MAHATMA GANDHI UNIVERSITY
Priyadarshini Hills P.O.
Kottayam–686560

SCHOOL OF DATA ANALYTICS
M. Sc. Data Science & Analytics

FACULTY OF SCIENCE

CURRICULUM AND SYLLABUS

UNDER CSS REGULATIONS 2020

(Revised According to OBE Scheme and 2021 Modifications)

EFFECTIVE FROM 2021-22 ACADEMIC YEAR ONWARDS

(EFFECTIVE FROM 2021 ADMISSIONS)

MAY 2021

Preface

Mahatma Gandhi University

Mahatma Gandhi University, Kottayam is an Indian collegiate public University based in Kerala, established in 1983, approved by UGC, and accredited by NAAC with "A" Grade, and CGPA 3.24 out of 4.00 . With its academic excellence, the University has bagged Chancellor's Award twice for the Best University (2015-16 and 2017-18) in Kerala. It has also secured 30th position in NIRF ranking (April 2019) and 11th position in India Today-MDRA ranking, 2018. CSIR has ranked the University 13th for its intellectual productivity and NISTADS has rated it as 19th in terms of h-index. It has been included among 401-500 band in Times Higher Education (THE) World University Rankings list.

At present, Mahatma Gandhi University offers research programs in forty disciplines through its own Schools and approved Research Centers. It has close collaboration for academic, research and extension programs with a number of national agencies and institutions including the UGC, DST-FIST, DRS, ISRO, COSIT, DIT, DST (Nano Mission), CSIR, DAAD, STEC, ICMR, BARC and MOEF. The University is also involved in active collaboration with research institutions of international reputation such as the Max Planck Institute of Technology, Germany; Brown University, USA; University of Nantes, France; California Institute of Technology, USA; University of Toronto, Canada; Catholic University, Belgium; Heidelberg University, Germany; the Institute of Political Studies, Rennes, France; Trent University, Canada; IPF Dresden, Germany; University of Paris and University of Strasbourg.

Mahatma Gandhi University has made immense strides in the fields of inter disciplinary teaching and research. The faculty comprises of outstanding scholars, many of whom have made original contributions in their respective fields of specialization. The faculty andresearch scholars of several departments have gained widespread recognition for the commendable quality of their research publications. The web enabled University library has large collection of books, journals, e-journals and online theses. The digital library provides open access to its enviable collection of digitized Ph.D dissertations. All these work in tandem with the academic business transacted by the University, making the whole experience a holistic one. The University has a well established instrumentation facility with many sophisticated equipments functioning at the various departments and also at the platform provided by the common Inter University Instrumentation Centre (IUIC).

The University has well established and internationally reputed facility and academic expertise in various areas like Nano Science, Environmental Science, Bioscience, Chemical Science, Physics, Arts and Humanities. The Centre for Nano Science and Nanotechnology focus on the enhancement of research and higher studies in the cutting edge areas of Nano Science and Nanotechnology. The Centre is motivated to thrust its research and development focusing on developing novel materials and devices prospering the outrage of Nano Science. With a vision to focus attention to research and interdisciplinary studies in the frontier areas of Environmental Science, the University has established the School of Environmental Sciences as a Centre of learning for advanced studies in different branches of environmental science. The major mandate of the School is to develop appropriate technologies and skilled human resource for sustainable utilization, management and conservation of natural resources. The school has established a Centralized Remote Sensing and GIS facility, the first of its kind in a University in the state, with the support of Indian Space Research Organization (ISRO). It has also established a Regional Center, the High range Environmental Research center (HERC) at Nedumkandam, Idukki district. The School has a live laboratory named as "Jeevaka" which consists of areas with rich biodiversity within the Mahatma Gandhi University Campus.

The University has recently established the School of Mathematics and Statistics as well as School of Data Analytics in 2019. A few more Schools namely School of Artificial Intelligence and Robotics; School of Food Science and Technology as well as School of Energy Sciences were established in 2019. Now the University is marching forward with various ambitious projects for fostering academics and industry collaborations through innovation and incubation centres as well as translational research centres.

Vision and Mission of MG University

Vision of Mahatma Gandhi University

"Mahatma Gandhi University envisions to excel in the field of higher education and cater to the scholastic and developmental needs of the individual, through continuous creation of critical knowledge base for the society's sustained and inclusive growth."

Mission of Mahatma Gandhi University

- **To conduct and support undergraduate, postgraduate and research-level programmes of quality in different disciplines**
- **To foster teaching, research and extension activities for the creation of new knowledge for the development of society**
- **To help in the creation and development of manpower that would provide intellectual leadership to the community**
- **To provide skilled manpower to the professional, industrial and service sectors in the country so as to meet global demands**
- **To help promote the cultural heritage of the nation and preserve the environmental sustainability and quality of life**
- **To cater to the holistic development of the region through academic leadership**

Preamble

OUTCOME BASED EDUCATION (OBE)FROM THE ACADEMIC YEAR 2020-21
MAHATMA GANDHI UNIVERSITY SCHOOL OF DATA ANALYTICS

Introduction

A high priority task in the context of education in India is improvement of quality of higher education for equipping young people with skills relevant for global and national standards and enhancing the opportunities for social mobility. Mahatma Gandhi University has initiated an Outcome Based Education (OBE) for enhancing employability of graduates through curriculum reforms based on a learning outcomes-based curriculum framework, upgrading academic resources and learning environment.

Learning outcomes specify what graduates completing a particular programme of study are expected to know, understand and be able to do at the end of their programme of study. The fundamental premise underlying the learning outcomes-based approach to curriculum development is that higher education qualifications are awarded on the basis of demonstrated achievement of outcomes, expressed in terms of knowledge, understanding, skills, attitudes and values. Outcomes provide the basis for an effective interaction among the various stakeholders. It is the results-oriented thinking and is the opposite of input-based education

where the emphasis is on the educational process.

OBE Framework

The OBE Framework is a paradigm shift from traditional education system into OBE system where there is greater focus on programme and course outcomes. It guarantees that curriculum, teaching and learning strategies and assessment tools are continuously enhancedthrough a continuous improvement process. All decisions including those related to curriculum,delivery of instruction and assessment are based on the best way to achieve the predetermined outcomes. Traditionally, educators have measured learning in terms of standardized tests. In contrast, outcome-based education defines learning as what studentscan demonstrate that they know.

Benefits of OBE:

*More directed & coherent curriculum.

*Graduates will be more "relevant" to industry & other stakeholders

*Continuous Quality Improvement is in place.

*OBE shifts from measuring input and process to include measuring the output (outcome)

Outcome Based Education (OBE) process

OBE is a comprehensive approach to organise and operate a curriculum that is focused onand defined by the successful demonstrations of learning sought from each learner. The term clearly means focusing and organising everything in an education system around "what is essential for all learners to be able to do successfully at the end of their learning experiences".

OBE is an approach to education in which decisions about the curriculum and instruction are driven by the exit learning outcomes that the students should display at the end of aprogramme or a course. By the end of educational experience, each student should have achievedthe outcomes.

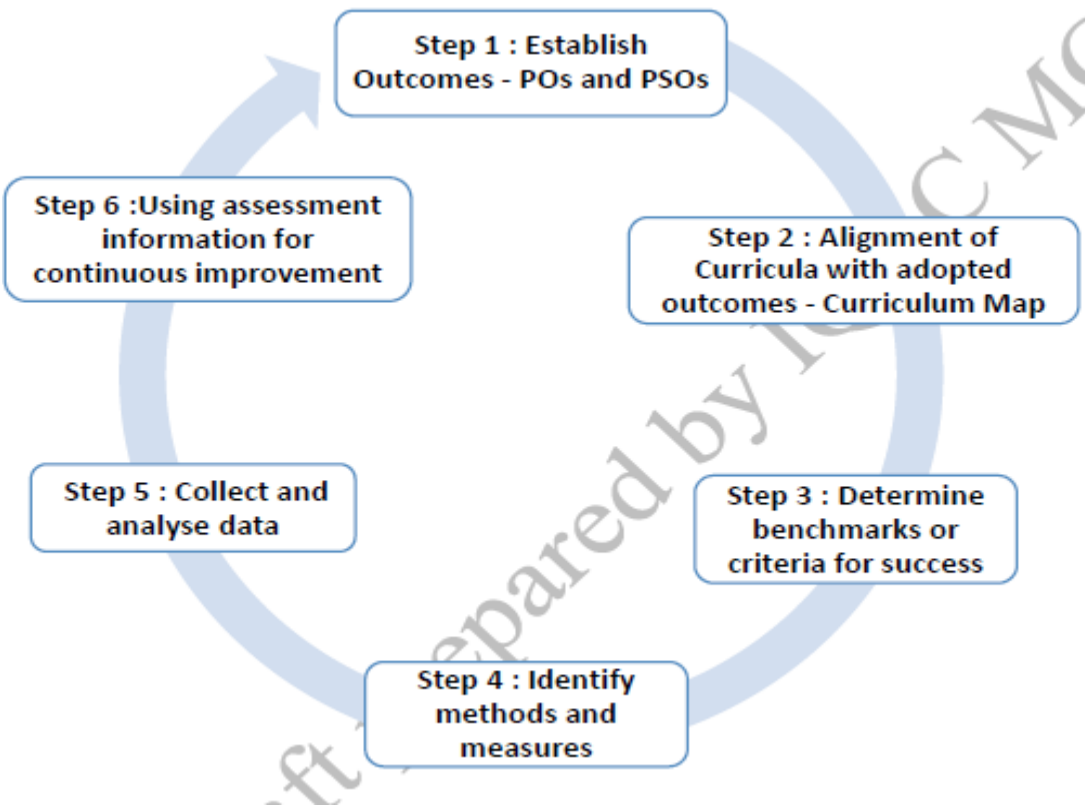
Learning Outcomes based Curriculum Framework (LOCF) for Post Graduate Programmes - IQAC MG University

One of the main objectives of OBE is to ensure continuous improvement of programmes in terms of maintaining the relevance in curriculum as well as responding to the requirements of the stakeholders. In other words, it ensures that Post graduate programme next year is better than Post graduate programme this year, offered by a department.

