed File System (HDFS)

How does MapReduce work in Hadoop, How Pig and Hive can be used in a Hadoop

### **Unit 2: Tableau**

Getting Started with Tableau, How to perform Sorting, Grouping, and Filtering of Data in Tableau, Types of Joins in Tableau

Visual Analytics with Tableau, Table Calculation in Tableau, Mapping, Dashboard and Story-telling with Data in Tableau

### **Unit 3: WEKA**

Getting started with Weka GUI, Different File Formats in Weka, How to Load Data, Preprocessing Data, Attribute Selection in Weka, Building Classifier and Regression model using Weka, Clustering Data using Weka

### **Unit 4: Excel**

Excel interface and Formatting in Excel, Data Types in Excel, Applying Sort, Filter, UPPER, LOWER, LEFT, RIGHT, TRIM, LEN, CONCAT functions, Finding the Outliers and Missing values in data

Bar, Line, Columns and Pie charts, Histogram and Scatter plot, Waterfall and Hierarchical chart, Pivot table and chart, Stock charts, Heatmaps

Regression Model using Excel

## **Reference Books:**

### **Hadoop**

1. Hadoop: The Definitive Guide By Tom White
2. Hadoop Beginner's Guide By Garry Turkington
3. Hadoop Real-World Solutions Cookbook By Jonathan Owens, Brian Femiano, Jon Lentz
4. Hadoop MapReduce v2 Cookbook By Thilina Gunarathne
5. Hadoop with Python by Zach Radtkia & Donald Miner, O'Reilly Media 2015

### **Excel**

1. Master business modeling and analysis techniques with Microsoft Excel 2019
2. Excel Data Analysis For Dummies
3. Managing Data Using Excel

### **Tableau**

1. Beautiful Visualization, Looking at Data Through the Eyes of Experts by Julie Steele, Noah Iliinsky
2. The Accidental Analyst: Show Your Data Who's Boss by Eileen and Stephen McDaniel
3. Information Dashboard Design: Displaying Data for At-a-glance Monitoring by Stephen Few
4. Learning Tableau by Joshua N Milligan

### **WEKA**

1. Machine Learning Mastering with WEKA by Jason Brownlee

## **CC-307 (P) Practical Based on CC-304**

### **Practical Based on CC-303 Machine Learning-I**

#### **Unit 1: Fundamentals of Machine Learning and Data Pre-processing**

Data Collection Process, Data Pre-processing, Data cleaning and wrangling using pandas, Features Generation and Selection, Feature scaling and Normalization, How to Test Overfitting and Underfitting, How to select algorithms for training purpose, Importance of Model Generalization

#### **Unit 2: Supervised Machine Learning - Part 1**

Develop various algorithms viz Linear regression, Logistic regression, Polynomial regression, Support Vector Machine etc. for both Classification and Regression methods. Pros and Cons of each algorithm

#### **Unit 3: Model Evaluation**

Evaluating the performance of machine learning model using Confusion matrix, Accuracy, Precision, Recall, F1 score, ROC (Receiver Operating Characteristics) curve. R Square/Adjusted R Square, Mean Square Error (MSE)/Root Mean Square Error (RMSE), Mean Absolute Error (MAE)

#### **Unit 4: Supervised Machine Learning - Part 2**

Develop more advanced algorithms viz Decision Tree including Ensemble Methods (Bagging and Boosting, Random Forest, AdaBoost etc.), Pros and Cons of each algorithm

#### **Reference Books:**

1. Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow 2, by Sebastian Raschka (Author), Vahid Mirjalili (Author)
2. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems by Aurélien Géron (Author)
3. Machine Learning with Python Cookbook: Practical Solutions from Preprocessing to Deep Learning by Chris Albon (Author)
4. Machine Learning using Python by U Dinesh Kumar and Manaranjan Pradhan

# Gujarat University

## School of Emerging Science and Technology

### M.Sc. (Integrated) Five Years Program

### Data Science

### SEMESTER - 6

Sr No.	Course No.	Subject	Credit	Exam
1	(#)FC-311	Personality Development	2	Theory
2	CC-311	Operations Research	4	Theory
3	CC-312	Research Methodology	3	Theory
4	CC-313	Machine Learning-II	4	Theory
5	CC-314 & CC-315	PROJECT - I: Mini - Project	6	Submission/ Presentation/ Viva
6				
7	CC-316 (P)	R Programming - II	3	Practical
8	CC-317 (P)	Practical Based on CC-313	3	Practical
		<b>Total</b>	<b>25</b>	

(#) Marks and credits of this subject will not be considered for class or total.



## **FC-311 Personality Development and GD-PI Training**

- Define Personality, Introvert, Extrovert & Ambivert person, Significance & Determinants of Personality Development, Perceptions, Perceptual Process, Self Esteem and Maslow, Self Esteem & Erik Erikson.
- Mind Mapping, Competency Mapping & 360 Degree Assessment
- Emotional intelligence and emotional quotient – Use and improvement
- Decision Making skills, Assertiveness Skills, Leadership & Qualities of Successful Leader, Attitude - Concept - Significance - Factors affecting attitudes
- Personal Branding; Golden rules of Personal branding

### **Reference Books:**

1. Hurlock, E.B (2006). Personality Development, 28th Reprint. New Delhi: Tata McGraw Hill.
2. Stephen P. Robbins and Timothy A. Judge (2014), Organizational Behavior 16th Edition: Prentice Hall.
3. Andrews, Sudhir. How to Succeed at Interviews. 21st (rep.) New Delhi. Tata McGraw-Hill 1988.
4. Heller, Robert. Effective leadership. Essential Manager series. Dk Publishing, 2002
5. Hindle, Tim. Reducing Stress. Essential Manager series. Dk Publishing, 2003

# CC-311 Operations Research

## Unit - 1: Linear Programming Problem and its solution techniques

Introduction to LPP, Formulation techniques of LP problems, Solution techniques of LP: Graphical Method and Simplex Method

## Unit - 2: Transportation Problem and Assignment Problem

Transportation Problem: Introduction, Mathematical formulation, Tabular representation, Methods for finding initial basic feasible solution: North West Corner Rule, Least Cost Method, Vogel's Approximation Method, Optimality test: MODI Method, Degeneracy in TP, Unbalanced TP

Assignment Problem: Introduction, Mathematical formulation, Method for solving AP: Hungarian Method, Unbalanced AP

## Unit - 3: PERT and CPM

Definition of a project, Job and events, Construction of project network, Rules for network construction, Critical Path Method (CPM), Concepts of float, total float and free float and its interpretations, Uncertain duration, PERT

## Unit - 4: Inventory Problems

Introduction, Types of inventories, Cost involved problems, Notations, EOQ model, Limitations of EOQ formula, EOQ model with finite replenishment rate, EOQ model with shortages

## Reference books:

1. Nita H. Shah, Ravi M. Gor, Hardik Soni, "Operations Research", Prentice-Hall India (2007).
2. J. K. Sharma, "Operations Research".
3. A. M. Natarajan, P. Balasubramani, A. Tamilarasi, "Operations Research", Pearson Education.
4. Hiller and Liberman, "Introduction to Operations Research", 9th edition, Tata McGraw-Hill
5. Hamdy A. Taha, "Operations Research: An Introduction", 8th edition, Prentice-Hall India (2007)
6. Wayne L. Winston, "Operations Research: Applications and Algorithms", 4th edition, Thomson
7. Anderson, Sweeney, Williams, "An Introduction to Management Science: Quantitative Approaches to Decision Making", 13th edition, South-Western College Publishing

