

Dr. N.G.P. ARTS AND SCIENCE COLLEGE

(An Autonomous Institution, Affiliated to Bharathiar University, Coimbatore)
Approved by Government of Tamil Nadu and Accredited by NAAC with 'A++' Grade (3rd Cycle-3.64 CGPA)
Dr. N.G.P. - Kalapatti Road, Coimbatore-641048, Tamil Nadu, India
Web: www.drngpasc.ac.in | Email: info@drngpasc.ac.in | Phone: +91-422-2369100

REGULATIONS 2026-27 for Post Graduate Programme (Outcome Based Education model with Choice Based Credit System)

Master of Science in Computer Science with Data Analytics Degree

(For the students admitted during the academic year 2026-27 and onwards)

Programme: M. Sc. (Computer Science with Data Analytics) Eligibility

Candidates for admission to the first year of the **Master of Science (Computer Science with Data Analytics)** Degree Programme shall be required to have passed in B.Sc. Computer Science/ B.C.A. / B.Sc. Computer Technology / B.Sc. Information Technology / B.Sc. Information Sciences / B.Sc. Information Systems / B.Sc. Software Systems / B.Sc. Software Sciences / B.Sc. Applied Sciences (Computer Science / Computer Technology) / B.Sc. Electronics / B.Sc. Mathematics of any University in Tamil Nadu or an Examination accepted as equivalent thereto by the Academic council, subject to conditions as may be prescribed are permitted to appear and qualify for the **Master of Computer Science with Data Analytics Degree Examination** of this College after a programme of study of two academic years.

Programme Educational Objectives

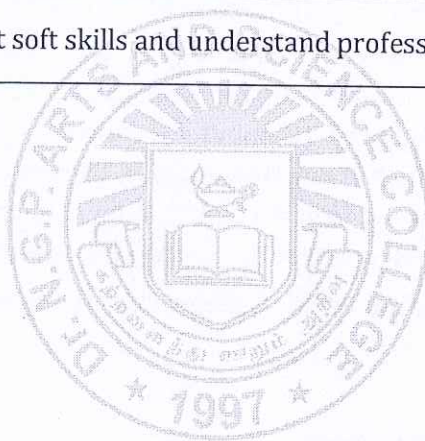
The Curriculum is designed to attain the following learning goals which students shall accomplish by the time of their graduation:

1. Exhibit technical proficiency in Data Analytics to solve real world problems.
2. Engage in successful careers in industry, research and public service.
3. Employ cutting edge tools and technologies for decision making and remain self-motivated and lifelong learners.
4. Practice profession with ethics, integrity, leadership and social responsibility
5. Apply knowledge in areas of Data Analytics for research and entrepreneurship

PROGRAMME OUTCOMES

On the successful completion of the program, the following are the expected outcomes.

PO Number	PO Statement
PO1	Ability to apply knowledge of Computer Science, Mathematics and Statistics to solve problem
PO2	Ability to model, analyze, design, visualize and realize physical systems or processes of increasing size and complexity
PO3	Ability to select appropriate methods and tools for data analysis in specific organizational contexts
PO4	Ability to analyze very large data sets in the context of real world problems and interpret results
PO5	Ability to exhibit soft skills and understand professional and social responsibilities



M.Sc. Computer Science with Data Analytics Credit Distribution

Part	Subjects	No. of Papers	Credit		Semester No.
III	Core	11	3 x 5 = 15 8 x 4 = 32	47	I - III
	Core Practical	06	6 x 2 = 12	12	I - III
	DSE	03	03 x 04 = 12		I - III
	EDC	01	01 x 04 = 04		I
	Industrial Training	01	01 x 02 = 02		III
	Core Project	01	01 x 15 = 15		IV
TOTAL CREDITS			92		

PG
CURRICULUM

M.Sc. Computer Science with Data Analytics
AY 26-27

Course Code	Course Category	Course Name	L	T	P	Duration		Exam (h)	Max Marks			Credits
						Week	Total		CIA	ESE	Total	
First Semester												
26DAP1CA	Core I	Principles of Data Science and Python	4	1	-	5	60	3	25	75	100	5
26MTP1EA	EDC I	Mathematical Foundations of Data Science	4	1	-	5	60	3	25	75	100	4
26CSP1CA	Core II	Design and Analysis of Algorithms	4	-	-	4	48	3	25	75	100	4
26DAP1CB	Core III	Data Engineering	4	-	-	4	48	3	25	75	100	4
26DAP1CP	Core Practical I	Python Programming	-	-	4	4	48	3	40	60	100	2
26DAP1CQ	Core Practical II	Data Engineering	-	-	4	4	48	3	40	60	100	2
26DAP1DA	DSE -I	Business Intelligence And Information Visualization	4	-	-	4	48	3	25	75	100	4
26DAP1DB		Information Retrieval										
26DAP1DC		Web Intelligence										
Total			20	2	8	30	360	-	-	-	700	25

Course Code	Course Category	Course Name	L	T	P	Duration		Exam (h)	Max Marks				
						Week	Total		CIA	ESE	Total	Credits	
Second Semester													
26DAP2CA	Core IV	Artificial Intelligence	4	1	-	5	60	3	25	75	100	5	
26DAP2CB	Core V	Data Mining	4	1	-	5	60	3	25	75	100	4	
26DAP2CC	Core VI	Customer Analytics	4	-	-	4	48	3	25	75	100	4	
26DAP2CD	Core VII	Exploratory Data Analysis	4	-	-	4	48	3	25	75	100	4	
26DAP2CP	Core Practical III	Data Mining	-	-	4	4	48	3	40	60	100	2	
26DAP2CQ	Core Practical IV	Data analysis with R	-	-	4	4	48	3	40	60	100	2	
26DAP2DA	DSE –II	Information and Network Security	4	-	-	4	48	3	25	75	100	4	
26DAP2DB		Natural Language Processing											
26DAP2DC		Advanced Statistics											
Total			20	2	8	30	360	-	-	-	700	25	

