Syllabus of 4 + 1 Year Integrated UG and PG Programme

w. e. f 2024-25 Academic Year

GRADUATE SCHOOL

Mahatma Gandhi University

P. D. Hills P O

Kottayam, Kerala

www.gs.mgu.ac.in

www.mgu.ac.in

Schools offering Majors

SL.No	School/Centre
1	School of Bio Sciences
2	School of Chemical Sciences
3	School of Computer Sciences
4	School of Environmental Sciences
5	School of Gandhian Thought and Development Studies
6	School of International Relations and Politics
7	School of Pure and Applied Physics
8	School of Social Sciences

Sl. No.	Major	Intake
SCIENCE		
1	Bio Sciences	6**
2	Chemistry	6
3	Computer Science	6
4	Environmental Science	6
5	Physics	6
SOCIAL SCIENCES		
1	Development Studies	5
2	Gandhian Studies	5
3	History	10
4	International Relations and Politics	10

Majors offered and Intake *1 seat shall be sanctioned over and above the intake in each major in the 3rd semester for students who opt for a change of major after two semesters.

**Progression to PG Shall be based on the specialization selected by students as Biochemistry (2 seats) Biotechnology (2 seats) and Microbiology (2 seats) based on merit.

Schools offering Minors/MDCs/AECs/VACs/SECs

SL.No	School/Centre
1	School of Artificial Intelligence And Robotics
2	School of Behavioural Sciences
3	School of Biosciences
4	School of Chemical Sciences
5	School of Computer Sciences
6	School of Data Analytics
7	School of Energy Materials
8	School of Environmental Sciences
9	School of Food Science And Technology
10	School of Gandhian Thought And Development Studies
11	School of Gender Studies
12	School of Indian Legal Thought
13	School of International Relations And Politics
14	School of Letters
15	School of Mathematics And Statistics
16	School of Nanoscience And Nano Technology
17	School of Pedagogical Sciences
18	School of Polymer Science And Technology
19	School of Pure And Applied Physics
20	School of Social Sciences
21	School of Tourism Studies
22	International and Inter University Centre for Nanoscience and Nanotechnology
23	K N Raj School of Economics

Scheme for 4 + 1 Integrated UG and PG Programme
Graduate School
Mahatma Gandhi University
School of Data Analytics

Course Code	Title	Credits	Hours per Week		Level	Type
			Theory	Practical		
SEMESTER I						
	Major	4			Foundation (100-199)	Not offered in 2024-25
MG1DSCUDA121	Minor A (Data Analytics): SQL FOR DATA ANALYTICS	4	4		"	Minor
	Minor B	4	4		"	Minor
MG1MDCUDA101	MDC (Data Analytics 1): INTRODUCTION TO DATA SCIENCE AND ANALYTICS	3	3		"	MDC
	AEC (Eng)	3			"	
	AEC (Mal)	3			"	
SEMESTER II						
	Major	4			"	Not offered in 2024-25
MG2DSCUDA121	Minor A (Data Analytics): Data Mining and Data Warehousing	4	4		"	Minor
	Minor B	4	4		"	Minor
MG2MDCUDA101	MDC (Data Analytics 2): Introduction to Natural Language Processing	3	3		"	MDC

	AEC (Eng)	3			“	
	AEC (Mal)	3			“	
SEMESTER III						
	Major	4			Intermediate (200-299)	
	Major	4			“	
	Major	4			“	
MG3DSCUDA221	Minor A (Data Analytics) Elements of Big Data Analytics	4			“	Minor
	MDC	3			“	
	VAC	3			“	
SEMESTER IV						
	Major	4			“	
	Major	4			“	
	Major	4			“	
	Minor B	4			“	
	SEC	3			“	
	VAC	3			“	
	Internship/ Fieldwork	2				
SEMESTER V						
	Major	4			Higher (300-399)	
	Major	4			“	
	Major	4			“	
	Major	4			“	
	SEC	3			“	

	VAC	3			“	
SEMESTER VI						
	Major	4			“	
	Major	4			“	
	Major	4			“	
	Major (E)	4			“	
	Major (E)	4			“	
	SEC	3			“	
Total Credits		133				

