The Curriculum Book

Master of Science

in

Computer Science

(Artificial Intelligence and Data Science)

2 YEAR-PROGRAMME

Choice Based Credit System and NEP-2020
w. e. f. July 2023

DEPARTMENT OF DATA SCIENCE

GURU JAMBHESHWAR UNIVERSITY OF SCIENCE & TECHNOLOGY

HISAR-125001, HARYANA

SEMESTER I

Sr. No.	Course Types	Course Codes	Nomenclature of the Courses	Hours per week			Credits	Internal	External
				L	T	P			
1.	Core Course (CC)	MDS-CC-111T	Data Structures & Algorithms	3	0	0	3	30	70
2.	Core Course (CC)	MDS-CC-112T	Database Management System	3	0	0	3	30	70
3.	Core Course (CC)	MDS-CC-113T	Fundamentals of Data Science	3	0	0	3	30	70
4.	Core Course (CC)	MDS-CC-114T	Probability and Statistics	3	0	0	3	30	70
5.	Skill Enhancement Course (SEC)	MDS-SEC-111T	Python and R Programming	3	0	0	3	30	70
6.	Core Course (CC)	MDS-CC-111P	Data Structures & Algorithms Lab.	0	0	4	2	50	50
7.	Core Course (CC)	MDS-CC-112P	Database Management System Lab.	0	0	4	2	50	50
8.	Skill Enhancement Course (SEC)	MDS-SEC-111P	Python and R Programming Lab.	0	0	4	2	50	50
9*	Open Elective (OEC)	MDS-OEC-111T	Introduction to Data Science	3	0	0	3	30	70
Total				15	0	12	21	300	500

* Open Elective Course offered by department to be opted by the students of other departments.

SEMESTER II

Sr. No.	Course Types	Course Codes	Nomenclature of the Courses	Hours per week			Credits	Internal	External
				L	T	P			
1.	Core Course (CC)	MDS-CC-121T	Data Analytics	3	0	0	3	30	70
2.	Core Course (CC)	MDS-CC-122T	Data Mining	3	0	0	3	30	70
3.	Core Courses(CC)	MDS-CC-123T	Artificial Intelligence	3	0	0	3	30	70
4.	Core Courses(CC)	MDS-CC-124T	Information Retrieval Systems	3	0	0	3	30	70
5.	Skill Enhancement Course (SEC)	MDS-SEC-121T	Python Tools for Data Science	3	0	0	3	30	70
6.	Core Course (CC)	MDS-CC-121P	Data Analytics Lab.	0	0	4	2	50	50
7.	Core Course (CC)	MDS-CC-122P	Data Mining Lab.	0	0	4	2	50	50
8.	Skill Enhancement Course (SEC)	MDS-SEC-121P	Python Tools for Data Science Lab.	0	0	4	2	50	50
Total				15	0	12	21	300	500

Note: Students exiting the programme after second semester and securing 48 credits including 6 credits of summer internship will be awarded PG Diploma in the relevant Discipline/Subject.

Data Structures and Algorithms

General Course Information

Course Code: MDS-CC-111T Course Credits: 3 Type: Core Course (CC) Contact Hours: 3 hours/week Mode: Lectures (L) Examination Duration: 3 hours	**Course Assessment Methods:** **Max. Marks: 100 (Internal: 30; External: 70)** Three minor tests, each of 20 marks, will be conducted. The third minor will be conducted by the Course Coordinator. No date sheet will be issued for the third minor at the level of the Departments. For the purpose of internal assessment, the highest marks obtained by a student in any of the minor examinations will be considered. Class Performance will be measured through percentage of lectures attended (4 marks) Assignments (4 marks) and class performance (2 marks). The end semester examination will be of 70 marks. For the end semester examination, nine questions are to be set by the examiner. All questions carry equal marks. Question number one will be compulsory and based on the entire syllabus. It will contain seven short answers type questions. The remaining eight questions are to be given by setting two questions from each of the four units of the syllabus. A candidate is required to attempt five questions in all, one compulsory and any other four questions by selecting one from each unit.

Pre-requisites: Programming in C

About the Course:

Data Structure and Algorithms is a core and an essential course for every graduate in Computer Science and Engineering. This course introduces data structures like arrays, linked lists, trees and graphs etc. and various operations to be implemented on these data structures for solving real world problems. It includes various sorting and searching algorithms as well. Further, it incorporates complexity analysis of algorithms implemented on various data structures.

Course Content

Unit I

Introduction to data structures and their types, Abstract data types, Linear lists: Arrays and linked lists: memory representations, implementing operations like traversing, searching, inserting and deleting etc. Applications of arrays and linked lists. Representing sets and polynomials using linked lists.

Unit II

Stack and Queue: Static and linked implementations, Operations and Applications. Circular queues, Tress, Binary trees and related terminology, Tree traversals (Recursive), Threaded Binary Trees, Binary Search Trees implementation and operations, Priority queues.

Unit III

Height Balanced or AVL trees and B trees. Graph definitions and related terminology, memory representations and related operations (traversal, insertion, deletion, search), Path Matrix, Warshall's Shortest path algorithm Hashing, Hash tables, hash function and collision resolution.

Unit IV

Sequential and binary search, Sorting algorithms: Bubble sort, Selection sort, Insertion sort, Quick sort, Merge sort, Count sort, Heap sort, Comparison of searching and sorting techniques based on their complexity analysis, Time and space complexity of algorithms: Asymptotic analysis, Big O, Omega, Theta notations.

Text and Reference Books:

1. Aho, A. V., Ullman, J. D., and Hopcroft, J. E., *Data Structures and Algorithms*, Addison-Wesley, 1983.
2. LangsamYedidyah, Augenstein J Moshe, Tenenbaum M Aaron, *Data Structures using C and C++*, 3rd edition, PHI, 2009.
3. Cormen, T. H., Leiserson, C. E., Rivest, R. L. and Stein, C., *Introduction to Algorithms*, MIT Press, 2009.
4. Robert L. Kruse, *Data Structure and Program Design in C*, Pearson Education India, 2007.
5. Weiss, M. A., *Data Structures and Algorithm Analysis in C++*, Addison-Wesley, 2007.
6. Sahni, S., *Data Structures, Algorithms, and Applications in C++*, WCB/McGraw-Hill, 2001.