Course Code	Course Category	Course Name	L	T	P	Duration		Exam (h)	Max Marks			Credits
						Week	Total		CIA	ESE	Total	
Third Semester												
26DAP3CA	Core VIII	Machine Learning	4	1	-	5	60	3	25	75	100	5
26DAP3CB	Core IX	Advanced Databases	4	1	-	5	60	3	25	75	100	4
26DAP3CC	Core X	Cloud Computing	4	-	-	4	48	3	25	75	100	4
26DAP3CD	Core XI	Big Data Analytics	4	-	-	4	48	3	25	75	100	4
26DAP3CP	Core Practical V	Machine Learning	-	-	4	4	48	3	40	60	100	2
26DAP3CQ	Core Practical VI	Big Data Analytics	-	-	4	4	48	3	40	60	100	2
26DAP3TA	IT	Industrial Training	-	-	-	-	-	3	40	60	100	2
26DAP3DA	DSE –III	Computer Vision	4	-	-	4	48	3	25	75	100	4
26DAP3DB		Large Language Models										
26DAP3DC		Internet of Things and Applications										
Total			20	2	8	30	360	-	-	-	800	27

Course	Course Category	Course Name	L	T	P	Duration		Exam (h)	Max Marks			Credits
						Week	Total		CIA	ESE	Total	
Fourth Semester												
26DAP4CV	Core - XII	Project Work and Viva Voce	-	-	-	-	-	3	80	120	200	15
Total			-	-	-	-	-	-	-	-	200	15
*Grand Total											2400	92

*Total Credits does not exceed 92 credits



DISCIPLINE SPECIFIC ELECTIVE

Students shall select the desired course of their choice in the listed elective course during Semesters I to IV

Semester I (Elective I)

List of Elective Courses

S. No.	Course Code	Name of the Course
1.	26DAP1DA	Business Intelligence and Information Visualization
2.	26DAP1DB	Information Retrieval
3.	26DAP1DC	Web Intelligence

Semester II (Elective II)

List of Elective Courses

S. No.	Course Code	Name of the Course
1.	26DAP2DA	Information and Network Security
2.	26DAP2DB	Natural Language Processing
3.	26DAP2DC	Advanced Statistics

Semester III (Elective III)

List of Elective Courses

S. No.	Course Code	Name of the Course
1.	26DAP3DA	Computer Vision
2.	26DAP3DB	Large Language Models
3.	26DAP3DC	Internet of Things and Applications

EXTRA CREDIT COURSES

The following are the courses offered under self-study to earn extra credits:

S. No.	Course Code	Course Title
1.	26DAPSSA	Business Analytics
2.	26DAPSSB	Professional Ethics

NP Sdse

BoS Chairman/HoD
Department of Computer Science with Data Analytics
Dr. N. G. P. Arts and Science College
Coimbatore – 641 048

 Dr.N.G.P. Arts and Science College		
APPROVED		
BoS- 14 th 02.04.26	AC -	GB -

Semester - I
CORE I: PRINCIPLES OF DATA SCIENCE AND PYTHON

Semester	Course Code	Course Name	Category	L	T	P	Credits
I	26DAP1CA	PRINCIPLES OF DATA SCIENCE AND PYTHON	CORE	48	12	-	5

Preamble	This course has been designed for students to learn and understand • Concepts of Data Science • Understand about Python Programming • Plotting and Visualization in Python	
Prerequisite	Basic Programming skills	
Course Outcomes (COs)		
CO Number	Course Outcomes (COs) Statement	Bloom's Taxonomy Knowledge Level
CO1	Understand the principles of data science	K2
CO2	Understand the techniques for Data Handling	K2
CO3	Apply Numpy and Pandas to perform numerical operations	K3
CO4	Apply the concepts of Python for Data Aggregation and Wrangling	K3
CO5	Design Interactive dashboards using Python Libraries	K6

Mapping with Program Outcomes: ★ 1997 ★					
COs / POs	PO1	PO2	PO3	PO4	PO5
CO1	S	S	S	S	M
CO2	S	S	S	S	M
CO3	S	S	S	S	S
CO4	S	S	S	S	S
CO5	S	S	S	S	S

26DAP1CA

PRINCIPLES OF DATA SCIENCE AND PYTHON

Syllabus

Unit	Content	Hours	E-Contents / Resources
I	Introduction to Data Science Process Introduction: Benefits of Data Science - Facets of Data –Big data eco system and data science - Data science process: Steps in data science process - Retrieving data – Data preparation – Data exploration – Data modeling – Presentation - Case Study	12	Text Book
II	Handling Large Data Problems when handling large data – General techniques for handling large data – General Programming dealing with Large Data Sets – Steps in big data – Distributing data storage and processing with Frameworks- Applications in Data Science - Case Study – Assessing risk when loaning money	12	Text Book
III	NumPy and Pandas Introduction to NumPy - Understanding the N - dimensional data structure -Creating NumPy arrays – Basic operations and manipulations on N-dimensional arrays – Indexing and Slicing-Advanced Indexing – Pandas: Mathematical Functions– Statistical Functions – Search, Sorting and Counting Functions –Matrix Library	12	Text Book
IV	Data Manipulation Introduction: GroupBy Mechanics – Data Aggregation – Groupwise Operations and Transformations – Pivot Tables and Cross Tabulations – Date and Time- Date Type tools – Time Series Basics – Data Ranges - Frequencies and Shifting - Combining and Merging DataSets – Reshaping and Pivoting – Data Transformation – String Manipulation, Regular Expressions	12	Text Book
V	Visualization Introduction: Data Acquisition by Scraping web applications – Submitting a form - Fetching web pages – CSS Selectors. Visualization: Visualization In Python: Matplotlib package – Plotting Graphs – Controlling Graph – Adding Text – More Graph Types – Getting and setting values - Plotting with Pandas and seaborn - Line plots - Bar Plots -Histogram and Density Plots - scatter or point plots - facet grids and categorical data	12	You Tube Videos
	Total	60	