SEMESTER VII						
	Major	4			Advanced (400-499)	
	Major (E)	4			“	
	Major (E)	4			“	
MG7DSCUDA421	Minor A/B Minor A (Data Analytics): ARTIFICIAL INTELLIGENCE	4			“	
	Minor A/B (E)	4			“	
	Minor A/B (E)	4			“	
SEMESTER VIII						
	Major	4			“	

	Major (E)	4			“	
	Research Project	12			“	
	Major*	4			“	
	Major*	4			“	
	Major*	4			“	
	Total Credits	44				
SEMESTER IX						
	Major	4			PG Level (500-599)	
	Major	4			“	
	Major	4			“	
	Major	4			“	
	Major	4			“	
SEMESTER X						
	Research Project	20			“	
	Major**	4			“	
	Major**	4			“	
	Major**	4			“	
	Major**	4			“	
	Major**	4			“	
Total Credits		40				

*Only for 4-Years Honours Students
**Only for students who opt for theory courses instead of Research Project

Note: General foundations courses shall be offered by different schools. Students can flexibly choose the courses across disciplines.

Level	Foundation (100-199	Intermediate (200-299)	Higher (300-399)	Advanced (400-499)	PG Level (500-599)

Type	Major	Minor	MDC	SEC	VAC	AEC

	MAHATMA GANDHI UNIVERSITY **Graduate School**
	4 + 1 Integrated UG and PG Programme

School	**SCHOOL OF DATA ANALYTICS**		
Programme	4 + 1 Integrated UG and PG Programme		
Course Title	**SQL FOR DATA ANALYTICS**		
Course Type	**Minor**		
Course Level	**100-199**		
Course Code	**MG1DSCUDA121**		
Course Overview	This course introduces the basics of relational databases and how to retrieve and manipulate data from one or more tables. The course also gives an overview of how to manipulate data with subqueries and aggregate functions and apply views and joins to manage database. The course helps the students to learn technologies like XML and NoSQL.		
Semester	**1**	**Credit**	**4**
Total Student Learning Time	**Instructional hours for theory**	**Instructional hours for practical/lab work/field work**	
	48	12	
Pre-requisite	PASS IN PLUS TWO/HIGHER SECONDARY		

COURSE OUTCOMES (CO)

CO No.	Expected Course Outcome	Learning Domains	PSO No.
	Upon completion of this course, students will be able to;		

1	Understand the basics of SQL and construct queries using SQL	E,R,U	1
2	Understand the relational database design principles and the basics of transaction processing.	C,S,U/ An	1,2
3	Understand database storage structures and access techniques	R,E	1,2
4	Understand different types of databases	U,S	1,2
5	Understand MongoDB and evaluate the NoSQL databases.	C,E,S	1,2,3

*(Learning Domains: Remember (R), Understand (U), Apply (A), Analyse (An), Evaluate (E) , Create (C), Skill (S))

COURSE CONTENT

Module 1	**Hours**	**CO No**
Data -Information- Databases- Introduction to File and Database systems- History of Databases-purpose of database systems- Types of Databases-RDBMS-Data Models-Keys -Introduction to SQL, Data types- **Data Definition Language Commands and Operations. Data Manipulation Language Commands and Operations- Data Control Language Commands- SQL Joins- Views- Triggers- Stored procedures-Functions in SQL-Group By and Order By-Sub queries in SQL.**	**20**	**1**
Module 2	**Hours**	**CO No**
Database Design –ER diagram –Database Design for Banking Enterprise –Functional Dependence –Normalization (1NF, 2NF, 3NF, BCNF, 4NF, 5NF).File Organization, types of file organization, SQL Transactions.	**20**	**2**
Module 3	**Hours**	**CO No**
Object-Oriented Databases-Distributed databases –characteristics, advantages, disadvantages, -Homogenous and Heterogeneous Distributed data Storage –XML –Structure of XML Data –XML Document. Introduction to Mongo DB , Overview of NoSQL.	**20**	**3**
Module 4	**Hours**	**Co No**
SQL Data Cleaning-Window Functions-Query Optimizations-Common Table Functions in SQL-accessing SQL from a Programming Language-	**20**	**4,5**

SQL Injection-SQL Projects for data analysis.		