Database Management System

General Course Information

Course Code: MDS-CC-112T Course Credits: 3 Type: Core Course (CC) Contact Hours: 3 hours/week Mode: Lectures (L) Exam Duration: 3 hours	**Course Assessment Methods:** **Max. Marks: 100 (Internal: 30; External: 70)** Three minor tests, each of 20 marks, will be conducted. The third minor will be conducted by the Course Coordinator. No date sheet will be issued for the third minor at the level of the Departments. For the purpose of internal assessment, the highest marks obtained by a student in any of the minor examinations will be considered. Class Performance will be measured through percentage of lectures attended (4 marks) Assignments (4 marks) and class performance (2 marks). The end semester examination will be of 70 marks. For the end semester examination, nine questions are to be set by the examiner. All questions carry equal marks. Question number one will be compulsory and based on the entire syllabus. It will contain seven short answers type questions. The remaining eight questions are to be given by setting two questions from each of the four units of the syllabus. A candidate is required to attempt five questions in all, one compulsory and any other four questions by selecting one from each unit.

Prerequisite: Knowledge of UNIX/ Windows, programming language and data structures

About the Course:

This course includes a detailed coverage of principles of database design and models. Students learn querying a database using SQL, normalization techniques, transaction processing etc.

Course Content

Unit I

Overview: Overview of File Systems and Database Systems, Characteristics of the Data Base Approach, Database users, Advantages and Disadvantages of a DBMS, Responsibility of Database Administrator.

Data Base Systems Concepts and Architecture: DBMS architecture and various views of Data, Data Independence, Database languages, Data Models: Relational Database Model, Hierarchical Data Model, Network Data Model, Schemas and Instances.

Unit II

E-R Model: Entity Types, Attributes & Keys, Relationships, Roles and Structural Constraints, E-R Diagrams, Reduction of an E-R Diagram to Tables. Relational Model and Query Language: Overview of Relational Database, Key Integrity Constraints, Relational Algebra, Relational Calculus, SQL fundamentals, Basic Operators, Missing information and NULL values, Advanced SQL features

Unit III

Relational Database Design: Overview of normalization, Database Anomalies, Candidate and Super Key, Functional Dependencies, Integrity Constraints, Decomposition, Normal forms: First, Second, Third Normal, Boyce Codd, Normal Form, Multi-valued Functional Dependencies and Fourth Normal Form, Join Dependencies and Fifth Normal Form, Denormalization.

Unit IV

Concurrency Control Techniques: Overview of database Transactions, Transaction states, ACID properties of a Transaction, Transaction Recovery, Concurrency Control, Locking Techniques, Time-stamp ordering, Multi-version Techniques, Deadlock, Recovery Techniques in centralized DBMS.

DDBMS Design: Replication and Fragmentation Techniques.

Text and Reference Books:

1. Elmasri, R., and Navathe, S. B., *Fundamentals of Database Systems*, 3rd Edition, Addison Wesley, 2002.
2. Silberschatz, A., Korth, H. F., and Sudarshan, S., Database *System Concepts,* McGraw Hill, 2011.
3. Pannerselvam R., *Database Management Systems*, 2nd Edition, PHI Learning, 2011.
4. Desai, B. C., *An Introduction to Database System*, Galgotia Publication, 2010.
5. Leon, A., and Leon, M., *Database Management Systems*, Ist Edition, Vikas Publishing, 2009.
6. Mata-Toledo, R., Cushman, P., Sahoo, D., *Database Management Systems*, Schaums' Outline series, TMH, 2007.

Fundamentals of Data Science

General Course Information

Course Code: MDS-CC-113T Course Credits: 3 Type: Core Course (CC) Contact Hours: 3 hours/week Mode: Lectures (L) Examination Duration: 3 hours	**Course Assessment Methods:** **Max. Marks: 100 (Internal: 30; External: 70)** Three minor tests, each of 20 marks, will be conducted. The third minor will be conducted by the Course Coordinator. No date sheet will be issued for the third minor at the level of the Departments. For the purpose of internal assessment, the highest marks obtained by a student in any of the minor examinations will be considered. Class Performance will be measured through percentage of lectures attended (4 marks) Assignments (4 marks) and class performance (2 marks). The end semester examination will be of 70 marks. For the end semester examination, nine questions are to be set by the examiner. All questions carry equal marks. Question number one will be compulsory and based on the entire syllabus. It will contain seven short answers type questions. The remaining eight questions are to be given by setting two questions from each of the four units of the syllabus. A candidate is required to attempt five questions in all, one compulsory and any other four questions by selecting one from each unit.

Pre-requisites: Knowledge about Fundamentals of Data Base Management System

About the Course:

This course is designed to introduce students to data science and its practice: how it works and how it can produce insights from social, political, business and economic data.

Course Content

Unit I

Data Science: Definition, Basic Terminology, Data science Venn diagram, Types of Data, Structured versus Unstructured data, Quantitative versus Qualitative data, The Four Levels of Data, Five steps of Data Science, Data Science Process Overview, Data science classification, Data Science Algorithms, Business Intelligence and Data Science, Components of Data Science, Introduction, Prior Knowledge, Data Preparation, Modeling, Applications, Objectives of Data Exploration, Datasets, Descriptive statistics

Unit II

Data Visualization: Introduction, Types of Data visualization, Technologies for visualization, Various visualization techniques, The Five Cs of Data Visualization, Data Science Methodology, Analytics for Data Science, Data Analytics Examples, Data Analytics Life Cycle, Data Discovery, Data preparation, Model Planning, Model Building, Operationalization.

Unit III

Feature Selection: Classifying feature selection methods, Anomaly Detection: Introduction, Distance and Density based outlier detection, Local Outlier Factor, Time series Forecasting, Decomposition, smoothing based methods, Regression based methods, Machine Learning methods.

Unit IV

Introduction to Data Science Tools: SAS, APACHE FLINK, BigML, Excel, Tableau, Matplotlib, TensorFlow, Weka

Applications: Hands on with Solving Data Problems, Collecting and Analyzing Twitter Data, Collecting and Analyzing YouTube Data.