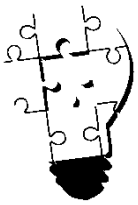
An OBE system has been proposed and to be implemented from 2021-22 at various

Departments of Mahatma Gandhi University, as a quality-assurance approach to improve teaching and learning outcomes and processes. This OBE plan incorporates the "outcomes assessment" process to be followed in the departments. OBE should be a key driver of the curriculum management in allthe departments of the university.

The OBE is a 6 step pr**oces**s as shown in the figure below. The process is presented as a cycle or a loop. The cycle represents the continuous nature of assessing learning outcomes.. As envisaged by the IQAC of Mahatma Gandhi university, an OBE based curricular framework has been proposed for the School of Data Analytics from the academic year 2020-2021 which is presented hereafter.



	Mahatma Gandhi University **Graduate attributes**

	**Critical thinking and analytical reasoning**	Capability to analyze, evaluate and interpret evidence, arguments, claims, beliefs on the basis of empirical evidence; reflect relevant implications to the reality; formulate logical arguments; critically evaluate practices, policies and theories to develop knowledge and understanding; able to envisage the reflective thought to the implication on the society.
	**Scientific reasoning and Problem solving**	Ability to analyze, discuss, interpret and draw conclusions from quantitative/qualitative data and experimental evidences; and critically evaluate ideas, evidence and experiences from an unprejudiced and reasoned perspective; capacity to extrapolate from what one has learned and apply their competencies to solve problems and contextualize into researchand apply one's learning to real life situations.
	**Multidisciplinary/ Interdisciplinary/ Transdisciplinary approach**	Acquire interdisciplinary /multidisciplinary/ transdisciplinary knowledge base as a consequence of the learning they engage with their programme of study; develop a collaborative-multidisciplinary/interdisciplinary/transdisciplinary-approach for formulate constructive arguments and rational analysis for achieving common goals and objectives.
	**Intra and Interpersonal skills**	Ability to work effectively and respectfully with diverse teams;facilitate collaborative and coordinated effort on the part of a group,and act together as a group or a team in the interests of a common cause and work efficiently as a member of a team; lead the team to guide people to the right destination, in a smooth and efficient way.
	**Digital literacy**	Capability to use ICT in a variety of learning situations, demonstrate ability to access, choose, collect and evaluate, and use a variety of relevant information sources; structure and evaluate those data for decision making.

	**Global Citizenship**	Building a sense of belonging to a common humanity and to become responsible and active global citizens. Appreciation and adaptation of different sociocultural setting and embrace and promote equity.
	**Social competency**	Possess knowledge of the values and beliefs of multiple cultures, appreciate and adapt to a global perspective; and capability to effectively engage in a multicultural society and interact respectfully, manage and lead with diverse groups.
	Equity, Inclusiveness and Sustainability	Appreciate and embrace equity, inclusiveness and sustainability and diversity; acquire ethical and moral reasoning and values of unity, secularism and national integration to enable to act as dignified citizens; able to understand and appreciate diversity
	**Lifelonglearning**	Continuous acquisition of knowledge and skills. Learn, unlearn and re-learn based on changing ecosystem. “Learning how to learn”, that are necessary for participating in learning activities throughout life, through self-paced and self-directed learning aimed at personal development, meeting economic, social and cultural objectives, and adapting to changing trades and demands of work place through knowledge/skill development/reskilling.

	Mahatma Gandhi University **Post Graduate Programme Outcome**

Programme Outcomes (PO)

PO 1: Critical Thinking and Analytical Reasoning
Capability to analyse, evaluate and interpret evidence, arguments, claims, beliefs on the basis of empirical evidence; reflect relevant implications to the reality; formulate logical arguments; critically evaluate practices, policies and theories to develop knowledge and understanding; able to envisage the reflective thought to the implication on the society.

PO 2 : Scientific Reasoning and Problem Solving
Ability to analyse, discuss, interpret and draw conclusions from quantitative/qualitative data and experimental evidences; and critically evaluate ideas, evidence and experiences from an unprejudiced and reasoned perspective; capacity to extrapolate from what one has learned and apply their competencies to solve problems and contextualise into researchand apply one's learning to real life situations.

PO 3: Multidisciplinary/Interdisciplinary/Transdisciplinary Approach
Acquire interdisciplinary /multidisciplinary/transdisciplinary knowledge base as a consequence of the learning they engage with their programme of study; develop a collaborative-multidisciplinary/interdisciplinary/transdisciplinary-approach for formulate constructive arguments and rational analysis for achieving common goals and objectives.

PO 4: Communication Skills
Ability to reflect and express thoughts and ideas effectively in verbal and nonverbal way; Communicate with others using appropriate channel; confidently share one's views and express herself/himself; demonstrate the ability to listen carefully, read and write analytically, and present complex information in a clear and concise manner and articulate in a specific context of communication.

PO 5: Leadership Skills
Ability to work effectively and lead respectfully with diverse teams; setting direction, formulating goal, building a team who can help achieve the goal, motivating and inspiring team members to engage with that goal, and using management skills to guide people to the right destination, in a smooth and efficient way.

PO 6: Social Consciousness and Responsibility
Ability to contemplate of the impact of research findings on conventional practices, and a clear understanding of responsibility towards societal needs and reaching the targets for attaining inclusive and sustainable development.

PO 7: Equity, Inclusiveness and Sustainability
Appreciate equity, inclusiveness and sustainability and diversity; acquire ethical and moral

reasoning and values of unity, secularism and national integration to enable to act as dignified citizens; able to understand and appreciate diversity, managing diversity and use of an inclusive approach to the extent possible.

PO 8: Moral and Ethical Reasoning
Ability to embrace moral/ethical values in conducting one's life, formulate a position/argument about an ethical issue from multiple perspectives, and use ethical practices in all work. Capable of demonstrating the ability to identify ethical issues related to one's work and living as a dignified person in the society.

PO 9: Networking and Collaboration
Acquire skills to be able to collaborate and network with scholars in an educational institutions, professional organizations, research organizations and individuals in India and abroad.

PO 10: Lifelong Learning
Ability to acquire knowledge and skills, including "learning how to learn", that are necessary for participating in learning activities throughout life, through self-paced and self-directed learning aimed atpersonal development, meeting economic, social and cultural objectives, and adapting to changing trades and demands of work place through knowledge/skill development/reskilling.

School of Data Analytics

In the emerging data driven world, Data Science and Data Analytics are receiving much acceptance in the national and international levels for making wise decisions. Data analytics is an essential field that brings together Data, technology, information, statistical analysis all in one platform. Every organization in private/ public sector creates a large volume of data from almost every area of daily activity. Analyzing such data has huge potential to predict the future of the organization. A good amount of knowledge and skill are very necessary in the field of data management, machine learning, natural language processing as they are the key factors in Data Science. The decision to establish a School of Data Analytics along with the School of Mathematics & Statistics in the Mahatma Gandhi University, Kottayam during 2020-21 is a most appropriate decision and will help the research and career opportunities in this emerging field for prospective students..

In the proposed curriculum for M.Sc. Data Science and Analytics, apart from teaching core subjects, the students are also offered training in programming languages and open softwares such as R, Python, etc to handle real life problems through the practical classes and case studies as well as industry visits.. As part of the program the students are also given training in computing softwares like SPSS, SAS etc. The program is so designed that on successful completion, the students would be able to get attractive jobs in the industry sector as data scientists, data analysts, data management specialists etc as well as pursue research or higher studies in the areas of Data Science, Data Analytics, Statistics, Mathematics, Computer Science, Economics, Management and allied fields. Moreover, emerging areas like Bayesian Inference and Computing, Stochastic Modeling, Time Series Forecasting, Internet of Things and Fraud Analytics are included in the curriculum. There will be an Open Course during third semester to encourage interdisciplinary studies and research. This can be selected from among courses offered by any other school/department/centre in the university. The following are some of the emerging areas of applications in this respect.

(i) Internet of Things (IoT): Analytics tools and techniques for dealing with the massive amounts of structured and unstructured data generated by IoT will continue to gain importance.

(ii) Finance and Banking: Creating newer business models or frameworks that leverages the available data to facilitate financial institutions to monetize data to deliver superior customer values.

(iii) Data Security: Fraud Security Analytics are already transforming intrusion detection, differential privacy, digital watermarking and malware countermeasures.

(iv) Health Care: Health Care Analytics and Predictive Analytics enable the examination of patterns in various healthcare data in order to determine how clinical care can be improved while limiting excessive spending.

There has been much interest in Bayesian Methods in Data Analytics in recent years. It is a way to get sharper predictions from the data, particularly when there is not much data available and when one want to utilize all information from it. During the last three decades, Actuarial Science has gone through revolutionary changes due to the implementation of high speed computers and modern theory. It applies mathematical and statistical methods to assess risk in insurance, finance and other industries. Official Statistics make information on economic and social development accessible to the public, allowing the impact of government polices to be assessed and thus improving accountability.

With the emergence of new diseases like AIDS, SARS, COVID 19 etc. more interest has come to analyze such epidemic data to study the dynamics of its spread. In the modern data driven world Data Science and Data Analytics are the principal tools for making wise decisions and drawing valid conclusions. We expect students to select the elective courses to ensure employment and research opportunities in emerging areas. The main thrust of the program will be to enable students to understand basic concepts and applications through real data sets using computer programs like R and Python in the Data Analytics lab rather than just imparting theory. Industry interaction, case studies and projects will be main components of all courses.

.

Our Vision

To realize the importance and power of quality data, achieve excellent standards of quality education using latest tools, nurturing collaborative culture and disseminating customer-oriented innovations to relevant areas of academia and industry towards serving the greater cause of society in a data driven world full of challenges.

Key points

- To become an Institution of Excellence in Data Science and Analytics
- To impart Professional competence, ambition and determination
- To develop competency for new challenges and new opportunities

- To enable well being and prosperity of nation and humanity
- To develop innovation, originality, creativity and positive attitude
- To utilize opportunities in research with global standards
- To inculcate leadership qualities and personality development

Our Mission

To provide advanced training in Data Science and Analytics principles and methodologies, and strive to develop highly trained, qualified data analyst and data scientist who can go on to pursue careers in industry, government or research with ethical and human values.