Mode of Transaction	**Classroom Procedure (Mode of transaction)** Direct Instruction: Brain storming lecture, Explicit Teaching, E-learning, Interactive Instruction: Active co-operative learning, Seminar, Group Assignments Authentic learning: Library work and Group discussion, Presentation by individual student/Group representative.
Mode of Assessment	A. **Continuous Comprehensive Assessment (CCA)-40 marks** 1. Two Internal Examinations: 2*10 = 20 marks 2. Assignments, Seminars, Case Studies, Presentations: 10 marks 3. Practical (Data Base Design, Practical Record): 10 marks B. **End Semester Evaluation (ESE)** **Theory & Practice: 60 marks**

Learning Resources

1.Steve Tale, "SQL: The Ultimate *Beginners Guide: Learn SQL Today"*, Create Space Independent Publishing Platform, 2016

2. .Abraham Silberschatz, Henry F. Korth, S. Sudharshan, "*Database System Concepts*", 6th Edition, Tata McGraw Hill, 2011

3.Brad Dayley, "*Teach Yourself NoSQL with MongoDB in 24 Hours*", Sams Publishing, First Edition, 2014.

Relevance of Learning the Course/ Employability of the Course
• Design a simple database with DDL and DML commands. • Write sub queries and join operations for retrieving data from various tables. • Enforce the security features in multiuser database environment. • Use NoSQL database systems and manipulate the data associated with it.

	MAHATMA GANDHI UNIVERSITY **Graduate School**
	4 + 1 Integrated UG and PG Programme

School	**School of Data Analytics**		
Programme	**4 + 1 Integrated UG and PG Programme**		
Course Title	**INTRODUCTION TO DATA SCIENCE AND ANALYTICS**		
Course Type	**MDC**		
Course Level	**100-199**		
Course Code	**MG1MDCUDA101**		
Course Overview	This course provides a comprehensive introduction to the field of data science and analytics covering fundamental concepts, tools, techniques, and applications. Students will learn how to handle, analyse, and visualize data, as well as how to build predictive models and understand ethical considerations in data science.		
Semester	**1**	**Credit**	**3**
Total Student Learning Time	**Instructional hours for theory**	**Instructional hours for practical/lab work/field work**	
Pre-requisite	Pass in Plus Two/Higher Secondary Mathematics		

COURSE OUTCOMES (CO)

CO No.	Expected Course Outcome	Learning Domains	PSO No.
	Upon completion of this course, students will be able to;		

1	Understand and Apply Data Science Principles	R	1,4
2	Perform basic Data Preparation Tasks	U,A	1,2,3
3	Perform basic Statistical and Mathematical Analysis and Data Visualization	U,An	1,3,4
4	Develop and Evaluate Simple Machine Learning Models.	C,E	1,2,3
5	Understand Big Data Technologies and Data Ethics	R,U	1,3,6

*(Learning Domains: Remember (R), Understand (U), Apply (A), Analyse (An), Evaluate (E) , Create (C), Skill (S))

COURSE CONTENT

Module 1	**Hours**	**CO No**
Fundamentals of Data Science: Definition and importance of data science, Data science workflow and lifecycle, Applications and case studies in various industries. **Data Types and Sources:** Structured, semi-structured, and unstructured data, Data collection methods: surveys, web scraping, APIs, Introduction to data storage systems: relational databases, NoSQL databases. **Tools for Data Science:** Overview of Python and R programming languages, Introduction to data science libraries: NumPy, pandas, Matplotlib, Scikit-learn. Setting up a development environment using Jupyter Notebooks.	**15**	**1,2**
Module 2	**Hours**	
Data Cleaning: Handling missing values, Data transformation and normalization, removing duplicates and managing inconsistent data **Data Analytics and Visualization:** Descriptive statistics: mean, median, mode, variance, standard deviation. Data visualization principles and techniques-creating visualizations using Matplotlib and Seaborn. **Exploratory Data Analysis (EDA):** Identifying patterns and trends in data, Correlation, Outlier detection and treatment.	**15**	**2,3**
Module 3	**Hours**	
Basics of statistical and mathematical concepts in data science. **Introduction to Machine Learning**: Definition and types of machine learning. Basics of AI, Natural language Processing and Computer Vision. Supervised Learning Algorithms-Linear regression and logistic regression, Classification techniques, Model evaluation metrics: accuracy, precision, recall, F1 score	**15**	**3,4**
Module 4	**Hours**	
Unsupervised Learning Algorithms-Clustering techniques: k-means,	**15**	**4,5**