Text and Reference Books:

1. Sanjeev J. Wagh, Manisha S. Bhende, Anuradha D. Thakare, *Fundamentals of Data Science*, 1st Edition, 2022
2. Daimi, Kevin, Ed. Hamid R. Arabnia, *Principles of Data Science*, Springer, 2020.
3. Vijay Kotu, Bala Deshpande, *Data Science: Concepts and Practices*, Morgan Kaufmann Publishers, Second edition, 2019
4. D J Patil, Hilary Mason, Mike Loukides, *Ethics and Data Science*, O' Reilly, 1st edition, 2018
5. Sinan Ozdemir, *Principles of Data Science*, Packt Publishing, December 2016.
6. Jure Leskovek, Anand Rajaraman and Jeffrey *Ullman, Mining of Massive Datasets* v2.1, Cambridge University Press, 2014.
7. Cielen, Davy, Arno DB Meysman, Mohamed Ali, *Introducing Data Science: Big Data, Machine Learning, and more, using Python Tools*, Manning Publications Co., 2016

Probability and Statistics

General Course Information

Course Code: MDS-CC-114T Course Credits: 3 Type: Core Course (CC) Contact Hours: 3 hours/week Mode: Lectures (L) Examination Duration: 3 hours	**Course Assessment Methods:** **Max. Marks: 100 (Internal: 30; External: 70)** Three minor tests, each of 20 marks, will be conducted. The third minor will be conducted by the Course Coordinator. No date sheet will be issued for the third minor at the level of the Departments. For the purpose of internal assessment, the highest marks obtained by a student in any of the minor examinations will be considered. Class Performance will be measured through percentage of lectures attended (4 marks) Assignments (4 marks) and class performance (2 marks). The end semester examination will be of 70 marks. For the end semester examination, nine questions are to be set by the examiner. All questions carry equal marks. Question number one will be compulsory and based on the entire syllabus. It will contain seven short answers type questions. The remaining eight questions are to be given by setting two questions from each of the four units of the syllabus. A candidate is required to attempt five questions in all, one compulsory and any other four questions by selecting one from each unit.

Pre-requisites: Basics of Mathematics and Statistics

About the Course:

Probability and statistics both are the most important concepts for Data Science and Machine Learning. Probability is about predicting the likelihood of future events, while statistics involves the analysis of the frequency of past events. This course will give an additional space for widening the applications of the knowledge in Data Science domain to the various fields of real life.

Course Content

Unit I

Raw Data , Graphical Plots and Charts - Frequency Distribution , Histogram and Frequency Polygons - Relative Frequency Distributions , Cumulative Frequency Distributions , Frequency Curves and Their Types - Measures of Central Tendency: Mean, Median, Mode, Trimmed Mean , Measures of Dispersion: Range, Standard Deviation, Quartile Deviation, Mean and Median Absolute Deviation , Moments - Measures of Skewness and Kurtosis , Notion of Linear Correlation and Linear Regression, Concept of Probability , Axioms of Probability - Conditional Probability , Simple Problems - Independent Events - Bayes' Rule (without proof) and Simple Applications.

Unit II

Discrete and Continuous Random Variables, Probability Distributions for Discrete and Continuous Random Variables , Distribution Functions for Discrete and Continuous Random Variables - Joint Distributions - Independent Random Variables - Probability Distributions of Functions of Random Variables , Marginal and Conditional Distributions , Mathematical Expectation, Notions of Binomial, Poisson Distribution and Normal Distributions , Properties , Relationship Between Binomial and Normal Distributions, Poisson and Normal Distributions , Uniform, Exponential, Gamma Distributions, t, Chi-square and F Distributions - Bivariate Normal Distribution , Simulation: Random Number Generation from Exponential, Gamma and Normal Distributions.

Unit III

Population and Sample - Random Samples, Sampling with and without Replacement, Sampling Distributions, Sampling distributions of Mean, Proportion and Difference of Means, Standard Error. Estimation of Parameters, Properties of Estimators: Unbiasedness, Consistency, Efficiency, Sufficiency. Point and Interval Estimates and Their Reliability, Confidence Interval Estimates of Population Parameters Based on Normal, t and Chi-square Distributions.

Unit IV

Statistical Decisions, Statistical Hypothesis, Tests of Hypothesis and Significance, One-tail and Two-tail Tests. Parametric Tests: Tests Involving Normal, t, Chi-square and F Distributions - Test for Goodness of Fit, Contingency Tables, Tests for Independence of Attributes, One-way and Two-way Analysis of Variance. Non-parametric Tests: Sign Test, Run Test, Wilcoxon Signed Rank Test, Mann-Whitney U test, Kruskal-Wallis Test.

Text and Reference Books:

1. Montgomery, D. C., and Runger, G. C. *Applied Statistics and Probability for Engineers*, Seventh Edition, John Wiley & Sons, 2018.
2. Bruce, P., Bruce, A., and Gedeck, P. *Practical Statistics for Data Scientists*, Second Edition, O'Reilly Media, 2020.
3. Spiegel, M. R., Schiller, J. J., and Alu Srinivasan, R. *Probability and Statistics*, Fourth Edition, Schaum's Outline Series, McGraw Hill Companies, 2013.
4. S.C. Gupta, V. K. Kapoor, *Fundamentals of Mathematical Statistics,* Sultan Chand & Sons, 2020.
5. S.C. Gupta, V. K. Kapoor, *Fundamentals of Applied Statistics,* Sultan Chand & Sons, 2014.

Python and R Programming

General Course Information

Course Code: MDS-SEC-111T Course Credits: 3 Type: Skill Enhancement Course (SEC) Contact Hours: 3 hours/week Mode: Lectures (L) Examination Duration: 3 hours	**Course Assessment Methods:** **Max. Marks: 100 (Internal: 30; External: 70)** Three minor tests, each of 20 marks, will be conducted. The third minor will be conducted by the Course Coordinator. No date sheet will be issued for the third minor at the level of the Departments. For the purpose of internal assessment, the highest marks obtained by a student in any of the minor examinations will be considered. Class Performance will be measured through percentage of lectures attended (4 marks) Assignments (4 marks) and class performance (2 marks). The end semester examination will be of 70 marks. For the end semester examination, nine questions are to be set by the examiner. All questions carry equal marks. Question number one will be compulsory and based on the entire syllabus. It will contain seven short answers type questions. The remaining eight questions are to be given by setting two questions from each of the four units of the syllabus. A candidate is required to attempt five questions in all, one compulsory and any other four questions by selecting one from each unit.

Pre-requisites: Basic knowledge on programming concepts and statistics.

About the Course:

R programming Language and Python are both used extensively for Data Sciences. Both are very useful and open-source languages as well. Python supports a very large community to general-purpose in data science. One of the most basic use for data analysis, primarily because of the fantastic ecosystem of data-centric Python packages. R Programming has a rich ecosystem to use in standard machine learning and data mining techniques. It works in statistical analysis of large datasets, and it offers a number of different options for exploring data and It makes it easier to use probability distributions, apply different statistical tests. The main objective of this course is to provide the basic understanding of both programming languages to students to solve real-life problems.