# **CC-312 Research Methodology**

## **Unit - 1: Introduction to Research Methodology**

Meaning of Research, Objectives of Research, Types of Research, Process of Research, Research Applications in Business Decisions, how to define Research Problem, Formulation of Research Hypotheses, Structuring of Research Proposal

## **Unit - 2: Research Design and Data Collection**

The Nature of Research Methodology, Formulation of the Research Design, Classification of Research Designs, Necessary Conditions for making Causal Inferences, Concepts used in Experiments, Classification of Data, Research Applications of Secondary Data, Benefits and Drawbacks of Secondary Data, Evaluation and Classification of Secondary Data

## **Unit - 3 Measurement and Scaling**

Distinguishing Qualitative from Quantitative Data Methods, Methods of Qualitative Research, Focus Group Method, Personal Interview Method, Scale Characteristics and Levels of Measurement, Primary Scales of Measurement, Single vs Multiple item Scale and Comparative vs Non-comparative Scales, Comparative Scaling Techniques, Noncomparative Scaling Techniques, graphic Rating Scale, Itemized Rating Scale, Nonitemized Rating Scale, Multi item Scale, Measurement Error

## **Unit - 4 Questionnaire Design**

Types of questionnaires, Determining the type of questions, Criteria for question designing, Questionnaire structure, Physical characteristics of Questionnaire, Pilot testing of the Questionnaire, Administering the Questionnaire

### **Reference Books:**

1. Research Methodology: Concepts and Cases, Deepak Chawla and Neena Sondhi, Vikas Publishing House, 2<sup>nd</sup> edition
2. Marketing Research and Applied Orientation, Naresh Malhotra, Daniel Nunan, David Briks, Pearson Education, 5<sup>th</sup> edition

# CC-313 Machine Learning-II

## Unit - 1: Introduction to Unsupervised Machine Learning

What is Unsupervised Machine Learning, why use Unsupervised Learning, Working of Unsupervised Learning, Types of Unsupervised Learning, Unsupervised Learning Algorithms, Advantages and Disadvantages of Unsupervised Learning, Application of Unsupervised Learning, Supervised vs. Unsupervised Machine Learning

## Unit - 2: Unsupervised Machine Learning - Part 1

Different Distance methods, Develop various algorithms viz K-means, Fuzzy C means and Mean Shift Clustering. Pros and Cons of each algorithm, Importance of Dimensionality Reduction, PCA-Dimensionality Reduction, Singular Value Decomposition (SVD), Factor Analysis

## Unit - 3: Unsupervised Machine Learning - Part 2

Implementation of Hierarchical Clustering algorithm, Density-Based Spatial Clustering (DBSCAN) and Gaussian Mixed Models (GMM). Pros and Cons of each algorithm

## Unit - 4: Machine Learning in Practice

Utilization of Unsupervised Machine Learning to solve real world problems:

1. Customer segmentation
2. Market research
3. Market segmentation
4. Inventory management
5. Data preparation and visualization

## Reference Books:

1. Machine Learning for Absolute Beginners: A Plain English Introduction by Oliver Theobald
2. Introduction to Machine Learning with Python: A Guide for Data Scientists by Andreas C. Müller (Author), Sarah Guido (Author)
3. Machine Learning by Tom M. Mitchell
4. Machine Learning using Python by U Dinesh Kumar and Manaranjan Pradhan

**CC-314 & CC-315**  
**PROJECT – I: Mini - Project**

The students will be given guided project from the subjects they study in the semester. They will be evaluated on the basis of the regularly, internal evaluation, project presentation and viva-voce & document submission.

# CC-316 (P) Programming with R-II

## Unit - 1: Testing of Hypothesis

**Construction of Hypothesis:** Null hypothesis, Alternative hypothesis, simple and composite hypothesis

**Inference about Mean** for ( $\sigma$ -known &  $\sigma$ -unknown): One-Tailed test and Two-Tailed test, Type-I & Type-II Errors, Power of Test

Inference about Proportion, Inference about Two-Population ( $\sigma$ -known &  $\sigma$ -unknown): Population mean between Two Matched Sample, Population between Two Independent Sample, Comparison of Two Sample

## Unit - 2: Inference about Population variance, Goodness of Fit Tests and Non-Parametric Tests

**Inference about Population variance:** Interval Estimation, Hypothesis Testing for Variance, Inference about Two Population Variance

**Goodness of Fit Tests:** Multinomial Population, Test of Independence, Distribution Fitting (Normal and Poisson Distribution)

**Non-Parametric Tests:** Sign Test, Wilcoxon Signed Rank Test, Mann-Whitney-Wilcoxon Test, Kruskal-Wallis Test

## Unit - 3: Analysis of Variance and Simple Linear Regression Model

**Analysis of Variance:** Completely Randomized Design, Randomized Block Design, Factorial Design

**Simple Linear Regression Model:** Estimated Simple Linear Regression Equation, Coefficient of Determination, Testing for Significance, Confidence Interval, Prediction Interval, Residual Plots, Correlation

## Unit - 4: Package 'deSolve' for differential equations

Ordinary Differential Equations (ODE): Initial Value Problems, Runge-Kutta Method, Euler Method, Model Diagnostics, Plotting

Differential Algebraic Equations (DAE): DAE solver in R, A Simple DAE (up to 3), Nonlinear Implicit ODE, Plotting

Partial Differential Equations (PDE): Methods for solving PDE, Heat Equation, Wave Equation, Laplace's Equation, Plotting

Delay Differential Equations (DeDE): DeDE solver in R, Events & Roots, Difference Equation, Plotting

## Reference Books:

1. Peter Dalgaard “Statistics and computing: Introductory Statistics with R”, Springer (2002)
2. Tilman M. Davies “The Book of R: A First Course in Programming and Statistics”, no starch press, San Francisco (2016)
3. Karlina Soetaert, Jeff Cash, Francesca Mazzia “Solving Differential Equation in R” Springer (2012)
4. Dr. Marcel Dettling “Applied Time Series Analysis” Institute for Data Analysis and Process Design, Zurich University of Applied Sciences.
5. Jonathan D. Cryer and Kung-Sik Chan “Time Series Analysis with Application in R” 2<sup>nd</sup> Edition Springer (2008)
6. Jared P. Lander, “R for Everyone”, 2<sup>nd</sup> edition, Addison Wesley Data & Analytics Series (2017)
7. Maindonald J. and Braum, J., “Data Analysis and Graphics Using R: An example-based approach”, 3<sup>rd</sup> edition, Cambridge Series in Statistical and Probabilistic Mathematics (2010)
8. Hey-Jahans, C., “An R Companion to Linear Statistical Models”, CRC Press (2012)

# CC-317(P) Practical Based on CC-313

## Practical Based on CC-313 Machine Learning-II

### Unit - 1: Introduction to Unsupervised Machine Learning

What is Unsupervised Machine Learning, why use Unsupervised Learning, Working of Unsupervised Learning, Types of Unsupervised Learning, Unsupervised Learning Algorithms, Advantages and Disadvantages of Unsupervised Learning, Application of Unsupervised Learning, Supervised vs. Unsupervised Machine Learning

### Unit - 2: Unsupervised Machine Learning - Part 1

Different Distance methods, Develop various algorithms viz K-means, Fuzzy C means and Mean Shift Clustering. Pros and Cons of each algorithm, PCA-Dimensionality Reduction