hierarchical clustering, Dimensionality reduction, Applications of unsupervised learning. **Introduction to Big Data and Cloud Computing**: Overview of big data technologies: Hadoop, Spark. Data processing in the cloud: AWS, Google Cloud, Azure**. Introduction to Data Ethics** -Privacy and Confidentiality, Bias and Fairness in Data Science, Accountability and Transparency, Ethical Data Use and Governance,The Five Cs.		

Mode of Transaction	**Classroom activities: Classroom activities:** Direct Instruction: Brainstorming lecture, E-learning, interactive Instruction, Seminar, Group Assignments, Authentic learning, Presentation by students group wise. **Field activities:** Case Studies and presentations **Lab based activities:** Data Analysis & Interpretation, Model Creation
Mode of Assessment	**A. Continuous Comprehensive Assessment (CCA): 40 marks** 1. Two Internal Examinations: 2*10 = 20 marks 2. Assignments, Seminars: 10 marks 3. Case Studies, Presentations: 10 marks **B. End Semester Evaluation (ESE): 60 marks**

Learning Resources

1. O'Neil,C.,& Schutt, R.(2013). Doing Data Science:Straight Talk from the Frontline .O'Reilly Media, Inc.
2. Davy Cielen, Arno D. B. Meysman, Mohamed Ali (2016) Introducing Data Science: Big Data, Machine Learning and More Using Python Tools .Manning Publications Co.
3. An Introduction to Statistical Learning :with Applications in R, Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, Springer, 1st edition,2013.
4. Deep Learning, Ian Goodfellow,Yoshua Bengio, Aaron Courville, MIT Press,1st edition,2016.
5. Ethics and Data Science, D. J. Patil, Hilary Mason, Mike Loukides,O'Reilly,1st edition,2018.
6. Python for Data Analysis-Wes McKinney.
7. Hands On Machine Learning with Scikit- Learn, Keras, and Tensor F low" by Aurélien Géron.

Relevance of Learning the Course/ Employability of the Course

1. Data science is used in finance, healthcare, e-commerce, marketing, and telecommunications to analyze customer behavior, predict maintenance needs, detect fraud, and personalize medical treatments.

2. Graduates with data science knowledge can pursue attractive professions as data analysts, data scientists, machine learning engineers, and business intelligence analysts, taking advantage of increasing demand and chances for professional advancement in data-intensive businesses.

3. The course integrates essential tools like Python, R, NumPy, pandas, and machine learning algorithms, equipping learners to manage large datasets, conduct advanced analyses, and build predictive models effectively.

	MAHATMA GANDHI UNIVERSITY **Graduate School**
	4 + 1 Integrated UG and PG Programme

School	**School of Data Analytics**		
Programme	4 + 1 Integrated UG and PG Programme		
Course Title	**Data Mining and Data Warehousing**		
Course Type	**Minor**		
Course Level	100-199		
Course Code	**MG2DSCUDA121**		
Course Overview	This course is an introductory course on data mining and data warehousing. It introduces the basic concepts, principles, methods, implementation techniques, and applications of data mining and data warehousing. This introductory data mining course helps to unravel the world of data mining.		
Semester	**2**	**Credit**	**4**
Total Student Learning Time	**Instructional hours for theory**	**Instructional hours for practical/lab work/field work**	
	48	12	
Pre-requisite	Basic understanding of algorithms, Database knowledge.		