Course Content

Unit I

Introduction to Python, Python Identifiers, Keywords and Indentation, Comments, Python Operators and Expressions, Function: print(), input(), eval(), Data Types: int, float, complex, Variables, Mutable vs Immutable variables, Namespaces, Decision Statements: Boolean Type, Boolean Operators, if statement, else statement, Nested Conditionals Statements, Multi-way Decision Statements (elif statement), Loop Control Statements: While loop, range() Function, For Loop, Nested Loops, Infinite Loop, Break Statement, Continue Statement, Pass Statement, Lists: Operations on List: Slicing, Inbuilt Functions for Lists, List Processing: Searching and Sorting, Dictionaries: Need of Dictionary, Operations on Directories: Creation, Addition, Retrieving Values, Deletion; Tuples, operations on Tuples, Inbuilt Functions for Tuples, Sets and operations on sets.

Unit II

Python Functions, Inbuilt functions, Main function, User Defined functions, Defining and Calling Function, Parameter Passing, Actual and Formal Parameters, Default Parameters, Global and Local Variables, Recursion, Lamda Function, Modules, Importing Own Module, Packages.

Python Object Oriented: Overview of OOP, Classes and objects, accessing attributes, Built-In Class Attributes, Methods, Class and Instance Variables, Polymorphism, Overlapping and Overloading of Operators, Class Inheritance: super (), Method Overriding, Exception Handling, Try-except-else clause, Python Standard Exceptions, User-Defined Exceptions

Unit III

Introduction to R, Installation of R, Installation of R-Studio, Types of Variables, Types of Datatypes: Real, Integer, Complex, Character, Strings, Vectors, Arrays, List, Matrices, Factors, Data Frames, Types of Operators: Arithmetic, Logical, Relational, Membership, Special Operators, If-else Flow Control, Loops in R (While, For, Break, Next), Nested Loops, Switch-Case

Functions in R, Function declaration with parameters, Function declaration without parameters

Unit IV

R Data Interface: Reading CSV files, Reading XML files, JSON files, Scraping data from the Web, SQL with R, Databases with R

R package: ggplot2 and dplyr

Data Visualization of R: Pie Chart, Bar graph, Line Graph, Scatter plot, Stack Plot, Box-Plot

Text and Reference Books:

1. R. Nageswara Rao, *Core Python Programming*, Dreamtech Press; Second edition (1 January 2018)
2. Allen B. Downey, *Think Python: How to Think Like a Computer Scientist*, 2nd edition, Updated for Python 3, Shroff/O'Reilly Publishers, 2016.
3. Jake Vanderplas, *Python Data Science Handbook: Essential Tools for Working with Data*, 1st Edition, O'Reilly Media, 2016.
4. Norman Matloff, *The Art of R Programming: A Tour of Statistical Software Design*, No Starch Press, First Edition, 2011.
5. Hadley Wickham, Garett Grolemund, *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*, O'Reily Publications, First Edition, Feb 2017
6. Reema, Thareja, *Python Programming: Using Problem Solving Approach*, Oxford University Press, June 2017
7. Garrett Grolemund, *Hands-on Programming with R: Write your own functions and simulations*, O'Reilly Publisher, 2014.
8. Saroj Dahiya Ratnoo and Himmat Singh Ratnoo, *Essentials of R for Data Analytics*, Wiley (1 January 2021)

Data Structures and Algorithms Lab.

General Course Information

Course Code: MDS-CC-111P Course Credits: 2 Type: Core Course (CC) Contact Hours: 4 hours/week Mode: Lab practice and assignments	**Course Assessment Methods:** **Total Marks: 100 (internal: 50; external: 50)** The internal and external assessment is based on the level of participation in lab. sessions and the timely submission of lab experiments/assignments, the quality of solutions designed for the assignments, the performance in VIVA VOCE, the quality of lab. file and ethical practices followed. The internal examination is conducted by the course coordinator. The external examination is conducted by external examiner appointed by the Controller of Examination in association with the internal examiner appointed by the Chairperson of the Department.

Pre-requisites: Programming in C language.

About the Course:

This lab course involves implementation of basic and advance data structures and various operations on these data structures. The objective of the lab course is to train the students to solve the problems related to data structures and choose the appropriate data structure for solving computational problem efficiently.

Practical Lab based on theory course MDS-CC-111T

Note:

The actual experiments/assignments will be designed by the course coordinator. One assignment should be designed to be done in groups of two or three students. The assignments must meet the objective of the course. The list of assignments and schedule of submission will be prepared by the course coordinator at the beginning of the semester.

Database Management System Lab.

General Course Information

Course Code: MDS-CC-112P Course Credits: 2 Type: Core Course (CC) Contact Hours: 4 hours/week Mode: Lab practice and assignments	**Course Assessment Methods:** **Total Marks: 100 (internal: 50; external: 50)** The internal and external assessment is based on the level of participation in lab. sessions and the timely submission of lab experiments/assignments, the quality of solutions designed for the assignments, the performance in VIVA VOCE, the quality of lab. file and ethical practices followed. The internal examination is conducted by the course coordinator. The external examination is conducted by external examiner appointed by the Controller of Examination in association with the internal examiner appointed by the Chairperson of the Department.

Pre-requisites: Exposure to programming language, MS Access.

About the Course:

This lab. course on DBMS involves a rigorous training on Oracle programming. It provides a strong formal foundation in database concepts, technology and practice to the students to groom them into well-informed database application developers. The objective of the lab course is to develop proficiency in the execution of commands of the database design and query using Oracle.

Practical Lab based on theory course MDS-CC-112T

Note:

The actual experiments/assignments will be designed by the course coordinator. One assignment should be designed to be done in groups of two or three students. The assignments must meet the objective of the course. The list of assignments and schedule of submission will be prepared by the course coordinator at the beginning of the semester.

Python and R Programming Lab.

General Course Information

Course Code: MDS-SEC-111P Course Credits: 2 Type: Skill Enhancement Course (SEC) Contact Hours: 4 hours/week Mode: Lab practice and assignments.	**Course Assessment Methods:** **Total Marks: 100 (internal: 50; external: 50)** The internal and external assessment is based on the level of participation in lab. sessions and the timely submission of lab experiments/assignments, the quality of solutions designed for the assignments, the performance in VIVA VOCE, the quality of lab. file and ethical practices followed. The internal examination is conducted by the course coordinator. The external examination is conducted by external examiner appointed by the Controller of Examination in association with the internal examiner appointed by the Chairperson of the Department.

Pre-requisites: Understanding of programming language.

About the Course:

The major objective of Python programming is to make the students solve real word problem including data science problems efficiently using python library. The understanding and knowledge of R programming helps the students to read the data and its manipulation using R.

Practical Lab based on theory course MDS-SEC-111T

Note:

The actual experiments/assignments will be designed by the course coordinator. One assignment should be designed to be done in groups of two or three students. The assignments must meet the objective of the course. The list of assignments and schedule of submission will be prepared by the course coordinator at the beginning of the semester.

