RAJIV GANDHI PROUDYOGIKI VISHWAVIDYALAYA, BHOPAL

New Scheme Based On AICTE Flexible Curricula

CSE-Data Science/Data Science, IV semester

CD401 Introduction to Discrete Structure & Linear Algebra

Unit 1: Set Theory, Relation, Function, Theorem Proving Techniques:Set theory: definition of sets, Venn Diagram, proofs of some general identities on set,Relation: Definition, Types of relation ,Composition of relation ,Equivalence relation, Partial ordering relation,POSET,Hasse diagram and Lattice .

Unit 2: Algebraic structure: Definition, Properties, types: Semi Group, Monoid, Groups, Abelian Group, Properties of group, cyclic group, Normal subgroup, Ring and Fields: definition and standard result, Introduction to Recurrence Relation and Generating Functions.

Unit 3: Propositional logic: Proposition, First order Logic, Basic logical operation, Truth tables, Tautologies and Contradiction, algebra of proposition, logical implication, logical equivalence ,predicates, Normal Forms, Quantifiers

Graph theory: Introduction and basic terminology of graph, types of graph, Path, Cycles, Shortest path in weighted graph, graph colorings.

Unit 4: Matrices: Determinant and Trace,Cholesky Decomposition, Eigen decomposition,Singular Value decomposition(SVD),Gradient of a matrix:Useful identities For computing Gradient.

Unit 5: Test of Hypothesis : Concept and Formulation ,Type-I and Type-II Errors,Time Series Analysis ,Analysis of Variance (ANOVA)

References:

1. C.L.Liu, "Elements of Discrete Mathematics" Tata Mc Graw-Hill Edition.

2. Trembley, J.P & Manohar; "Discrete Mathematical Structure with Application CS", McGraw Hill.

3. Kenneth H. Rosen, "Discrete Mathematics and its applications", McGraw Hill.

4. Bisht, "Discrete Mathematics",Oxford University Press

5. Biswal,"Discrete Mathematics & Graph Theory", PHI

6. Mathematics For Machine Learning-Marc Peter Deisenroth,A. Aldo Faisal,Cheng soon ong

7.Statistical Method- S.P. Gupta

RAJIV GANDHI PROUDYOGIKI VISHWAVIDYALAYA, BHOPAL

CD402 Analysis &Design of Algorithms

Unit I : Definitions of algorithms and complexity, Time and Space Complexity; Time space tradeoff, various bounds on complexity, Asymptotic notation, Recurrences and Recurrences solving techniques, Introduction to divide and conquer technique, example: binary search, merge sort, quick sort, heap sort, strassen's matrix multiplication etc, Code tuning techniques: Loop Optimization, Data Transfer Optimization, Logic Optimization, etc.

Unit II : Study of Greedy strategy, examples of greedy method like optimal merge patterns, Huffman coding, minimum spanning trees, knapsack problem, job sequencing with deadlines, single source shortest path algorithm etc. Correctness proof of Greedy algorithms.

Unit III : Concept of dynamic programming, problems based on this approach such as 0/1 knapsack, multistage graph, reliability design, Floyd-Warshall algorithm etc.

Unit IV : Backtracking concept and its examples like 8 queen's problem, Hamiltonian cycle, Graph colouring problem etc. Introduction to branch & bound method, examples of branch and bound method like travelling salesman problem etc. Meaning of lower bound theory and its use in solving algebraic problem, introduction to parallel algorithms.

Unit V : Advanced tree and graph algorithms, NP-hard and NP-complete problems, Approximations Algorithms, Data Stream Algorithms, Introduction to design and complexity of Parallel Algorithms

References:

1. Coremen Thomas, Leiserson CE, Rivest RL, Introduction to Algorithms, Third edition, PHI.
2. Horowitz &Sahani, Analysis & Design of Algorithm, Fourth Edition Computer Science Press.
3. Dasgupta, algorithms, Fifth Edition, TMH
4. Ullmann; Analysis & Design of Algorithm, Addison-wesley publishing company,
5. Michael T Goodrich, RobartoTamassia, Algorithm Design, Wiely India
6. Rajesh K Shukla: Analysis and Design of Algorithms: A Beginner's Approach; Wiley

List of Experiments:
1. Write a program for Iterative and Recursive Binary Search.
2. Write a program for Merge Sort.
3. Write a program for Quick Sort.
4. Write a program for Strassen's Matrix Multiplication.
5. Write a program for optimal merge patterns.
6. Write a program for Huffman coding.
7. Write a program for minimum spanning trees using Kruskal's algorithm.
8. Write a program for minimum spanning trees using Prim's algorithm.
9. Write a program for single sources shortest path algorithm.
10. Write a program for Floye-Warshal algorithm.
11. Write a program for traveling salesman problem.
12. Write a program for Hamiltonian cycle problem.