COURSE OUTCOMES (CO)

CO No.	Expected Course Outcome	Learning Domains	PSO No.
	Upon completion of this course, students will be		

	able to;		
1	Understand the concepts of data warehouse, architecture, schema designs, OLAP operations and servers.	U, R, A	1,2,3
2	To know the Architecture of a Data Mining system.	U, R, An	1,2,3
3	To understand the various Data pre-processing methods.	U, A, S	1,2,3
4	To be familiar with important pattern discovery concepts.	U, C, S	1,2,3,4
5	To perform classification and clustering	U, A, C, S	1,2,3

*(Learning Domains: Remember (R), Understand (U), Apply (A), Analyse (An), Evaluate (E) , Create (C), Skill (S))

COURSE CONTENT

Module 1	**Hours**	**CO No**
Introduction to Data warehousing: Data warehousing Components - Data Warehouse Architecture – DBMS vs Data warehouse – Data Mart - Online Analytical Processing, Characteristics of OLAP - Difference between OLAP and OLTP – OLAP operations.	**15**	1
Module 2	**Hours**	
Data Mining: Introduction - Techniques, Issues and challenges, application - Functionalities - Knowledge representation - Various risks in Data Mining- Advantages and disadvantages of Data Mining - Ethical issues in Data Mining.	**15**	2
Module 3	**Hours**	
Data Pre-processing: Data Cleaning – Data Integration and Transformation – Data Reduction – Data Discretization and Concept Hierarchy Generation, Association Rule mining.	**15**	3,4
Module 4	**Hours**	
Introduction to classification and clustering – Advanced techniques: Web mining, Text mining, Spatial mining.	**15**	5

Mode of Transaction	**Classroom activities:** Direct Instruction: Brainstorming lecture, E-learning, interactive Instruction,

	Seminar, Group Assignments, Authentic learning, Presentation by students group wise. **Lab activities:** Problem Solving, Data Analysis & Interpretation **Field activities:** Case Studies and presentations Students have to submit a practical record on problems/case studies associated with the topics covered in various modules.
Mode of Assessment	**MODE OF ASSESSMENT** **A. Continuous Comprehensive Assessment (CCA)-40 marks** 1. Two Internal Examinations: 2*10 = 20 marks 2. Assignments, Seminars, Case Studies, Presentations : 10 marks 3. Practical (Database Mining and Warehousing Skills, Practical Record): 10 marks
	B. End Semester Evaluation (ESE) **Theory & Practice: 60 marks**

Learning Resources

1. Jiawei Han, Micheline Kamber and Jian Pei: *"Data Mining Concepts and Techniques"*, Third Edition, Elsevier, 2011.
2. Alex Berson and Stephen J. Smith: *"Data Warehousing, Data Mining & OLAP"*, Tata McGraw – Hill Edition, Tenth Reprint 2007.
3. Pang-Ning Tan, Michael Steinbach and Vipin Kumar: *"Introduction to Data Mining"*, Pearson Education, 2007.

Relevance of Learning the Course/ Employability of the Course
In the digital age, data volumes are exploding from web, mobile, sensors, IoT devices, etc. Mining value from data has become critical for success. Data mining powers every function - marketing, risk, operations, finance, etc. After this course, students are able to learn the technical know-how of Data mining principles and techniques for real-time applications.

	MAHATMA GANDHI UNIVERSITY **Graduate School**
	4 + 1 Integrated UG and PG Programme

School	**School of Data Analytics**		
Programme	**4 + 1 Integrated UG and PG Programme**		
Course Title	**Introduction to Natural Language Processing**		
Course Type	**MDC**		
Course Level	**100-199**		
Course Code	**MG2MDCUDA101**		
Course Overview	This course introduces students to the fundamental concepts, techniques, and applications of NLP, providing them with the knowledge and skills necessary to understand, process, and analyse natural language data.		
Semester	2	**Credit**	3
Total Student Learning Time	**Instructional hours for theory**	**Instructional hours for practical/lab work/field work**	
	40	**20**	
Pre-requisite	PASS IN PLUS TWO/HIGHER SECONDARY		

COURSE OUTCOMES (CO)