Text Book	1.	Davy Cielen, Arno D.B. Meysmen, Mohamed Ali, 2020, "Introducing Data Science", Dream Tech Press (UNITS I,II)
	2.	Wes Mc Kinney, 2020, "Python for Data Analysis", 5th Edition, O'Reilly (UNITS III, IV, V)
Reference Books	1.	John V Guttag, 2016, "Introduction to Computation and Programming Using Python", 2nd Edition., MIT press]
	2.	Gypsy Nandi, Rupam Kumar Sharma, 2020, "Data Science Fundamentals and Practical Approach, BPB
	3.	Zed Shaw, 2014, "Learn Python the Hard Way", 3rd Edition, Addison-Wesley, USA,
	4.	Fabio Nelli, 2018, "Python Data Analytics", Second Edition, Apress, NewYork,

Journal and Magazines	https://www.pythonpapers.com/ https://www.pythonweekly.com/
E-Resources and Website	https://www.python.org/ https://www.linkedin.com/learning/paths/master-python-for-data-science-16361344 https://www.edx.org/learn/python

Learning Method	Chalk and Talk/Assignment/Seminar/Online Compiler
Focus of the Course	Skill Development/ Employability

Semester – I							
EDC I: MATHEMATICAL FOUNDATIONS OF DATA SCIENCE							
Semester	Course Code	Course Name	Category	L	T	P	Credits
I	26MTP1EA	MATHEMATICAL FOUNDATIONS OF DATA SCIENCE	EDC	48	12	-	4

Preamble	This course has been designed for students to learn and understand • Data numerically and visually • The knowledge of testing of hypothesis for small and large samples which plays an important role in real life applications • Data-based claims and quantitative arguments
Prerequisite	Knowledge on Basic Mathematics

Course Outcomes (Cos)		
CO Number	Course Outcomes (COs) Statement	Bloom's Taxonomy Knowledge Level
CO1	Explain fundamental concepts of probability, including definitions, rules, and importance	K2
CO2	Apply probability distributions to compute probabilities in real-life problems	K3
CO3	Analyze relationships between variables using regression lines and equations	K4
CO4	Evaluate hypotheses based on test statistics and significance levels	K5
CO5	Develop analytical models using ANOVA and F-test for complex decision-making problems	K6

Mapping with Program Outcomes:					
Cos / POs	PO1	PO2	PO3	PO4	PO5
CO1	S	M	M	M	M
CO2	S	M	S	S	S
CO3	S	S	S	S	S
CO4	S	S	S	S	S
CO5	S	S	S	S	S

26MTP1EA	EDC: MATHEMATICAL FOUNDATIONS OF DATA SCIENCE
----------	---

Syllabus

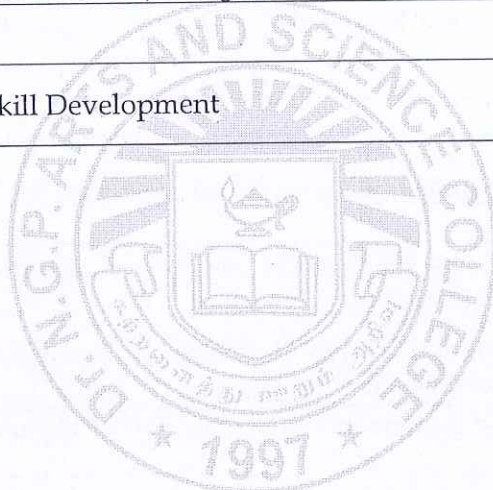
Unit	Content	Hrs	Resources
I	Probability Theory Introduction - Probability Defined - Importance of the Concept of Probability - Calculation of Probability - Theorems of Probability - Addition Theorem - Multiplication Theorem - Conditional Probability - Bayes Theorem -Mathematical Expectation.	10	Text Book
II	Probability Distributions Introduction - Binomial Distribution-Fitting a Binomial Distribution- Poisson Distribution - Fitting a Poisson Distribution - Normal Distributions - Fitting a Normal Curve.	10	Text Book
III	Correlation and Regression Correlation - Scatter Diagram Method - Graphic Method- Karl Pearson's Coefficient of Correlation -Spearman's coefficient of Correlation - Regression Analysis - Regression Lines - Regression Equations -Regression Equation of Y on X - Regression Equation of X on Y.	12	Text Book
IV	Testing of Hypothesis Introduction - Hypothesis Testing - Standard Error and Sampling Distribution - Estimation - Tests of Significance for Large Samples - Difference between small and large samples - Two tailed test for difference between the means of two samples -Standard Error of the difference between two standard deviations - Tests of significance for small samples - Assumption of Normality - Student's t distribution - Application of the t Distribution.	14	NPTEL
V	F- Test and ANOVA Introduction- Chi-Square test- F-Test -Applications of F-Test - Analysis of Variance - Assumptions -Technique of Analysis of Variance - One-Way Classification - Analysis of Variance in Two-Way Classification Model.	14	You Tube Videos
Total		60	

Text book	1.	Gupta S.P,2017, "Statistical Methods", 45 th Edition, Sultan Chand and Sons, New Delhi.
Reference Books	1.	Ronald E. Walpole,2018, "Probability and Statistics", 9 th Edition, Pearson Education South Asia.
	2.	Sheldon M. Ross, 2017, "Introductory Statistics", 4 th Edition, Academic Press,United States.
	3.	Vijay K. Rohatgi A.K, MD. Ehsanes Saleh, 2015, " An introduction toProbability and Statistics", 3 rd Edition, John Wiley and Sons, New Delhi.
	4.	Sheldon M. Ross, 2017, "A first course in Probability", 5 th Edition, PHI, NewJersey University Press.