### Unit - 3: Unsupervised Machine Learning - Part 2

Implementation of Hierarchical Clustering algorithm, Density-Based Spatial Clustering (DBSCAN) and Gaussian Mixed Models (GMM). Pros and Cons of each algorithm

### Unit - 4: Machine Learning in Practice

Utilization of Unsupervised Machine Learning to solve real world problems:

1. Customer segmentation
2. Market research
3. Market segmentation
4. Inventory management
5. Data preparation and visualization

### Practical Manual:

1. Implementation of K-means Clustering on self-made dataset
2. Implement K-means Clustering on Iris dataset to cluster different flower species based on their features
3. Implementation of K-means Clustering using Titanic dataset
4. Implementation of Fuzzy-C means Clustering on self-made dataset
5. Implement Fuzzy-C means Clustering using Iris dataset to cluster different flower species based on their features
6. Implement Mean Shift Clustering
7. Implement Density-Based Spatial Clustering (DBSCAN)
8. Implement Gaussian Mixed Models (GMM)
9. Solving the Wholesale Customer Segmentation problem using Hierarchical Clustering



10. Find out how much a customer has spent in the mall based on Annual Income using Hierarchical Clustering
11. Find out relations between purchased product using Association Rules
12. How to use Principal Component Analysis (PCA)

**Reference Books:**

1. Machine Learning for Absolute Beginners: A Plain English Introduction by Oliver Theobald
2. Introduction to Machine Learning with Python: A Guide for Data Scientists by Andreas C. Müller (Author), Sarah Guido (Author)
3. Machine Learning by Tom M. Mitchell
4. Machine Learning using Python by U Dinesh Kumar and Manaranjan Pradhan

# Gujarat University

## School of Emerging Science and Technology

### M.Sc. (Integrated) Five Years Program

#### Data Science

#### SEMESTER - 7

<b>Sr No.</b>	<b>Course No.</b>	<b>Subject</b>	<b>Credit</b>	<b>Exam</b>
1	CC-401	Advanced Algorithms	4	Theory
2	CC-402	Text, Image & Video Analytics	4	Theory
3	CC-403	Deep Learning	4	Theory
4	CC-404 (P)	Practical Based on CC-403	4	Practical
5	CC-405 (P)	Power BI & Tableau	4	Practical
6	CC-406	PROJECT – II: Project	5	Project
		<b>Total</b>	<b>25</b>	

# CC-401 Advanced Algorithms

## Unit - 1:

**Introduction:** Role of Algorithms in computing, Order Notation, Recurrences, Probabilistic Analysis and Randomized Algorithms. Sorting and Order Statistics: Heap sort, Quick sort and Sorting in Linear Time.

**Advanced Design and Analysis Techniques:** Dynamic Programming- Matrix chain Multiplication, longest common Subsequence and optimal binary Search trees.

## Unit - 2:

**Greedy Algorithms:** Huffman Codes, Activity Selection Problem. Amortized Analysis. Graph Algorithms: Topological Sorting, Minimum Spanning trees, Single Source Shortest Paths, Maximum Flow algorithms.

## Unit - 3:

**Sorting Networks:** Comparison Networks, Zero-one principle, Bitonic Sorting Networks, Merging Network, Sorting Network.

**Matrix Operations:** Strassen's Matrix Multiplication, Inverting matrices, Solving system of linear Equations

## Unit - 4:

**String Matching:** Naive String Matching, Rabin-Karp algorithm, matching with finite Automata, Knuth- Morris – Pratt algorithm

**NP-Completeness and Approximation Algorithms:** Polynomial time, polynomial time verification, NP-Completeness and reducibility, NP-Complete problems. Approximation Algorithms- Vertex cover Problem, Travelling Sales person problem.

## Reference Books:

1. Introduction to Algorithms," T.H. Cormen, C.E. Leiserson, R.L. Rivest, and C. Stein, Third Edition, PHI.
2. Fundamentals of Computer Algorithms, Ellis Horowitz, Satraj Sahni and Rajasekharam, Galgotia publications pvt. Ltd.
3. Design and Analysis Algorithms – Parag Himanshu Dave, Himanshu Bhalchandra Dave Publisher: Pearson
4. Algorithm Design: Foundations, Analysis and Internet examples, M.T. Goodrich and R. Tomassia, John Wiley and sons.
5. Data structures and Algorithm Analysis in C++, Allen Weiss, Second edition, Pearson education

# CC-402 Text, Image & Video Analytics

## Unit - 1: Introduction

Fundamental Steps in Image Processing, How does it works? Relationship between a digital image and a signal Components of an Image Processing System, Sampling and Quantization, Representing Digital Images (Data structure), Some Basic Relationships Between Pixels-Neighbors and Connectivity of pixels in image, Applications of Image Processing

## Unit - 2: Image Enhancement in The Spatial Domain and Frequency Domain

**In The Spatial Domain:** Some Basic Gray Level Transformations, Histogram Processing, Enhancement Using Arithmetic/Logic Operations, Basics of Spatial Filtering, Smoothing Spatial Filters, Sharpening Spatial Filters, Combining Spatial Enhancement Methods.

**In Frequency Domain:** Introduction, Fourier Transform, Discrete Fourier Transform (DFT), properties of DFT, Discrete Cosine Transform (DCT), Image filtering in frequency domain

## Unit - 3: Image Segmentation

Introduction, Detection of isolated points, line detection, Edge detection, Edge linking, Region based segmentation- Region growing, split and merge technique, local processing, regional processing, Hough transform, Segmentation using Threshold

## Unit - 4: Image Compression

Introduction, coding Redundancy, Inter-pixel redundancy, image compression model, Lossy and Lossless compression, Huffman Coding, Arithmetic Coding, LZW coding, Transform Coding, Sub-image size selection, blocking, DCT implementation using FFT, Run length coding.

## Reference Books:

1. Rafael C G., Woods R E. and Eddins S L, Digital Image Processing, Prentice Hall, 3rd edition, 2008
2. Milan Sonka, "Image Processing, analysis and Machine Vision", Thomson Press India Ltd, Fourth Edition.
3. Fundamentals of Digital Image Processing- Anil K. Jain, 2nd Edition, Prentice Hall of India.
4. S. Sridhar, Digital Image Processing, Oxford University Press, 2nd Ed, 2016.