CD403 SOFTWARE ENGINEERING

RATIONALE:
The purpose of this subject is to cover the underlying concepts and techniques used in Software Engineering & Project Management. Some of these techniques can be used in software design & itsimplementation.

PREREQUISITE:-
The students should have at least one year of experience in programming a high-level language and databases. In addition, a familiarity with software development life cycle will be useful in studying thissubject.

Unit I : The Software Product and Software Process
Software Product and Process Characteristics, Software Process Models: Linear Sequential Model, Prototyping Model, RAD Model, Evolutionary Process Models like Incremental Model, Spiral Model, Component Assembly Model, RUP and Agile processes. Software Process customization and improvement, CMM, Product and ProcessMetrics

Unit II : Requirement Elicitation, Analysis, and Specification
Functional and Non-functional requirements, Requirement Sources and Elicitation Techniques, Analysis Modeling for Function-oriented and Object-oriented software development, Use case Modeling, System and Software Requirement Specifications, Requirement Validation,Traceability

Unit III : Software Design
The Software Design Process, Design Concepts and Principles, Software Modeling and UML, Architectural Design, Architectural Views and Styles, User Interface Design, Function- oriented Design, SA/SD Component Based Design, DesignMetrics.

Unit IV : Software Analysis and Testing
Software Static and Dynamic analysis, Code inspections, Software Testing, Fundamentals, Software Test Process, Testing Levels, Test Criteria, Test Case Design, Test Oracles, Test Techniques, Black-Box Testing, White-Box Unit Testing and Unit, Testing Frameworks, Integration Testing, System Testing and other Specialized, Testing, Test Plan, Test Metrics, Testing Tools. , Introduction to Object-oriented analysis, design and comparison with structured SoftwareEngg.

Unit V : Software Maintenance & Software Project Measurement
Need and Types of Maintenance, Software Configuration Management (SCM), Software Change Management, Version Control, Change control and Reporting, Program Comprehension Techniques, Re-engineering, Reverse Engineering, Tool Support. Project Management Concepts, Feasibility Analysis, Project and Process Planning,Resources Allocations, Software efforts, Schedule, and Cost estimations, Project Scheduling and Tracking, Risk Assessment and Mitigation, Software Quality Assurance (SQA). Project Plan, ProjectMetrics.

Practical and Lab work :
Lab work should include a running case study problem for which different deliverable sat the end of each phase of a software development life cycle are to be developed. This will include modeling the requirements, architecture and detailed design. Subsequently the design models will

be coded and tested. For modeling, tools like Rational Rose products. For coding and testing, IDE like Eclipse, Net Beans, and Visual Studio can beused.

References :

1. Pankaj Jalote ,"An Integrated Approach to Software Engineering", NarosaPub, 2005
2. RajibMall, "Fundamentals of Software Engineering" Second Edition, PHI Learning
3. R S. Pressman,"Software Engineering: A Practitioner's Approach", Sixth edition 2006, McGraw-Hill.
4. Sommerville,"Software Enginerring",PearsonEducation.
5. Richard H.Thayer,"SoftwareEnginerring& Project Managements",WileyIndia
6. Waman S.Jawadekar,"Software Enginerring",TMH
7. BobHughes,M.Cotterell,RajibMall"SoftwareProjectManagement",McGrawHill

RAJIV GANDHI PROUDYOGIKI VISHWAVIDYALAYA, BHOPAL

New Scheme Based On AICTE Flexible Curricula

CD404 INTRODUCTION TO DATA SCIENCE

Unit – I: Introduction

Introduction to Data Science – Evolution of Data Science – Data Science Roles – Stages in a Data Science Project – Applications of Data Science in various fields – Data Security Issues.

Unit – II: Data Collection and Data Pre-Processing

Data Collection Strategies – Data Pre-Processing Overview – Data Cleaning – Data Integration and Transformation – Data Reduction – Data Discretization.

Unit – III: Exploratory Data Analytics

Descriptive Statistics – Mean, Standard Deviation, Skewness and Kurtosis – Box Plots – Pivot Table – Heat Map – Correlation Statistics – ANOVA.