Introduction to Data Science

General Course Information

Course Code: MDS-OEC-111T Course Credits: 3 Type: Open Elective (OEC) Contact Hours: 3 hours/week Mode: Lectures (L) Examination Duration: 3 hours	**Course Assessment Methods:** **Max. Marks: 100 (Internal: 30; External: 70)** Three minor tests, each of 20 marks, will be conducted. The third minor will be conducted by the Course Coordinator. No date sheet will be issued for the third minor at the level of the Departments. For the purpose of internal assessment, the highest marks obtained by a student in any of the minor examinations will be considered. Class Performance will be measured through percentage of lectures attended (4 marks) Assignments (4 marks) and class performance (2 marks). The end semester examination will be of 70 marks. For the end semester examination, nine questions are to be set by the examiner. All questions carry equal marks. Question number one will be compulsory and based on the entire syllabus. It will contain seven short answers type questions. The remaining eight questions are to be given by setting two questions from each of the four units of the syllabus. A candidate is required to attempt five questions in all, one compulsory and any other four questions by selecting one from each unit.

Pre-requisites: Probability and Statistics

About the Course:

The major goals of this course are to learn how to use tools for acquiring, cleaning, analyzing, exploring, and visualizing data; making data-driven inferences and decisions; and effectively communicating results.

Course Content

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 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.

Unit IV

Model Development: Simple and Multiple Regression, Model Evaluation using Visualization, Residual Plot, Distribution Plot.

Model Evaluation: Generalization Error, Out-of-Sample Evaluation Metrics, Cross Validation, Overfitting, Under Fitting and Model Selection.

Text and Reference Books:

1. Vijay Kotu, Bala Deshpande, *Data Science: Concepts and Practices*, Morgan Kaufmann Publishers, Second edition, 2019.
2. Jojo Moolayil, *Smarter Decisions: The Intersection of IoT and Decision Science*, PACKT, 2016.
3. Sanjeev J. Wagh, Manisha S. Bhende, Anuradha D. Thakare, *Fundamentals of Data Science*, 1st Edition, 2022
4. Cathy O'Neil and Rachel Schutt, *Doing Data Science*, O'Reilly, 2015.
5. Daimi, Kevin, Ed. Hamid R. Arabnia, *Principles of Data Science*, Springer, 2020.
6. David Dietrich, Barry Heller, Beibei Yang, *Data Science and Big Data Analytics*, EMC 2013
7. Raj, Pethuru, *Handbook of Research on Cloud Infrastructures for Big Data Analytics*, IGI Global, 2014.
8. Sinan Ozdemir, *Principles of Data Science*, Packt Publishing, December 2016.
9. Cielen, Davy, Arno DB Meysman, Mohamed Ali, *Introducing Data Science: Big Data, Machine Learning, and more, using Python Tools*, Manning Publications Co., 2016.

Data Analytics

General Course Information

Course Code: MDS-CC-121T Course Credits: 3 Type: Core Course (CC) Contact Hours: 3 hours/week Mode: Lectures (L) Examination Duration: 3 hours	**Course Assessment Methods:** **Max. Marks: 100 (Internal: 30; External: 70)** Three minor tests, each of 20 marks, will be conducted. The third minor will be conducted by the Course Coordinator. No date sheet will be issued for the third minor at the level of the Departments. For the purpose of internal assessment, the highest marks obtained by a student in any of the minor examinations will be considered. Class Performance will be measured through percentage of lectures attended (4 marks) Assignments (4 marks) and class performance (2 marks). The end semester examination will be of 70 marks. For the end semester examination, nine questions are to be set by the examiner. All questions carry equal marks. Question number one will be compulsory and based on the entire syllabus. It will contain seven short answers type questions. The remaining eight questions are to be given by setting two questions from each of the four units of the syllabus. A candidate is required to attempt five questions in all, one compulsory and any other four questions by selecting one from each unit.

Pre-requisites: Basic programming skills, Probability and Statistics

About the Course:

In this course, the learners will be able to develop expertise in R programming for manipulating, exploring, visualizing, applying descriptive and inferential statistics. In addition, they will learn to implement predictive modelling.

Course Content

Unit I

Data analytics preliminaries: Introduction to data analytics, scales of measurements (Data types) and their implementation in R. Working with vectors, matrices and tabular data (data frames), reading and writing tabular data from and to files (text and CSV). Describing data with statistical summaries (mean, median, mode, variance and standard deviation). Discriminating between sample and population, Quantile-Quantile plot. writing user-defined functions in R.

Unit II

Manipulating tabular data: Sorting, filtering cases, selecting variables, deriving new variables, grouping and summarizing data. working with packages (tidyverse) for data manipulations and transformations.

Exploratory data analysis: random and normally distributed variables, skewed normal distribution, z-score, detecting outliers in data, handling missing values.

Visualizing data through various plots and charts: bar charts, histogram, frequency polygon, density plots, scatter plots, box & whisker plots, heat and contour plots, plotting the above graphs in R, plotting with package ggplot2.

Unit III

Predictive modelling: what is predictive modelling, estimating a function, the trade-off between model accuracy and prediction accuracy and model interpretability, regression versus classification, measuring the quality of fit, The bias and variance trade- off.

Simple and multiple linear regression modelling: estimating the coefficients, assessing the accuracy of the coefficient estimates, assessing the accuracy of the model. Building regression models in R.

Unit IV

Classification Modeling: The process of classification, decision tree, Bayesian, k-nearest neighbor, support vector machine classification models and their implementation in R. evaluating a classification model: confusion matrix, accuracy, sensitivity, specificity, f-measure, kappa statistics, ROC and area under curve. accuracy and interpretability of classification models.

Evaluating the accuracy of a classifier: holdout or random sampling methods, cross-validation, bootstrap methods.

Text and Reference Books:

1. W. N. Venables, D. M. Smith and the R core Team, *An introduction to R, Notes on R: A Programming Environment for Data Analysis and Graphics*, version 3.3.2, 2016.
2. Saroj Dahiya Ratnoo and Himmat Singh Ratnoo, *Essentials of R for Data Analytics*, Wiley, 2021.
3. Hadley Wickham and Garrett Grolemund, *R for Data Science Import, Tidy, Transform and model Data*, O'Reilly, 2017.
4. Paul Teeter, *R Cookbook*, O'Reilly, 2011.
5. Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, *An Introduction to Statistical Learning with Applications in R*, Springer, 2013.
6. Han, J., Kamber, M, Pei, J., *Data Mining Concepts and Techniques*, Third edition, Morgan Kaufmann, 2012.