# CC-403 Deep Learning

## Unit - 1: Basics of artificial neural networks (ANN)

Biological Neural Network and Artificial Neural Networks, Computational models of neurons, Architectures of ANN, Adjustment of weights to ANN Learning, Activation functions, Types of Artificial Neural Network, Advantages of Artificial Neural Network (ANN), Dis-advantages of Artificial Neural Network (ANN), Perceptron, Feedforward network, Multilayer feedforward neural networks (MLFFNNs), Linear and non-linear separable problems, AND, OR, XOR gates, Applications of ANN, Backpropagation algorithm

## Unit - 2: Deep neural networks (DNNs)

Overview of DNN, Advantages of DNN over ANN, Difficulty of training DNNs, Greedy layer wise training, Optimization for training DNNs, Newer optimization methods for neural networks (AdaGrad, RMSProp, Adam), Second order methods for training, Regularization methods (dropout, drop connect, batch normalization), Recurrent Network and L1 and L2 Regularization, Autoencoders, Simulation of AND or XOR gate through ANN, Generalization

## Unit - 3: Convolution neural networks (CNNs)

Introduction to Convolution neural networks (CNNs), convolution, pooling, Deep CNNs, Different deep CNN architectures – LeNet, AlexNet, VGG, Training a CNNs: weights initialization, batch normalization, hyperparameter optimization, Understanding and visualizing CNNs

## Unit - 4: Recurrent neural networks (RNNs)

Overview of Recurrent Neural Network, Architecture of RNN, Sequence modeling using RNNs, Backpropagation through time, Long Short-Term Memory (LSTM), Bidirectional LSTMs, Bidirectional RNNs, Gated RNN Architecture. Recursive Neural Network, Difference between RNN and Recursive Neural Network, Applications of RNN and ReNN. Transformers

## Reference Books:

1. Ian Goodfellow, Yoshua Bengio and Aaron Courville, Deep learning, In preparation for MIT Press, Available online: <http://www.deeplearningbook.org>, 2016
2. Neural Network Methods in Natural Language Processing by Yoav Goldberg, Graeme Hirst (Published on April 17, 2017)
3. S. Haykin, Neural Networks and Learning Machines, Prentice Hall of India, 2010
4. Satish Kumar, Neural Networks - A Class Room Approach, Second Edition, Tata McGraw-Hill, 2013
5. <http://cs231n.stanford.edu/> CS231n: Convolutional Neural Networks for Visual Recognition

# CC-404 (P) Practical Based on CC-403

## Practical Based on CC-403 Deep Learning

### Unit - 1: Introduction to Deep Learning

Overview of Deep Learning Applications, Understanding Neural Networks, Introduction to Deep Learning Frameworks: TensorFlow, Keras, PyTorch, Caffe, MXNet

### Unit - 2 Data Preparation for Deep Learning

Data Preprocessing Techniques, Data Augmentation, Handling Missing Data, Data Scaling and Normalization

### Unit - 3: Building Neural Networks

Building Simple Neural Networks with Keras, Understanding Activation Functions, Regularization Techniques for Neural Networks, Optimizers and Loss Functions

### Unit - 4: CNN, RNN, CV & NLP

Understanding Convolutional Layers, Building a CNN for Image Classification, Transfer Learning with CNNs, Understanding Recurrent Layers, Building a Simple RNN for Text Classification, Building a LSTM for Text Generation, Image Preprocessing for CV, Building an Object Detection Model, Building a Segmentation Model, Text Preprocessing for NLP, Building a Text Classification Model, Building a Text Generation Model

### Reference Books:

1. "Hands-On Deep Learning Algorithms with Python" by Sudharsan Ravichandiran
2. "Deep Learning with Python" by Francois Chollet
3. "Deep Learning" by Ian Goodfellow, Yoshua Bengio, and Aaron Courville
4. "Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow" by Aurélien Géron
5. "Deep Learning with PyTorch" by Eli Stevens, Luca Antiga, and Thomas Viehmann
6. "Deep Learning for Computer Vision with Python" by Adrian Rosebrock
7. "Applied Deep Learning" by Umberto Michelucci
8. "Deep Learning for Time Series Forecasting" by Jason Brownlee
9. "Natural Language Processing with Python-Analyzing Text with the Natural Language Toolkit" Steven Bird, Ewan Klein, and Edward Loper.

# **CC-405 (P) Power BI & Tableau**

## **Unit - 1: Excel Essentials and Tableau Introduction**

Text-to-columns, Remove duplicates, Lookup functions, Conditional formulas, Nested formulas, Formula auditing, Conditional formatting, Data validation, Pivot Table, imp keyboard shortcuts

Introduction to Business Intelligence, Tableau Products, Install & set-up Tableau Public, Tableau UI and Navigation, connect to csv and excel files, Dimensions vs Measures, Tableau Generated Fields, Organize Data in Groups, Create Hierarchy, Modify Metadata, Discrete vs Continuous, Change Default Properties

Bar & Column charts, Line & Area charts, Dual-axis & Combined axis, Pie charts, Treemaps, Manual & computed sorting, Filter on views, customize filter cards, Relative Date Filters, Top N View with Context Filter, Add Labels & Tooltips

## **Unit - 2: Tableau Fundamentals**

Text & Highlight tables, assign a geographic role, modify ambiguous locations, create symbol & choropleth (Filled) Maps, Navigating maps, Scatterplots, Bins & Histograms, View Underlying Data, Aggregation Options, Quick Table Calculations, Simple Custom Calculations, Aggregation & Granularity, Groups & Sets, Use Parameters, Format Fonts and Numbers, Format Color, Size & Shape, Format Tooltip & Legends, Format Axis, Titles & Captions, create a basic Dashboard, Add Interactivity, Export Data, Export Views, Save to Tableau Public

## **Unit - 3: Power BI - Power Query and Visualizations**

Introduction to Power BI, Building Blocks of Power BI, Install & set-up Power BI Desktop, Import data from different sources, Data transform using Query Editor, Profile the data, Data cleaning operations, Preparing data for analysis, Conditional and custom column, Handling errors, Leverage Advanced Editor, Create a common date table, Bar and Column charts, Lines, Area & Combo charts, Sorting and Filtering, Tables and Matrix, Pie, Donut and Treemap charts, Interacting with hierarchy, Maps, KPI and Gauges, Cards & Multi-row cards, Slicers, Format tooltip

## **Unit - 4: Power BI - DAX, Data Modelling and Power BI Service**

Fact vs Dim tables, manage data relationship, Introduction to DAX, Create calculated columns, Basic Operators, Aggregation functions, Logical functions, Text functions, Date and Time functions, Calculate and Filter, Create hierarchies, Define quick measures, Change aggregation, apply analytics on charts, AI visuals, Import a custom visual, Insert text and image, Design the background, Buttons & Actions, Edit interactions, Sync slicers, Introduction of Power BI Services, Interact with Report and Dataset in Power BI Service, Publish to web

## **Reference Books:**

1. Documentation of Power BI & Tableau

## **CC-406 PROJECT-II: Project**

The students will be given guided project from the subjects they study in the semester. They will be evaluated on the basis of the regularly, internal evaluation, project presentation and viva-voce & document submission.



# Gujarat University

## School of Emerging Science and Technology

### M.Sc. (Integrated) Five Years Program

### Data Science

### SEMESTER - 8

<b>Sr No.</b>	<b>Course No.</b>	<b>Subject</b>	<b>Credit</b>	<b>Exam</b>
1	CC-411	Multivariate Analysis	4	Theory
2	CC-412	Cloud Computing	4	Theory
3	CC-413	Big Data Analytics	4	Theory
4	CC-414	Advanced NLP	4	Theory
5	CC-415 (P)	Practical Based on CC-414	4	Practical
6	CC-416	PROJECT – III: Project	5	Project
		<b>Total</b>	<b>25</b>	

# CC-411 Multivariate Analysis

## Unit - 1: Bivariate and Factor Analysis

**Univariate and Bivariate Analysis of Data:** Descriptive vs Inferential Analysis, Descriptive Analysis of Univariate and Bivariate data, Analysis of Multiple Responses, Cross-tabulation, Spearman's Rank order correlation coefficient, Calculating Rank order, Data transformation

**Factor analysis:** Factor analysis model, Statistics associated with factor analysis, Establishing the Strength of the Factor Analysis, The Factor Score Coefficient Matrix, Factor Loadings and Computation of Eigenvalues, Extracted Factors, Rotation of Factors, Labelling or Naming the Factors.

