TKR COLLEGE OF ENGINEERING AND TECHNOLOGY
(AUTONOMOUS)
(Sponsored by TKR Educational Society , Approved by AICTE, Affiliated by JNTUH,
Accredited by NBA & NAAC with 'A' Grade)

B.TECH – COMPUTER SCIENCE & ENGINEERING (DATA SCIENCE)
Course Structure R-20

SEMESTER III

S.No.	Class	Course Code	Name of the Subject	L	T	P	C
1	BS	CBSM5	Statistical Methods	3	0	0	3
2	ES	CESOP1	Introduction to Object-Oriented Programming & Data Structures using Java	3	0	0	3
3	PC	C83PC1	Database Management Systems	3	0	0	3
4	PC	C83PC2	Operating Systems	3	0	0	3
5	PC	C83PC3	R Programming	3	0	0	3
6	PC	C83PC4	Formal Language & Automata Theory	3	0	0	3
7	ES	CESOP2	Introduction to Object-Oriented Programming & Data Structures using Java Lab	0	0	2	1
8	PC	C83PC5	Database Management Systems Lab	0	0	2	1
9	PC	C83PC6	R Programming Lab	0	0	2	1
10	MC	MC003	Cultural Activity	0	0	0	Satisfactory
Total Credits				**18**	**0**	**6**	**21**

Mandatory Course: Cultural Activity
The student should participate in culture activity (Music/Dance/Singing/etc.) conducted by the College, student should produce the participation certificate for clearing this course.

SEMESTER IV

S.No.	Class	Course Code	Name of the Subject	L	T	P	C
1	BS	CBSM6	Probability & Algebra	3	0	0	3
2	PC	C84PC1	Data Warehousing & Data Mining	3	0	0	3
3	PC	C84PC2	Information Security	3	0	0	3
4	PC	C84PC3	Design and Analysis of Algorithms	3	0	0	3
5	PC	C84PC4	Data Visualization	3	0	0	3
6	PC	C84PC5	Python Programming	3	0	0	3
7	PC	C84PC6	Data Warehousing & Data Mining Lab	0	0	2	1
8	PC	C84PC7	Data Visualization Lab	0	0	2	1
9	PC	C84PC8	Python Programming Lab	0	0	2	1
10	MC	MC004	Video with Social Messages	0	0	0	Satisfactory
Total Credits				**18**	**0**	**6**	**21**

Mandatory Course: Video with Social Messages
Student should make video with social messages. This has to be uploaded in the youtube.com, by maintaining the terms and conditions of youtube.com. Student should produce youtube.com link with screen shot for clearing this mandatory course.

TKR COLLEGE OF ENGINEERING AND TECHNOLOGY
(AUTONOMOUS)
(Sponsored by TKR Educational Society , Approved by AICTE, Affiliated by JNTUH,
Accredited by NBA & NAAC with 'A' Grade)

CSE (DATA SCIENCE)

B.Tech III Semester **L/T/P/C**
3 /0/ 0/ 3

STATISTICAL METHODS (CBSM5)

Course Objectives: To learn:

1. To introduce the concepts of Statistics.
2. To understand the basic statistical tools for analysis & interpretation of qualitative and quantitative data.
3. To understand the Sampling techniques of population, and to estimate the parameters.
4. Testing of Hypothesis for Large and small samples.
5. To understand the linear relationship between two variables in correlation.
6. To understand the mathematical measure of the average relationship in regression analysis.

Course Outcomes:

After learning the contents of this course, the student will be able to
CO1: Apply Statistical logic for solving the problems.
CO2: Analyse the qualitative & quantitative data.
CO3: Apply the sampling techniques.
CO4: Find the error in sampling distributions.
CO5: Test the hypothesis and give the inference to the given data.
CO6: Predict the value of dependent variable by regression analysis.

UNIT I
Introduction of Statistics

Functions of statistics, Collection of data, Classification of data, Tabulation of data, diagramatic and Graphical representation of data. Measures of Central Tendency- Mean, Median, Mode, Geometric Mean and Harmonic Mean.

UNIT II
Measures of Dispersion

Range, Quartile deviation, Mean Deviation, Standard deviation and Coefficient of variation, Skewness: Karl Pearson's co-efficient of skewness, Bowley's co-efficient of skewness, Kelleys co-efficient of skewness, Kurtosis.

UNIT III
Sampling Distributions

Sampling distributions of means (σknown and unknown), Method of Estimation, Point estimations, Maximum Error, Interval estimations.

STATISTICAL METHODS (CBSM5)

UNIT IV
Testing of Hypothesis

Null hypothesis-alternative hypothesis Type-1, Type-2 errors, critical region, testing of single mean and two means, (large and small samples). Test of Hypothesis for the single proportion and difference between the two proportions.

UNIT V
Analysis of Variance

One Way and Two Way ANOVA ,Correlation Analysis-Scatter diagram, Karl Pearson's oefficient of correlation, Spearman's Rank correlation, Regression Analysis-Concept, least square fit of a linear regression, two lines of regression, Properties of regression coefficients.

Text Books:

1. Gupta S.C&V.K. Kappor., Fundamentals of Mathematical Statistics, S Chand Publishers.
2. Introduction to Statistics by Wolfgang Karl Hardle, Sigbert Klinke, Bernd Ronz, Springer (e- BOOK).

Reference Books:

1. Probability & Statistics for Engineers by G.S.S. Bhisma Rao, Sci Tech Publications.
2. Business Statistics, 1e, Tata McGraw Hill, 2015.

TKR COLLEGE OF ENGINEERING AND TECHNOLOGY
(AUTONOMOUS)
(Sponsored by TKR Educational Society , Approved by AICTE, Affiliated by JNTUH,
Accredited by NBA & NAAC with 'A' Grade)

CSE (DATA SCIENCE)

B.Tech III Semester **L/T/P/C**
3 /0/ 0/ 3

INTRODUCTION TO OBJECT-ORIENTED PROGRAMMING & DATA STRUCTURES USING JAVA (CESOP1)

Course Objective:

To understand the features of object-oriented paradigm and Data Structure concepts using JAVA programming Language.

Course Outcomes:

After completion of course, the students will be able to

1. Understand fundamentals of object-oriented programming in Java which includes defining classes, invoking methods
2. Implement the inheritance concept
3. Solve the exceptions in programs and recursion
4. Implement the Basic data structures and operations
5. Apply the ADTs and use the collections in Java

UNIT I
Object-Oriented Thinking

A way of viewing world – Agents and Communities, messages and methods, Responsibilities, Classes and Instances, Class Hierarchies- Inheritance, Method binding, Overriding and Exceptions, Summary of Object-Oriented concepts. History of object-oriented programming, overview of java, Object oriented design, Structure of java program, Java buzzwords, Data types, Variables and Arrays, operators, expressions, control statements, Introducing classes, Methods and Classes, String handling.
Stream based I/O (java.io): The Stream classes-Byte streams and Character streams, Reading console Input and Writing Console Output

UNIT II
Inheritance

Inheritance concept, Inheritance basics, Member access, Constructors, Creating Multilevel hierarchy, super uses, using final with inheritance, Polymorphism-ad hoc polymorphism, pure polymorphism, method overriding, abstract classes, Object class, forms of inheritance- specialization, specification, construction, extension, limitation, combination, benefits of inheritance, costs of inheritance.
Packages
Defining a Package, CLASSPATH, Access protection, importing packages. Interfaces Defining an interface, implementing interfaces, Nested interfaces, applying interfaces, variables in interfaces and extending interfaces.

INTRODUCTION TO OBJECT-ORIENTED PROGRAMMING & DATA STRUCTURES USING JAVA (CESOP1)

UNIT III
Exception Handling

Fundamentals of exception handling, Exception types, Termination or presumptive models, Uncaught exceptions, using try and catch, multiple catch classes, nested try statements, throw, throws and finally, built- in exceptions, creating own exception sub classes.
Recursion: Analyzing Recursion Algorithms, Designing of Recursive Algorithms.

UNIT IV
Sorting, Searching & Data Structure

Introduction to Sorting, Bubble Sort, Insertion Sort, Introduction to Searching, Linear Search and Binary Search.
Data Structures Fundamentals: Using Arrays, Singly Linked Lists, Circularly Linked Lists, Doubly Linked Lists.

Unit V
Stacks, Queues, and Deques

Stack, Queue, Double – ended queues.
List and Iterator ADTs: The List ADT, Array List, Positional Lists, Iterators, Java Collection Frameworks

Text Books:

1. Java The complete reference, 9 th edition, Herbert Schildt, McGraw Hill Education (India) Pvt. Ltd.
2. Data Structures & Algorithms in Java 6th Edition, Michale T. Goodrich, Roberto Tamssia, Michale H. Goldwasser, WILEY.