CO No.	Expected Course Outcome	Learning Domains	PSO No.
	Upon completion of this course, students will be able to;		
1	Familiarize in essential NLP tasks like text pre-processing and	R,U	1,2

	statistical analysis.		
2	Implement basic algorithms for text classification, named entity recognition, sentiment analysis, and language modelling.	A	1,2,3
3	Acquire skills in sequence labelling, word embeddings, and basics deep learning for NLP tasks.	A,S	1,4
4	Analyze ethical issues in NLP, including biases and privacy concerns.	An,E	2,5,6
5	Showcase a real-world NLP project that demonstrates how to apply NLP technology responsibly and effectively.	A,C	2,4

*(Learning Domains: Remember (R), Understand (U), Apply (A), Analyse (An), Evaluate (E) , Create (C), Skill (S))

COURSE CONTENT

Module 1	**Hours**	**CO No**
Definition of NLP, its applications, and importance**.** Overview of key tasks in NLP, Text Pre-processing, Basics of probability and statistics in NLP, TF-IDF. Introduction to Python programming language, Libraries for NLP.	**20**	**1,2**
Module 2	**Hours**	
Syntax and Semantics-Part-of-speech tagging, Parsing techniques. Word Embeddings-Introduction to word vectors, Word2Vec and GloVe embeddings. Text Classification-Supervised learning approaches, building a text classifier using machine learning algorithms. Named Entity Recognition.	**20**	**2,3**
Module 3	**Hours**	
Sequence Labeling - Language Models and Text Generation-language modelling, Topic Modelling, Text Summarization, Sentiment Analysis Techniques, Challenges.	**20**	**2,3,5**

Module 4	Hours	
NLP Applications- Information retrieval and search engines, Chatbots and conversational agents. Ethical Considerations in NLP, Deep learning in NLP, Multilingual and cross-lingual NLP, Practical implementation of an NLP project.	20	2,3,4,5

Mode of Transaction	**Classroom activities:** Direct Instruction: Brain storming lecture, Explicit Teaching, E-learning, interactive Instruction: Active co-operative learning, Seminar, Group Assignments Authentic learning, Library work and Group discussion, Presentation by individual student/ Group representative. **Field activities**: Case Studies **Lab based activities:** Problem Solving, Code Generation
Mode of Assessment	**A. Continuous Comprehensive Assessment (CCA)-40 marks** 1. Two Internal Examinations: 2*10 = 20 marks 2. Assignments, Seminars, Case Studies, Presentations: 10 marks 3. Practical (Text Analytics, Practical Record): 10 marks **B. End Semester Evaluation (ESE)** **Theory & Practice: 60 marks**

Learning Resources

1. Dipanjan Sarkar (2016): *Text Analytics with Python*, Apress / Springer,
2. Bird, Steven, Ewan Klein, and Edward Loper (2009): *Natural Language Processing,* Oreilly Media Inc.,
3. Daniel Jurafsky and James H. Martin(2024). Speech and Language Processing,
4. Steven Bird, Ewan Klein, and Edward Loper(2009). *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit,* Oreilly Media Inc.
5. Christopher D. Manning and Hinrich Schütze (1999): *Foundations of Statistical Natural Language Processing,* MIT Press
6. Jacob Eisenstein (2019): *Natural Language Processing: A Concise Introduction*, MIT Press

Relevance of Learning the Course/ Employability of the Course
1. High Demand in Industry, Business 2. NLP is essential to AI applications that drive productivity and innovation in organizations,

such as chatbots, virtual assistants, sentiment analysis tools, and automated content creation.

3. Graduates with NLP knowledge can work as NLP engineers, data scientists specialized in text analysis, AI researchers, or consultants for firms implementing AI-driven solutions.
4. Interactions between NLP and domains such as data science, machine learning, and computational linguistics allow for multidisciplinary job paths and cooperative opportunities.

	MAHATMA GANDHI UNIVERSITY **Graduate School**
	4 + 1 Integrated UG and PG Programme

School	**School of Data Analytics**		
Programme	**4 + 1 Integrated UG and PG Programme**		
Course Title	**Elements of Big Data Analytics**		
Course Type	**Minor**		
Course Level	**200-299**		
Course Code	**MG3DSCUDA221**		
Course Overview	The course enables the students to understand Big Data processing used in different business intelligence applications and provide an in-depth coverage of MapReduce analytics using Hadoop Eco system tools. The student will gain knowledge in Pig, Hive, Spark and they will get exposure in blooming Big Data technologies.		
Semester	**3**	**Credit**	**4**
Total Student Learning Time	**Instructional hours for theory**	**Instructional hours for practical/lab work/field work**	
	48	12	
Pre-requisite	Knowledge in basics of SQL (queries and sub queries), exposure to Linux Environment.		