Data Mining

General Course Information

Course Code: MDS-CC-122T Course Credits: 3 Type: Core Course (CC) Contact Hours: 3 hours/week Mode: Lectures (L) Examination Duration: 3 hours	**Course Assessment Methods:** **Max. Marks: 100 (Internal: 30; External: 70)** Three minor tests, each of 20 marks, will be conducted. The third minor will be conducted by the Course Coordinator. No date sheet will be issued for the third minor at the level of the Departments. For the purpose of internal assessment, the highest marks obtained by a student in any of the minor examinations will be considered. Class Performance will be measured through percentage of lectures attended (4 marks) Assignments (4 marks) and class performance (2 marks). The end semester examination will be of 70 marks. For the end semester examination, nine questions are to be set by the examiner. All questions carry equal marks. Question number one will be compulsory and based on the entire syllabus. It will contain seven short answers type questions. The remaining eight questions are to be given by setting two questions from each of the four units of the syllabus. A candidate is required to attempt five questions in all, one compulsory and any other four questions by selecting one from each unit.

Pre-requisites: Knowledge of database systems, elementary knowledge of statistics and probability.

About the Course:

Today's era is the era of information. Data is growing exponentially day by day. There is a need to process and analyse the data to extract knowledge from it, so that one can use that knowledge for decision making. This course provides introductory concepts of data mining and data warehousing. The course will be taught with a database as well as machine learning perspectives. The objective of the course is to provide a comprehensive understanding of data mining tasks and evaluation of results obtained out of data mining processes.

Course Content

Unit I

Introduction to Data Mining: Kind of data to be mined, Data Mining Functionalities, Technologies used in Data Mining, Applications of data Mining, Major Issues in Data Mining.

Data Warehouse: Introduction, Data Warehouse and Database Systems, Data Warehouse Architecture, Data Warehouse Models, Data Cube and OLAP, Multidimensional data Model, Concept Hierarchies, OLAP operations

Pattern Mining: Mining Frequent Patterns, Associations and Correlations, Frequent Itemset Mining using Apriori Algorithm, Generating Association Rules from Frequent Itemsets, Pattern Growth Approach for Mining Frequent Itemsets, Pattern evaluation Methods

Unit II

Classification: Introduction, Classification using Decision Tree Induction, Bayesian Classification Methods, Rule Based Classification, Model Evaluation and Selection, Techniques to Improve Classification Accuracy.

Introduction to advanced classifiers: k-Nearest Neighbor, Support Vector Machine, Artificial Neural Network.

Unit III

Cluster Analysis: Introduction, overview of Basic Clustering Methods,
Partitioning Methods: k-mean, k-medoids,
Hierarchical Methods: Agglomerative versus Divisive Hierarchical Clustering, Distance Measures in Algorithmic Methods, Balanced Iterative Reducing and Clustering using Hierarchies (BIRCH), Chameleon: Multiphase Hierarchical Clustering Using Dynamic Modeling, Probabilistic Hierarchical Clustering,
Density-based methods: DBSCAN, OPTICS, DENCLUE,
Grid-based Methods: STING, CLIQUE, **Evaluation of Clustering.**

Unit IV

Outlier Detection: Introduction, types of outliers, challenges of outlier detection.
Outlier detection methods: statistical approaches, proximity-based approaches, clustering based approaches, classification-based approaches, Outlier detection in high dimensional data.

Text and Reference Books:

1. Jiawei Han, Micheline Kamber and Jian Pei, *Data Mining Concepts and Techniques*, Morgan Kaufmann Publishers, Third Edition, July 2011.
2. Alex Berson, Stephen J. Smith, *Data Warehousing, Data Mining & OLAP*, Tata McGraw Hill, 2004.
3. Pang-Ning Tan, Michael Steinbach and Vipin Kumar, *Introduction to Data Mining*, Pearson Education, 2014.
4. K. P. Soman, Shyam Diwakar and V. Ajay, *Insight into Data Mining Theory and Practice*, Easter Economy Edition, Prentice Hall of India, 2009.
5. G. K. Gupta, *Introduction to Data Mining with Case Studies*, Prentice Hall of India, 2006.
6. Daniel T. Larose, *Data Mining Methods and Models*, Wiley, 2006.
7. W. H. Inman, *Building the Data Warehouse*, Wiley India, 2005

Artificial Intelligence

General Course Information

Course Code: MDS-CC-123T Course Credits: 3 Type: Core Course (CC) Contact Hours: 3 hours/week Mode: Lectures (L) Examination Duration: 3 hours	**Course Assessment Methods:** **Max. Marks: 100 (Internal: 30; External: 70)** Three minor tests, each of 20 marks, will be conducted. The third minor will be conducted by the Course Coordinator. No date sheet will be issued for the third minor at the level of the Departments. For the purpose of internal assessment, the highest marks obtained by a student in any of the minor examinations will be considered. Class Performance will be measured through percentage of lectures attended (4 marks) Assignments (4 marks) and class performance (2 marks). The end semester examination will be of 70 marks. For the end semester examination, nine questions are to be set by the examiner. All questions carry equal marks. Question number one will be compulsory and based on the entire syllabus. It will contain seven short answers type questions. The remaining eight questions are to be given by setting two questions from each of the four units of the syllabus. A candidate is required to attempt five questions in all, one compulsory and any other four questions by selecting one from each unit.

Pre-requisites: Basic Knowledge of Algorithms and Probability.

About the Course:

Artificial Intelligence is a core and an essential course for every graduate in Computer Science and Engineering. This course introduces the concepts of Artificial Intelligence and challenges inherent in building intelligent systems. It includes the role of knowledge representation in problem solving and how these are used in making intelligent machine. Further it incorporates the concepts of expert system and its applications.

Course Content

Unit I

Introduction to AI: Introduction, Turing Test, AI problems, State Space Search, production system

Problem Solving Using Search: Blind search techniques - Breadth first search, Depth first search. Heuristic search techniques - Generate and test, Hill Climbing, Best first search, A* Algorithm, AO* Algorithm, The Minimax Search Procedure, Adding Alpha-Beta Cut-offs.

Unit II

Knowledge Representation: Introduction, Knowledge Representation- Representation and Mappings, Symbolic Logic - Propositional logic, Predicate logic- Representing simple facts in logic, Representing Instances and ISA Relationship, Computable functions and Predicates, Unification, Resolution.

Representing Knowledge Using Rules: Procedural versus Declarative Knowledge, Logic Programming, Forward versus Backward Reasoning, Matching, Control Knowledge.

Unit III

Reasoning Under Uncertainty: Introduction to Nonmonotonic Reasoning, Probability and Baye's Theorem, Certainty Factors and Rule-based Systems, Bayesian Networks.