Reference Books:

1. An Introduction to programming and OO design using Java, J. Nino and F.A. Hosch, John Wiley & sons.
2. Introduction to Java programming, Y. Daniel Liang, Pearson Education.
3. Object Oriented Programming through Java, P. Radha Krishna, and Universities Press.

TKR COLLEGE OF ENGINEERING AND TECHNOLOGY
(AUTONOMOUS)
(Sponsored by TKR Educational Society , Approved by AICTE, Affiliated by JNTUH,
Accredited by NBA & NAAC with 'A' Grade)

CSE (DATA SCIENCE)

B.Tech III Semester **L/T/P/C**
3 /0/ 0/ 3

DATABASE MANAGEMENT SYSTEMS (C83PC1)

Course Objective:

It emphasizes the understanding of the fundamentals of relational systems including data models, databases.

Course Outcomes:

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

1. Demonstrate the basic elements of a relational database management system, and identify the data models for relevant problems.
2. Design entity relationship model and convert entity relationship diagrams into RDBMS and formulate SQL queries on the data.
3. Apply normalization for the development of application software.
4. Understand transaction processing, concurrency control and recovery techniques.
5. Understand the indexing data structures and hashing.

UNIT I
Introduction and Basic Concepts

File organization for conventional data management system, Higher-level file organization for DBMS, Database System Applications, Purpose of Database Systems, View of Data, Database Languages – DDL, DML, Relational Databases, Database Design, Data Storage and Querying, Transaction Management, Database Architecture, Data Mining and Information Retrieval, Specialty Databases, Database Users and Administrators. Relational Model-Introduction to the Relational Model, Integrity Constraints over Relations, Enforcing Integrity, constraints, Querying relational data, Logical data base Design: ER to Relational, Introduction to Views, Destroying /Altering Tables and Views.

UNIT II
Relational Algebra and Calculus

Preliminaries, Relational Algebra, Relational calculus, Tuple relational Calculus, Domain relational calculus, Expressive Power of Algebra and calculus.
Introduction to Database design: Database Design and ER diagrams, Entities, Attributes and Entity sets, Relationships and Relationship sets, Additional features of ER Model, Conceptual Design with the ER Model, Conceptual Design for Large enterprises.

DATABASE MANAGEMENT SYSTEMS (C83PC1)

UNIT III
SQL

SQL data definition and Data types, Schema and catalog concepts in SQL, Queries, Constraints, Triggers: Form of Basic SQL Query, UNION, INTERSECT, and EXCEPT, Nested Queries, Aggregate Operators, NULL values, Complex Integrity Constraints in SQL, Triggers and Active Data bases, Designing Active Databases, NoSQL database (MongoDB introduction).
Schema Refinement and Normal Forms: Introduction to Schema Refinement, Functional Dependencies - Reasoning about FDs, Normal Forms - 1NF, 2NF, 3NF, Properties of Decompositions, Normalization, Schema Refinement in Database Design, Other Kinds of Dependencies.

UNIT IV
Transaction Management

Transactions, Transaction Concept, A Simple Transaction Model, Storage Structure, Transaction Atomicity and Durability, Transaction Isolation, serializability and Atomicity, Transaction Isolation Levels, Implementation of Isolation Levels. Concurrency Control, Lock–Based Protocols, Multiple Granularity, Timestamp-Based Protocols, Validation-Based Protocols, Multi-version Schemes. Recovery System-Failure Classification, Storage, Recovery and Atomicity, Recovery Algorithm, Buffer Management, Failure with loss of non-volatile storage, Early Lock Release and Logical Undo Operations, Remote Backup systems.

UNIT V
Indexing

Index Data Structures, and Comparison with File Organizations. Tree-Structured Indexing, Intuition for tree Indexes, Indexed Sequential Access Method (ISAM), B+ Trees: A Dynamic Index Structure, Search, Insert, Delete., Hash- Based Indexing, Static Hashing, Extendible hashing, Linear Hashing, Extendible vs. Linear Hashing.

Text Books:

1. Database Management Systems, Raghu Ramakrishnan, Johannes Gehrke, McGraw Hill Education (India) Private Limited, 3rd Edition.
2. Database System Concepts, A. Silberschatz, Henry. F. Korth, S. Sudarshan, McGraw Hill Education(India) Private Limited l, 6th edition.

Reference Books:

1. Database Systems, 6th edition, R Elmasri, Shamkant B.Navathe, Pearson Education.
2. Database System Concepts, Peter Rob & Carlos Coronel, Cengage Learning.
3. Introduction to Database Management, M. L. Gillenson and others, Wiley, Student Edition.
4. Database Development and Management, Lee Chao, Auerbach publications, Taylor & Francis Group.
5. Introduction to Database Systems, C. J. Date, Pearson Education.

TKR COLLEGE OF ENGINEERING AND TECHNOLOGY
(AUTONOMOUS)
(Sponsored by TKR Educational Society , Approved by AICTE, Affiliated by JNTUH,
Accredited by NBA & NAAC with 'A' Grade)

CSE (DATA SCIENCE)

B.Tech III Semester **L/T/P/C**
3 /0/ 0/ 3

OPERATING SYSTEMS (C83PC2)

Course Objective:

Analyze the basic components of a computer operating system, and the interactions among the various components. This course will cover an introduction on the policies for scheduling, deadlocks, memory management, synchronization, system calls, and file systems.

Course Outcomes:

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

1. Understand the basic concepts of operating system.
2. Understand the CPU scheduling and process scheduling.
3. Detect deadlocks and recovery the deadlocks using different mechanisms.
4. Understand the virtual memory management and storage file management system.
5. Implementing the file system.

UNIT I
Overview

Introduction-Operating system objectives, User view, System view, Operating system definition, Computer System Architecture, OS Structure, OS Operations, Process Management, Memory Management, Storage Management, Protection and Security, Computing Environments. Operating System services, User and OS Interface, System Calls, Types of System Calls, System Programs, Operating System Design and Implementation, OS Structure.

UNIT II
CPU Scheduling Process Concepts

The Process, Process State, Process Control Block, Threads, Process Scheduling, Scheduling Queues, Schedulers, Context Switch, Operations on Processes, System calls fork(),exec(),wait(),exit(), Inter-process communication-ordinary pipes and named pipes, message queues, shared memory, in Unix.
Process Scheduling-Basic concepts, Scheduling Criteria, Scheduling algorithms, Multiple-Processor Scheduling, Real Time Scheduling, Thread scheduling, Linux scheduling and Windows scheduling. Process Synchronization, Background, The Critical Section Problem, Peterson’s solution, Synchronization Hardware, Semaphores, Classic Problems of Synchronization, Monitors, Synchronization in Linux and Windows.

UNIT III
Deadlocks-System

Deadlocks-System Model, Deadlock Characterization, Methods for Handling Deadlocks, Deadlock Prevention, Deadlock Avoidance, Deadlock Detection, and Recovery from Deadlock.

OPERATING SYSTEMS (C83PC2)

Memory Management: Memory Management Strategies, Background, Swapping, Contiguous Memory Allocation, Segmentation, Paging, Structure of Page Table.

UNIT IV
Virtual Memory

Virtual Memory Management Background, Demand Paging, Copy-on-Write, Page Replacement, Page Replacement Algorithms, Allocation of Frames, Thrashing, Virtual memory in Windows. Storage Management File System- Concept of a File, System calls for file operations - open (), read(), write (), close (), seek (), unlink (), Access methods, Directory and Disk Structure, File System Mounting, File Sharing.

UNIT V
File System

File System Structure, File System Implementation, Directory Implementation, Allocation methods, Free-space Management, Efficiency, and Performance. Overview of Mass Storage Structure. System Protection, Goals of Protection, Principles of Protection, Domain of Protection, Access Matrix, Implementation of Access Matrix, Access Control, Revocation of Access Rights, Capability-Based Systems, Language-Based Protection.

Text Book:

1. Operating System Concepts, Abraham Silberschatz, Peter B. Galvin, Greg Gagne, 8th Edition, Wiley, 2016 India Edition.
2. Operating Systems – Internals and Design Principles, W. Stallings, 7th Edition, Pearson.

Reference Books:

1. Modern Operating Systems, Andrew S Tanenbaum, 3rd Edition, PHI
2. Operating Systems: A concept-based Approach, 2nd Edition, D.M. Dhamdhere, TMH.
3. Principles of Operating Systems, B. L. Stuart, Cengage learning, India Edition.
4. An Introduction to Operating Systems, P.C.P. Bhatt, PHI.
5. Principles of Operating systems, Naresh Chauhan, Oxford University Press.

TKR COLLEGE OF ENGINEERING AND TECHNOLOGY
(AUTONOMOUS)
(Sponsored by TKR Educational Society , Approved by AICTE, Affiliated by JNTUH,
Accredited by NBA & NAAC with 'A' Grade)

CSE (DATA SCIENCE)

B.Tech III Semester **L/T/P/C**
3 /0/ 0/ 3

R PROGRAMMING (C83PC3)

Course Objectives:

Gain knowledge on statistical data manipulation and analysis.