COURSE OUTCOMES (CO)

CO No.	Expected Course Outcome	Learning Domains	PSO No.

	Upon completion of this course, students will be able to;		
1	Provide an overview of Apache Hadoop and HDFS	U, An	1,2
2	Understand the Big Data Platform and its Use cases	U, R, A	1,3,4
3	Familiarize with MapReduce	U, S	1,2,3
4	Demonstrate Hive and Pig.	U, R, An	1,2,3
5	Demonstrate Spark programming.	U, R, An	1,2

*(Learning Domains: Remember (R), Understand (U), Apply (A), Analyse (An), Evaluate (E) , Create (C), Skill (S))

COURSE CONTENT

Module 1	**Hours**	**CO No**
Types of Digital Data, Introduction to Big Data, Big Data Analytics, Introduction to Big Data – Introduction to Data Analytics, Type of Data Analytics – Descriptive, Predicative, Prescriptive, definition & importance of Big Data - four dimensions of Big Data - volume, velocity, variety, veracity – industry examples – terminologies - NoSQL.		2
Module 2	**Hours**	
Hadoop – Requirement of Hadoop Framework - Hadoop Ecosystem - Design principle of Hadoop –Comparison with other system - Hadoop Components – Hadoop Daemon's – HDFS - HDFS Architecture.		1
Module 3	**Hours**	
Anatomy of a Map Reduce Job Run, Failures, Job Scheduling, Shuffle and Sort, Task Execution, Map Reduce Types and Formats, Map Reduce Features.		3
Module 4	**Hours**	
Introduction to Hive – Features, Architecture, Pig – Features, Pig Latin, Spark – features.		4,5

Mode of Transaction	**Classroom activities:** Direct Instruction: Brainstorming lecture, E-learning, interactive Instruction, Seminar, Group Assignments, Authentic learning, Presentation by students group wise.

	Lab activities: Problem Solving, Data Analysis & Interpretation **Field activities:** Case Studies and presentations Students have to submit a practical record related to problems/case studies on topics covered in various modules.
Mode of Assessment	**MODE OF ASSESSMENT** **A. Continuous Comprehensive Assessment (CCA)-40 marks** 1. Two Internal Examinations: 2*10 = 20 marks 2. Assignments, Seminars, Case Studies, Presentations : 10 marks 3. Practical (Big Data Analytics Skills, Practical Record): 10 marks
	B. End Semester Evaluation (ESE) **Theory & Practice: 60 marks**

Learning Resources

1. Tom White " *Hadoop: The Definitive Guide*" Third Edit on, O'reily Media, 2012.
2. Seema Acharya, Subhasini Chellappan: "*Big Data Analytics*" Wiley 2015.

3. Michael Berthold, David J. Hand: "*Intelligent Data Analysis*", Springer, 2007.

Relevance of Learning the Course/ Employability of the Course
In line with the trends of Big Data in general, the request for skilled Big Data professionals is growing rapidly. Studying Big Data is a rewarding and (at times) fun investment of your time.

	MAHATMA GANDHI UNIVERSITY **Graduate School**
	4 + 1 Integrated UG and PG Programme

School	**SCHOOL OF DATA ANALYTICS**		
Programme	**DATA ANALYTICS**		
Course Title	**ARTIFICIAL INTELLIGENCE**		
Course Type	**Minor**		
Course Level	**200-299**		
Course Code	**MG4DSCUDA221**		
Course Overview	• To equip students with fundamental knowledge of thinking and intelligence in ways that enable the construction of computer systems that are able to reason in uncertain environments • To analyze different types of data representation. • To compare different reasoning methods and Bayesian networks • To explain the principles, components, operations and other technological advancements in machine learning, deep learning, and Natural language processing.		
Semester	4	**Credit**	4
Total Student Learning Time	**Instructional hours for theory**	**Instructional hours for practical/lab work/field work**	
	65	15	