Fuzzy logic system: Introduction, Crisp Set, Fuzzy Sets, Fuzzy Membership Functions, Operations on Fuzzy Sets, Fuzzy Relations.

Unit IV

Planning: Introduction, Components of Planning System, Goal Stack Planning, Nonlinear Planning using Constraint Posting, Hierarchical Planning.

Expert System and Applications: Introduction, Architecture, Rule based Expert Systems, Applications of Expert Systems.

T**ext and Reference Books:**

1. Elaine Rich, Kevin Knight and Shivashankar B Nair, *Artificial intelligence*, McGraw Hill Education. 3rd edition, 2009.
2. Stuart Russel and Peter Norvig, *Artificial intelligence: A modern Approach*, Pearson Education, 3rd edition, 2015.
3. Dan W. Patterson, *Introduction to Artificial Intelligence and Expert System*, Pearson Education.1st edition, 2007.
4. Deepak Khemani, *A first course in Artificial Intelligence*, McGraw Hill Education. 3rd edition, 1st edition, 2013.
5. George F. Luger, *Artificial Intelligence: Structures and Strategies for Complex Problem Solving*, Pearson Education, 5th edition, 2009.

Information Retrieval Systems

General Course Information:

Course Code: MDS-CC-124T Course Credits: 3 Type: Core Course (CC) Contact Hours: 3 hours/week Mode: Lectures (L) Examination Duration: 3 hours	**Course Assessment Methods:** **Max. Marks: 100 (Internal: 30; External: 70)** Three minor tests, each of 20 marks, will be conducted. The third minor will be conducted by the Course Coordinator. No date sheet will be issued for the third minor at the level of the Departments. For the purpose of internal assessment, the highest marks obtained by a student in any of the minor examinations will be considered. Class Performance will be measured through percentage of lectures attended (4 marks) Assignments (4 marks) and class performance (2 marks). The end semester examination will be of 70 marks. For the end semester examination, nine questions are to be set by the examiner. All questions carry equal marks. Question number one will be compulsory and based on the entire syllabus. It will contain seven short answers type questions. The remaining eight questions are to be given by setting two questions from each of the four units of the syllabus. A candidate is required to attempt five questions in all, one compulsory and any other four questions by selecting one from each unit.

Pre-requisites: Data Structures, Data Base Management Systems

About the Course:

This course would enable the students to understand the various aspects of an information retrieval system and its evaluation and to be able to design. The main aim of this course is to give students an understanding about data/file structures that are necessary to design, and implement information retrieval (IR) systems, IR principles to locate relevant information large collections of data, different document clustering algorithms, information retrieval systems for web search tasks etc.

Course content

Unit I

Introduction to Information Retrieval Systems: Definition of Information Retrieval System, Objectives of Information Retrieval Systems, Functional Overview, Relationship to Database Management Systems, Digital Libraries and Data Warehouses. Information Retrieval System Capabilities: Search Capabilities, Browse Capabilities, Miscellaneous Capabilities.

Unit II

Cataloging and Indexing: History and Objectives of Indexing, Indexing Process, Automatic Indexing, Information Extraction. Data Structure: Introduction to Data Structure, Stemming Algorithms, Inverted File Structure, N-Gram Data Structures, PAT Data Structure, Signature File Structure, Hypertext and XML Data Structures, Hidden Markov Models.

Unit III

Automatic Indexing: Classes of Automatic Indexing, Statistical Indexing, Natural Language, Concept Indexing, Hypertext Linkages. Document and Term Clustering: Introduction to Clustering, Thesaurus Generation, Item Clustering, Hierarchy of Clusters.

User Search Techniques: Search Statements and Binding, Similarity Measures and Ranking, Relevance Feedback, Selective Dissemination of Information Search, Weighted Searches of Boolean Systems, Searching the INTERNET and Hypertext.

Unit IV

Information Visualization: Introduction to Information Visualization, Cognition and Perception, Information Visualization Technologies.

Text Search Algorithms: Introduction to Text Search Techniques, Software Text Search Algorithms, Hardware Text Search Systems. Multimedia Information Retrieval: Spoken Language Audio Retrieval, Non-Speech Audio Retrieval, Graph Retrieval, Imagery Retrieval, Video Retrieval.

Text and References Books:

1. Kowalski & Maybury, *Information storage and retrieval systems: theory and implementation* (Vol. 8). Springer Science & Business Media, 2002.
2. Frakes & Baeza-Yates (Eds)., Information *retrieval: data structures and algorithms*. Prentice-Hall, Inc., 1992.
3. Korfhage, *Information Retrieval and Storage*, John Wiley & Sons, 1997
4. Baeza-Yates & Ribeiro-Neto (1999), *Modern information retrieval* (Vol. 463), New York: ACM press, 1999.

Python Tools for Data Science

General Course Information

Course Code: MDS-SEC-121T Course Credits: 3 Type: Skill Enhancement Course (SEC) Contact Hours: 3 hours/week Mode: Lectures (L) Examination Duration: 3 hours	**Course Assessment Methods:** **Max. Marks: 100 (Internal: 30; External: 70)** Three minor tests, each of 20 marks, will be conducted. The third minor will be conducted by the Course Coordinator. No date sheet will be issued for the third minor at the level of the Departments. For the purpose of internal assessment, the highest marks obtained by a student in any of the minor examinations will be considered. Class Performance will be measured through percentage of lectures attended (4 marks) Assignments (4 marks) and class performance (2 marks). The end semester examination will be of 70 marks. For the end semester examination, nine questions are to be set by the examiner. All questions carry equal marks. Question number one will be compulsory and based on the entire syllabus. It will contain seven short answers type questions. The remaining eight questions are to be given by setting two questions from each of the four units of the syllabus. A candidate is required to attempt five questions in all, one compulsory and any other four questions by selecting one from each unit.

Pre-requisites: Python Programming Language and Database Management System

About the Course:

Python as a programming language has become very popular in recent times. It has been used in data science, IoT, AI, and other technologies, which has added to its popularity. Python is used as a programming language for data science because it contains costly tools from a mathematical or statistical perspective. It is one of the significant reasons why data scientists around the world use Python. This course focus on study of python tools for data scientists. In this course, powerful python tools -NumPy, Pandas and Matplotlib are to be studied. After studying this course, a student is expected to apply python tool for solving data science problems efficiently.

Course content

Unit I

Datasets: Datasets, Data Preprocessing, Preparing Datasets, Missing Data, Anomalies, and Outliers, Imbalanced Classification.