Key points

1. Provide advanced knowledge and technological knowhow
2. To utilize best quality data and the expertise of faculty.
3. Benefitting the students in achieving their career goals.
4. Pursue cutting-edge research with innovation for applications
5. Extend the knowledge gained from lab to land and bench top to bedside

Programme Outcomes of M.Sc. Program (PO)

1. Having a clear understanding of the subject related concepts and contemporary issues.
2. Having problem-solving ability- to assess social issues (societal, health, safety, legal and cultural) and engineering problems.
3. Having a clear understanding of professional and ethical responsibilities.
4. Having cross-cultural competency exhibited by working as a member or in teams.
5. Having an ability to be socially intelligent with good SIQ (Social Intelligence Quotient) and EQ (Emotional Quotient)
6. Having computational thinking (Ability to translate vast data into abstract concepts and to understand database reasoning
7. Having ability to make wise decisions using data analytics techniques
8. Having Virtual Collaborating ability and dynamic leadership qualities and team spirit
9. Having critical thinking and innovative skills with a good digital footprint
10. Having global standards in a competitive data driven world.

Programme Specific Outcomes (PSO) of M. Sc. Data Science & Analytics

1.PSO1: After undergoing this program, students will get advanced knowledge in theory and applications in all areas of Data Science, Data Analytics, Statistical Learning, Machine Learning, Data Base Management, Artificial Intelligence, etc.

2.PSO2: Students have secured practical skills in statistical methods and computer programming to plan and execute projects and decision making using Data Science, Data Analytics, Machine Learning, Visualization Techniques etc

3.PSO3: Students are well equipped to undertake any work involving exploratory data analysis, fraud analytics, deep learning, text mining etc. as future entrepreneurs.

4.PSO4: Students have developed skills in advanced computing softwares like R and Python for big data analytics, computing and interpretation of results.

5.PSO5: Students are well trained to take up jobs in reputed industries, business firms and MNCs etc as Data Scientists, Data Analysts, Data Engineers, Risk Analysts, Business Analysts, Financial Analysts, Decision Makers, Entrepreneurs etc.

6.PSO6: Students are equipped with practical knowledge and on the job training through case studies, assignments, presentations, industrial visits, internships, projects etc

7.PSO6: Students are motivated to pursue teaching and research in all emerging areas of research in theoretical and applied branches of Data Science, Data Analytics and related areas.

8.PSO7: To become expert decision makers utilizing big data in the data driven world

The Program Structure: M. Sc. Data Science & Analytics

Table of Courses and Credits

Course Code	Course Title	Teaching hours per week L+T+P	Credits
SEMESTER I	**Total Credits 21**		
DA M 21 C01	INTRODUCTION TO DATA SCIENCE & DATA ANALYTICS	3 +0+2	3
DA M 21 C02	LINEAR ALGEBRA AND MATRICES	3 +0+2	3
DA M 21 C03	PROBABILITY AND DISTRIBUTION THEORY	3 +0+2	3
DA M 21 C04	DATA BASE MANAGEMENT SYSTEMS	3+0+2	3
DA M 21 C05	PROGRAMMING IN PYTHON FOR DATA SCIENCE	3+0+2	3
DA M 21 C06	STATISTICAL INFERENCE	3 +0+2	3
DA M 21 C07	PRACTICAL - DATA SCIENCE LAB	2+0+ 3	3
SEMESTER II	**Total Credits 24**		
DA M 21 C08	SAMPLING METHODS & DESIGN OF EXPERIMENTS	3+0+ 2	3
DA M 21 C09	APPLIED MULTIVARIATE ANALYSIS	3 +0+2	3
DA M 21 C10	APPLIED REGRESSION ANALYSIS	3 +0+2	3
DA M 21 C11	PROGRAMMING IN R FOR DATA ANALYTICS	3 +0+2	3
DA M 21 C12	DATA VISUALIZATION AND PRESENTATION	3+0+2	3
	ELECTIVE 1	3 +0+2	3
	ELECTIVE 2	3+0+ 2	3
DA M 21 C13	PRACTICAL – DATA ANALYTICS LAB	2+0+ 3	3
SEMESTER III	**Total Credits 22**		
DA M 21 C14	MACHINE LEARNING	3+0+2	3
DA M 21 C15	STOCHASTIC MODELING & TIME SERIES FORECASTING	3+0+2	3
DA M 21 C16	ARTIFICIAL INTELLIGENCE	3 +0+2	3
	ELECTIVE 3	3 +0+2	3
	ELECTIVE 4	3 +0+2	3
DA M 21 C17	PRACTICAL – COMPUTER LAB	2+ 0+3	3
	OPEN COURSE	4 +0+0	4
SEMESTER IV	**Total Credits 21**		
	ELECTIVE 5	3 +0+2	3
	ELECTIVE 6	3+0+2	3
DA M 21 C18	PROJECT / DISSERTATION	24	15
	Grand Total of Credits		**88**

Thus in the present Program Structure there are 18 Core Courses with a total of 66 credits, 6 Elective courses with a total of 18 credits and an Open Course of 4 credits so that the grand total of credits is 88 for the whole program.

TABLE OF ELECTIVE / OPEN COURSES:

The students have to select any 6 Elective Courses in the table below in consultation with the Faculty Coordinator/ Head/ Director of the School. More Elective Courses may be included in the list as per the specialization and expertise of faculty members subject to approval by the Faculty Council of the School.

Course Code	**Elective Course Title**	Teaching hours per week L+T+P	Credits
DA M 21 E01	NEURAL NETWORKS AND DEEP LEARNING	3+0+2	3
DA M 21 E02	BAYESIAN INFERENCE & COMPUTING	3+0+2	3
DA M 21 E03	BIG DATA AND HADOOP	3+0+2	3
DA M 21 E04	INFORMATION RETRIEVAL TECHNIQUES	3+0+2	3
DA M 21 E05	DATA WAREHOUSING & DATA MINING	3+0+2	3
DA M 21 E06	WEB SCRAPING AND TEXT MINING	3+0+2	3
DA M 21 E07	NATURAL LANGUAGE PROCESSING WITH PYTHON	3+0+2	3
DA M 21 E08	FRAUD ANALYTICS	3+0+2	3
DA M 21 E09	INTERNET OF THINGS IN THE CLOUD	3+0+2	3
DA M 21 E10	OPERATIONS RESEARCH	3+0+2	3
DA M 21 E11	CLOUD COMPUTING	3+0+2	3
DA M 21 E12	BUSINESS INTELLIGENCE & ANALYTICS	3+0+2	3
DA M 21 E13	DATA ANALYTICS COMPUTING	3+0+2	3
DA M 21 E14	COMPLEX NETWORK ANALYSIS	3+0+2	3
DA M 21 E15	BIOSTATISTICS AND EPIDEMIOLOGY	3+0+2	3
DA M 21 E16	BIOINFORMATICS AND COMPUTATIONAL BIOLOGY	3+0+2	3
DA M 21 E 17	ADVANCED RESAMPLING TECHNIQUES	3+0+2	3
DA M 21 E 18	TIME SERIES ANALYSIS & FORECASTING	3+0+2	3
DA M 21 E 19	DEMOGRAPHY & POPULATION DYNAMICS	3+0+2	3
DA M 21 E 20	INDUSTRIAL STATISTICS & QUALITY CONTROL	3+0+2	3
OPEN COURSE DURING SEMESTER III FOR STUDENTS OF OTHER SCHOOLS			
DA M 21 O 01	ADVANCED STATISTICAL TECHNIQUES FOR DATA ANALYTICS	4+0+1	4

N.B. 1 Open Course is any course offered by another Department / School / Inter University Centre of this University other than the parent Department / School / Centre. Students can select one open course subject to permission from both Departments / Schools to encourage interdisciplinary studies and research in emerging areas.

N.B. 2 In case students wish to undergo online MOOC Courses in SWAYAM PORTAL or offered by IITs and other reputed institutes of national / international importance, they can choose them as electives during Semester 3 and 4 with the permission of the Head of the Department / Director of the School in accordance with the CSS regulations of the university.

NB. 3 During Semester IV a Major Project / Dissertation shall be carried out in a reputed research institute / department or industry or software company under the joint supervision of an internal faculty and external guide / expert approved by the Head of the Dept / Director of the School. (This can be started from the end of the second semester with initial discussions and literature collection and review.) Each student has to submit a bound copy and soft copy of the Project Report (documented in LaTex) of about 50 to 75 pages certified by the supervisors, at least 7 days before the conduct of the Viva Voce and Project Presentation.

N.B. 4 Along with each course necessary lab training and demonstration of applications of all topics is to be provided in the computer lab using case studies and real data sets as part of Practical. Records are to be prepared by the students and submitted before lab examinations to the Faculty in charge / Head of the Department / School. Case study reports and presentations and paper publications shall be part of internal assessment.