Unit – IV: Model Development

Simple and Multiple Regression – Model Evaluation using Visualization – Residual Plot – Distribution Plot – Polynomial Regression and Pipelines – Measures for In-sample Evaluation – Prediction and Decision Making.

Unit – V: Model Evaluation

Generalization Error – Out-of-Sample Evaluation Metrics – Cross Validation – Overfitting – Under Fitting and Model Selection – Prediction by using Ridge Regression – Testing Multiple Parameters by using Grid Search.

REFERENCES:

1. JojoMoolayil, "Smarter Decisions : The Intersection of IoT and Data Science",PACKT, 2016.

2. Cathy O'Neil and Rachel Schutt , "Doing Data Science", O'Reilly, 2015.

3. David Dietrich, Barry Heller, Beibei Yang, "Data Science and Big data Analytics",EMC 2013

4. Raj, Pethuru, "Handbook of Research on Cloud Infrastructures for Big DataAnalytics", IGI Global.

List of Experiments:

1. READING AND WRITING DIFFERENT TYPES OF DATASETS using Python

a. Reading different types of data sets (.txt, .csv) from web and disk and writing in file in specific disk location.

b. Reading Excel data sheet in python.

c. Reading XML dataset in python.

2. VISUALIZATIONS:

a. Find the data distributions using box and scatter plot.

b. Find the outliers using plot.

c. Plot the histogram, bar chart and pie chart on sample data

3. EXPLORATORY DATA ANALYSIS (EDA): Perform EDA on Credit Card Fraud Detection Dataset (open source dataset) for analyzing the data.

4. LINEAR REGRESSION MODEL FOR PREDICTION: Apply Regression Model techniques to predict the future values of data on the open source available datasets.

5. LOGISTIC REGRESSION MODEL: Import the Red-Wine dataset from the UCI Machine Learning Repository having three qualities of wines. Apply logistic regression model for multi-class classification of the wine categories.

6. MODEL EVALUATION USING RESIDUAL PLOT: Plotting Accuracy and Error Metrics against number of iterations for evaluation of model performance.

7. EVALUATING UNDER-FITTING AND OVER-FITTING: Plotting Learning curves for model evaluation for Under-fitting and Over-fitting

RAJIV GANDHI PROUDYOGIKI VISHWAVIDYALAYA, BHOPAL

New Scheme Based On AICTE Flexible Curricula

CSE-Data Science/Data Science, IV semester

CD405 OPERATING SYSTEMS

RATIONALE: The purpose of this subject is to cover the underlying concepts Operating System. This syllabus provides a comprehensive introduction of Operating System, Process Management, Memory Management, File Management and I/O management.

UNIT 1 **Introduction to Operating Systems**: Function, Evolution, Desirable Characteristics and features of an O/S, Operating Systems Services: Types of Services, Different ways of providing these Services – Utility Programs, System Calls.

UNIT 2 **Process Management**: Concept of a process, Process State Diagram, Process based kernel, Dual mode of process execution, CPU scheduling algorithms, deterministic modeling, and System calls for Process Management, Concept of Threads: User level & Kernel level Threads. Process Management in UNIX & Windows

Inter Process Communication: Real and Virtual Concurrency, Mutual Exclusion, Synchronization, Critical Section Problem, Solution to Critical Section Problem : Semaphores and their Operations and their implementation.Deadlocks: Deadlock Problems, Characterization, Prevention, Avoidance, Recovery. IPC in UNIX & Windows

UNIT 3 **Memory Management**: Different Memory Management Techniques – Partitioning, Swapping, Segmentation, Paging, Paged Segmentation, Comparison of these techniques, Techniques for supporting the execution of large programs: Overlay, Dynamic Linking and Loading, Virtual Memory – Concept, Implementation by Demand Paging etc. Memory management in UNIX & Windows

UNIT 4 **File Systems Management**: File Concept, User's and System Programmer's view of File System, Disk Organization, Tape Organization, Different Modules of a File System, Disk Space Allocation Methods – Contiguous, Linked, Indexed. Directory Structures, File Protection, System Calls for File Management, Disk Scheduling Algorithms. File Systems in UNIX & Windows.