Journal and Magazines	https:// www.worldscientific.com/worldscinet/bms
E-Resources and Website	https://resources.nu.edu/statsresources/Chi-Square https://nptel.ac.in

Learning Method	Chalk and Talk/ Assignment/Seminar
-----------------	------------------------------------

Focus of the Course	Skill Development
---------------------	-------------------



Semester – I CORE II: DESIGN AND ANALYSIS OF ALGORITHMS							
Semester	Course Code	Course Name	Category	L	T	P	Credits
I	26CSP1CA	DESIGN AND ANALYSIS OF ALGORITHMS	CORE	48	-	-	4

Preamble	This course has been designed for students to learn and understand - To analyse the asymptotic performance of algorithms - To demonstrate a familiarity with major algorithms and data structures - To synthesize efficient algorithms in common engineering design situations
Prerequisite	Basic Knowledge in Data Structures and Programming.

Course Outcomes (Cos)		
CO Number	Course Outcomes (COs) Statement	Bloom's Taxonomy Knowledge Level
CO1	Understand the fundamental concepts of algorithm analysis and basic data structures.	K2
CO2	Apply Divide and Conquer and Greedy strategies to solve basic computational problems.	K3
CO3	Analyze the performance and efficiency of algorithmic techniques.	K4
CO4	Evaluate the suitability of different algorithmic paradigms.	K5
CO5	Design efficient algorithmic solutions for Real time problems.	K6

Mapping with Program Outcomes:					
Cos / POs	PO1	PO2	PO3	PO4	PO5
CO1	S	M	M	S	S
CO2	S	S	S	M	M
CO3	S	S	M	S	S
CO4	M	M	S	M	M
CO5	S	S	M	M	S

26CSP1CA	DESIGN AND ANALYSIS OF ALGORITHMS
----------	-----------------------------------

Syllabus

Unit	Content	Hrs	Resources
I	Introduction to Algorithms Algorithm-Algorithm Specification - Performance Analysis- Space Complexity - Time Complexity - Asymptotic Notations -Performance Measurement - Elementary Data Structures - Stacks and Queues - Trees - Dictionaries - Priority Queues - Graphs.	9	Text Book
II	Divide and Conquer General Method - Binary Search - Finding Maximum and Minimum - Merge Sort - Quick Sort - Strassen's Matrix Multiplication - Convex Hull.	9	Text Book
III	The Greedy Method The General Method - Knapsack Problem - Tree Vertex Splitting - Job Sequencing with Deadlines - Minimum-Cost Spanning Trees - Prim's Algorithm - Kruskal's Algorithm - Single-Source Shortest Paths.	10	Text Book
IV	Dynamic Programming The General Method - Multistage Graphs - All-Pairs Shortest Paths - Single-Source Shortest Paths: General Weights - Optimal Binary Search Trees - 0/1 Knapsack - The Traveling Salesperson Problem - Flow Shop Scheduling.	10	Text Book
V	Backtracking, Branch and Bound The General Method -The 8-Queens problem - Sum of Subsets - Graph Coloring - Hamiltonian Cycles - Knapsack Problem. Branch and Bound - The method - 0/1 Knapsack Problem - Traveling Salesperson problem	10	Text Book
Total		48	

	Note: Case studies related to the above topics to be discussed (Examined Internal only)
--	---

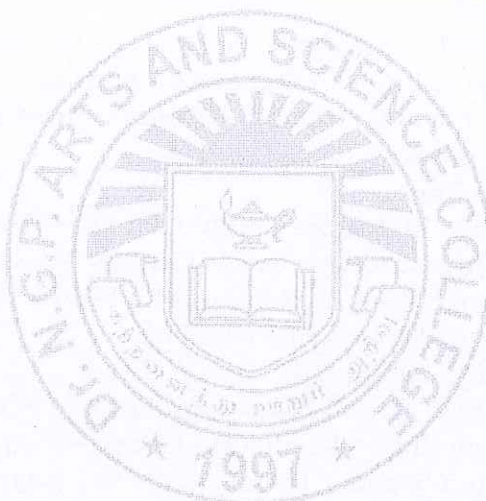
Text book	1.	Ellis Horowitz, Sartaj Sahni, Sanguthevar Rajasekaran, 2025, "Fundamentals of Computer Algorithms", 2nd Edition, Universities Press Private Limited, India.
Reference Books	1.	Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, Clifford Stein, 2022, "Introduction to Algorithms", Fourth Edition, The MIT Press.
	2.	M. H. Alsuwaiyel, 2016, "Algorithms Design Techniques and Analysis" Revised Edition, World Scientific Publishing Co. Pvt. Ltd.

3.	Kleinberg, Jon, and Eva Tardos, 2014 "Algorithm Design", First Edition, Pearson new international Edition.
4.	Priya Sen, 2017, "Design and Analysis of Algorithm", Tutorial Point(I) Pvt. Ltd.