## Unit - 2: Discriminant and Logit analysis

Discriminant and Logit analysis: Basic concept, Relationship to regression and ANOVA, Statistics associated with discriminant analysis: Descriptive Statistics, Tests for Differences in Group Means, Correlation Matrix, Unstandardized Discriminant Function, Standardized Discriminant Function, Structural Coefficients, Assessing Classification Accuracy, Logit Model and Binary Logit Analysis.

## Unit - 3: Cluster Analysis

Cluster analysis: Basic concept, Statistics associated with cluster analysis, process of clustering, Cluster analysis: metric data and non-metric data, Stabilizing the cluster assumptions, Type of clustering, Hierarchical clustering, Fuzzy clustering and k-means clustering, model-based cluster analysis.

## Unit - 4: Multidimensional scaling and conjoint analysis

Multidimensional scaling and conjoint analysis: Basic concept of Multidimensional scaling, Statistics associated with multidimensional scaling, Creating Spatial Maps Using Multidimensional Scaling, Conducting MDS with Similarity Data, Conducting MDS with Preference Data.

## Reference Books:

1. Research Methodology: Concepts and Cases, Deepak Chawla and Neena Sondhi, Vikas Publishing House, 2<sup>nd</sup> edition
2. Marketing Research and Applied Orientation, Naresh Malhotra, Daniel Nunan, David Briks, Pearson Education, 5<sup>th</sup> edition
3. Business Research Methods, 8<sup>th</sup> Edition by Willian G. Zikmund, Barry J. Babin, Jon C. Carr, Mitch Griffin
4. Business Research Methods by Donald R. Cooper, Pamela S. Schindler

# CC-412 Cloud Computing

## Unit - 1: Fundamentals of Cloud Computing

Cloud Computing, Adoption of cloud-based IT resources, Service Models: Infrastructure-as-a-Service (IaaS), Platform-as-a-Service (PaaS), Software-as-a-Service (SaaS), Deployment models: Public Cloud, Private Cloud, Hybrid Cloud, Community Cloud, Cloud Computing Characteristics, Challenges of cloud computing, Virtualization concept, Types of virtualizations, Demo of virtualization, Virtualization Merits, Role of virtualization in cloud computing, Virtualization Demerits, VM Placement, VM Migration, VM Migration Demo, VM clustering, Design Issues in VM Clustering, Need of Dockers and Containers, Docker Eco-System, Hypervisor vs Docker.

## Unit - 2: Cloud Services and Task Scheduling Algorithms

Microservices, Service-Oriented Architecture, REST API, IP Addressing, Subnetting, Supernetting, Designing of Virtual Private Cloud, Demo of VPC, VPC Peering, VPC Case Study, Cloud Storage, Serverless Computing, Cloud API Gateway, Cloud Databases, Resource Provisioning, Time shared and space shared, Efficient VM Consolidation on cloud server, Task/DAG Scheduling Algorithms, Min-Min, Max-Min, MET, B-level Demo, T-level Demo, Task-VM Mapping, Auto Scaling, Load Balancing.

## Unit - 3: Cloud Security and Big Data handling on Cloud

Case Study: Cloud Market analysis, Security and Compliances, Shared security model in IAAS/PAAS/SAAS, Shared technology issues, Data loss or leakage, Account or service hijacking, Implementation of cloud security, Security Groups, Network Access Control Lists, Cloud databases, Parallel Query Execution with NoSQL Database, Big Data, Handling Big Data on Cloud Platform, Map- Reduce framework for large clusters using Hadoop, Design of data applications based on Map Reduce in Apache Hadoop.

## Unit - 4: Cloud Computing way Forward

Comparative study/analysis of public clouds, Edge Computing, Fog Computing, Data Offloading, Cloud-Based DevOps Tools, Task Partitioning, Data Partitioning, Data Synchronization, Distributed File System, Data center, Ongoing Research Topics.

### Reference Books:

1. Lizhe Wang, Rajiv Ranjan, Jinjun Chen and Boualem Benatallah, Cloud Computing (1 ed.), CRC Press, 2017. ISBN 978-1351833097.
2. Judith S. Hurwitz and Daniel Kirsch, Cloud Computing For Dummies (1 ed.), Hoboken: John Wiley & Sons, 2020. ISBN 978-1119546658.
3. Prerna Sharma, Moolchand Sharma and Mohamed Elhoseny, Applications of Cloud Computing (1 ed.), missing, 2020. ISBN 9780367904128.

# CC-413 Big Data Analytics

## Unit - 1: Introduction to Big Data Analytics

What is Big Data?, Source of Big Data, 5 V's of Big Data, Data Life Cycle, SEMA Methodology, Big Data Life Cycle, Methodology, Core Deliverables, Big Data Overview, Big Data Analytics, Types of Big Data Analytics, Data Structures, Analyst Perspective on data repositories, state of practice in analytics, Bi VS Data Science, Current Analytical Architecture, Drivers of Big Data, Emerging Big Data Ecosystem & New Approach To analytics, Key role of new big data ecosystem, Example of Big data analytics, Case Study of Big Data Solutions

## Unit - 2: Hadoop and Spark

Overview of Distributed database Systems, Hadoop eco-system, Hadoop core components, Hadoop distributions, Developing enterprise applications with Hadoop. Storing Data in Hadoop: Moving data in and out of Hadoop, HDFS architecture, HDFS files, Hadoop specific file types, HDFS federation and high availability, working with HDFS Commands, Fundamentals of HBASE, Zookeeper concepts and methods to build applications with Zookeeper.

## Unit - 3: Spark, HIVE, HIVEQL and PIG

Introduction to Data Analysis with Spark; Downloading Spark and Getting Started, Programming with RDDs, Machine Learning with MLlib. HIVE: Architecture and installation, Comparison with traditional database, HIVEQL querying data, Sorting and aggregating, Joins & sub queries, HIVE Vs PIG, PIG: Architecture and installation, Execution Mechanisms, load/store operator, Pig scripts.

## Unit - 4: NoSQL

What is NoSQL? NoSQL business drivers; NoSQL case studies; NoSQL data architecture patterns: Key-value stores, Graph stores, Column family (Bigtable) stores, Document stores, Variations of NoSQL architectural patterns; Using NoSQL to manage big data: What is a big data NoSQL solution? Understanding the types of big data problems; Analyzing big data with a shared-nothing architecture; Choosing distribution models: master-slave versus peer-to-peer; Four ways that NoSQL systems handle big data problems; MongoDB: MongoDB Support for dynamic queries, Replications, Sharding, Create Database and Drop Database, Collections and Documents, MongoDB Query Language.