NumPy: Introduction to NumPy, importing NumPy in Python code, NumPy useful features, NumPy Arrays, working with Loops, Appending Elements to Arrays, Multiplying Lists and Arrays, Lists and Exponents, Arrays and Exponents, Math Operations and Arrays, Arrays and Vector Operations, reshape () Method, Mean and Standard Deviation, NumPy and Matplotlib, Linear Regression, Mean Squared Error (MSE) formula, MSE by Successive Approximation.

Unit II

Pandas: Introduction to Pandas, importing pandas in Python code, Pandas useful features, Pandas Data Frames, Pandas Data Frames and Data Cleaning, Boolean Data Frames, Pandas Data Frames and Random Numbers, Reading CSV Files in Pandas, loc() and iloc() Methods, Converting Categorical Data to Numeric Data, Merging and Splitting Columns in Pandas, Combining Pandas Data Frames, Pandas Data Frames and CSV Files, Managing Columns and rows in Data Frames, Handling Missing Data in Pandas, groupby(), apply() and mapapply() in Pandas, Handling Outliers in Pandas, Pandas Data Frames and Simple Statistics, Pandas Method Chaining, Pandas Profiling.

Unit III

Data Cleaning: Data Cleaning, Data Cleaning in SQL, Replace Multiple Values with a Single Value, Handle Mismatched Attribute Values, Convert Strings to Date Values, working with Variable Column Counts, Truncating Rows in CSV Files, Converting Numeric Date Formats, Converting Alphabetic Date Formats, Data Cleaning on a Kaggle Dataset.

Unit IV

Data Visualization: Data Visualization, Types of Data Visualization, Matplotlib, Diagonal Lines in Matplotlib, Colored Grid in Matplotlib, Plotting Multiple Lines in Matplotlib, Trigonometric Functions in Matplotlib, Plot Best-Fitting Line in Matplotlib, SkLearn, Pandas, and Iris Dataset (Introduction only), Seaborn and its Features.

Text and Reference Books:

1. Campesato, Oswald. *Python Tools for Data Scientists Pocket Primer*. Stylus Publishing, LLC, 2022.
2. VanderPlas, Jake. *Python data science handbook: Essential tools for working with data*. " O'Reilly Media, Inc.", 2016.
3. Cielen, Davy, and Arno Meysman. *Introducing data science: big data, machine learning, and more, using Python tools*. Simon and Schuster, 2016.
4. Müller, Andreas C., and Sarah Guido. *Introduction to machine learning with Python: a guide for data scientists*. " O'Reilly Media, Inc.", 2016.
5. Boschetti, Alberto, and Luca Massaron. *Python data science essentials.* Packt Publishing Ltd, 2015.
6. Grus, Joel. *Data science from scratch: first principles with python*. O'Reilly Media, 2019.
7. Boschetti, Alberto, and Luca Massaron. *Python data science essentials: A practitioner's guide covering essential data science principles, tools, and techniques*. Packt Publishing Ltd, 2018.
8. Scavetta, Rick J., and Boyan Angelov. *Python and R for the Modern Data Scientist*. " O'Reilly Media, Inc.", 2021.

Data Analytics Lab.

General Course Information

Course Code: MDS-CC-121P Course Credits: 2 Type: Core Course (CC). Course Contact Hours: 4 hours/week Mode: Lab practice and assignments	**Course Assessment Methods:** **Total Marks: 100 (internal: 50; external: 50)** The internal and external assessment is based on the level of participation in lab. sessions and the timely submission of lab experiments/assignments, the quality of solutions designed for the assignments, the performance in VIVA VOCE, the quality of lab. file and ethical practices followed. The internal examination is conducted by the course coordinator. The external examination is conducted by external examiner appointed by the Controller of Examination in association with the internal examiner appointed by the Chairperson of the Department.

Pre-requisites: Basic programming skills.

About the Course:

The objective of this lab is to enable students to apply advanced data analytics tools for manipulating data, applying statistics, regression and classification.

Practical Lab based on theory course MDS-CC-121T

Note:

The actual experiments/assignments will be designed by the course coordinator. One assignment should be designed to be done in groups of two or three students. The assignments must meet the objective of the course. The list of assignments and schedule of submission will be prepared by the course coordinator at the beginning of the semester.

Data Mining Lab.

General Course Information

Course Code: MDS-CC-122P Course Credits: 2 Type: Core Course (CC). Course Contact Hours: 4 hours/week Mode: Lab practice and assignments	**Course Assessment Methods:** **Total Marks: 100 (internal: 50; external: 50)** The internal and external assessment is based on the level of participation in lab. sessions and the timely submission of lab experiments/assignments, the quality of solutions designed for the assignments, the performance in VIVA VOCE, the quality of lab. file and ethical practices followed. The internal examination is conducted by the course coordinator. The external examination is conducted by external examiner appointed by the Controller of Examination in association with the internal examiner appointed by the Chairperson of the Department.

Pre-requisites: Basic programming skills.

About the Course:

The course helps the students to learn how to perform data mining tasks using a data mining toolkit (such as open-source WEKA), understand the data sets and data pre-processing, demonstrate the working of algorithms for data mining tasks such as association rule mining, classification, clustering and regression, and exercise the data mining techniques with varied input values for different parameters.

Practical Lab based on theory course MDS-CC-122T

Note:

The actual experiments/assignments will be designed by the course coordinator. One assignment should be designed to be done in groups of two or three students. The assignments must meet the objective of the course. The list of assignments and schedule of submission will be prepared by the course coordinator at the beginning of the semester.

Python Tools for Data Science Lab.

General Course Information

Course Code: MDS-SEC-121P Course Credits: 2 Type: Skill Enhancement Course (SEC) Course Contact Hours: 4 hours/week Mode: Lab practice and assignments	**Course Assessment Methods:** **Total Marks: 100 (internal: 50; external: 50)** The internal and external assessment is based on the level of participation in lab. sessions and the timely submission of lab experiments/assignments, the quality of solutions designed for the assignments, the performance in VIVA VOCE, the quality of lab. file and ethical practices followed. The internal examination is conducted by the course coordinator. The external examination is conducted by external examiner appointed by the Controller of Examination in association with the internal examiner appointed by the Chairperson of the Department.

Pre-requisites: Core Python Programming.

About the Course:

In this course, students will learn how to use Python tools -NumPy and Pandas for data exploration, preparation, and analysis.

Practical Lab based on theory course MDS-SEC-121T

Note:

The actual experiments/assignments will be designed by the course coordinator. One assignment should be designed to be done in groups of two or three students. The assignments must meet the objective of the course. The list of assignments and schedule of submission will be prepared by the course coordinator at the beginning of the semester.