Course Outcomes:

After competition of this course, the students will be able to

1. Understand the basic functions of R and Create vectors in R.
2. Gain knowledge on creation of matrices and arrays in R.
3. Gain knowledge on creation of Factors and Data frames in R.
4. Understand and implement the searching and sorting techniques in R. and the file concepts in R.
5. Automate analyses and create new functions that extend the existing language features. Incorporates features found in object-oriented and functional programming languages.

UNIT I
Introduction to R

Introduction, Functions, Preview of Some Important R Data Structures, Regression Analysis of Exam Grades, Startup and Shutdown, Getting Help, The help() Function, The example() Function. Vectors, Scalars, Vectors, Arrays, and Matrices, Declarations, Common Vector Operations, Using all() and any(), Vectorized Operations, NA and NULL Values, Filtering, Vectorized if-then-else.

UNIT II
Matrices and Arrays

Creating Matrices, General Matrix Operations, Applying Functions to Matrix Rows and Columns, More on the Vector/Matrix Distinction, Avoiding Unintended Dimension Reduction, Naming Matrix Rows and Columns, Higher-Dimensional Arrays.
Lists: Creating Lists, General List Operations, Accessing List Components and Values Applying Functions to Lists, Recursive Lists.

UNIT III
Data Frames

Creating Data Frames, Other Matrix-Like Operations, Merging Data Frames, Applying Functions to Data Frames**.**
Factors and Tables: Factors and Levels, Common Functions Used with Factors, Working with Tables, Other Factor-and Table-Related Functions.

R PROGRAMMING (C83PC3)

UNIT IV
R Programming Structures

Control Statements, Arithmetic and Boolean Operators and Values, Default Values for Arguments, Return values, Functions Are Objects, Environment and Scope Issues, No Pointers in R, Writing Upstairs, Recursion, Replacement Functions, Anonymous Functions.
Math and Simulations in R: Math Functions, Functions for Statistical, Sorting, Set Operations.

UNIT V
Files

Accessing the Keyboard and Monitor, Reading and Writing Files, Accessing the Internet.
String Manipulation: String-Manipulation Functions.
Graphics: Creating Graphs, Customizing Graphs.

Text Books:

1. The Art of R Programming by Norman Matloff-No Starch Press.

Reference Books:

1. R Programming for Bioinformatics by Robert Gentleman—CRC Press.
2. Data Analytics using R by Seema Acharya-TMH.
3. Hands-On Programming with R by Grrett Grolemund-OREILLY.
4. Beginners guide for Data Analytics using R by Jeeva Jose-Khanna Publications.
5. R for Beginners by Sandip Bakshit-TMH.

TKR COLLEGE OF ENGINEERING AND TECHNOLOGY
(AUTONOMOUS)
(Sponsored by TKR Educational Society , Approved by AICTE, Affiliated by JNTUH,
Accredited by NBA & NAAC with 'A' Grade)

CSE (DATA SCIENCE)

B.Tech III Semester **L/T/P/C**
3 /0/ 0/ 3

FORMAL LANGUAGES & AUTOMATA THEORY (C83PC4)

Course Objective:

To provide introduction to some of the central ideas of theoretical computer science from the perspective of formal languages.

Course Outcomes:

After completion of the course the student will be able to

1. Understand the concept of abstract machines and build up the ability to recognize the formal languages.
2. Employ finite state machines for modeling and solving computing problems.
3. Design context free grammars for formal languages.
4. Normalizing the context Free Grammar and design Turing Machines.
5. Distinguish between decidability and intractable problems.

UNIT I
Introduction

Introduction to Finite Automata, Structural Representations, Automata and Complexity, the Central Concepts of Automata Theory – Alphabets, Strings, Languages, Problems.
Deterministic Finite Automata, Nondeterministic Finite Automata, an application: Text Search, Finite Automata with Epsilon-Transitions, Finite automata with output Epsilon Mealy and Moore machines, Equivalence of Mealy and Moore machines.

UNIT II
Regular Expressions

Finite Automata and Regular Expressions, Applications of Regular Expressions, Algebraic Laws for Regular Expressions, Automata and Regular expressions, Converting DFA's to Regular Expressions, Converting Regular Expressions to DFA, Properties of Regular Languages-Pumping Lemma for Regular Languages, Applications of the Pumping Lemma, Closure Properties of Regular Languages, Decision Properties of Regular Languages, Equivalence and Minimization of Automata.

UNIT III
Context-Free Grammars

Definition of Context-Free Grammars, Derivations Using a Grammar, Leftmost and Rightmost Derivations, the Language of a Grammar, Sentential Applications of Context-Free Grammars, Ambiguity in Grammars and Languages.
Push Down Automata:
Definition of the Pushdown Automaton, the Languages of a PDA, Equivalence of PDA's and CFG's, Deterministic Pushdown Automata, non-deterministic pushdown automata, power of Deterministic Pushdown Automata and Non-Deterministic Pushdown Automata.

FORMAL LANGUAGES & AUTOMATA THEORY (C83PC4)

UNIT IV
Normal Forms for Context- Free Grammars

The Pumping Lemma for Context-Free Languages, Closure Properties of Context-Free Languages. Decision Properties of CFL's - Complexity of Converting among CFG's and PDA's, Running time of conversions to Chomsky Normal Form. Introduction to Turing Machines: Problems That Computers Cannot Solve, The Turing Machine, Programming Techniques for Turing Machines, Extensions to the basic Turing machine, Restricted Turing Machines, Turing Machines, and Computers.

UNIT V
Undesirability

A Language that is Not Recursively Enumerable, An Undecidable Problem That is RE, Undecidable Problems about Turing Machines, Post's Correspondence Problem, Other Undecidable Problems, Intractable Problems: Polynomial time and space, Some NP-complete problems, The Classes P and NP, NP-Complete Problem.

Text Book:

1. Introduction to Automata Theory, Languages, and Computation, 3nd Edition, John E. Hopcroft, Rajeev Motwani, Jeffrey D. Ullman, Pearson Education.
2. Introduction to the Theory of Computation, Michael Sipser, 3rd edition, Cengage Learning.

Reference Books:

1. Introduction to Languages and the Theory of Computation, John C Martin, TMH.
2. Introduction to Computer Theory, Daniel I.A. Cohen, John Wiley.
3. A Text book on Automata Theory, P. K. Srimani, Nasir S. F. B, Cambridge University Press.
4. Introduction to Formal languages Automata Theory and Computation, Kamala Krithivasan, Rama R, Pearson.
5. Theory of Computer Science – Automata languages and computation, Mishra and Chandrashekaran, 2nd edition, PHI.

TKR COLLEGE OF ENGINEERING AND TECHNOLOGY
(AUTONOMOUS)
(Sponsored by TKR Educational Society , Approved by AICTE, Affiliated by JNTUH,
Accredited by NBA & NAAC with 'A' Grade)

CSE (DATA SCIENCE)

B.Tech III Semester **L/T/P/C**
0 /0/ 2/ 1

INTRODUCTION TO OBJECT-ORIENTED PROGRAMMING & DATA STRUCTURES USING JAVA LAB (CESOP2)

Course Objective:

To implement Object Oriented Data Structures concepts using Java programming in real time applications.

Course Outcomes:

After completion of course the student will be able to

1. Write the Java programs to implement object-oriented concepts.
2. Implement the Data Structures using Java programming.
3. Understand the usage of Java Collections.

List of Experiments to be performed during the Course:

1. Use Eclipse or Net bean platform and acquaint with the various menus. Create a test project, adda test class, and run it. See how you can use auto suggestions, auto fill. Try code formatter and code refactoring like renaming variables, methods, and classes. Try debug step by step with a small program of about 10 to 15 lines which contains at least one if else condition and a for loop.
2. Write a Java program to create an abstract class named Shape that contains two integers and an empty method named print Area (). Provide three classes named Rectangle, Triangle, and Circle such that each one of the classes extends the class Shape. Each one of the classes. contains only the method print Area () that prints the area of the given shape.
3. Write a Java program that demonstrate division by zero error exception handling.
4. Write a Java program to demonstrate recursions (ex: Factorial, X power N etc…).
5. Write a Java program to demonstrate all Stack operations.
6. Write a Java program to demonstrate all Queue operations .
7. Write a Java program to Demonstrate all DeQueue operations.
8. Write a Java program to Demonstrate all Singly Linked List operations.
9. Write a Java program to Demonstrate all Doubly Linked List operations.
10. Write a Java program to Demonstrate all Circularly Linked List operations.
11. Write a Java program to Demonstrate Java Collections.
12. Write a Java program to demonstrate Linear Search, Binary Search and Bubble Sort algorithm.