### Text Book:

Data Science & Big Data Analytics - John Willey

## Reference Books:

1. Bill Franks, "Taming The Big Data Tidal Wave: Finding Opportunities In Huge Data Streams With Advanced Analytics", Wiley
2. Michael Minelli, Michele Chambers, Ambiga Dhiraj, "Big Data Big Analytics: Emerging Business Intelligence And Analytic Trends For Today's Businesses", Wiley India
3. Boris lublinsky, Kevin t. Smith, Alexey Yakubovich, "Professional Hadoop Solutions", Wiley.
4. Chris Eaton, Dirk derooset al., "Understanding Big data", McGraw Hill, 2012.
5. BIG Data and Analytics , Seema Acharya, Subhashini Chhellappan, Willey
6. Tom White, "HADOOP: The Definitive Guide", O Reilly 2012.
7. Vignesh Prajapati, "Big Data Analytics with R and Hadoop", Packet Publishing 2013.
8. Learning Spark: Lightning-Fast Big Data Analysis Paperback by Holden Karau

## Case Study (UNIT 1):

### 1. Hospitality

- a. Airbnb focuses on growth by analysing customer voice using data science
- b. Qantas uses predictive analytics to mitigate losses

### 2. Healthcare

- a. Novo Nordisk is Driving innovation with NLP
- b. AstraZeneca harnesses data for innovation in medicine

### 3. Covid 19

- a. Johnson and Johnson uses data science to fight the Pandemic

### 4. Ecommerce

- a. Amazon uses data science to personalize shopping experiences and improve customer satisfaction

### 5. Supply chain management

- a. UPS optimizes supply chain with big data analytics

### 6. Meteorology

- a. IMD leveraged data science to achieve a record 1.2m evacuation before cyclone "Fani"

### 7. Entertainment Industry

- a. Netflix uses data science to personalize the content and improve recommendations
- b. Spotify uses big data to deliver a rich user experience for online music streaming

### 8. Banking and Finance

- a. HDFC utilizes Big Data Analytics to increase income and enhance the banking experience

- **Apart from case studies students will perform at the following programming exercises:**
  1. Implement following using Map- Reduce
    - Matrix multiplication
    - Sorting
    - Indexing
  2. Distributed Cache & Map Side Join, Reduce side Join Building and Running a Spark Application Word count in Hadoop and Spark Manipulating RDD
  3. Implementation of Matrix algorithms in Spark Sql programming
  4. Implementing K-Means Clustering algorithm using Map-Reduce
  5. Implementing any one Frequent Item set algorithm using Map-Reduce
  6. Create A Data Pipeline Based On Messaging Using PySpark And Hive - Covid-19 Analysis

# CC-414 Advanced NLP

## Unit - 1: Introduction to Natural Language Processing (NLP)

What is NLP? Why NLP is difficult? History of NLP, Advantage of NLP, Disadvantages of NLP, Components of NLP, Applications of NLP, NLP Pipeline, Phases of NLP, NLP APIs, NLP Libraries

## Unit - 2: Language Modeling and Part of Speech Tagging

Unigram Language Model, Bigram, Trigram, N-gram, Advanced smoothing for language modeling, Empirical Comparison of Smoothing Techniques, Applications of Language Modeling, Natural Language Generation (NLG), Parts of Speech Tagging (POS), Morphology, Named Entity Recognition (NER)

## Unit - 3: Words and Word Forms

Bag of words, skip-gram, Continuous Bag-Of-Words, Embedding representations for words Lexical Semantics, Word Sense Disambiguation, Knowledge Based and Supervised Word Sense Disambiguation, Text Analysis, Summarization and Extraction

## Unit - 4: Machine Translation

Need of Machine Translation, Problems of Machine Translation, Machine Translation Approaches, Direct Machine Translations, Rule-Based Machine Translation, Knowledge Based Machine Translation System, Statistical Machine Translation (SMT), Parameter learning in SMT (IBM models) using EM, Encoder-decoder architecture, Neural Machine Translation

## Reference Books:

1. Neural Network Methods in Natural Language Processing by Yoav Goldberg, Graeme Hirst (Published on April 17, 2017)
2. Deep Learning in Natural Language Processing by Li Deng, Yang Liu (Published on May 23, 2018)
3. Deep Learning for NLP and Speech Recognition by Uday Kamath, John Liu, James Whitaker (Published on August 14, 2020)
4. Introduction to Natural Language Processing by Jacob Eisenstein (Published on October 1, 2019)
5. Speech and Language Processing: An introduction to Natural Language Processing, Computational Linguistics and Speech Recognition Jurafsky, David, and James H. Martin, PEARSON
6. Foundations of Statistical Natural Language Processing, Manning, Christopher D., and Hinrich Schutze, Cambridge, MA: MIT Press

7. Natural Language Understanding, James Allen. The Benjamin/Cummings Publishing Company Inc..
8. Natural Language Processing with Python-Analyzing Text with the Natural Language Toolkit Steven Bird, Ewan Klein, and Edward Loper.



# CC-415 (P) Practical Based on CC-414

## Practical Based on CC-414 Advanced NLP

### Unit - 1: Text Preprocessing and Feature Engineering Level-1

NLP libraries, Install NLTK, Tokenize text using pure Python, Count word frequency, Remove stop words using NLTK, Punctuation removal, Tokenize text using NLTK, Tokenize non-English languages text, Get synonyms from WordNet, Get antonyms from WordNet, NLTK word stemming, Stemming non-English words, Lemmatizing words using WordNet, Stemming and lemmatization difference

### Unit - 2: Text Preprocessing and Feature Engineering Level-2

Tagging Parts of Speech (POS), Chunking, Chinking, Using Named Entity Recognition (NER) Getting Text to Analyze, Using a Concordance, Making a Dispersion Plot, Making a Frequency Distribution, Finding Collocations. Advanced level Text Cleaning: Normalization, Correction of Typos, etc.

### Unit - 3: Text Preprocessing and Feature Engineering Level-3

Implementation of Hierarchical Clustering algorithm, Density-Based Spatial Clustering (DBSCAN) and Gaussian Mixed Models (GMM). Pros and Cons of each algorithm, Word Cloud, Properties of Word Cloud, Advantage and Disadvantage of Word Cloud, Named Entity Recognition (NER)

Wordnet:

- a. We can check how many different definitions of a word are available in Wordnet.
- b. We can also check the meaning of those different definitions.
- c. All details for a word.
- d. All details for all meanings of a word.
- e. Hypernyms: Hypernyms gives us a more abstract term for a word.
- f. Hyponyms: Hyponyms gives us a more specific term for a word.
- g. Get a name only.
- h. Synonyms.
- i. Antonyms.
- j. Synonyms and antonyms.
- k. Finding the similarity between words.

### Unit - 4: Current NLP Technologies

Word Embedding, What is Prediction-based Embedding? Bag of words (BOW), Term frequency Inverse Document Frequency (TFIDF), Unigram, Bigram, and Ngrams, Word2vec, Average word2vec.

Different Model Architectures for Word representation: FeedForward Neural Net Language Model (NNLM), Recurrent Neural Net Language Model (RNNLM), Word 2 Vec, Encoders and Decoders, Self-attention models, Transformers.

**Reference Books:**

1. Natural Language Processing with Python-Analyzing Text with the Natural Language Toolkit Steven Bird, Ewan Klein, and Edward Loper.
2. Practical Natural Language Processing: A Comprehensive Guide to Building Real-World NLP Systems by Sowmya Vajjala, Bodhisattwa Majumder, Anuj Gupta, Harshit Surana
3. Natural Language Processing with PyTorch: Build Intelligent Language Applications Using Deep Learning by Delip Rao, Brian McMahan (Published on February 19, 2019)
4. Natural Language Processing in Action: Understanding, analyzing, and generating text with Python by Hobson Lane, Hannes Hapke, Cole Howard (Published on April 14, 2019)
5. Deep Learning for Coders with fastai and PyTorch: AI Applications Without a PhD by Jeremy Howard, Sylvain Gugger (Published on August 4, 2020)

## **CC-416 PROJECT-III: Project**

The students will be given guided project from the subjects they study in the semester. They will be evaluated on the basis of the regularly, internal evaluation, project presentation and viva-voce & document submission.