SEMESTER I COURSES TOTAL CREDITS: 21

Course Code	Course Title	Teaching L+T+P	Credits
DA M 21 C01	**INTRODUCTION TO DATA SCIENCE & DATA ANALYTICS**	3 +0+2	3
DA M 21 C02	**LINEAR ALGEBRA AND MATRICES**	3 +0+2	3
DA M 21 C03	PROBABILITY AND DISTRIBUTION THEORY	3 +0+2	3
DA M 21 C04	DATA BASE MANAGEMENT SYSTEMS	3+0+2	3
DA M 21 C05	PROGRAMMING IN PYTHON FOR DATA SCIENCE	3+0+2	3
DA M 21 C06	STATISTICAL INFERENCE	3 +0+2	3
DA M 21 C07	**PRACTICAL - DATA SCIENCE LAB**	2+0+3	3
	Total Credits		**21**

MAHATMA GANDHI UNIVERSITY

DA M 21 C01: INTRODUCTION TO DATA SCIENCE & DATA ANALYTICS

School Name	**School of Data Analytics**					
Programme	**M.Sc. Data Science & Analytics**					
Course Name	**INTRODUCTION TO DATA SCIENCE & DATA ANALYTICS**					
Type of Course	Core					
Course Code	**DA M 21 C01**					
Course Summary & Justification	The course is designed • To introduce basic concepts in Data Science, Data Analytics Data types, steps in data science etc. • To develop creative thinking for information extraction from a given data set and gaining. • To understand big data and its management insights and to apply standard data analytics procedures. .					
Semester	First					
Total Student Learning Time (SLT)	Learning Approach	Lecture	Tutorial	Practical	Others	Total Learning Hours
	Authentic learning Collaborative learning Independent learning	70	10	30	10	120
Pre - requisite	Basic understanding of Data Processing					

O No.	Expected Course Outcome	Learning Domains	PSO No.
1	To get the overall idea of Data Science & Data Analytics.	U	1,2
2	To understood the different roles of a Data Scientist, Data Analyst, methodology and important steps to deal with different types of data sets including big data	U,A	1,2,4
3	To Apply data pre-processing techniques	A,S	1,2, 4,5
4	To Use appropriate tools and technology to collect,	A,E	1, 3,5

	process, transform, summarize, and visualize data		
5	To Explain the importance of and be able to formulate a data analysis problem statement that is clear, concise, and measurable	R,U,A	1,4,5
Remember (R), Understand (U), Apply (A), Analyse (An), Evaluate (E), Create (C), Skill (S), Interest (I) and Appreciation (Ap)			

COURSE CONTENT

Module No	Module Content	C O	Hr
1	Origins of Data Science-Development-Popularization- Definition of Data Science Academic programs-Professional Organizations-Case Study- Mesh up of Disciplines- data Engineering-Acquiring – Ingesting –Transforming – Metadata –Storing – Retrieving – Scientific method-Reasoning Principles – Empirical Evidence –Hypothesis Testing – Repeatable Experiments	1,2	10
2	Data Scientist-Thinking Like a Mathematician- Quantity –Structure – Space –Change – Thinking Like a Statistician – Collection – Organization – Analysis – Interpretation – Thinking Like a Programmer – Software Design – Programming Language – Source Code – Thinking Like a Visual Artist – Creative Process – Data Abstraction – Informational Interesting	1,2,3	20
3.	What is Data-Data Point-Data Set-Data types- Data types in Mathematics, Statistics, Computer Science- Data types in R- Objects, Variables, Values and Vectors in R- Data sets and Data frames-Creating a data set in R- Talking to subject matter experts- looking for exception-exploring risks and uncertainty, Big data definition, enterprise / structured data, social / unstructured data, unstructured data needs for analytics, Big data programming.	1,3,4	20
4	Doing Data Science-Steps –Acquire-Parse-Filter-Mine-Represent-Refine-Interact–Define the question- Define the ideal data set-Determine what data you can access- Obtain the data- Clean the data-Exploratory data analysis- Statistical prediction/modeling- Interpret results- Challenge results- Synthesize/write up results- Create reproducible codes- Distributing results to other people. Data Science Ethics – Doing good data science – Owners of the data – Valuing different aspects of privacy – Getting informed consent – The Five Cs – Diversity – Inclusion – Future Trends	1,4,5	20
	Total Credits of the Course	3	70

Books for Reference

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

TeachingandLearningApproach	**Classroom Procedure (Mode of transaction)** Direct Instruction, 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
Assessment Types	**Mode of Assessment** **A. Continuous Internal Assessment (CIA)-40 marks** Internal Tests -20 marks Assignment – Every student needs to write assignments on a given topic based on the available published literature – 10 marks Seminar Presentation – A topic needs to be presented and discussed with the class- 10 marks **B. Semester End examination – 60 marks**

Approval Date	
Version	
Approval by	
Implementation Date	

MAHATMA GANDHI UNIVERSITY

DA M 21 C02:LINEAR ALGEBRA AND MATRICES

SchoolName	**School of Data Analytics**
Programme	**M.Sc. Data Science & Analytics**
Course Name	**LINEAR ALGEBRA AND MATRICES**
Type of Course	Core
Course Code	**DA M 21 C02**
Course Summary & Justification	The objective of the course content is to To make students familiar with various concepts in linear algebra and matrices and their applications in data science, data analytics and pattern recognition

Total Student Learning Time (SLT)	Learning Approach	Lecture	Tutorial	Practical	Others	Total Learning Hours
	Authentic learning Collaborative learning Independent learning	80	10	20	10	120
Pre - requisite	Basics of Matrix Theory and Vector Spaces					

COURSE OUTCOMES (CO)

CO No.	**Expected Course Outcome**	**Learning Domains**	**PSO No.**
1	Students get awareness about vector spaces, independence, dimension, rank, inverse and are able to solve systems of linear equations.	R, U,Ap	1
2	Students get exposure to Linear transforms and their applications in Data Analytics	R,U,I	2
3	Students get exposure to egen values, eigen vectors, Eigen decomposition, Orthogonal decomposition, Quadratic forms, diagonal reduction etc and their applications in Data Science and Data Analytics	A, An	3
4	Students get deeper understanding of Spectral Decomposition of symmetric matrices, Jordan Canonical Form, LU Decomposition, Singular value decomposition etc	A, An	4
5	Students are now aware of handwritten digits and simple algorithm, Classification of handwritten digits using SVD bases , Tangent distance etc and their applications in Text Mining.	E,C,A	5
Remember (R), Understand (U), Apply (A), Analyse (An), Evaluate (E), Create (C), Skill (S), Interest (I) and Appreciation (Ap)			

COURSE CONTENT

Module No	**Module Content**	**CO**	**Hrs**
1	Introduction to Vector Spaces: Group, Ring, Field, Vector spaces, Subspaces, sums of Subspaces, Direct Sums, Span and Linear Independence, basis, dimension. Examples, row subspace, column subspace, row rank, column rank, rank of a matrix, inverse of a nonsingular matrix, generalized inverse computation and applications, solution of system of linear equations, null space, nullity.	1,2	20
2	Linear transformation and matrices: Definition of Linear Maps- Algebraic Operations on L (V,W)- Null spaces and Injectivity – Range and Surjectivity – Fundamental Theorems of Linear Maps-	1,2,3	20

	Representing a Linear Map by a Matrix-Invertible Linear Maps-Isomorphic Vector spaces-Linear transformation as Matrix Multiplication – Operators – Products of Vector Spaces – Product of Direct Sum – Quotients of Vector spaces.		
3.	Eigen values and Eigenvectors – Eigenvectors and Upper Triangular matrices – Eigen spaces and Diagonal Matrices – Inner Products and Norms – Linear functional on Inner product spaces. Quadratic forms, diagonal reduction, canonical forms, nature of definiteness, orthogonal reduction.	4	20
4	Application to Data analytics: Spectral Decomposition of symmetric matrices, Jordan Canonical Form, LU Decomposition,Singular value decomposition – Handwritten digits and simple algorithm – Classification of handwritten digits using SVD bases – Tangent distance – Text Mining.	5	20
	Total Credits of the Course	3	80

Teaching and Learning Approach	**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
Assessment Types	**Mode of Assessment** **Continuous Internal Assessment (CIA) -40 marks** 1. Internal Tests of maximum 20 marks 2. Seminar Presentation – a theme is to be discussed and identified to prepare a paper and present in the seminar Maximum marks 10 3. Write a detailed report on a given topic based on research findings and literature search – 10 marks **Semester End examination – 60 marks**

References

1. Sheldon Axler, Linear Algebra Done Right, Springer, 2017.
2. Eldén Lars, Matrix Methods in Data Mining and Pattern Recognition, Society for Industrial and Applied Mathematics (SIAM) Publications, 2007.

Recommended Reading

3. E. Davis, Linear Algebra and Probability for Computer Science Applications, CRC Press, 2012.
4. J. V. Kepner and J. R. Gilbert, Graph Algorithms in the Language of Linear Algebra, Society for Industrial and Applied Mathematics, 2011.
5. D. A. Simovici, Linear Algebra Tools for Data Mining, World Scientific Publishing, 2012.
6. P. N. Klein, Coding the Matrix: Linear Algebra Through Applications to Computer Science, Newtonian Press, 2015.

7. Rao A.R. and Bhimasankaram P. (2000) Linear Algebra, Second edition, Hindustan Book Agency.
8. Rao, C.R. (2009) Linear Statistical Inference and its Applications, Wiley Asia, Second Edition.

Approval Date	
Version	
Approval by	
Implementation Date	

MAHATMA GANDHI UNIVERSITY

DA M 21 C03: PROBABILITY AND DISTRIBUTION THEORY

SchoolName	**School of Data Analytics**					
Programme	**M.Sc Data Science & Analytics**					
Course Name	PROBABILITY AND DISTRIBUTION THEORY					
Type of Course	Core					
Course Code	**DA M 21 C03**					
Course Summary & Justification	The objective of the course content is to To make the students aware of basic concepts in probability theory and distribution theory for applications in data science and data analytics					
Semester	First					
Total StudentLearning Time (SLT)	Learning Approach	Lecture	Tutorial	Practical	Others	Total LearningHours
	Authentic learning Collaborative learning Independent learning	80	10	20	10	120
Pre-requisite	Basics of Probability Theory and Distributions					

COURSE OUTCOMES (CO)

CO No.	Expected Course Outcome	Learning Domains	PSO No.
1	**Now the students are aware of basic concepts in distribution theory for applications in data science and data analytics.**	R,U	1

2	**After undergoing this course students have understood various concepts in probability theory and are able to compute probability.**	A, Ap	2
3	**They can use various distributions as models for real data sets.**	An, E	3
4	**They have understood the importance of central limit theorem and its applications**	C, S	4
5	**Students are aware of order statistics, life time distributions etc and can apply these for modeling**	A,Ap	5
Remember (R), Understand (U), Apply (A), Analyse (An), Evaluate (E), Create (C), Skill (S), Interest (I) and Appreciation (Ap)			