TKR COLLEGE OF ENGINEERING AND TECHNOLOGY
(AUTONOMOUS)
(Sponsored by TKR Educational Society , Approved by AICTE, Affiliated by JNTUH,
Accredited by NBA & NAAC with 'A' Grade)

CSE (DATA SCIENCE)

B.Tech III Semester **L/T/P/C**
0 /0/ 2/ 1

DATABASE MANAGEMENT SYSTEMS LAB (C83PC5)

Course Objective:

To emphasize on designing, developing and querying a database in the context of example database "Roadway travels".

Course Outcomes:

After Completion of this course the student will be able to

1. Design and implement a database schema for a given problem.
2. Apply the normalization techniques for development of application software to realistic problems.
3. Formulate queries using SQL DML/DDL/DCL commands.
4. Develop application programs using PL/SQL.

Roadway Travels

"Roadway Travels" is in business since 1997 with several buses connecting different places in India. Its main office is located in Hyderabad. The company wants to computerize its operations in the following areas:

Reservation & Cancellation

Reservations are directly handled by booking office, Reservations can be made 30 days in advance and tickets issued to passenger. One Passenger/person can book many tickets (to his/her family). Cancellations are also directly handed at the booking office.

In the process of computerization of Roadway Travels you have to design and develop a Database which consists the data of Buses, Passengers, Tickets, and Reservation and cancellation details. You should also develop queries using SQL to retrieve the data from the database.

The above process involves many steps like

1.Analyzing the problem and identifying the Entities and Relationships,
2.E-RModel
3.Relational Model
4.Normalization
5.Creating the database
6.Querying Students are supposed to work on these steps week wise and finally create a complete "Database System" to Roadway Travels. Examples are given at every experiment for guidance to students.

Experiment 1: E-R Model

Analyse the problem carefully and come up with the entities in it. Identify what data has to be persisted in the database. This contains the entities, attributes etc.

Identify the primary keys for all the entities. Identify the other keys like candidate keys, partial keys, if any.

Example:

Entities: 1. Bus 2. Ticket 3. Passenger

Relationships 1. Reservation. 2. Cancellation.

Primary Key Attributes: 1. Ticket ID (Ticket Entity) 2. Passport ID (Passenger Entity)
3. Bus_No. (Bus Entity)

Apart from the above mentioned entities you can identify more. The above mentioned are few.

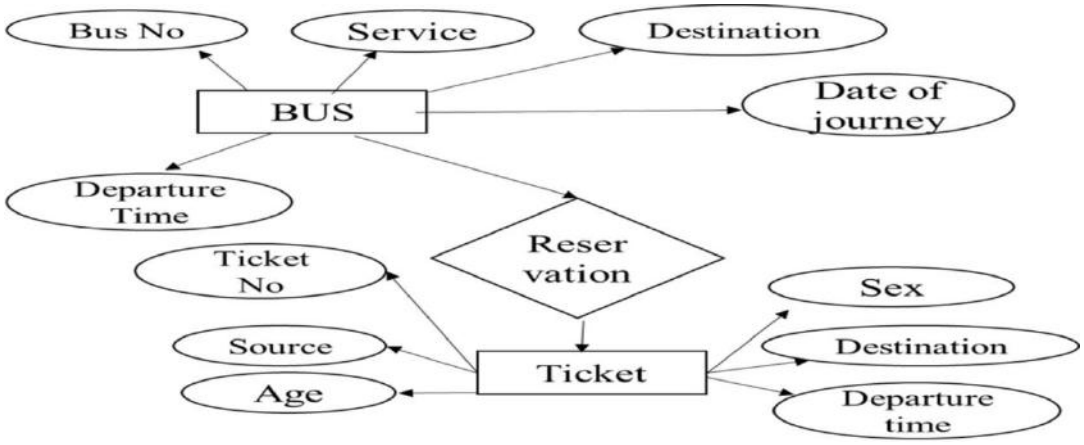
DATABASE MANAGEMENT SYSTEMS LAB (C83PC5)

Experiment 2: Concept design with E-R Model

Relate the entities appropriately. Apply cardinalities for each relationship. Identify strong entities and weak entities (if any). Indicate the type of relationships (total / partial). Try to incorporate generalization, aggregation, specialization etc wherever required.

Example: E-R diagram for bus

Experiment 3: Relational Model



Represent all the entities (Strong, Weak) in tabular fashion. Represent relationships in a tabular fashion. There are different ways of representing relationships as tables based on the cardinality. Represent attributes as columns in tables or as tables based on the requirement. Different types of attributes (Composite, Multi-valued, and Derived) have different way of representation.

Example: The passenger tables look as below. This is an example. You can add more attributes based on your E-R model. This is not a normalized table.

Passenger Name	Age	Sex	Address	PassportID	Ticket_ID

Experiment 4: Normalization

Database normalization is a technique for designing relational database tables to minimize duplication of information and, in so doing, to safeguard the database against certain types of logical or structural problems, namely data anomalies. For example, when multiple instances of a given piece of information occur in a table, the possibility exists that these instances will not be keptconsistentwhenthedatawithinthetableisupdated,leadingtoalossofdataintegrity.Atable that is sufficiently normalized is less vulnerable to problems of this kind, because its structure reflects the basic assumptions for when multiple instances of the same information should be representedbyasingleinstanceonly.FortheabovetableintheFirstnormalizationwecanremove the multi valued attribute Ticket_id and place it in another table along with the primary key of passenger.

First Normal Form: The above table can be divided into two tables as shown below.

Passenger Name	Age	Sex	Address	PassportID

PassportID	Ticket_ID

You can do the second and third normal forms if required. Any, how Normalized tables are given at the end.

DATABASE MANAGEMENT SYSTEMS LAB (C83PC5)

Experiment 5: Installation of Mysql and practicing DDL commands

Installation of MySql. In this week student will learn Creating databases, How to create tables, altering the database, dropping tables and databases if not required. Student will also try truncate, rename commands etc.

Example for creation of a normalized "Passenger" table.

CREATE TABLE Passenger (Passport_id INTEGER PRIMARY KEY, Name VARCHAR (50) Not NULL, Age Integer Not NULL, Sex Char, Address VARCHAR (50) Not NULL); Similarly create all other tables.

Experiment 6: Practicing DML commands

DML commands are used to for managing data within schema objects. Some examples:

SELECT - retrieve data from the a database INSERT - insert data into a table

UPDATE - updates existing data within a table

DELETE - deletes all records from a table, the space for the records remain

Inserting values into "Bus" table:

Insert into Bus values (1234,'hyderabad', 'tirupathi');
Insert into Bus values (2345,'hyderabd','Banglore');
Insert into Bus values (23,'hyderabd','Kolkata');
Insert into Bus values (45,'Tirupathi,'Banglore');
Insert into Bus values (34,'hyderabd','Chennai');

Inserting values into "Passenger" table:

Insert into Passenger values (1, 45,'ramesh', 45,'M','abc123');
Insert into Passenger values (2, 78,'geetha', 36,'F','abc124');
Insert into Passenger values (45, 90,'ram', 30,'M','abc12');
Insert into Passenger values (67, 89,'ravi', 50,'M','abc14');
Insert into Passenger values (56, 22,'seetha', 32,'F','abc55');

Few more Examples of DML commands:

Select * from Bus; (selects all the attributes and display)

UPDATE BUS SET Bus No = 1 WHERE BUS NO=2;

Experiment 7: Querying

In this week student are going to practice queries (along with sub queries) using ANY, ALL, IN, Exists, NOT EXISTS, UNION, INTERSECT, Constraints etc.

Practice the following Queries

1. Display unique PNR_No of all passengers.
2. Display all the names of male passengers.
3. Display the ticket numbers and names of all the passengers.
4. Find the ticket numbers of the passengers whose name start with 'r' and ends with'h'.
5. Find the names of passengers whose age is between 30 and45.
6. Display all the passengers names beginning with'A'
7. Display the sorted list of passengers names

Experiment 8: and Experiment 9: Querying (continued…)

Student are going to practice queries using Aggregate functions (COUNT, SUM, AVG, and MAX and MIN), GROUP BY, HAVING and Creation and dropping of Views.

Write a Query to display the Information present in the Passenger and cancellation tables.

1. Display the number of days in a week on which the 9W01 bus is available.
2. Find number of tickets booked for each PNR_no using GROUP BYCLAUSE.
3. Find the distinct PNR numbers that are present.
4. Find the number of tickets booked by a passenger where the number of seats is greater than1.

Hint: Use GROUP BY, WHERE and HAVING CLAUSES.

5.Find the total number of cancelled seats.

DATABASE MANAGEMENT SYSTEMS LAB (C83PC5)

Experiment 10: PL/SQL

1. Write a PL/SQL block for Addition of Two Numbers
2. Write a PL/SQL block for IF Condition
3. Write a PL/SQL block for IF and else condition
4. Write a PL/SQL block for greatest of three numbers using IF ANDELSEIF
5. Write a PL/SQL block for summation of odd numbers using for LOOP.