Journal and Magazines	https://link.springer.com/chapter/10.1007/978-1-4612-4400-4_1
E-Resources and Website	https://nptel.ac.in/courses/106105157

Learning Method	Chalk and Talk/ Assignment/Seminar/Case Study
-----------------	---

Focus of the Course	Skill Development/ Employability/ Innovations
---------------------	---



Semester - I CORE III: DATA ENGINEERING							
Semester	Course Code	Course Name	Category	L	T	P	Credits
I	26DAP1CB	DATA ENGINEERING	CORE	48	-	-	4

Preamble	This course has been designed for students to learn and understand • The principles and practices of data engineering, focusing on the design, development, and management of data systems. • The data engineering lifecycle including data generation, ingestion, storage, transformation and serving. • Modern architectures, data pipelines, real-time processing and data security for building scalable analytics solutions.	
Prerequisite	Fundamentals of Database Management Systems, Data analytics concepts	
Course Outcomes (COs)		
CO Number	Course Outcomes (COs) Statement	Bloom's Taxonomy Knowledge Level
CO1	Explain the concepts of data engineering lifecycle and architecture in the context of modern data-driven applications.	K2
CO2	Apply data generation, storage, and ingestion processes to collect and manage data efficiently from diverse sources.	K3
CO3	Apply appropriate querying and data transformation techniques to extract, process and manipulate data.	K3
CO4	Analyze data models and evaluate pipeline performance to ensure efficiency and reliability.	K4
CO5	Design scalable data engineering solutions by integrating appropriate architectures and data pipelines.	K6

Mapping with Program Outcomes:

COs / POs	PO1	PO2	PO3	PO4	PO5
CO1	S	M	S	S	M
CO2	S	S	S	S	S
CO3	S	S	S	S	S
CO4	M	S	S	S	M
CO5	S	S	S	S	S

26DAP1CB

DATA ENGINEERING

Syllabus

Unit	Content	Hours	E-Contents/ Resources
I	Introduction Data Engineering: Definition – Skills - Data Engineering Life Cycle: Generation – Storage – Ingestion Transformation – Serving data. Major undercurrents across Data Engineering: Security – Data Management – Data Architecture – Orchestration. Principles of Good Data Architecture – Major Architecture Concepts: Domains and Services – Tight vs Loose coupling – User access – Event drive architecture- Types of Data Architecture.	10	Text Book
II	Data Generation, Storage Sources of Data – Source Systems - Practical Details: Databases – APIs - Data Sharing- Third Party Data Sources- Message Queues and Event Streaming platforms. Data Storage : Ingredients of Data Storage - Data Storage Systems – Data Engineering storage abstractions – Trends in storage.	10	Text Book
III	Ingestion Ingestion : Key considerations for Ingestion phase – Batch Ingestion: Snapshot- File based Export- ETL Versus ELT – Inserts, updates and batch size – Data Migration – Message and stream ingestion: Schema Evolution – Late Arriving Data- Ordering and Multiple Delivery-Replay- Time to Live- Message Size- Error Handling- Consumer Pull and Push, Ways to Ingest data	10	Text Book
IV	Queries, Modeling and Transformation Queries: Life of a Query- Query Optimizer- Improving Query Performance – Queries on Streaming Data – Data Modeling: Definition – Types- Normalization – Techniques for modelling batch data-Modeling Streaming data-Transformation : Batch – Materialized Views, Federation and Query Virtualization – Streaming Transformations and Processing.	10	Text Book
V	Security and Privacy Ways to serve Data for analytics: File exchange – Databases – Streaming – Query Federation- Data sharing – Semantic and Metrics Layers- Serving data in notebooks. Security: Processes- Technology- Cloud Scale Data OS – Enterprise Data Engineering – Live Data Stack-Streaming Pipelines and Real time analytical Databases – Fusion of Data with Applications.	8	Reference Book
	Total	48	

Text Book	1.	Joe Reis, Matt Housley, Fundamentals of Data Engineering, O'Reilly, 2022
Reference Books	1.	Zhaolong Liu, Data Engineering Fundamentals: Building scalable data solutions with ETL pipelines and strategic data architecture design, BPB Publications, 2025.
	2.	Data Engineering with Python – Paul Crickard (2020)
	3.	Bas P. Harenslak & Julian de Ruiter Data Pipelines with Apache Airflow – (2021)
	4.	Data Engineering on Azure – Vlad Riscutia (2021)

Journal and Magazines	Journal of Data Engineering and Knowledge Discovery International Journal of Data Engineering Research and Development IEEE Transactions on Knowledge and Data Engineering IEEE Data Engineering Bulletin
E-Resources and Website	https://dataengineeringcommunity.com https://www.thedataengineering.com https://kafka.apache.org https://www.pyspark.in

Learning Method	Chalk and Talk/Assignment/Seminar/Collaborative Learning/Demonstration
-----------------	--

Focus of the Course	Employability/ Skill Development
---------------------	----------------------------------

Semester – I CORE PRACTICAL I: PYTHON PROGRAMMING							
Semester	Course Code	Course Name	Category	L	T	P	Credits
I	26DAP1CP	PYTHON PROGRAMMING	CORE PRACTICAL		-	48	2

Preamble	This course has been designed for students to learn and understand • Data Aggregation and Wrangling in Python • Working with structured data formats using built-in and external libraries • Data visualization using Matplotlib and Seaborn libraries
Prerequisite	Programming Concepts, Logic and Problem Solving Skills

Course Outcomes (Cos)		
CO Number	Course Outcomes (COs) Statement	Bloom's Taxonomy Knowledge Level
CO1	Understand and implement basic aggregation operations and sequential search	K2
CO2	Apply string manipulation techniques and built-in string functions for text processing	K3
CO3	Demonstrate the ability to perform basic file operations and understand file handling mechanisms	K3
CO4	Apply Pandas to analyze, process, and visualize structured data effectively	K3
CO5	Create effective data visualizations using Matplotlib and Seaborn for graphical representation of data	K6

Mapping with Program Outcomes:					
Cos / POs	PO1	PO2	PO3	PO4	PO5
CO1	S	S	S	S	S
CO2	S	M	S	S	M
CO3	S	S	S	M	S
CO4	S	S	S	S	S
CO5	S	M	M	S	M

26DAP1CP	PYTHON PROGRAMMING
----------	--------------------

S.No	List of Programs
1	Programs to perform aggregation operations.
2	Programs to Implement a sequential search.
3	Programs to Explore string functions.
4	Programs to Read and Write into a file
5	Programs to Demonstrate use of List.
6	Programs to Demonstrate use of Dictionaries.
7	Programs to Demonstrate use of Tuples
8	Programs to Create Comma Separate Files (CSV), Load CSV files into internal Data
9	Programs using Pandas: Extract items at given positions from a series
10	Programs to implement correlation and covariance.
11	Program to plot graphs using Matplotlib and seaborn packages.
12	Programs to Perform Analysis for given data set using Pandas

Text Books	1.	Davy Cielen, Arno D.B. Meysmen, Mohamed Ali, 2020, "Introducing Data Science", Dream Tech Press)..
	2.	Wes Mc Kinney, 2020, "Python for Data Analysis", 5th Edition, O'Reilly.