COURSE CONTENT

Module No	Module Content	CO	Hrs
1	Basics of Probability, Different Approaches- classical, frequency, axiomatic, subjective, – simple properties – conditional probability, independence, Bayes' Theorem and applications, - Random variables and Random vectors – PDF and CDF, Distribution functions – Decomposition of distribution functions, independence of rvs, Expectation and Conditional Expectation and properties	1,2,6	20
2	Characteristic function – definition and properties – Inversion formula. Convergence of a sequence of random variables – convergence in distribution – convergence in probability almost sure convergence and convergence in quadratic mean; Chebycheff's inequality, Khintchin's weak law of large numbers, Kolmogorov strong law of large numbers (statement only) – Central Limit Theorem – Lindberg – Levy form, Linderberg – Feller form, Liapounov form. (statements only) Illustration through numerical problems.	1,2,6	20
3.	Distribution – Binomial, Poisson, Negative Binomial, Hypergeometric distributions, Power series distributions, Logarithmic series distribution- Multinomial distribution; exponential, Gamama, Weibull, Pareto, normal, Laplace, logistic, Cauchy, Inverse Gaussian, Lognormal,– Bivariate Normal – Bivariate Exponential – Compound, truncated and mixture of distributions, Concept of convolution – Sampling distributions: chi – square, t and F distributions and their properties	3,6	20
4	Order statistics and their distributions– Joint and marginal distributions of order statistics – Distribution of range and mid range (uniform and exponential cases only) –Extreme values and their asymptotic distributions (concepts only) –Empirical distribution function and its properties – Kolmogorov – Smirnov statistics – Life time distributions and reliability concepts – Exponential and Weibull distributions classified by hazard rate.	4,5,6	20
	Total Credits of the Course	3	80

Teaching and Learning Approach	**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
Assessment Types	**Mode of Assessment** A. Continuous Internal Assessment (CIA)-40 marks 1. Internal Tests of maximum 20 marks 2. Seminar Presentation – a theme is to be discussed and identified to prepare a paper and present in the seminar Maximum marks 10 3. Write a detailed report on a given topic based on research findings and literature search – 10 marks B. **Semester End examination – 60 marks**

REFERENCES

1. B.R Bhat (2014). Modern Probability Theory, New Age International, 4th Edition,
2. V.K Rohatgi and Saleh M. (2015) An Introduction to Probability and Statistics, , 3rd Edition,
3. A.M Mood, F.A Graybill and D.C Boes (2017) Introduction to the Theory of Statistics, , Tata McGraw-Hill, 3rd Edition.
4. H.A David and H.N Nagaraja (2003) Order Statistics, , John Wiley & Sons, 3rd Edition.
5. Gupta, S .C., Kapoor, V.K. (2016) Fundamentals of Mathematical Statistics, Sultan Chand & Co.

Approval Date	
Version	
Approval by	
Implementation Date	

MAHATMA GANDHI UNIVERSITY

DA M 21 C04: DATABASE MANAGEMENT SYSTEMS

SchoolName	**School of Data Analytics**					
Programme	**M.Sc. Data Science & Analytics**					
Course Name	DATA BASE MANAGEMENT SYSTEMS					
Type of Course	Core					
Course Code	**DA M 21 C04**					
Course Summary & Justification	The objective of the course is to understand the basic concepts and the applications of database systems.					
Total Student Learning Time (SLT)	Learning Approach	Lecture	Tutorial	Practical	Others	Total Learning Hours
	Authentic learning Collaborative learning Independent learning	80	10	20	10	120
Pre-requisite	Basics of Statistics and Computer Science					

COURSE OUTCOMES (CO)

CO No.	Expected Course Outcome	Learning Domains	PSO No.
1	Students understood the basics of SQL and can construct queries using SQL	E,R,U	1
2	Understood the relational database design principles and the basic issues of transaction processing and concurrency control.	C,S,U/ An	1,2
3	Understood database storage structures and access techniques.	R,E	1,2
4	Understood object oriented databases, data warehousing and OLAP tools.	U,S	1,2
5	Understood MongoDB and can evaluate the nosql databases.	C,E,S	1,2,3

Remember (R), Understand (U), Apply (A), Analyse (An), Evaluate (E), Create (C), Skill (S), Interest (I) and Appreciation (Ap)

COURSE CONTENT

Module No	Module Content	CO	Hrs
1	Introduction to File and Database systems- History- Advantages, disadvantages- Data views – Database Languages – DBA – Database Architecture – Data Models- Keys – Mapping Cardinalities, Relational Algebra and calculus – Query languages – SQL – Data definition – Queries in SQL – Updates– Views – Integrity and Security – triggers, cursor, functions, procedure – Embedded SQL – overview of QUEL, QBE.	1	20
2	Design Phases – Pitfalls in Design – Attribute types –ER diagram – Database Design for Banking Enterprise – Functional Dependence – Normalization (1NF, 2NF, 3NF, BCNF, 4NF, 5NF).File Organization – Organization of Records in files – Indexing and Hashing. Transaction concept – state serializability – Recoverability- Concurrency Control – Locks- Two Phase locking – Deadlock handling – Transaction Management in Multi Databases.	2	20
3.	Object-Oriented Databases- OODBMS- rules – ORDBMS- Complex Data types – Distributed databases –characteristics, advantages, disadvantages, rules- Homogenous and Heterogeneous Distributed data Storage – XML – Structure of XML Data – XML Document. Introduction to Mongo DB , Overview of NoSQL.	3	20
4	Introduction to data warehousing, evolution of decision support systems –Modeling a data warehouse,granularity in the data warehouse – Data warehouse life cycle, building a data warehouse, Data Warehousing Components, Data Warehousing Architecture – On Line Analytical Processing, Categorization of OLAP Tools	4,5	20
	Total Credits of the Course	3	80

Teachingand LearningApproach	**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
Assessment Types	**Mode of Assessment** **A. Continuous Internal Assessment (CIA)-40 marks** 1. Internal Tests of maximum 20 marks 2. Seminar Presentation – a theme is to be discussed and identified to prepare a paper and present in the seminar Maximum marks 10 3. Write a detailed report on a given topic based on research findings and literature search – 10 marks B. **Semester End examination – 60 marks**

REFERENCES

1. Silberschatz, A., Korth, H. F., & Sudarshan, S. (1997). Database System Concepts
   (Vol. 4). NewYork: McGraw-Hill.

Additional References

2. Pratt, P. J. & Adamski, J. J. (2011). Database Systems: Management and Design. Boyd & Fraser Pub. Co.
3. James R Groff and Paul N Weinberg (2003) The Complete Reference SQL –, Second Edition, Tata McGraw Hill,
4. Shamkant, R. E., & Navathe, B. (2009). Fundamentals of Database Systems. Addison-Wesley Publishing Company.
5. Elmasri, R.& Navathe, S. (2010). Fundamentals of Database Systems. Addison-Wesley Publishing Company.)
6. Majumdar, A. K., & Bhattacharyya, P. (1996). Database Management Systems. McGraw-Hill.
7. ISRD group, Introduction to Database Management Systems, TMH, 2008
8. Ramakrishnan, R., &Gehrke, J. (2000). Database Management Systems. McGraw Hill
9. Chodorow, K. (2013). Mongo D B: The Definitive Guide: Powerful and Scalable Data Storage. "O'Reilly Media, Inc.".
10. Harrison, G. (2015). Next Generation Databases: NoSQL and Big Data. Apress.

Approval Date	
Version	
Approval by	
Implementation Date	

MAHATMA GANDHI UNIVERSITY

DA M 21 C05: PYTHON PROGRAMMING FOR DATA SCIENCE

SchoolName	**School of Data Analytics**					
Programme	**M. Sc. Data Science & Analytics**					
Course Name	**PYTHON PROGRAMMING FOR DATA SCIENCE**					
Type of Course	Core					
Course Code	**DA M 21 C05**					
Course Summary & Justification	This course is designed to develop a thorough understanding of Data Analytics and Data Science techniques using Python					
Total Student LearningTime (SLT)	Learning Approach	Lecture	Tutorial	Practical	Others	Total Learning Hours
	Authentic learning Collaborative learning Independent learning	60	10	40	10	120
Pre-requisite	**Basics of Computer Programming**					

COURSE OUTCOMES (CO)

CO No.	Expected Course Outcome	Learning Domains	PSO No.
1	Demonstrate the usage of built-in objects in Python	U,A,S	1
2	Analyze the significance of python program development environment by working on real world examples	U,A,C	1,2
3	Implement numerical programming, data handling and visualization through NumPy.	U,A	1,3
4	Implement numerical programming, data handling and visualization through Pandas .	C,U,A	3,4
5	Implement numerical programming, data handling and visualization through MatplotLib modules.	C,U,A,S	5

Remember (R), Understand (U), Apply (A), Analyse (An), Evaluate (E), Create (C), Skill (S), Interest (I) and Appreciation (Ap)

COURSE CONTENT

Module No	Module Content	CO	Hrs
1	Installation and maintenance of Linux platform for open softwares. Structure of Python Program-Underlying mechanism of Module Execution-Branching and Looping-Problem Solving Using Branches and Loops-Functions – Lists and Mutability- Problem Solving Using Lists and Functions. Sequences, Mapping and Sets-Dictionaries- -Classes: Classes and Instances-Inheritance-Exceptional Handling-Introduction to Regular Expressions using "re" module.	1,2	15
2	Basics of NumPy-Computation on NumPy-Aggregations-Computation on Arrays-Comparisons, Masks and Boolean Arrays-Fancy Indexing-Sorting Arrays-Structured Data: NumPy's Structured Array.	2,3	15
3.	Introduction to Pandas Objects- Data indexing and Selection-Operating on Data in Pandas- Handling Missing Data-Hierarchical Indexing – Combining Data Sets. Aggregation and Grouping-Pivot Tables-Vectorized String Operations –Working with Time Series-High Performance Pandas- and query ()	2,4	15
4	Basic functions of matplotlib –Simple Line Plot, Scatter Plot-Density and Contour Plots- Histograms, Binnings and Density-Customizing Plot Legends, Colour Bars- Three- Dimensional Plotting in Matplotlib.	2,5	15
	Total Credits of the Course	3	60

Teachingand LearningApproach	**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
Assessment Types	**Mode of Assessment** **A. Continuous Internal Assessment (CIA)-40 marks** 1. Internal Tests of maximum 20 marks 2. Seminar Presentation – a theme is to be discussed and identified to prepare a paper and present in the seminar Maximum marks 10 3. Write a detailed report on a given topic based on research findings and literature search – 10 marks B. **Semester End examination – 60 marks**

REFERENCES

1.. Jake Vander Plas, Python Data Science Handbook – Essential Tools for Working with Data, O'Reilly Media Inc, 2016

2. Zhang.Y. , An Introduction to Python and Computer Programming, Springer Publications, 2016
3. Wes McKinney, (2017) Python for Data Analysis: Data Wrangling with Pandas, NumPy, and Ipython, 2nd Edition, O'Reilly Media.
4. Haslwanter, T.(2015) An Introduction to Statistics with Python, Springer

Essential Reading / Recommended Reading

5. Joel Grus (2016) Data Science from Scratch First Principles with Python, O'Reilly Media.
6. T.R.Padmanabhan(2016) Programming with Python, Springer Publications.