Experiment 11: Triggers

Creation of insert trigger, delete trigger, update trigger. Practice triggers using the above database.
Eg: CREATE TRIGGER updcheck BEFORE UPDATE ON passenger
FOR EACH ROW
BEGIN
IF NEW.TickentNO> 60 THEN SET New.Tickent no = Ticket no; ELSE
SET New.Ticketno = 0; END
IF; END;

Experiment 12: Procedures

Learn creation of stored procedure, Execution of procedure and modification of procedure. Practice procedures using the above database.

Eg: CREATE PROCEDURE myProc() BEGIN
SELECT COUNT (Tickets) FROM Ticket WHERE age>=40; End;

Experiment 13 Cursors

Declare a cursor that defines a result set. Open the cursor to establish the result set. Fetch the data into local variables as needed from the cursor, one row at a time. Close the cursor when done.

CREATE PROCEDURE myProc(in_customer_id INT) BEGIN
DECLARE v_idINT;
DECLARE v_name VARCHAR (30);
DECLARE c1 CURSOR FOR SELECT stdId, stdFirstname FROM students WHERE stdId=in_customer_id;
OPEN c1;
FETCH c1 into v_id, v_name; Close c1; END;

Tables BUS
Bus No: Varchar: PK (public key) Source : Varchar Destination : Varchar
Passenger
PPNO: Varchar (15)) : PK Name: Varchar (15) Age : int (4) Sex:Char (10) : Male / Female Address: VarChar (20)
Passenger_Tickets
PPNO: Varchar (15)) : PK Ticket_No: Numeric (9)
Reservation
PNR_No:Numeric(9):FKJourney_date:datetime(8)No_of_seats:int(8)Address:Varchar(50) Contact_No: Numeric (9) --> Should not be less than 9 and Should not accept any other character other than Integer Status: Char (2) : Yes /No
Cancellation
PNR_No: Numeric(9) : FK Journey_date : datetime(8) No_of_seats : int (8) Address : Varchar
(50) Contact_No: Numeric (9) --> Should not be less than 9 and Should not accept any other character other than Integer Status: Char (2) : Yes / No
Ticket
Ticket_No: Numeric (9): PK Journey_date : datetime(8) Age : int (4) Sex:Char(10) : Male / Female Source : Varchar Destination : Varchar Dep_time : Varchar

TKR COLLEGE OF ENGINEERING AND TECHNOLOGY
(AUTONOMOUS)
(Sponsored by TKR Educational Society , Approved by AICTE, Affiliated by JNTUH,
Accredited by NBA & NAAC with 'A' Grade)

CSE (DATA SCIENCE)

B.Tech III Semester **L/T/P/C**
0 /0/ 2/ 1

R PROGRAMMING LAB (C83PC6)

Course Objective:

This course takes student from having no previous experience in programming to an intermediate level in R. Student will learn the basic toolkit of the data-oriented professional, and learn where and how to learn more advanced skills when needed.

Course Outcomes:

At the end of this course, student has all the computational tools to:

1. Implement empirical economic analyses.
2. Participate in online data science challenges.
3. Student learns on his own further R, or other programming languages.

List of Programs:

1. R and R studio set-up.
2. R types, vectors and writing functions.
3. More R types, vectorization and efficient R code.
4. Importing data, time series data and Google Trends.
5. A data science project – a first look.
6. Data manipulation with tidyr.
7. Plotting with ggplot2.
8. R style guide, git integration, projects in R Studio.
9. Maps, plots and Amazon’s choice for a next office building.
10. 10.A data science project – more advanced.

TKR COLLEGE OF ENGINEERING AND TECHNOLOGY
(AUTONOMOUS)
(Sponsored by TKR Educational Society , Approved by AICTE, Affiliated by JNTUH,
Accredited by NBA & NAAC with 'A' Grade)

CSE (DATA SCIENCE)

B.Tech IV Semester **L/T/P/C**
3 /0/ 0/ 3

PROBABILITY & ALGEBRA (CBSM6)

Course Objectives:

To learn:

1. Concepts of Basic probability.
2. Random variables that describe randomness or an uncertainty in certain realistic situation.
3. The study of discrete and continuous distributions predominantly describes important probability distributions.
4. To relate practical examples to the appropriate set, function or relation model, and interpret the associated operations and terminology in context.
5. Introduce the concepts of semi groups, monoids, groups, sub-groups, abelian groups, Isomorphism and homomorphism of groups.

Course Outcomes:

After learning the contents of this course, the student will be able to
CO1: Learn the concept of basic probability to solve the real life problems.
CO2: To solve problems on discrete and continuous random variables.
CO3: Learn various discrete and continuous probability distribution and their properties.
CO4: Solve problems based on area properties of standard normal distribution.
CO5: Illustrate the basic terminology of functions, relations, sets, and demonstrate knowledge of their associated operations.
CO6: Understand the importance of algebraic properties with regard to working within various number systems.

UNIT I
Probability

Basic concepts of probability, Axiomatic definition of probability, Addition theorem, conditional probability, multiplication theorem, Independent events, Baye's theorem.

UNIT II
Random Variables

Random variables –discrete and continuous, Mathematical expectation, Variance, co-variance, joint and marginal probability density function, statistical independence.

PROBABILITY & ALGEBRA (CBSM6)

UNIT III
Distributions

Probability mass function, density function of Binomial, Poisson and Normal distributions related properties.

UNIT IV
Relations

Properties of Binary relations, equivalence, transitive closure, compatibility and partial ordering Relations, Hasse diagram.
Functions: Inverse function, composition of functions, recursive functions.

UNIT V
Groups

Algebraic structures, examples and general properties, Semi groups and monoids, Groups, Sub groups, cosets and Lagranges theorem, homomorphism, and isomorphism of groups, cyclic groups, permutation groups.

Text Books:

1. Probability & Statistics for Engineers by G.S.S. Bhisma Rao, SciTech Publications.
2. Discrete Mathematics for Computer scientists & Mathematicians, J. L. Mott, A. Kandel, T.P.Baker.

Reference Books:

1. W.Feller- An introduction to probability theory and its applications- Vol.1- 3rd edition Wiley-1968.
2. Probability & Statistics for Engineers, Millers and John E. Freund, Prentice Hall of India.
3. Discrete mathematical structures theory and applications- malik & Sen Cengage.

TKR COLLEGE OF ENGINEERING AND TECHNOLOGY
(AUTONOMOUS)
(Sponsored by TKR Educational Society , Approved by AICTE, Affiliated by JNTUH,
Accredited by NBA & NAAC with 'A' Grade)

CSE (DATA SCIENCE)

B.Tech IV Semester **L/T/P/C**
3 /0/ 0/ 3

DATA WAREHOUSING & DATA MINING (C84PC1)

Course Objectives:

Study data warehouse principles and its working learn data mining concepts understand association rules mining. Discuss classification algorithms learn how data is grouped using clustering techniques.

Course Outcomes:

1. Be familiar with mathematical foundations of data mining tools.
2. Understand and implement classical models and algorithms in data warehouses and data mining.
3. Characterize the kinds of patterns that can be discovered by association rule mining, classification and clustering
4. Master data mining techniques in various applications like social, scientific and environmental context
5. Develop skill in selecting the appropriate data mining algorithm for solving practical problems.

UNIT I
Data Warehouse

Introduction to Data warehouse, Difference between operational database systems and data warehouses. Data warehouse Characteristics, Data warehouse Architecture and its Components, Extraction – Transformation – Loading, Logical (Multi – Dimensional), Data Modelling, Schema Design, Star, Snow Flake Schema and Fact Constellation, Fact Table, Fully Addictive, Semi – Addictive, Non Addictive Measures; Fact Constellation, Fact Table, Fully Addictive, Semi – Addictive, Non Addictive Measures; Fact – Less – Facts, Dimension Table Characteristics; OLAP Cube, OLAP Operations, OLAP Server Architecture – ROLAP, MOLAP and HOLAP

UNIT II
Introduction to Data Mining

Introduction, What is Data Mining, Definition, KDD, Challenges, Data Mining Tasks, Data Preprocessing, Data Cleaning, Missing data, Dimensionality Reduction, Feature Subset Selection, Discretization and Binarization, Data Transformation; Measures of Similarity and Dissimilarity- Basics.

Unit III
Association Rules

Problem Definition, Frequent Item Set Generation, The APRIORI Principle, Support and Confidence Measures, Association Rule Generation; APRIOIRI Algorithm, The Partition Algorithms, FP-Growth Algorithms, Compact Representation of Frequent Item Set, Maximal Frequent Item Set, Closed Frequent Item Set.

DATA WAREHOUSING & DATA MINING (C84PC1)

UNIT IV
Classification

Problem Definition, General Approaches to solving a classification problem, Evaluation of Classifiers, Classification techniques, Decision Trees-Decision tree Construction, Methods for Expressing attribute test conditions, Measures for Selecting the Best Split, Algorithm for Decision tree Induction; Naïve – Bayes Classifier, Bayesian Belief Networks; K- Nearest neighbor classification-Algorithm and Characteristics.