Learning Method	Demonstration/ Hands on Experiments
-----------------	-------------------------------------

Focus of the Course	Skill Development/ Employability
---------------------	----------------------------------

Semester – I CORE PRACTICAL - II: DATA ENGINEERING							
Semester	Course Code	Course Name	Category	L	T	P	Credits
I	26DAP1CQ	DATA ENGINEERING	CORE PRACTICAL		-	48	2

Preamble	This course has been designed for students to learn and understand • Core concepts of data engineering, including data ingestion, transformation, and storage techniques. • Data pipeline orchestration using ETL tools and workflow management systems. • ETL processes, data integration and data validation for data engineering tasks.
Prerequisite	Database, SQL.

Course Outcomes (Cos)		
CO Number	Course Outcomes (COs) Statement	Bloom's Taxonomy Knowledge Level
CO1	Explain the concepts of ETL pipelines, data ingestion techniques, and workflow orchestration in data engineering.	K2
CO2	Develop and implement data pipelines to extract data from files and APIs, transform it, and load it into databases.	K3
CO3	Build and manage workflow pipelines to automate tasks including data transformation, batch processing and failure handling.	K3
CO4	Analyze data by performing operations such as filtering, aggregation, joins, and data format conversions (CSV, JSON, Parquet).	K4
CO5	Evaluate and optimize data pipelines by integrating multiple data sources, implementing incremental loading, and ensuring data validation and quality checks.	K6

Mapping with Program Outcomes:					
Cos / POs	PO1	PO2	PO3	PO4	PO5
CO1	S	S	S	M	M
CO2	S	S	S	M	M
CO3	M	S	S	S	S
CO4	S	S	S	M	M
CO5	S	S	S	S	S

26DAPI1CQ	DATA ENGINEERING
------------------	-------------------------

S.No	List of Programs
------	------------------

- 1 Create a simple ETL pipeline that extracts data from a CSV file, transform it and load it into a database.
- 2 Build a basic Directed Acyclic Graph (DAG) with extract, transform, and load tasks.
- 3 Fetch data from a public API, process it, and store it in a database.
- 4 Design a batch ingestion pipeline for processing daily data files.
- 5 Perform data transformations such as filtering, aggregation and joins using Python/SQL.
- 6 Implement retry logic and failure handling in a DAG.
- 7 Convert data from CSV to Parquet format and store it.
- 8 Integrate data from multiple sources (API) and store the unified dataset.
- 9 Implement incremental data loading. Load only new records and update the existing records.
- 10 Perform data validation checks such as nulls, duplicates, schema validation before loading.

Text Book	1.	Joe Reis, Matt Housley, Fundamentals of Data Engineering, O'Reilly, 2022.
Reference Book	1.	Zhaolong Liu, Data Engineering Fundamentals: Building scalable data solutions with ETL pipelines and strategic data architecture design, BPB Publications, 2025.

Learning Method	Demonstration/ Hands on Experiments
-----------------	-------------------------------------

Focus of the Course	Skill Development/ Employability
---------------------	----------------------------------

Semester –I							
DSE I: BUSINESS INTELLIGENCE AND INFORMATION VISUALIZATION							
Semester	Course Code	Course Name	Category	L	T	P	Credits
I	26DAP1DA	BUSINESS INTELLIGENCE AND INFORMATION VISUALIZATION	DSE	48	-	-	4

Preamble	This course has been designed for students to learn and understand • Methodologies used by enterprises for data analysis and decision making. • Application of visualization techniques to support various data analysis tasks. • Visualization methods for interpreting and analyzing multivariate data.
Prerequisite	Knowledge on Python, statistics and data handling

Course Outcomes (Cos)		
CO Number	Course Outcomes (COs) Statement	Bloom's Taxonomy Knowledge Level
CO1	Understand the basic concepts of business intelligence and its application	K1
CO2	Design and process the data for Visualization.	K2
CO3	Apply appropriate visualization techniques for spatial and geospatial data	K3
CO4	Evaluate the effectiveness of visualization techniques for multivariate data	K5
CO5	Develop a visualization system using interaction concepts and techniques	K6

Mapping with Program Outcomes:					
Cos / POs	PO1	PO2	PO3	PO4	PO5
CO1	S	S	S	M	M
CO2	S	S	S	S	S
CO3	S	S	S	S	S
CO4	S	S	S	S	S
CO5	S	S	S	M	M