Approval Date	
Version	
Approval by and Date	

MAHATMA GANDHI UNIVERSITY

DA M 21 C06: STATISTICAL INFERENCE

SchoolName	**School of Data Analytics**					
Programme	**M.Sc. Data Science & Analytics**					
Course Name	**STATISTICAL INFERENCE**					
Type of Course	Core					
Course Code	**DA M 21 C06**					
Course Summary & Justification	This course is designed to introduce the concepts of theory of estimation and testing of hypothesis. This paper also deals with the concept of parametric tests for large and small samples. It also provides knowledge about non-parametric tests and its applications. It is also expected to give lab illustration of the concepts through original data sets.					
Total StudentLearning Time (SLT)	Learning Approach	Lecture	Tutorial	Practical	Others	Total Learning Hours
	Authentic learning Collaborative learning Independent learning	70	10	30	10	120
Pre-requisite	Basic concepts in Statistics, estimation and testing of hypotheses					

COURSE OUTCOMES (CO)

CO No.	Expected Course Outcome	Learning Domains	PSO No.
1	Students have understood various estimation methods	Ap	1,2,4,5
2	Students have studied Neyman-Pearson methodology for constructing Most Powerful Tests	Ap/S	1,2.4.5
3	They are aware of various properties for comparing estimators and tests.	S	1,2.4.5
4	They are aware of Sequential estimation and testing methods	S	1,2,4.5
5	They are aware of various Nonparametric Tests and can apply these for testing of hypotheses.	S	1,2,4,5
Remember (R), Understand (U), Apply (A), Analyse (An), Evaluate (E), Create (C), Skill (S), Interest (I) and Appreciation (Ap)			

COURSE CONTENT

Module No	Module Content	CO	Hours
1	Sufficient statistics- Neyman – Fisher Factorization theorem –the existence and construction of minimal sufficient statistics – Minimal sufficient statistics and exponential family – sufficiency and completeness – sufficiency and invariance.Minimum variance unbiased estimation – locally minimum variance unbiased estimators – Rao Blackwell theorem – Completeness: Lehmann Scheffe theorems – Necessary and sufficient condition for unbiased estimators – Cramer- Rao inequality and lower bound – Bhattacharya system of lower bounds in the one parameter regular case – Chapman –Robbins inequality.	1	15
2	Computational routines – strong consistency of maximum likelihood estimators – Asymptotic Efficiency of maximum likelihood estimators – Best Asymptotically Normal estimators – Method of moments – Bayes' and minimax estimation: The structure of Bayes' rules – Bayes' estimators for SELF, AELF, 0-1, quadratic and convex loss functions – minimax estimation – interval estimation, confidence intervals, credibility intervals	2	20
3.	Basic ideas of testing of hypotheses, significance level, power, p-value – Neyman-Pearson fundamental Lemma – Distributions with monotone likelihood ratio – Problems – Generalization of the fundamental lemma to randomized tests, Uniformly most powerful tests two sided hypotheses – testing the mean and variance of a normal distribution, testing equality of means and variances of two normal distributions. Likelihood ratio tests – locally most powerful tests – the concept of confidence sets and uniformly most accurate confidence sets. Unbiased tests – similarity and completeness – UMP unbiased tests.	3,4	20

4	SPRT procedures –ASN and O.C. functions, exponential, normal and binomial cases – Non parametric tests- sign test, signed rank tests, matched pair signed rank test, Chi-square test for goodness of fit, homogeneity and independence of attributes, Kolmogrov Smirnov one sample and two sample tests, Fishers exact test, Mann-Whitney U test, Wilcoxon Sum of Ranks test, Kruskal-Wallis test, Runs tests for randomness, Wald-Wolfowitz run test.	5	15
	Total Credits of the Course	3	70

Books for Reference

1. An Introduction to Probability and Statistics, V.K Rohatgi and Saleh, 3rd Edition, 2015.
2. Rajagopalan M. and Dhanavanthan P, Statistical Inference, PHI Learning (P) Ltd, New Delhi, 2012.

Essential Reading / Recommended Reading

3. Introduction to the Theory of Statistics, A.M Mood, F.A Graybill and D.C Boes, Tata McGraw-Hill, 3rd Edition (Reprint), 2017.
4. Linear Statistical Inference and its Applications, Rao C.R, Willy Publications, 2nd Edition, 2001

Teachingand LearningApproach	**Classroom Procedure (Mode of transaction)** Direct Instruction: Explicit Teaching, interactive Instruction: Active co-operative learning and skill development, Demonstrations, Group Assignments, Authentic learning, Library work and Group discussion, Preparation of experiment design and reports
Assessment Types	**Mode of Assessment** **A. Continuous Internal Assessment (CIA)-40 marks** 1. Internal Tests of maximum 20 marks 2. Seminar Presentation – a theme is to be discussed and identified to prepare a paper and present in the seminar Maximum marks 10 3. Write a detailed report on a given topic based on research findings and literature search – 10 marks **B. Semester End examination – 60 marks**

Approval Date	
Version	
Approval by	
Implementation Date	

MAHATMA GANDHI UNIVERSITY

DA M 21 C07: PRACTICAL (DATA SCIENCE LAB)

SchoolName	**School of Data Analytics**					
Programme	**M. Sc. Data Science & Analytics**					
Course Name	**PRACTICAL (DATA SCIENCE LAB)**					
Type of Course	Core					
Course Code	**DA M 21 C07**					
Course Summary & Justification	The purpose of this laboratory course is to enhancing their skills in data analysis, data science and data analytics dealing with original data sets and big data					
Total StudentLearning Time (SLT)	Learning Approach	Lecture	Tutorial	Practical	Others	Total LearningHours
	Authentic learning Collaborative learning Independent learning	30	10	70	10	120
Pre-requisite	Basics of Python, Statistics, Probability, Distribution Theory					

COURSE OUTCOMES (CO)

CO No.	Expected Course Outcome	Learning Domains	PSO No.
1	Enhancing their skills in data analysis, data science and data analytics dealing with original data sets and big data. Familiarity in Case Studies, Presentations, Viva Voce etc.	U,A,C,S	1,2,3

Remember (R), Understand (U), Apply (A), Analyse (An), Evaluate (E), Create (C), Skill (S), Interest (I) and Appreciation (Ap)

Teachingand LearningApproach	**Laboratory Procedure (Mode of transaction)** Direct Instruction: Lecture, Explicit Teaching, Demonstration, Hands on experimental sections, Skill acquisition by laboratory training, Applications and illustrations of all concepts studied in the semester using Python
Assessment Types	**Mode of Assessment** **Continuous Internal Assessment (CIA)-100** 1. Internal Laboratory Skill Tests -40 marks

	2. Records, Accuracy, Lab discipline- 20 marks 3. Presentations, Viva-Voce -20 marks 4. Regularity, Punctuality, Attendance—20 marks

REFERENCES

1.Wes McKinney, (2017) Python for Data Analysis: Data Wrangling with Pandas, NumPy, and Ipython, 2nd Edition, O'Reilly Media. 2. Haslwanter, T.(2015) An Introduction to Statistics with Python, Springer
3.Joel Grus (2016) Data Science from Scratch First Principles with Python, O'Reilly Media. 4. T.R.Padmanabhan(2016) Programming with Python, Springer Publications.

Approval Date	
Version	
Approval by	

SEMESTER II COURSES TOTAL CREDITS: 24

Course Code	Course Title	Teaching L+T+P	Credits
DA M 21 C08	SAMPLING METHODS & DESIGN OF EXPERIMENTS	3+0+ 2	3
DA M 21 C09	APPLIED MULTIVARIATE ANALYSIS	3 +0+2	3
DA M 21 C10	APPLIED REGRESSION ANALYSIS	3 +0+2	3
DA M 21 C11	PROGRAMMING IN R FOR DATA ANALYTICS	3 +0+2	3
DA M 21 C12	DATA VISUALIZATION AND PRESENTATION	3+0+2	3
	ELECTIVE 1	3 +0+2	3
	ELECTIVE 2	3+0+ 2	3
DA M1 21 C13	PRACTICAL – DATA ANALYTICS LAB	2+0+3	3
	Total Credits		24

MAHATMA GANDHI UNIVERSITY

DA M 21 C08: SAMPLING METHODS & DESIGN OF EXPERIMENTS

SchoolName	School of Data Analytics					
Programme	M.Sc. Data Science & Analytics					
Course Name	**SAMPLING METHODS & DESIGN OF EXPERIMENTS**					
Type of Course	Core					
Course Code	**DA M 21 C08**					
Course Summary & Justification	To make the students familiar with different sampling schemes and their advantages as well as their applications in estimating population mean, total, proportion etc. It is also expected to impart knowledge on basic principles of experimentation, different designs like CRD,RBD, LSD, BIBD, Factorial experiments etc.					
Total Student Learning Time (SLT)	Learning Approach	Lecture	Tutorial	Practical	Others	Total Learning Hours

	Authentic learning Collaborative learning Independent learning	80	10	20	10	120
Pre-requisite	Basic Statistics,Estimation and Testing, Linear Algebra					

COURSEOUTCOMES (CO)