UNIT V
Clustering

Problem Definition, Clustering Overview, Evaluation of Clustering Algorithms, Partitioning Clustering-K-Means Algorithm, K-Means Additional issues, PAM Algorithm; Hierarchical Clustering-Agglomerative Methods and divisive methods, Basic Agglomerative Hierarchical Clustering Algorithm, Specific techniques, Key Issues in Hierarchical Clustering, Strengths and Weakness; Outlier Detection.

Text Books:

1. Data Mining- Concepts and Techniques- Jiawei Han, Micheline Kamber, Morgan Kaufmann Publishers, Elsevier 2 Edition, 2006.
2. Introduction to Data Mining, Pang-Ning Tan, Vipin Kumar, Michael Steinbanch, Pearson Education.

Reference Books:

1. Data Mining Techniques, Arun K Pujari, 3rd Edition, Universities Press.
2. Data Warehousing Fundamentals, Paulraj Ponnaiah, Wiley Student Edition.
3. The Data Warehousing Life Cycle Toolkit – Ralph Kimbal. Wiley Student Edition.
4. Data Mining, Vikaram Pudi, P Radha Krishna, Oxford University Press.

TKR COLLEGE OF ENGINEERING AND TECHNOLOGY
(AUTONOMOUS)
(Sponsored by TKR Educational Society , Approved by AICTE, Affiliated by JNTUH,
Accredited by NBA & NAAC with 'A' Grade)

CSE (DATA SCIENCE)

B.Tech IV Semester **L/T/P/C**
3 /0/ 0/ 3

INFORMATION SECURITY (C84PC2)

Course Objective:

To understand and learn the objectives of Network security, Cryptographic algorithms.

Course Outcomes:

After completion of the course student will be able to

1. Understand the security concepts and classical encryption techniques.
2. Understand the symmetric and asymmetric key algorithms
3. Understand the authentication and hash algorithms.
4. Understand cryptographic algorithms for web, E-mail and security issues.
5. Understand the IP security and system security.

UNIT I
Security Concepts

Introduction, security trends, OSI Architecture, security attacks, security services, security mechanisms, A Model for Network Security.
Cryptography Concepts and Techniques: Introduction, Plain Text and cipher text, substitution techniques(Caesar cipher, Playfair cipher, Hill cipher), transposition techniques, steganography.

UNIT II
Symmetric Key Ciphers

Block Cipher principles, DES, AES, Block Cipher Modes of Operation, Stream ciphers, RC4.
Asymmetric Key Ciphers: Principles of public key cryptosystems, RSA algorithm, Difffie-Hellman Key Exchange, Elliptic Curve Cryptography and Arithmetic.

UNIT III
Cryptographic Hash Functions

Authentication requirements and Functions, Message Authentication Code, Secure Hash Algorithm (SHA-512), Message authentication codes: HMAC, CMAC, Digital signatures,
AUTHENTICATION APPLICATIONS: Kerberos, X.509 Authentication Service, Public – Key Infrastructure.

INFORMATION SECURITY (C84PC2)

UNIT IV
Web Security

Web security considerations, Secure Socket Layer, and Transport Layer Security, Secure Electronic Transaction.
E-Mail Security: Pretty Good Privacy, S/MIME.

UNIT V
IP Security

IP Security overview, IP Security architecture, Authentication Header, Encapsulating Security Payload, Combining Security Associations, Key Management.
System Security: Intruders, Intrusion Detection, Password Management.

Text Books:

1. Cryptography and Network Security - Principles and Practice: William Stallings, Pearson Education, 4th Edition
2. Cryptography and Network Security: Atul Kahate, Mc Graw Hill, 3rd Edition

Reference Books:

1. Cryptography and Network Security: C K Shyamala, N Harini, Dr T R Padmanabhan, Wiley India, 1st Edition.
2. Cryptography and Network Security: Forouzan, Mukhopadhyay, Mc Graw Hill,3rd Edition.
3. Information Security, Principles, and Practice: Mark Stamp, Wiley India.
4. Principles of Computer Security: WM, Arthur Conklin, Greg White, TMH.
5. Introduction to Network Security: Neal Krawetz, CENGAGE Learning.
6. Network Security and Cryptography: Bernard Menezes, CENGAGE Learning.

TKR COLLEGE OF ENGINEERING AND TECHNOLOGY
(AUTONOMOUS)
(Sponsored by TKR Educational Society , Approved by AICTE, Affiliated by JNTUH,
Accredited by NBA & NAAC with 'A' Grade)

CSE (DATA SCIENCE)

B.Tech IV Semester **L/T/P/C**
3 /0/ 0/ 3

DESIGN AND ANALYSIS OF ALGORITHMS (C84PC3)

Course Objective:

To understand the design paradigms for developing an algorithm and analyzing it for a given problem.

Course Outcomes:

1. Argue the correctness of algorithms using inductive proofs and invariants.
2. Apply important algorithmic design paradigms and methods of analysis.
3. Synthesize efficient algorithms in common engineering design situations such as the greedy, divide and conquer, dynamic programming, backtracking and branch-bound.
4. Explain the different ways to analyze randomized algorithms (expected running time, probability of error)
5. Differentiate between tractable and intractable problems.

UNIT I
Introduction

Algorithm definition, Algorithm Specification, Performance Analysis-Space complexity, Time complexity, Randomized Algorithms. Divide and conquer- General method, applications – Binary search, Merge sort, Quick sort, Strassen's Matrix Multiplication.

UNIT II
Disjoint Set Operations

Disjoint set operations, union and find algorithms, AND/OR graphs, Connected Components and Spanning trees, Bi-connected components Backtracking-General method, applications the 8-queen problem, sum of subsets problem, graph coloring, Hamiltonian cycles.

UNIT III
Greedy Method

General method, applications- Knapsack problem, Job sequencing with deadlines, Minimum cost spanning trees, Single source shortest path problem.

UNIT IV
Dynamic Programming

General Method, applications- Chained matrix multiplication, All pairs shortest path problem, Optimal binary search trees, 0/1 knapsack problem, Reliability design, Travelling sales person problem.

DESIGN AND ANALYSIS OF ALGORITHMS (C84PC3)

UNIT V
Branch and Bound

General Method, applications-0/1 Knapsack problem, LC Branch and Bound solution, FIFO Branch and Bound solution, Traveling sales person problem. NP-Hard and NP Complete problems- Basic concepts, Non-deterministic algorithms, NP – Hard and NP-Complete classes, Cook's theorem.

Text Books:

1. Fundamentals of Computer Algorithms, 2nd Edition, Ellis Horowitz, Sartaj Sahni and S.Rajasekharan, Universities Press.
2. Design and Analysis of Algorithms, P. H. Dave, H. B. Dave, 2nd edition, Pearson Education.

Reference Books:

1. Algorithm Design: Foundations, Analysis and Internet examples, M. T. Goodrich and R. Tomassia, John Wiley and sons.
2. Design and Analysis of Algorithms, S. Sridhar, Oxford Univ. Press
3. Design and Analysis of algorithms, Aho, Ullman and Hopcroft, Pearson Education.
4. Foundations of Algorithms, R. Neapolitan and K. Naimipour, 4th edition, Jones and Bartlett Student edition.
5. Introduction to Algorithms, 3rd Edition, T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein PHI.

TKR COLLEGE OF ENGINEERING AND TECHNOLOGY
(AUTONOMOUS)
(Sponsored by TKR Educational Society , Approved by AICTE, Affiliated by JNTUH,
Accredited by NBA & NAAC with 'A' Grade)

CSE (DATA SCIENCE)

B.Tech IV Semester **L/T/P/C**
3 /0/ 0/ 3

DATA VISUALIZATION (C84PC4)

Course Objectives:

To understand the visual representation of structured and un structured data.

Course Outcomes:

After completion of course, the students will be able to

1. Understand the visualization and Data basics
2. Understand the Visualization process and know the representation of Spatial & Geo spatial data
3. Analyze various Visualization techniques for Multivariate data and other structures of data
4. Interacting the different operators and different data spaces
5. Design effective visualization of modern toolkits

UNIT I
Introduction

What is Visualization, History, Relationship visualization with other fields, The visualization Process, Pseudocode Conventions, The Scatter plot
Data Foundations: Types of Data, Structure within and between the records, Data Processing.

UNIT II
Visualization Foundations

The Visual Process, Semiology of Graphical Symbols, The Eight Visual Variables, Historical Perspective, Taxonomies.
Visualization Techniques for Spatial Data: One-Dimensional Data, Two-Dimensional Data, Three-Dimensional Data, Dynamic Data, Combining Techniques.
Visualization Techniques for Geospatial Data: Visualizing Spatial Data, Visualization of Point Data, Visualization of Line Data, Visualization of Area Data.