26DAP1DA	BUSINESS INTELLIGENCE AND INFORMATION VISUALIZATION
----------	--

Syllabus

Unit	Content	Hrs	Resources
I	Business Intelligence Changing Business Environments and Computerized Decision Support – Managerial Decision Making – An Early Framework for Computerized Decision Support – The Concept of Decision Support Systems (DSS) – A Framework for Business Intelligence(BI) – Business Analytics Overview.	8	Text Book
II	Foundations of Visualization Visualization stages - Semiology of Graphical Symbols - The Eight Visual Variables - Historical Perspective - Taxonomies - Experimental Semiotics based on Perception Gibson's Affordance theory – A Model of Perceptual Processing.	10	Text Book
III	Visualization of Spatial, Geospatial Data Spatial Data: One-Dimensional Data - Two-Dimensional Data – Three Dimensional Data - Dynamic Data - Combining Techniques. Geospatial Data: Visualizing Spatial Data - Visualization of Point Data -Visualization of Line Data - Visualization of Area Data - Other Issues in Geospatial	10	Text Book
IV	Visualization of Multivariate Data, Text and Document Visualization Data Visualization Multivariate Data: Point-Based Techniques – Line Based Techniques - Region-Based Techniques - Combinations of Techniques – Trees Displaying Hierarchical Structures – Graphics and Networks- Displaying Arbitrary Graphs/Networks -Text and Document Visualization: Introduction - Levels of Text Representations - The Vector Space Model - Single Document Visualizations -Document Collection Visualizations	10	Text Book
V	Interaction Concepts Extended Text Visualizations Interaction Concepts: Interaction Operators - Interaction Operands and Spaces - A Unified Framework. Interaction Techniques: Screen Space - Object-Space -Data Space -Attribute Space- Data Structure Space - Visualization Structure - Animating Transformations - Interaction Control	10	Text Book
	Total	48	

Note: Case studies related to the above topics to be discussed (Examined Internal only)

Text books	1.	Efraim Turban, Ramesh Sharda, Dursun Delen, 2018, "Business Intelligence and Analytics", 10 th Edition, Person (Unit - I)
	2.	Matthew Ward, Georges Grinstein and Daniel Keim, 2010, "Interactive Data Visualization Foundations, Techniques, Applications", A. K. Peters, Ltd.
Reference Books	1.	S. Christian Albright, Wayne L. Winston, 2020, "Business Analytics Data Analysis and Decision Making", 6 th Edition
	2.	Alexandru C. Telea, 2008, "Data Visualization: Principles and Practice," A. K. Peters Ltd.
	3.	Ben Fry, 2008, "Visualizing data: Exploring and explaining data with the processing environment", O'Reilly,
	4.	Scott Murray, 2013, "Interactive data visualization for the web", O'Reilly Media, Inc..
	5.	Robert Spence, 2007, "Information visualization – Design for interaction", 2nd Edition, Pearson Education.

Journal and Magazines	–
E-Resources and Website	Infosys Springboard: Digital Learning and Reskilling Programs Principles of Management Coursera

Learning Method	Chalk and Talk/Assignment/Seminar/Group Discussion/Case Study
-----------------	---

Focus of the Course	Skill Development/ Employability/ Entrepreneurial Development/ Innovations
---------------------	--

Semester - I							
DSE I: INFORMATION RETRIEVAL							
Semester	Course Code	Course Name	Category	L	T	P	Credits
I	26DAP1DB	INFORMATION RETRIEVAL	DSE	48	-	-	4

Preamble	This course has been designed for students to learn and understand • The concepts of information retrieval techniques • The techniques focused on document classification, tolerant retrieval and evaluation • The methods of developing an information retrieval system	
Prerequisite	Basic Knowledge on Data structures, Algorithms and Databases	
Course Outcomes (COs)		
CO Number	Course Outcomes (COs) Statement	Bloom's Taxonomy Knowledge Level
CO1	Understand the concepts of the standard models of Information Retrieval	K2
CO2	Understand the methods for handling wild card queries and spelling correction	K2
CO3	Apply appropriate methods for scoring and evaluating IR systems	K3
CO4	Apply text classification to locate relevant information from large collections of text data	K3
CO5	Design an Information Retrieval System for search tasks involving XML and web data	K6

Mapping with Program Outcomes:					
COs / POs	PO1	PO2	PO3	PO4	PO5
CO1	S	S	S	S	M
CO2	S	S	S	S	M
CO3	S	S	S	S	S
CO4	S	S	S	S	S
CO5	S	S	S	S	S

26DAP1DB

INFORMATION RETRIEVAL

Syllabus

Unit	Content	Hours	E-Contents / Resources
I	Introduction Introduction to Information Retrieval- Building an inverted Index - Processing Boolean Queries - Boolean Model vs Ranked Retrieval - Term Vocabulary and Postings: Tokenization - Stop words - Normalization- Stemming - Skip pointers -Phrase queries: Biword indexes- Positional indexes	09	Text Book
II	Wild card queries, Spelling correction Search Structures for Dictionaries- Wild card queries - General wild card queries-k gram indexes for wild card queries - Spelling correction - Forms- Edit distance -k gram indexes for spelling correction - Phonetic Correction - Index construction-Distributed indexing- Statistical properties of terms: Heaps' Law- Zipf's Law	10	Text Book/ You Tube Videos
III	Evaluation of IR-systems Term frequency and weighting -Inverse document frequency TF-IDF weighting -Vector space model for scoring -Efficient scoring and ranking -Evaluation: Information retrieval system evaluation- Evaluation of unranked retrieval sets-Evaluation of ranked retrieval sets- Case study	10	Text Book
IV	Text classification Text classification and Naive Bayes- The text classification problem- Naive Bayes text classification - Feature selection - Mutual information- Vector space classification: Document representations and measures of relatedness in vector spaces k nearest neighbour - Linear versus nonlinear classifiers - Case study	10	Text Book
V	XML Indexing Web Search XML Indexing and Search: Basic XML concepts - Challenges in XML retrieval- A vector space model for XML retrieval - Data vs. Text-centric XML- Web search basics-Web characteristics-Web crawling - Features of web crawler Architecture-Distributing indexes - Machine learning methods in ad hoc information retrieval - Case study	09	Text/ Reference Book
	Total	48	

Text Book	1.	Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schuetze, 2009, "Introduction to Information Retrieval", First Edition, Cambridge University Press
Reference Books	1.	Baeza -Yates Ricardo and Berthier Ribeiro - Neto, 2011, "Modern Information Retrieval", 2nd edition, Addison-Wesley
	2.	Gerald Kowalski, 2010, "Information Retrieval Architecture and Algorithms ", First Edition, Berlin, Heidelberg: Springer-Verlag
	3.	G.G. Chowdhury, 2010, "Introduction to Modern Information Retrieval", 3rd Edition, Facet Publishing.
	4.	Bruce Croft, Donald Metzler, and Trevor Strohman, 2009, "Search Engines: Information Retrieval in Practice" Pearson Education