COURSE OUTCOMES (CO)

CO No.	Expected Course Outcome	Learning Domains	PSO No.
	Upon completion of this course, students will be able to;		

1	On completing this course, the student will be able to articulate and exemplify the basic knowledge of artificial intelligence, machine learning , deep learning and natural language processing	E,R,U	1
2	Understand the basics of knowledge representation.	C,S,U/ An	1,2
3	can use AI programming languages and the methods of AI implementation	R,E	1,2
4	Understand the basics of reasoning techniques .	U,S	1,2
5	Understand machine learning, deep learning and text analytics	C,E,S	1,2,3

*(Learning Domains: Remember (R), Understand (U), Apply (A), Analyse (An), Evaluate (E) , Create (C), Skill (S))

COURSE CONTENT

Module 1	**Hours**	**CO No**
Artificial Intelligence –History of AI, Branches of Artificial Intelligence, Applications of Artificial Intelligence, AI Programming Languages**.** Defining the Problem as a State Space Search, Problem Characteristics, Production Systems, Problem Solving by Search: Uninformed and Informed Search Strategies, BFS, DFS; Heuristic Search Techniques: Generate-And-Test, Hill Climbing, Best-First Search, A*Algorithm, Problem Reduction, AO*Algorithm, Constraint Satisfaction, Means-Ends Analysis.	**20**	**1**
Module 2	**Hours**	**CO No**
Knowledge Representation: Procedural Vs Declarative Knowledge, Representations & Approaches to Knowledge Representation, Types of Knowledge, First Order logic, Frames, Conceptual Dependency, Scripts, Semantic network, Matching Techniques, Partial Matching, Fuzzy Matching Algorithms.	**20**	**2**
Module 3	**Hours**	**CO No**
Reasoning - Types of reasoning, Non-monotonic reasoning, Bayes' Theorem, Bayesian networks, Dempster -Shafer Theory, Fuzzy Logic: Crisp Sets ,Fuzzy Sets, Fuzzy Logic Control ,Experts Systems, Different Types of Expert Systems, Case studies in expert systems, CYC	**20**	**3**
Module 4	**Hours**	**Co No**

Machine Learning, Types of machine learning, machine learning algorithms, applications of machine learning,Deep learning Fundamentals – Common Neural Networks architectures,Types of neural networks, Overview of deep learning libraries, applications of deep learning, Introduction to Natural Language Processing, **NLP Libraries and Tools**, Processing and Understanding Text: Text Tokenization, **Word Embeddings**, Named entity recognition	**20**	**4,5**

Mode of Transaction	**Classroom Procedure (Mode of transaction)** Direct Instruction: Brain storming lecture, Explicit Teaching, E-learning, Interactive Instruction: Active co-operative learning, Seminar, Group Assignments Authentic learning: Library work and Group discussion, Presentation by individual student / Group representative.
Mode of Assessment	**MODE OF ASSESSMENT** **C. Continuous Comprehensive Assessment (CCA)-40 marks** 2. Two Internal Examinations: 2*10 = 20 marks 2. Assignments, Seminars, Case Studies, Presentations : 10 marks 3. Practical (AI/ML Skills, Practical Record): 10 marks
	D. End Semester Evaluation (ESE) **Theory & Practice: 60 marks**

Learning Resources

1. Elaine Rich and Knight (2017) *Artificial Intelligence,* Mc Graw-Hill Publications
2. Patterson, D.W.(2005) *Introduction to Artificial Intelligence & Expert Systems,* Prentice Hall of India

3) Aurélien Géron's, *"Hands-On Machine Learning with Scikit- Learn and Tensor Flow",* O'Reilly Media, Inc.,2017.

4) Dipanjan Sarkar, *Text Analytics with Python,* Apress /Springer, 2016

Relevance of Learning the Course/ Employability of the Course

- **Technological Literacy**
- **Innovation and Creativity**.
- **Global Competitiveness**
- Research and development
- High demand for those proficient in AI & ML.