UNIT III
Visualization Techniques for Multivariate Data

Point-Based Techniques, Line-Based Techniques, Region-Based Techniques, Combinations of Techniques, Visualization Techniques for Trees, Graphs, and Networks: Displaying Hierarchical Structures, Displaying Arbitrary Graphs/Networks.

DATA VISUALIZATION (C84PC4)

UNIT IV
Text and Document Visualization

Levels of Text Representation, The Vector Space Model, Single Document Visualizations, Document Collection Visualizations.
Interaction Concepts: Interaction Operators, Interaction Operands and Spaces, A Unified Framework.
Interaction Techniques: Object Space, Data Space, Attribute Space, Data Structure Space, Visualization Structure Space, Animating Transformations, Interaction Control.

UNIT V
Designing Effective Visualizations

Steps in Designing Visualizations, Problems in Designing Effective Visualizations
Comparing and Evaluating Visualization Techniques: User Tasks, User Characteristics, Data Characteristics, Visualization Characteristics, Structures for Evaluating Visualizations.
Visualization Systems: Systems Based on Data Type, Systems Based on Analysis Type, Text Analysis and Visualization, Modern Integrated Visualization Systems, Toolkits

Text Books:

1. Interactive Data Visualization Foundations, Techniques, and Applications by Grinstein, Georges Keim, Daniel Ward, Matthew O , CRC Press Taylor & Francis Group.
2. Digital Image. Processing. Third Edition. Rafael C. Gonzalez. University of Tennessee. Richard E. Woods. NledData Interactive. Pearson International Edition.

TKR COLLEGE OF ENGINEERING AND TECHNOLOGY
(AUTONOMOUS)
(Sponsored by TKR Educational Society , Approved by AICTE, Affiliated by JNTUH,
Accredited by NBA & NAAC with 'A' Grade)

CSE (DATA SCIENCE)

B.Tech IV Semester **L/T/P/C**
3 /0/ 0/ 3

PYTHON PROGRAMMING (C84PC5)

Course Objective:

Enable the student to do Python Programming which includes Regular Expressions and GUI

Course Outcomes:

After completion of course the student will be able to

1. Examine Python syntax and semantics and be fluent in the use of Python flow control and functions.
2. Demonstrate proficiency in handling Strings and File Systems.
3. Create, run and manipulate Python Programs using core data structures like Lists, Dictionaries and use Regular Expressions.
4. Interpret the concepts of Object-Oriented Programming as used in Python.
5. Implement exemplary applications related to Network Programming, Web Services and Databases in Python.

UNIT I
Introduction

Introduction to Python, History, Need of Python Programming, features Applications, python environment setup, Basic syntax, Variables, Data Types, Keywords, Input-Output, Indentation, script structure, Running Python Scripts.
Operators: Arithmetic Operators, Comparison (Relational) Operators, Assignment Operators, Logical Operators, Bitwise Operators, Membership Operators, Identity Operators, Expressions and order of evaluations, Conditional statements if, if-else Looping Control Structures for, while Control Statements: Break, Continue, Pass.

UNIT II
Functions

Defining Functions, Calling Functions, Passing Arguments, Keyword Arguments, Default Arguments, Variable-length arguments, Anonymous Functions, Fruitful Functions (Function Returning Values), Scope of the Variables in a Function - Global and Local Variables.
Data Structures : Lists, Tuples, dictionaries, sets, Sequences, Comprehensions.

PYTHON PROGRAMMING (C84PC5)

UNIT III
Regular Expressions

Introduction/Motivation , Special Symbols and Characters, REs and Python.
OBJECT ORIENTED PROGRAMMING IN PYTHON
Classes, 'self-variable', Methods, Constructor Method, Inheritance, Overriding Methods, Data hiding. ERROR AND EXCEPTIONS Difference between an error and Exception, Handling exceptions, try, except block, Raising Exceptions and User Defined Exceptions.

UNIT IV
Files

File input/output, Text processing file functions.
MODULES and Introduction to Packages, Creating modules, import statement, from. Name spacing, Packages, using packages, implementing packages: numpy, iterator tools, scipy, matplotlib.

UNIT V
GUI Programming

Introduction, Tkinter and Python Programming, Brief Tour of other GUIs, Related Modules and other GUIs.
Database Programming: Introduction, Python Database, Application Programmer's Interface (DB-API), Object Relational Managers (ORMs), Related Modules.

Text Book:

1. Core Python Programming, Wesley J. Chun, Second Edition, Pearson.

Reference Books:

1. Allen Downey, "Think Python", Second Edition , Green Tea Press.
2. Introduction to Computation & Programming Using Python, Spring 2013 Edition, By John V.Guttag.
3. 3. Programming in Python 3: A Complete Introduction to the Python Language (Developer's Library), by Mark Summerfield, 2nd Edition.

TKR COLLEGE OF ENGINEERING AND TECHNOLOGY
(AUTONOMOUS)
(Sponsored by TKR Educational Society , Approved by AICTE, Affiliated by JNTUH,
Accredited by NBA & NAAC with 'A' Grade)

CSE (DATA SCIENCE)

B.Tech IV Semester **L/T/P/C**
0 /0/ 2/ 1

DATA WAREHOUSING AND DATA MINING LAB (C84PC6)

Course Objective:

Learn how to build a data warehouse and query it, perform data mining tasks using a data mining toolkit and understand the data sets and data preprocessing.

Course Outcomes:

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

1. Ability to understand the various kinds of tools
2. Demonstrate the classification, clustering and etc. in large data sets.
3. Ability to add mining algorithms as a component to the exiting tools.
4. Ability to apply mining techniques for realistic data

List of experiments and Tasks

Experiment-1: Build Data Warehouse and Explore WEKA

A. Build Data Warehouse/Data Mart (using open source tools like Pentaho Data Integration Tool, Pentaho Business Analytics; or other data warehouse tools like Microsoft-SSIS, Informatica, Business Objects, etc.,)

(i) Identify source tables and populate sample data.

(ii) Design multi-dimensional data models namely Star, Snowflake and Fact Constellation schemas for any one enterprise (ex. Banking, Insurance, Finance, Healthcare, manufacturing, Automobiles, sales etc).

(iii) Write ETL scripts and implement using data warehouse tools

(iv) Perform Various OLAP operations such as slice, dice, roll up, drill up and pivot.

(v). Explore visualization features of the tool for analysis like identifying trends etc.

B. Explore WEKA Data Mining/Machine Learning Toolkit.

(i) Downloading and/or installation of WEKA data mining toolkit.

(ii) Understand the features of WEKA tool kit such as Explorer, Knowledge flow interface, Experimenter, command-line interface.

(iii) Navigate the options available in the WEKA(ex.select attributes panel, preprocess panel, classify panel, cluster panel, associate panel and visualize)

(iv) Study the ARFF file format

(v) Explore the available data sets in WEKA.

(vi) Load a data set (ex.Weather dataset,Iris dataset,etc.)

(vii) Load each dataset and observe the following:

(vii.i) List attribute names and they types

(vii.ii) Number of records in each dataset.

(vii.iii) Identify the class attribute (if any)

(vii.iv) Plot Histogram

(vii.v) Determine the number of records for each class

(vii.vi) Visualize the data in various dimensions

DATA WAREHOUSING AND DATA MINING LAB (C84PC6)

Experiment-2: Perform data preprocessing tasks and Demonstrate performing association rule mining on data sets

- **A.** Explore various options in Weka for preprocessing data and apply (like Discretization Filters, Resample filter, etc.) in each dataset.
- **B.** Load each dataset into Weka and run Apriori algorithm with different support and confidence values. Study the rules generated.
- **C.** Apply different discretization filters on numerical attributes and run the Apriori association rule algorithm. Study the rules generated. Derive interesting insights and observe the effect of discretization in the rule generation process.

Experiment-3: Demonstrate performing classification on data sets.

- **A.** Load each dataset into Weka and run id3, j48 classification algorithm, study the classifier output. Compute entropy values, Kappa statistic.
- **B.** Extract if-then rues from decision tree generated by classifier, Observe the confusion matrix and derive Accuracy, F- measure, TPrate, FPrate , Precision and recall values. Apply cross-validation strategy with various fold levels and compare the accuracy results.
- **C.** Load each dataset into Weka and perform Naïve-bayes classification and k-Nearest Neighbor classification, Interpret the results obtained.
- **D.** Plot RoC Curves.
- **E.** Compare classification results of ID3,J48, Naïve-Bayes and k-NN classifiers for each dataset , and reduce which classifier is performing best and poor for each dataset and justify.

Experiment-4: Demonstrate Performing Clustering on Data Sets Clustering Tab

- **A.** Load each dataset into Weka and run simple k-means clustering algorithm with different values of k(number of desired clusters). Study the clusters formed. Observe the sum of squared errors and centroids, and derive insights.
- **B.** Explore other clustering techniques available in Weka.
- **C.** Explore visualization features of weka to visualize the clusters. Derive interesting insights and explain.

