BSc. (Physical Sciences/ Mathematical Sciences) with Computer Science as one of the Core Disciplines

Category II

DISCIPLINE SPECIFIC CORE COURSE (DSC-2): Data Structures using C++

CREDIT DISTRIBUTION, ELIGIBILITY AND PRE-REQUISITES OF THE COURSE

Course title & Code	Credits	Credit distribution of the course			Eligibility criteria	Pre-requisite of the course (if any)
		Lecture	Tutorial	Practical/ Practice		
DSC02: Data Structures using C++	4	3	0	1	Class XII pass with Mathematics	Nil

Learning Objectives

The course aims at developing the ability to use basic data structures like arrays, stacks, queues, lists, trees to solve problems. C++ is chosen as the language to understand implementation of these data structures.

Learning outcomes

On successful completion of the course, students will be able to:

- Compare two functions for their rates of growth.
- Understand abstract specification of data-structures and their implementation.
- Compute time and space complexity of operations on a data-structure.
- Identify the appropriate data structure(s) for a given application and understand the trade-offs involved in terms of time and space complexity.
- Apply recursive techniques to solve problems.

SYLLABUS OF DSC-2 UNIT – I (06 Hours)

Growth of Functions, Recurrence Relations. Functions used in analysis, asymptotic notations, asymptotic analysis, solving recurrences using recursion tree, Master Theorem.

UNIT – II (12 Hours)

Arrays, Linked Lists, Stacks, Queues, Deques. Arrays: array operations, applications, sorting, two-dimensional arrays, dynamic allocation of arrays; Linked Lists: singly linked lists, doubly linked lists, circularly linked lists, Stacks: stack as an ADT, implementing stacks using arrays, implementing stacks using linked lists, applications of stacks; Queues: queue as an ADT, implementing queues using arrays, implementing queues using linked lists, double-ended queue as an ADT. Time complexity analysis of operations on all data structures.

UNIT – III (06 Hours)

Sorting: Insertion Sort, Count Sort and their complexity analysis.

UNIT – IV (03 Hours)

Recursion: Recursive functions, linear recursion, binary recursion.

UNIT – V (06 Hours)

Trees, Binary Trees. Trees: definition and properties, binary trees: definition and properties, traversal of binary trees and their time complexity analysis.

UNIT – VI (09 Hours)

Binary Search Trees, Balanced Search Trees: Binary Search Trees: insert, delete (by copying), search operations, time complexity analysis of these operations; Balanced Search Trees and (2,4) Trees: motivation and introduction.

UNIT – VII (03 Hours)

Binary Heap, Priority Queue: Binary Heaps: motivation and introduction, application of heaps - Priority Queues.

Practical component (if any) – 30 Hours

1. Perform matrix addition and multiplication.
2. Implement following recursive functions:
    a. Factorial of a number
    b. Nth fibonacci number
    c. Power function: x^y
3. Implement singly linked lists.
4. Implement doubly linked lists.
5. Implement circular linked lists.
6. Implement stack data structure and its operations using arrays.
7. Implement stack data structure and its operations using linked lists.
8. Convert Prefix expression to Infix and Postfix expressions, and evaluate.
9. Implement queue data structure and its operations using arrays.
10. Implement queue data structure and its operations using linked lists.
11. Implement Binary Trees and its traversals.

Essential/recommended readings

1. Goodrich, M., Tamassia, R., & Mount, D., *Data Structures and Algorithms Analysis in C++,* 2nd edition. Wiley, 2011.
2. Cormen, T.H., Leiserson, C.E., Rivest, R. L., Stein C., *Introduction to Algorithms*, 3rd edition, Prentice Hall of India, 2010.
3. Drozdek, A., *Data Structures and Algorithms in C++*, 4th edition, Cengage Learning, 2012.

Suggestive readings

(i) Sahni, S. *Data Structures, Algorithms and applications in C++*. 2nd Edition. Universities Press, 2011.

(ii)Tanenbaum, A. M., Augenstein, M. J., & Langsam Y., *Data Structures Using C and C++*. 2nd edition. Prentice Hall of India, 2009.

Common Pool of Generic Electives (GE) Courses
Offered by Department of Computer Sciences
Category-IV

GENERIC ELECTIVES (GE-2a): Data Analysis and Visualization

Credit distribution, Eligibility and Pre-requisites of the Course

Course title & Code	Credits	Credit distribution of the course			Eligibility criteria	Pre- requisite of the course
		Lecture	Tutorial	Practical/ Practice		
GE2a Data Analysis and Visualization using Python	**4**	**3**	**0**	**1**	**Class XII pass with Mathematics**	knowledge of Python

Learning Objectives

This course is designed to introduce the students to real-world data analysis problems, their analysis and interpretation of results in the field of exploratory data science using Python.