CO No.	Expected Course Outcome	Learning Domains	PSO No.
1.	Students get good exposure to various sample survey techniques	R,U,I	1,2
2.	Students can develop estimators of population mean and its variance under different sampling schemes and can compare their efficiencies.	A,An,E,C	3,4,5
3.	Students get exposure to linear estimation and BLUE including Gauss-Markov Theorem and its applications..	R,U,A,I	1,2
4.	They are aware of various designs including CRD,RBD,LSD, BIBD, Factorial Designs and can plan experiments suitably and test hypotheses and conduct ANOVA, ANCOVA etc.	A,An,E,C	3,4,5
**Remember (R), Understand (U), Apply (A), Analyse (An), Evaluate (E), Create (C), Skill (S), Interest (I) and Appreciation (Ap)*			

COURSE CONTENT

Module No	**Module Content**	**CO**	**Hrs**
1	Principal steps in a sample survey, preparation of questionnaires, probability sampling and non probability sampling, sampling and non sampling errors, bias, variance and MSE, simple random sampling with and without replacement – estimation of population mean, total and proportions, estimation of sample size, confidence intervals, determination of the sample size, Stratified random sampling, estimation of the population mean, various methods of allocation of a sample, comparison of the precisions of estimators under proportional allocation, optimum allocation and SRS. Systematic sampling – Linear and Circular, estimation of the mean and its variance, comparison of systematic sampling, SRS and stratified random sampling for a population with a linear trend.	1	20

2	Ratio method of estimation, estimation of population ratio, mean and total, Bias and relative bias of ratio estimator, comparison with SRS estimation. Regression method of estimation. Comparison of ratio and regression estimators with mean per unit method, Cluster sampling, single stage cluster sampling with equal and unequal cluster sizes, estimation of the population mean and its standard error. Multistage and Multiphase sampling (Basic Concepts), estimation of the population mean and its standard error.	1,2	20
3.	Linear Estimation and estimability of parametric functions, Standard Gauss-Markov set up and Gauss Markov Theorem. Planning of experiments: Basic principles of experimental design, Analysis of variance (ANOVA), Completely randomized design (CRD), Randomized block design (RBD), Latin square design (LSD) , Greco-Latin square designs, missing plot technique.	2,3,4	20
4	Analysis of covariance (ANCOVA), ANCOVA with one concomitant variable in CRD and RBD. Incomplete block design: Balanced incomplete block design (BIBD); Intra-block analysis of BIBD, Basic ideas of partially balanced incomplete block design (PBIBD), Factorial experiments, Yates procedure, 2^n and 3^n factorial experiments, Confounding in factorial experiments, Construction of confounded scheme in 2^n factorial experiments, Basics of split plot experiments and their analysis.	2,4,5	20
	Total Credits	**3**	80

Teaching and Learning Approach	**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

REFERENCES

1. Mukhopadhyay P. (2009) Theory and Methods of Survey Sampling, Second Edition, PHI Learning (P) Ltd
2. Arnab, R. (2017). Survey Sampling: Theory and Applications. Academic Press.
3. Montgomery, C.D. (2012) Design and Analysis of Experiments, John Wiley, New

York.

4. Das, M.N. and Giri, N.C. (1994) Design and Analysis of Experiments, Wiley Eastern Ltd

Books for additional reading

5. Sampath S. C. (2001) Sampling Theory and Methods, Alpha Science International Ltd.,
6. Thomas Lumley (1996) Complex Surveys. A Guide to Analysis Using R, Wiley eastern Ltd.
7. Des Raj (1967) Sampling Theory. Tata McGraw Hill ,NewDelhi
8. Dean, A. and Voss, D. (1999) Design and Analysis of Experiments, Springer Texts in Statistics

Approval Date	
Version	
Approval by	
Implementation Date	

MAHATMA GANDHI UNIVERSITY

DA M 21 C09: APPLIED MULTIVARIATE ANALYSIS

SchoolName	**School of Data Analytics**
Programme	**Msc Data Science & analytics**
Course Name	**APPLIED MULTIVARIATE ANALYSIS**
Type of Course	Core
Course Code	**DA M 21 C09:**

Course Summary & Justification	1. To introduce multivariate data and associated concepts, distributions, testing and estimation, and the theory and applications of discriminant function and classification rules, principal components, canonical correlations, factor analysis etc.					
Total StudentLearning Time (SLT)	Learning Approach	Lecture	Tutorial	Practical	Others	Total Learning Hours
	Authentic learning Collaborative learning Independent learning	80	10	20	10	120
Pre-requisite	Univariate Theory of random variables and distributions, Matrix Theory, Spectral Decomposition, generalized inverses, quadratic forms					

COURSE OUTCOMES (CO)

CO No.	Expected Course Outcome	Learning Domains	PSO No.
1	Students are exposed to multivariate random vectors and their importance in data science and analytics.		
2	They are aware of various multivariate distributions and tests of hypotheses, MANOVA, Profile Analysis etc.		
3	They have studied various dimension reduction techniques like Principal Component Analysis, Canonical Correlation Analysis, Factor Analysis etc.		
4	They are aware of different discrimination as well as classification techniques, cluster analysis etc and can apply these for data analytics and data science as well as deep learning.		
Remember (R), Understand (U), Apply (A), Analyse (An), Evaluate (E), Create (C), Skill (S), Interest (I) and Appreciation (Ap)			

COURSE CONTENT

Module No	Module Content	CO	Hrs
1	Basic concepts on multivariate variable. Concept of random vector: Its expectation and dispersion (Variance-Covariance) matrix. Marginal and joint distributions. Conditional distributions and Independence of random vectors. Multinomial distribution. Multivariate normal distribution, Marginal and conditional distribution, characteristic function, additive property, mles of mean and dispersion matrix	1,5,6	20
2	Sample mean vector and its distribution, Hotelling's T^2 and Mahalanobis' D^2 statistics and applications. Tests of hypotheses about the mean vectors and covariance matrices for multivariate normal populations. Wishart distribution, Rao's U, Pillai's Trace statistics; Independence of sub vectors and sphericity test.	1,2,5,6	20

3.	Bayes' minimax and Fisher's criteria for discrimination between two multivariate normal populations. Sample discriminant function. Tests associated with discriminant functions. Probabilities of misclassification and their estimation. Discrimination for several multivariate normal populations. Multivariate analysis of variance (MANOVA) of one and two- way classified data. Multivariate analysis of covariance, illustrative numerical examples.	2,5,6	20
4	Principal components, sample principal components, asymptotic properties. Canonical variables and canonical correlations: definition, estimation, computations. Factor analysis: Orthogonal factor model, factor loadings, estimation of factor loadings, factor scores, cluster analysis-agglomerative and divisive techniques. Applications to real data sets and problems.	3,4,5.6	20
	Total Credits of the Course	3	80

Teachingand LearningApproach	**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 .
Assessment Types	**Mode of Assessment** **A. Continuous Internal Assessment (CIA);40 marks** 1. Two Internal Tests of maximum 20 marks 2. Seminar Presentation – a theme is to be discussed and identified to prepare a paper and present in the seminar Maximum marks 10 3. Write a detailed report on a given topic based on research findings and literature search – 10 marks **B.Semester End examination – 60 marks**

REFERENCES

1.Rencher, A. C. (2012) Methods of Multivariate Analysis.(3rd ed.) John Wiley.
2.Johnson R.A. and Wichern D.W. (2008) Applied Multivariate Statistical Analysis. 6th Edition, Pearson Education.
3. Anderson, T.W. 2009. An Introduction to Multivariate Statistical Analysis, 3rd Edition, John Wiley.
4. Everitt B, Hothorn T, 2011. An Introduction to Applied Multivariate Analysis with R, Springer.
5. Barry J. Babin, Hair, Rolph E Anderson, and William C. Blac, 2013, Multivariate Data Analysis, Pearson New International Edition,
6.Giri, N.C. 1977. Multivariate Statistical Inference. Academic Press.
7. Chatfield, C. and Collins, A.J. 1982. Introduction to Multivariate analysis. Prentice Hall
8.Srivastava, M.S. and Khatri, C.G. 1979. An Introduction to Multivariate Statistics. North Holland

Approval Date	
Version	
Approval by	
Implementation Date	

MAHATMA GANDHI UNIVERSITY

DA M 21 C10: APPLIED REGRESSION ANALYSIS

SchoolName	**School of Data Analytics**					
Programme	**M.Sc. Data Science & Analytics**					
Course Name	**APPLIED REGRESSION ANALYSIS**					
Type of Course	Core					
Course Code	**DA M 21 C10**					
Course Summary & Justification	To introduce the various concepts and techniques in regression analysis for modeling data and forecasting future values					
Total Student Learning Time (SLT)	Learning Approach	Lecture	Tutorial	Practical	Others	Total Learning Hours
	Eg. Authentic learning Collaborative learning Independent learning	80	10	20	10	120
Pre-requisite	Basic understanding of Regression					

COURSE OUTCOMES (CO)

CO No.	Expected Course Outcome	Learning Domains	PSO No.
1	The students have studied simple linear regression, multiple regression, residual analysis for fitting a suitable model to a given data and to check the suitability.	U/A	1,2. 4,5

2	They have studied necessary transformations and modifications to be made when model assumptions are violated.	A	1,2,4,5
3	They are capable of fitting logistic and Poisson models, orthogonal and polynomial models.	A/An	1,2,4,5
4	They have understood ridge regression, kernel regression, non-parametric regression etc.	A	1,2,4,5
Remember (R), Understand (U), Apply (A), Analyse (An), Evaluate (E), Create (C), Skill (S), Interest (I) and Appreciation (Ap)			