# Gujarat University

## School of Emerging Science and Technology

### M.Sc. (Integrated) Five Years Program

### Data Science

### SEMESTER -9

<b>Sr No.</b>	<b>Course No.</b>	<b>Subject</b>	<b>Credit</b>	<b>Exam</b>
1	CC-501	Cyber Security	4	Theory
2	CC-502	Project Deployment on Cloud AWS	4	Theory
3	CC-503	Blockchain Technology	4	Theory
4	CC-504	Reinforcement Learning	4	Theory
5	CC-505 (P)	Practical Based on CC-504	4	Practical
6	CC-506	PROJECT – IV: Project	5	Project
		<b>Total</b>	<b>25</b>	

# **CC-501 Cyber security**

## **Unit - 1: Introduction to Cyber security**

Defining Cyberspace and Overview of Computer and Web-technology, Architecture of cyberspace, Communication and web technology, Internet, World wide web, Advent of internet, Internet infrastructure for data transfer and governance, Internet society, Regulation of cyberspace, Concept of cyber security, Issues and challenges of cyber security

## **Unit - 2: Systems Vulnerability Scanning**

Introduction, Computer Security, Threats, Harm, Vulnerabilities, Controls, Authentication, Access Control and Cryptography. • Web attack: Browser Attacks, Web Attacks Targeting Users, Obtaining User or Website Data, Email Attacks. • Network Vulnerabilities: Overview of vulnerability scanning, Open Port / Service Identification, Banner /Version Check, Traffic Probe, Vulnerability Probe, Vulnerability Examples, OpenVAS, Metasploit. Networks Vulnerability Scanning (Ncat, Socat), Network Sniffers and Injection tools.

## **Unit - 3: Network Defense tools**

Firewalls and Packet Filters: Firewall Basics, Packet Filter Vs Firewall, How a Firewall Protects a Network, Packet Characteristic to Filter, Stateless Vs Stateful Firewalls, Network Address Translation (NAT) and Port Forwarding, the basic of Virtual Private Networks, Linux Firewall, Windows Firewall, Snort: Introduction Detection System

## **Unit - 4: Cyber Crime & Cyber Law**

Classification of cyber crimes, Common cyber crimes- cyber crime targeting computers and mobiles, cyber crime against women and children, financial frauds, social engineering attacks, malware and ransomware attacks, zero day and zero click attacks, Cybercriminals modus-operandi , Reporting of cyber crimes, Remedial and mitigation measures, Legal perspective of cyber crime, IT Act 2000 and its amendments, Cyber crime and offences, Organisations dealing with Cyber crime and Cyber security in India, Case studies.

## **Reference Books:**

1. Anti-Hacker Tool Kit (Indian Edition) by Mike Shema, Publication Mc Graw Hill.
2. Cyber Security Understanding Cyber Crimes, Computer Forensics and Legal Perspectives by Nina Godbole and Sunit Belpure, Publication Wiley

# CC-502 Project Deployment on Cloud AWS

## Unit - 1: AWS Basics

AWS — Cloud Computing - What is Cloud Computing? - Types of Clouds - Cloud Service Models - Advantages of Cloud Computing - AWS — Basic Architecture – Key Considerations for Web Hosting in AWS - AWS — Management Console - How to Access AWS? - Customizing the Dashboard - AWS — Console Mobile App - AWS — Account - How to Use AWS Account? - AWS Account Identifiers - Account Alias - Multi Factor Authentication - AWS Identity & Access Management (IAM)

## Unit - 2: Amazon Storage Services

AWS — Amazon S3 - How to Configure S3? - How to Move S3 Objects? - How to Delete an Object? - How to Empty a Bucket? - Amazon S3 Features - AWS — Elastic Block Store - EBS Volume Types - Amazon EBS Benefits - How to Set Up Amazon EBS? - AWS — Storage Gateway - Volume Gateways - Gateway-Virtual Tape Library (VTL) - AWS — CloudFront - How AWS CloudFront Delivers the Content? - Features of CloudFront - How to Set Up AWS CloudFront? - Test the Links

## Unit - 3: Amazon Database Services

AWS — Relational Database Service - Features of Amazon RDS - How to Set Up Amazon RDS? - How to Connect Database to MySQL DB Instance? - How to Delete a DB Instance? - Cost of Amazon RDS - AWS — DynamoDB - How to Run DynamoDB on Computer? - Benefits of Amazon DynamoDB - AWS — Redshift - How to Set Up Amazon Redshift? - Features of Amazon Redshift

## Unit - 4: Amazon Analytics Services

AWS — Amazon Kinesis - How to Use Amazon KCL? - Limits of Amazon Kinesis? - How to Use Amazon Kinesis? - Features of Amazon Kinesis - AWS — Elastic MapReduce - How to Set Up Amazon EMR? - Benefits of Amazon EMR - AWS — Data Pipeline - How to Set Up Data Pipeline? - How to Delete a Pipeline? - Features of AWS Data Pipeline - AWS — Machine Learning - Types of Tasks Performed by Amazon Machine Learning - How to Use Amazon Machine Learning? - Features of Amazon Machine Learning

## Reference Books:

1. Learning Amazon Web Services (AWS): A Hands-On Guide to the Fundamentals of AWS Cloud 1st Edition by Mark Wilkins
2. AWS: The Ultimate Guide From Beginners To Advanced For The Amazon Web Services by Theo H. King, 2020
3. Amazon Web Services in Action by Andreas Wittig and Michael Wittig.

# CC-503 Blockchain Technology

## Unit - 1: Blockchain Fundamentals

**Blockchain Fundamentals:** Introduction, Origin of Blockchain, Traditional Vs Blockchain transactions, Key Blockchain concepts, Component of Blockchain, Consensus Protocol, Double-spending problem, Layers of Block chain, Advantages and Limitations of Block-chain technology

**Blockchain Types and Consensus Mechanism:** Key Terms: Decentralization, Transparency, Privacy, Security, Distributed Ledger Technology, Types of Blockchain, Consensus Protocols: Byzantine General Problem, Proof of Work and Proof of Stack

## Unit - 2: Mining in Blockchain and Cryptocurrency – Bitcoin, altcoin and Token

**Mining in Blockchain:** Introduction, Importance of mining block in Blockchain, Concept of Block, Nonce, Time-Stamp, SHA256 algorithm, Hash and Previous Hash fields. Process of mining in detail.

**Cryptocurrency – Bitcoin, altcoin and Token:** Introduction, Evolution of Cryptocurrency, Birth of Bitcoin, what is cryptocurrency? characteristics of Cryptocurrency, Cryptocurrency Wallets: Hot wallet, Cold wallet, Types of Cryptocurrencies: Altcoin, Tokens, Utility Tokens, Security Token, Popular Coins and Token: Ethereum, Ripple(XRP), Bitcoin Cash(BCH), Litecoin(LTC)

## Unit - 3: Crypto-mining and Public Blockchain

**Crypto-mining:** Solo mining, Pool mining, concept of Mempool, Hard and Soft fork, Airdrop, Types of Airdrop and benefits of Airdrop. Transaction in Cryptocurrency and Token or Coin burning.