Learning outcomes

On successful completion of the course, students will be able to:

- Apply descriptive statistics to obtain a deterministic view of data
- Apply basic and advanced level statistical function on data
- Perform data handling using Numpy arrays
- Do data cleaning and transformation before extracting useful information
- Visualize data for ease of understanding the revealed information

SYLLABUS OF GE-2a

UNIT – I & II (09 Hours)

Introduction to basic statistics and analysis: Fundamentals of Data Analysis, Statistical foundations for Data Analysis, Types of data, Descriptive Statistics, Python Libraries: NumPy, Pandas, Matplotlib

Array manipulation using NumPy: NumPy array: Creating NumPy arrays, various data types of NumPy arrays

UNIT – I & II (09 Hours)

Introduction to basic statistics and analysis: contd..

Correlation and covariance, Linear Regression, Statistical Hypothesis Generation and Testing

Unit 2 Array manipulation using Numpy: contd..

Indexing and slicing, swapping axes, transposing arrays, data processing using Numpy arrays

UNIT – III (15 Hours)

Data Manipulation using Pandas: Data Structures in Pandas: Series, Data Frame, Index objects, loading data into Panda's data frame, Working with Data Frames: Arithmetics, Statistics, Binning, Indexing, Reindexing, Filtering, Handling missing data, Hierarchical indexing, Data wrangling: Data cleaning, transforming, merging and reshaping

UNIT – IV (12 Hours)

Plotting and Visualization: Using Matplotlib to plot data: figures, subplots, markings, color and line styles, labels and legends, Plotting functions in Pandas: Lines, bar, Scatter plots, histograms, stacked bars, Heatmap

Practical component (if any) – 30 Hours

Use data set of your choice from Open Data Portal (https:// data.gov.in/, UCI repository) or load from scikit, seaborn library for the following exercises to practice the concepts learnt.

1. Load a Pandas data frame with a selected dataset. Identify and count the missing values in a data frame. Clean the data after removing noise as follows
   a. Drop duplicate rows.
   b. Detect the outliers and remove the rows having outliers
   c. Identify the most correlated positively correlated attributes and negatively correlated attributes
2. Import iris data using sklearn library or (Download IRIS data from: https://archive.ics.uci.edu/ml/datasets/iris or import it from sklearn.datasets)
   a. Compute mean, mode, median, standard deviation, confidence interval and standard error for each feature
   b. Compute correlation coefficients between each pair of features and plot heatmap
   c. Find covariance between length of sepal and petal
   d. Build contingency table for class feature
3. Load Titanic data from sklearn library , plot the following with proper legend and axis labels:
   a. Plot bar chart to show the frequency of survivors and non-survivors for male and female passengers separately
   b. Draw a scatter plot for any two selected features
   c. Compare density distribution for features age and passenger fare

   d. Use a pair plot to show pairwise bivariate distribution

4. Using Titanic dataset, do the following
   a. Find total number of passengers with age less than 30
   b. Find total fare paid by passengers of first class
   c. Compare number of survivors of each passenger class

Project students are encouraged to work on a good dataset in consultation with their faculty and apply the concepts learned in the course.

Essential/recommended readings

1. McKinney W. *Python for Data Analysis: Data Wrangling with Pandas, NumPy and IPython*. 2nd edition, O'Reilly Media, 2018.
2. Molin S. *Hands-On Data Analysis with Pandas*, Packt Publishing, 2019.
3. Gupta S.C., Kapoor V.K., *Fundamentals of Mathematical Statistics*, Sultan Chand & Sons, 2020.

Suggestive readings

(i) Chen D. Y, *Pandas for Everyone: Python Data Analysis*, Pearson, 2018.

(ii) Miller J.D. *Statistics for Data Science,* Packt Publishing, 2017.

GENERIC ELECTIVES (GE-2b): Data Analysis and Visualization using Spreadsheet

Credit distribution, Eligibility and Pre-requisites of the Course

Course title & Code	Credits	Credit distribution of the course			Eligibility criteria	Pre-requisite of the course
		Lecture	Tutorial	Practical/ Practice		
GE2b **Data Analysis and Visualization using Spreadsheet**	**4**	**3**	**0**	**1**	**Class XII pass with Mathematics**	**Nil**

Learning Objectives