UNIT 5 **Input / Output Management** : Principles and Programming, Input/Output Problems, Different I/O operations: Program Controlled, Interrupt Driven, Concurrent I/O, Asynchronous Operations, Logical structure of I/O function, I/O Buffering,Kernel I/o Subsystem.Introduction to Network, Distributed and Multiprocessor Operating Systems. I/O management in UNIX & Windows

TEXT BOOKS RECOMMENDED:

1. Silberschatz, Galvin, Gagne, "Operating System Concepts'', Wiley, 9/E

2. William Stalling, "Operating Systems", Pearson Education

REFERENCE BOOKS:

1. Andrew S. Tanenbaum, "Modern Operating Systems", 3/e, Prentice Hall

2. Maurice J. Bach, " The Design of Unix Operating System", Prentice Hall of India,

3. Bovet &Cesati, "Understanding the Linux Kernel", O'Reily, 2/E.

List of Experiments :

1. Write a program to implement FCFS CPU scheduling algorithm.

2. Write a program to implement SJF CPU scheduling algorithm.

3. Write a program to implement Priority CPU Scheduling algorithm.

4. Write a program to implement Round Robin CPU scheduling algorithm.

5. Write a program to compare various CPU Scheduling Algorithms over different
Scheduling Criteria.

6. Write a program to implement classical inter process communication problem(producer
consumer).

7. Write a program to implement classical inter process communication problem(Reader
Writers).

8. Write a program to implement classical inter process communication
problem(Dining Philosophers).

9. Write a program to implement & Compare various page replacement algorithms.

10. Write a program to implement & Compare various Disk & Drum scheduling algorithms

11. Write a program to implement Banker's algorithms.

12. Write a program to implement Remote ProcedureCall(RPC).

13. Write a Devices Drivers for any Device or peripheral.

RAJIV GANDHI PROUDYOGIKI VISHWAVIDYALAYA, BHOPAL

New Scheme Based on AICTE Flexible Curricula

CSE-Data Science/Data Science, IV semester

DS406 PYTHON FOR DATA SCIENCE

Unit – I: Python Concepts, Data Structures and OOPs in Python

Interpreter – Program Execution – Statements – Expressions – Flow Controls – Functions –
Numeric Data Types – Sequences – Strings – Tuples – Lists – Dictionaries – Class Definition –
Constructors – Object Creation – Inheritance.

Unit – II: Numpy and Pandas Libraries of Python

Numerical operations with Numpy– Pandas Series and Dataframes– Data Manipulation with Pandas
– Overloading – Text Filesand Binary Files – Reading and Writing.

Unit – III: Data Wrangling

Combining and Merging Data Sets – Reshaping and Pivoting – Data Transformation – String
manipulations – Regular Expressions.

Unit – IV: Data Aggregation and Group Operations

GroupBy Mechanics – Data Aggregation – GroupWise Operations – Transformations – Pivot
Tables – Cross Tabulations – Date and Time data types.

Unit – V: Visualization in Python

Matplotlib Package – Plotting Graph - Controlling Graphs – Adding Text – More Graph
Types – Getting and Setting Values – Patches.

REFERENCES:

1. Mark Lutz, "Programming Python", O'Reilly Media, 4th edition, 2010.

2. Joel Grus, "Data Science from scratch", O'Reilly, 2015.

3. Tim Hall and J-P Stacey, "Python 3 for Absolute Beginners", Apress, 1st edition, 2009.

4. Magnus Lie Hetland, "Beginning Python: From Novice to Professional", Apress, Second Edition, 2005.

5. Shai Vaingast, "Beginning Python Visualization Crafting Visual Transformation Scripts", Apress, 2nd edition, 2014.

6. Wes Mc Kinney, "Python for Data Analysis", O'Reilly Media, 2012.

List of Experiments :

1. Write a python program to reverse a string.
2. **Write a python program to perform following operation using lists**:

      a. <u>append element in the list</u>

      b. <u>compare two lists</u>

      c. <u>convert list to dictionary</u>

3. Write a Program to transpose a table/pandas data frame.

4. Write a NumPy program to create a 3x3 matrix with values ranging from 2 to 10.

5. Write a python program to perform following operation on Data Frame:

      a. Create two different Data Frames and perform the merging operations on it.

      b. Create two different Data Frames and perform the grouping operations on it.

      c. Create two different Data Frames and perform the concatenating operations on it

6. Program to check regular expression pattern is matching with string or not in Python

7. Create a sample dataset and apply the following aggregation function on it:

      mean(), median()Mean and median

      min(), max()   Minimum and maximum

      std(), var()   Standard deviation and variance

      sum()   Sum of all items

8. Write a python program to get row wise proportion using crosstab () function.

9. Write a python program to display a bar chart of the popularity of programming languages.

10. Write a python program to create bar plot of scores by group and gender. Use multiple X values on the same chart for men and women.