Journal and Magazines	https://dl.acm.org/journal/infre
E-Resources and Website	https://nlp.stanford.edu/IR-book/information-retrieval-book.html https://people.ischool.berkeley.edu/~hearst/irbook/

Learning Method	Chalk and Talk/ Assignment/Seminar/Problem Solving
-----------------	--

Focus of the Course	Skill Development/Employability
---------------------	---------------------------------

Semester – I DSE I: WEB INTELLIGENCE							
Semester	Course Code	Course Name	Category	L	T	P	Credits
I	26DAP1DC	WEB INTELLIGENCE	DSE	48	-	-	4

Preamble	This course has been designed for students to learn and understand • The concepts of web mining and crawling • The techniques in opinion mining and sentiment analysis • The concepts of social network Analysis	
Prerequisite	Knowledge on Web Technologies	
Course Outcomes (COs)		
CO Number	Course Outcomes (COs) Statement	Bloom'sTaxonomy Knowledge Level
CO1	Understand the concepts of web mining	K2
CO2	Analyze social networks and web crawling	K4
CO3	Experiment with opinion mining and sentiment analysis	K5
CO4	Understand Google Analytics	K2
CO5	Design Applications using web intelligence	K6

Mapping with Program Outcomes:					
COs / POs	PO1	PO2	PO3	PO4	PO5
CO1	S	S	S	M	M
CO2	S	S	S	S	M
CO3	S	S	S	S	S
CO4	S	S	S	S	M
CO5	S	S	S	S	S

26DAP1DC

WEB INTELLIGENCE

Syllabus

Unit	Content	Hours	E-Contents/ Resources
I	Web Mining Introduction - Information Retrieval and Web Search - Basic Concepts of Information Retrieval - Information Retrieval Models - Relevance Feedback - Evaluation Measures - Text and Web Page Pre- Processing - Web Search - Meta Search: Combining Multiple Rankings - Web Spamming	10	Text Book
II	Social Network Analysis Introduction - Co-Citation and Bibliographic Coupling - Page Rank - Semantic web - Web Intelligence: Levels - Goals - Characteristics - Challenges and issues Tools for web crawling - Web Crawling: Basic Crawler Algorithm - Implementation Issues - Universal Crawlers - Focused Crawlers - Topical Crawlers: Topical Locality and Cues - Best-First Variations - Adaptation - Evaluation - Crawler Ethics and Conflicts	10	Reference Book
III	Opinion Mining The Problem of Opinion Mining - Document Sentiment Classification - Sentence Subjectivity and Sentiment Classification - Opinion Lexicon Expansion - Aspect- Based Opinion Mining - Mining Comparative Opinions - Opinion Search and Retrieval - Case study	08	Text Book
IV	Google Analytics Introduction - Cookies - Accounts vs Property - Tracking Code -Tracking Unique Visitors - Demographics - Page Views and Bounce Rate Acquisitions - Custom Reporting - Case study	10	You Tube Videos
V	Applications Filters - Ecommerce Tracking - Real Time Reports - Customer Data-Alert - Adwords Linking - Adsense Linking - Attribution Modeling - Segmentation -Campaign Tracking - Multi-Channel Attribution - Case Study - Recommendation engines based on users, items and contents	10	You Tube Videos
Total		48	

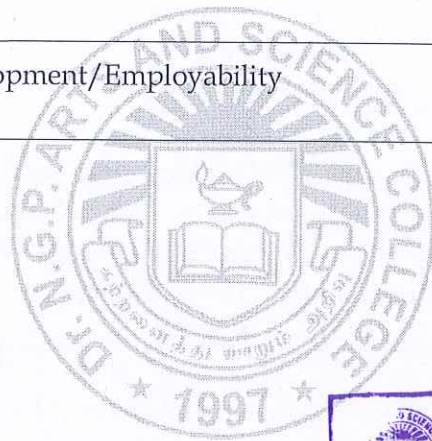
Text Book	1.	Bing Liu, 2011, "Web Data Mining Exploring Hyperlinks, Contents, and Usage Data", 2nd Edition, Springer (Unit I-III)
	2.	Ning Zhong, Jiming Liu and Yiyu Yao, 2010, "Web Intelligence", Springer (Unit IV, V)
Reference Books	1.	Ricardo Baeza -Yates and Berthier Ribeiro-Neto, 2011, "Information Retrieval: The Concepts and Technology behind Search", 2nd Edition, ACM Press

2.	Juan D. Velasquez, Lakhmi C. Jain (Eds.), 2010, "Advanced Techniques in Web Intelligence - 1", 1st Edition, Springer
3.	Mark Levene, 2010, "An Introduction to Search Engines and WebNavigation", 2nd Edition, Wiley
4.	Eric Fettman, Shiraz Asif, FerasAlhlou , 2016 "Google Analytics Breakthrough", Wiley

Journal and Magazines	https://www.emeraldgrouppublishing.com/journal/ijwis
E-Resources and Website	https://www.youtube.com/user/googleanalytics https://www.youtube.com/watch?v=CISLdoNKZMI

Learning Method	Chalk and Talk/ Assignment/Seminar/Brainstorming
-----------------	--

Focus of the Course	Skill Development/Employability
---------------------	---------------------------------



100-80se
 BoS Chairman/HoD
 Department of Computer Science with Data Analytics
 Dr. N. G. P. Arts and Science College
 Coimbatore – 641 048

 Dr.N.G.P. Arts and S. College		
APPROVED		
BoS- <i>14th</i>	AC -	GB -
<i>02.04.26</i>		