Experiment-5: Demonstrate Performing Regression on Data Sets

- **A.** Load each dataset into Weka and build Linear Regression model. Study the cluster formed. Use training set option. Interpret the regression model and derive patterns and conclusions from the regression results.
- **B.** Use options cross-validation and percentage split and repeat running the Linear Regression Model. Observe the results and derive meaningful results.
- **C.** Explore Simple linear regression techniques that only looks at one variable.

Experiment-5: Sample Programs using German Credit Data.

Task 1: Credit Risk Assessment

Description: The business of banks is making loans. Assessing the credit worthiness of an applicant is of crucial importance. You have to develop a system to help a loan officer decide whether the credit of a customer is good or bad. A bank's business rules regarding loans must consider two opposing factors. On the one hand, a bank wants to make as many loans as possible.

Interest on these loans is the banks profit source. On the other hand, a bank cannot afford to make too many bad loans. Too many bad loans could lead to the collapse of the bank. Th**e** bank's loan policy must involved a compromise. Not too strict and not too lenient. To do the assignment, student first and foremost need some knowledge about the world of credit. Student can acquire such knowledge in a number of ways.

DATA WAREHOUSING AND DATA MINING LAB (C84PC6)

1. Knowledge engineering: Find a loan officer who is willing to talk. Interview him/her and try to represent him/her knowledge in a number of ways.
2. Books: Find some training manuals for loan officers or perhaps a suitable textbook on finance. Translate this knowledge from text form to production rule form.
3. Common sense: Imagine yourself as a loan officer and make up reasonable rules which can be used to judge the credit worthiness of a loan applicant.
4. Case histories: Find records of actual cases where competent loan officers correctly judged when and not to. Approve a loan application.

The German Credit Data

Actual historical credit data is not always easy to come by because of confidentiality rules. Here is one such data set. Consisting of 1000 actual cases collected in Germany.

In spite of the fact that the data is German, you should probably make use of it for this assignment(Unless you really can consult a real loan officer!)

There are 20 attributes used in judging a loan applicant (ie., 7 Numerical attributes and 13 Categorical or Nominal attributes). The goal is to classify the applicant into one of two categories. Good or Bad.

Subtasks:

1. List all the categorical (or nominal) attributes and the real valued attributes separately.

2. What attributes do you think might be crucial in making the credit assessment? Come up with some simple rules in plain English using your selected attributes.

3. One type of model that you can create is a Decision tree . train a Decision tree using the complete data set as the training data. Report the model obtained after training.

4. Suppose you use your above model trained on the complete dataset, and classify credit good/bad for each of the examples in the dataset. What % of examples can you classify correctly?(This is also called testing on the training set) why do you think can not get 100% training accuracy?

5. Is testing on the training set as you did above a good idea? Why or why not?

6. One approach for solving the problem encountered in the previous question is using cross-validation? Describe what is cross validation briefly. Train a decision tree again using cross validation and report your results. Does accuracy increase/decrease? Why?

7. Check to see if the data shows a bias against “foreign workers” or “personal-status”. One way to do this is to remove these attributes from the data set and see if the decision tree created in those cases is significantly different from the full dataset case which you have already done. Did removing these attributes have any significantly effect? Discuss.

8. Another question might be, do you really need to input so many attributes to get good results? May be only a few would do. For example, you could try just having attributes 2,3,5,7,10,17 and 21. Try out some combinations.(You had removed two attributes in problem 7. Remember to reload the arff data file to get all the attributes initially before you start selecting the ones you want.)

DATA WAREHOUSING AND DATA MINING LAB (C84PC6)

9. Sometimes, The cost of rejecting an applicant who actually has good credit might be higher than accepting an applicant who has bad credit. Instead of counting the misclassification equally in both cases, give a higher cost to the first case (say cost 5) and lower cost to the second case. By using a cost matrix in weak. Train your decision tree and report the Decision Tree and cross validation results. Are they significantly different from results obtained in problem 6.

10. Do you think, it is a good idea to predict simple decision trees instead of having long complex decision tress? How does the complexity of a Decision Tree relate to the bias of the model?

11. You can make your Decision Trees simpler by pruning the nodes. One approach is to use Reduced Error Pruning. Explain this idea briefly. Try reduced error pruning for training your Decision Trees using cross validation and report the Decision Trees you obtain? Also Report your accuracy using the pruned model Does your Accuracy increase?

12. How can you convert a Decision Tree into “if-then-else rules”. Make up your own small Decision Tree consisting 2-3 levels and convert into a set of rules. There also exist different classifiers that output the model in the form of rules. One such classifier in weka is rules, train this model and report the set of rules obtained. Sometimes just one attribute can be good enough in making the decision, yes, just one! Can you predict what attribute that might be in this data set? OneR classifier uses a single attribute to make decisions (it chooses the attribute based on minimum error).Report the rule obtained by training a oneR classifier. Rank the performance of j48, OneR.

TKR COLLEGE OF ENGINEERING AND TECHNOLOGY
(AUTONOMOUS)
(Sponsored by TKR Educational Society , Approved by AICTE, Affiliated by JNTUH,
Accredited by NBA & NAAC with 'A' Grade)

CSE (DATA SCIENCE)

B.Tech IV Semester **L/T/P/C**
0 /0/ 2/ 1

DATA VISUALIZATION LAB (C84PC7)

Course Objectives:

To obtain practical experience using Tableau public or similar tools.

Course Outcomes:

After completion of the lab student will be able to:
Visualize the different types of data.

List of Sample Problems:

1. Tableau/ Power BI or similar tools setup for Data Visualization (Importing packages etc.…).
2. Extracting and operations of data from different sources.
3. Working on worksheets.
4. Applying the Calculations (Operators, Functions, Numerical Calculations, String, Date, Table).
5. Usage of Different types of Filter & Sort.
6. Construction of Charts (Line, BAR, etc.).
7. Creation of Dashboard (Optional).

TKR COLLEGE OF ENGINEERING AND TECHNOLOGY
(AUTONOMOUS)
(Sponsored by TKR Educational Society , Approved by AICTE, Affiliated by JNTUH,
Accredited by NBA & NAAC with 'A' Grade)

CSE (DATA SCIENCE)

B.Tech IV Semester **L/T/P/C**
0 /0/ 2/ 1

PYTHON PROGRAMMING LAB (C84PC8)

Course Objectives:

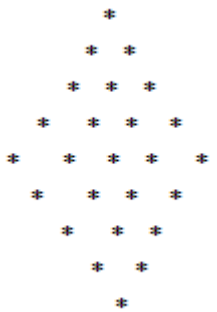
1. To be able to introduce core programming basics and program design with functions using Python programming language.
2. To understand a range of Object-Oriented Programming, as well as in-depth data and information processing techniques.
3. To understand the high-performance programs designed to strengthen the practical expertise.

Course Outcomes:

1. Student should be able to understand the basic concepts scripting and the contributions of scripting language.
2. Ability to explore python especially the object-oriented concepts, and the built-in objects of Python.
3. Ability to create practical and contemporary applications such as TCP/IP network programming, Web applications, discrete event simulations.

List of Programs:

1. Write a program to demonstrate different number data types in Python.
2. Write a program to perform different Arithmetic Operations on numbers in Python.
3. Write a program to create, concatenate and print a string and accessing sub-string from a given string.
4. Write a python script to print the current date in the following format “Sun May 29 02:26:23 IST 2017”
5. Write a program to create, append, and remove lists in python.
6. Write a program to demonstrate working with tuples in python.
7. Write a program to demonstrate working with dictionaries in python.
8. Write a python program to find largest of three numbers.
9. Write a Python program to convert temperatures to and from Celsius, Fahrenheit. [Formula: c/5 = f-32/9]
10. Write a Python program to construct the following pattern, using a nested for loop



PYTHON PROGRAMMING LAB (C84PC8)

11. Write a Python script that prints prime numbers less than 20.
12. Write a python program to find factorial of a number using Recursion.
13. Write a program that accepts the lengths of three sides of a triangle as inputs. The program output should indicate whether or not the triangle is a right angled triangle (Recall from the Pythagorean Theorem that in a right angled triangle, the square of one side equals the sum of the squares of the other two sides).
14. Write a python program to define a module to find Fibonacci Numbers and import the module to another program.
15. Write a python program to define a module and import a specific function in that module to another program.
16. Write a script named copyfile.py. This script should prompt the user for the names of two text files. The contents of the first file should be input and written to the second file.
17. Write a program that inputs a text file. The program should print all of the unique words in the file in alphabetical order.
18. Write a Python class to convert an integer to a roman numeral.
19. Write a Python class to implement pow(x, n).
20. Write a Python class to reverse a string word by word.
21. Write a python program to demonstrate GUI form.
22. Write a python program to Create a database and perform SQL commands.