**Public Blockchain:** Introduction, Blockchain Layers, concept of Mempool, Common mining technologies like Mining, Block frequency, Industrial Mining, Mining Pools, Block, Block header, Orphaned block, Timestamp Nonce, WEGWIT, POW. Bitcoin Mining and types of Mining. Monetary policy of Bitcoin (Halving and Block frequency policies), Transaction and UTXOs in Bitcoin.

## Unit - 4: Ethereum Blockchain, Smart Contracts, Private Block chain System and Block chain Development:

**Ethereum Blockchain:** Introduction to smart contract, Code is Law and Ethereum Classic, Ethereum components, Miner and mining node in Ethereum network, EVM, Ether, Gas, Transaction in Ethereum, How mining works in Ethereum? Bitcoin Vs Ethereum.

**Smart Contracts:** Introduction, Smart Contracts, Examples, How smart contract works? Characteristics of smart contract, DApps, DAO, Smart contract with IOT, Types of Oracles, Smart contract in Industry.

**Private Block chain System:** Introduction, Private block chain examples, Problem with the centralized server, Smart contract in private environment. DOS attack.

**Block chain Development:**

- Introduction to Python in Blockchain
- Introduction to Solidity in Ethereum blockchain and Smart contract

**Reference Books:**

1. Blockchain Technology By – Chandramaouli Subramaniam, Asha A George, Abhilash K A and Meena Karthikeyan by Universities Press ISBN 978 93 89211 63 4
2. Arvind Narayanan, Joseph Bonneau, Edward Felten, Andrew Miller and Steven Goldfeder, Bitcoin and Cryptocurrency Technologies: A Comprehensive Introduction, Princeton University Press, July 2016
3. Blockchain Blueprint for a New Economy, By Melanie Swan, O'Reilly Publication
4. Blockchain For Dummies By Tiana Laurence, Wiley Publication



# CC-504 Reinforcement Learning

## Unit - 1: Fundamentals of Reinforcement Learning

An introduction to Sequential Decision-Making, The K-Armed Bandit Problem: Sequential Decision Making with Evaluative Feedback, Estimating Action Values, Learning Action Values Estimating Action-Values Incrementally, Exploration vs. Exploitation Tradeoff: What is the trade-off?, Optimistic Initial Values, Upper-Confidence Bound (UCB) Action Selection, Introduction to Markov Decision Processes: Markov Decision Process, Examples of MDPs, Goal of Reinforcement Learning: What is the Goal of Reinforcement Learning?, Reward Hypothesis Continuing and Episodic Tasks: Types of tasks: Continuing Tasks and Episodic Tasks, Examples of Episodic and Continuing Tasks

## Unit - 2: Value Functions & Bellman Equations

Policies and Value Functions: What is Policy?, What are Value Functions?, Bellman Equations: What is Bellman Equation Derivation, Why is Bellman Equations important?, Optimality (Optimal Policies & Value Functions): Optimal Policies, Optimal Value Functions, Dynamic Programming: Policy Evaluation (Prediction), Policy Iteration (Control)

## Unit – 3: Sample-based Learning Methods

Introduction to Monte-Carlo Methods: What is the Monte Carlo method?, Types of Policies: On-Policy and Off-Policy, Introduction to Temporal Difference Learning: What is Temporal Difference (TD) learning?, Advantage of temporal difference learning, Comparing TD and Monte Carlo, Q-Learning and SARSA: What is Q-Learning and SARSA Learning?, What is a Model?, Random Tabular Q-planning

## Unit - 4: Prediction and Control with Function Approximation

What is Function approximation?, Non-Linear Approximation with Neural Network, Actor-Critic for Continuing Tasks

## Reference Books:

1. Reinforcement Learning: An Introduction, Richard S. Sutton & Andrew G. Barto

## **CC-505 (P) Practical based on CC-504**

### **Practical Based on CC-504 Reinforcement Learning**

#### **Unit - 1: Basics of Reinforcement Learning**

This practical introduces the basics of reinforcement learning and its important components, namely agent and environment. It includes a simple Python code for the environment class, which defines variables for maximum steps and functions for providing observation, actions, and checking completion. The practical also discusses the agent class and how it observes the environment, decides on actions, submits them and receives rewards.

#### **Unit - 2: Q-Learning from Scratch with OpenAI**

This practical involves basic steps of Q Learning with the help of a self-learning cab/dog to find a strategy and improve the strategy. We are using OpenAI to simulate the environment for this practical.

#### **Unit - 3: Snake Game with Reinforcement Learning**

This practical involves the famous game of snake to make the agent (snake) to play the game on its own.

#### **Unit - 4: Autonomous Driving Car OpenAI**

It uses OpenAI's gym for making simulation and train the agent. It teaches agent to learn the road environment to drive on its own.

#### **Reference Books:**

1. Reinforcement Learning: An Introduction, Richard S. Sutton & Andrew G. Barto
2. [https://youtu.be/g\\_8gw2POOYE](https://youtu.be/g_8gw2POOYE)
3. <https://www.learndatasci.com/tutorials/reinforcement-q-learning-scratch-python-openai-gym/>
4. <https://towardsdatascience.com/snake-played-by-a-deep-reinforcement-learning-agent-53f2c4331d36>
5. <https://youtu.be/L8ypSXwyBds>
6. <https://github.com/PacktPublishing/Hands-On-Intelligent-Agents-with-OpenAI-Gym>

## **CC-506 PROJECT-IV: Project**

The students will be given guided project from the subjects they study in the semester. They will be evaluated on the basis of the regularly, internal evaluation, project presentation and viva-voce & document submission.

# Gujarat University

## School of Emerging Science and Technology

### M.Sc. (Integrated) Five Years Program

#### Data Science

#### SEMESTER - 10

<b>Sr No.</b>	<b>Course No.</b>	<b>Subject</b>	<b>Credit</b>	<b>Exam</b>
1	CC-511	Dissertation/ Project Work	16	Project
2	CC-512	Seminar/ Symposium/ Conference	4	Seminar/ Conference
3	CC-513	Research Paper/ MOOCs/ Internship/ OJT	5	Certificate
		<b>Total</b>	<b>25</b>	

## CC-511 Dissertation/Project Work

- Marking Component of Dissertation/Project Work:

Components		Internal	External
Regular Involvement	Understanding of Topic, Knowledge of Literature, Knowledge of Methodology, understanding different relevant Techniques/Tools, Result / Conclusion, efforts put in during the project work	60	--
Addressing Mentor Queries and Requirement		30	--
Student behaviour, Academic Discipline & Regularity		30	--
Relevance of Topic/Motivation of choosing		--	20
Originality of Approach		--	20
Literature Review/ Review of best practices/applied cases		--	20
Methodology / originality of different method if any		--	20
Usage of relevant tools and techniques / software		--	20
Result / Conclusions / applicability of the project		--	20
Viva Voce		--	100
PowerPoint Presentation	Quality, layout, content and design and flow of work	--	30
Submission of Project	Quality, layout, content and design and flow of work [Spiral bound Document (MS Word/PDF), soft and hard copy both]	--	30
<b>Total</b>		<b>120</b>	<b>280</b>

**CC-512**  
**Seminar/ Symposium/ Conference**

**1. Conference Presentations**

- Marking Component of Presentations:

<b>Components</b>	<b>Marks</b>
Title Selection	
Content	
Presentation & Understanding	
Defence	
Layout	
<b>Total</b>	

**Note:** Conference presentation certificate will be compulsory to fetch marks.

**CC-513**

**Research Paper/ MOOCs/ Internship/ OJT**