COURSE CONTENT

Module No	Module Content	CO	Hours
1	Introduction to regression analysis: overview and applications of regression analysis, major steps in regression analysis. Simple linear regression (Two variables): assumptions, estimation and properties of regression coefficients, significance and confidence intervals of regression coefficients, measuring the quality of the fit. Residual analysis, various types of residuals, Departures from underlying assumptions, Departures from normality, Diagnostics and remedies,	1	20
2	Multiple linear regression model: assumptions, ordinary least square estimation of regression coefficients, interpretation and properties of regression coefficient, significance and confidence intervals of regression coefficients. Mean Square error criteria, coefficient of determination, criteria for model selection; Need for transformation of variables; power transformation, Box-Cox transformation;, removal of heteroscedasticity and serial correlation, Leverage and influence. Effect of outliers.	1,2	20
3.	Generalized least squares and Weighted least squares.Polynomial regression models, Forward, Backward and Stepwise procedures. Nonparametric regression, Kernel regression, Loess, ridge regression,orthogonal polynomials, indicator variables**,** subset regression, stepwise regression, variable selection Robust regression.	3	20
4	Introduction to nonlinear regression, linearity transformations, logarithmic transformation, Least squares in the nonlinear case and estimation of parameters, Models for binary response variables, generalized linear models, estimation and diagnosis methods for Logistic and Poisson regressions. Prediction and residual analysis, Multinomial logistic regression, Random and mixed effect models, Multi-collinearity, sources, effects, tests.	3	20
	Total Credits of the Course	3	80

Books for Reference

Compulsory Reading:

1. D. C Montgomery, E.A Peck and G.G Vining (2003). Introduction to Linear Regression Analysis, John Wiley and Sons, Inc. NY,
2. S. Chatterjee and A. Hadi (2013) Regression Analysis by Example, 5th Ed., Wiley.
3. Seber, A.F. and Lee, A.J. (2003) Linear Regression Analysis, John Wiley,
4. Iain Pardoe (2012) Applied Regression Modeling, John Wiley and Sons, Inc,.
5. P. McCullagh, J.A. Nelder,(1989) Generalized Linear Models, Chapman & Hall,. John O. Rawlings, 6. Sastry G. Pantula, David A. Dickey (1998) Applied Regression Analysis, Second Edition, Springer.
7. Draper, N. and Smith, H. (2012) Applied Regression Analysis – John Wiley & Sons.

Teachingand LearningApproach	**Classroom Procedure (Mode of transaction)** Direct Instruction, 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
Assessment Types	**Mode of Assessment** **C. Continuous Internal Assessment (CIA): 40 marks** Internal Test -20 marks Assignment – Every student needs to write an assignment on a given topic based on the available published literature – 10 marks Seminar Presentation – A topic needs to be presented and discussed with the class- 10 marks **D. Semester End examination – 60 marks**

Approval Date	
Version	
Approval by	
Implementation Date	

MAHATMA GANDHI UNIVERSITY

DA M 21 C11: PROGRAMING IN R FOR DATA ANALYTICS

SchoolName	**School of Data Analytics**					
Programme	**M.Sc. Data Science & analytics**					
Course Name	**PROGRAMING IN R FOR DATA ANALYTICS**					
Type of Course	Core					
Course Code	**DA M 21 C11**					
Course Summary & Justification	To make students aware of R commands and programming and to impart training in R for Data Analytics using various techniques.					
Total StudentLearning Time (SLT)	Learning Approach	Lecture	Tutorial	Practical	Others	Total Learning Hours
	Authentic learning Collaborative learning Independent learning	60	10	40	10	120
Pre-requisite	**Basics of** R Programing					

COURSE OUTCOMES (CO)

CO No.	Expected Course Outcome	Learning Domains	PSO No.
1	Understood the various commands in R and can apply these for data analysis and analytics	E,U	1,3,4,5
2	Able to write programs in R for Data Analytics	U/ An,C	2,3,4,5
3	Experienced the importance of R in Data Analytics and can apply this for Data Analytics	R,U,S	2.3.4.5

Remember (R), Understand (U), Apply (A), Analyse (An), Evaluate (E), Create (C), Skill (S), Interest (I) and Appreciation (Ap)

COURSE CONTENT

Module No	Module Content	CO	Hrs

1	Download and install R – R IDE environments – Why R – Getting started with R – Vectors and Data Frames – Loading Data Frames – Data analysis with summary statistics and scatter plots – Summary tables - Working with Script Files Linear Regression – Introduction – Regression model for one variable regression – Selecting best model – Error measures SSE, SST, RMSE, R^2 – Interpreting R^2 – Multiple linear regression – Loess and ridge regression – Correlation – Recitation – A minimum of 3 data sets for practice. Logistic Regression – The Logit – Confusion matrix – sensitivity, specificity – ROC curve – Threshold selection with ROC curve – Making predictions – Area under the ROC curve (AUC) - Recitation – A minimum of 3 data sets for practice	1,2	20
2	Approaches to missing data – Data imputation – Multiple imputation – Classification and Regression Tress (CART) – CART with Cross Validation – Predictions from CART – ROC curve for CART – Random Forests – Building many trees – Parameter selection – K-fold Cross Validation – Recitation – A minimum of 3 data sets for practice	1,3	20
3.	Using text as data – Text analytics – Natural language processing – Bag of words – Stemming – word clouds – Recitation – min 3 data sets for practice – Time series analysis – Clustering – k-mean clustering – Random forest with clustering – Understanding cluster patterns – Impact of clustering – Heatmaps – Recitation – min 3 data sets for practice	1,3	20
4	Support Vector Machines – Gradient Boosting – Naïve Bayes – Bayesian GLM – GLMNET – Ensemble modeling – Experimenting with all of the above approaches with and without data imputation and assessing predictive accuracy – Recitation – minimum 3 data sets for practice.	1,2,3	20
	Total Credits of the Course	3	80

Teachingand LearningApproach	**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
Assessment Types	**Mode of Assessment** **A. Continuous Internal Assessment (CIA): 40 marks** 1. Internal Tests of maximum 20 marks 2. Seminar Presentation – a theme is to be discussed and identified to prepare a paper and present in the seminar Maximum marks 10 3. Write a detailed report on a given topic based on research findings and literature search

	– 10 marks B. **Semester End examination – 60 marks**

REFERENCES

1. Statistics : An Introduction Using R, Michael J. Crawley, WILEY, Second Edition, 2015.
2. Hands-on programming with R, Garrett Grolemund, O'Reilley, 1st Edition, 2014
3. R for everyone, Jared Lander, Pearson, 1st Edition, 2014

Approval Date	
Version	
Approval by	
Implementation Date	

MAHATMA GANDHI UNIVERSITY

DA M 21 C12 : DATA VISUALIZATION AN PRESENTATION

SchoolName	**School of Data Analytics**					
Programme	**M.Sc. Data Science & Analytics**					
Course Name	**DATA VISUALIZATION AND PRESENTATION**					
Type of Course	Core					
Course Code	**DA M 21 C12**					
Course Summary & Justification	To make students familiar with data visualization techniques for displaying data.					
Total Student Learning Time (SLT)	Learning Approach	Lecture	Tutorial	Practical	Others	Total Learning Hours

	Authentic learning Collaborative learning Independent learning	80	10	20	10	120
Pre- requisites	Basic understanding of Data Visualization					

COURSE OUTCOMES (CO)

CO No.	Expected Course Outcome	Learning Domains	PSO No.
1	Able to use visualization techniques for multidimensional visualization.	R,U	1,2
2	Able to use visualization techniques for information visualization applications and systems	A,U	1,3
3	Understand visualization packages, grammar of graphics using R etc.	U,A,S	1,2,3
4	Understand different visualization tools and methods especially tableau.	U,R,A	1,2,3
Remember (R), Understand (U), Apply (A), Analyse (An), Evaluate (E), Create (C), Skill (S), Interest (I) and Appreciation (Ap)			

COURSE CONTENT

Module No	Module Content	CO	Hrs
1	Purpose of visualization, visual perception, cognitive issues – evaluation as well as other theory and design principles behind information visualization, Multidimensional visualization, tree visualization, graph visualization. Time series data visualization techniques.	1	20
2	Understanding analytics output and their usage, basic interaction techniques such as selection and distortion, evaluation, Examples of information visualization applications and systems, user tasks and analysis- visualization packages	2	20
3.	Grammar of graphics using R-Construct/Deconstruct a graphic into a data- order of accuracy of perceptual tasks and its impact and Case study presentations and lab based on R package of Data Visualizations.	1,3	20
4	Introduction to Tableau, installing tableau, connecting data with tableau and creating data visualizations, Charts and graphs, Working with calculations and expressions, Dashboard and stories.	3,4	20
	Total Credits of the Course	3	80

Teaching and Learning Approach	**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	
Assessment Types	**Mode of Assessment** **A. Continuous Internal Assessment (CIA):40 marks** 1. Internal Tests of maximum 20 marks 2. Seminar Presentation – a theme is to be discussed and identified to prepare a paper and present in the seminar Maximum marks 10 3. Write a detailed report on a given topic based on research findings and literature search – 10 marks **B. Semester End examination – 60 marks**	

REFERENCES

1. Wickham, H. (2016). Ggplot2: Elegant Graphics for Data Analysis. Springer.2nd Edition
2. Keen, K. J. (2010). Graphics for Statistics and Data Analysis with R. CRC Press.
3. Buja, A., Swayne, D. F. & Cook, D., (2007). Interactive and Dynamic Graphics for Data Analysis: with R and Ggobi. Springer Science & Business Media.
4. Dalgaard, P. (2008). Introductory Statistics with R. Springer
6. Marleen Meier ; David Baldwin (2021) Mastering Tableau 2021: Implement Advanced Business Intelligence Techniques and Analytics with Tableau, Edition 3.
7. S. Acharya ;S. Chellappan (2019) Pro Tableau: A Step by step Guide, Apress
6. Murrell, P. (2016). R graphics. CRC Press.
7. Cleveland, W. S. (1993). Visualizing data. Hobart Press.
9. Cheshire, CT: Graphics press.
10. Tufte, E., & Graves-Morris, P. (2014). The visual display of quantitative information.; 1983.
11.Verzani, J. (2014). Using R for introductory statistics. CRC Press.
12.Murrell, P. (2016). R graphics. CRC Press.

Approval Date	
Version	
Approval by	
Implementation Date	

ELECTIVE-1

ELECTIVE-2

MAHATMA GANDHI UNIVERSITY

DA M 21 C13 : PRACTICAL (DATA ANALYTICS LAB)

SchoolName	**School of Data Analytics**					
Programme	**M.Sc. Data Science & Analytics**					
Course Name	**PRACTICAL (DATA ANALYTICS LAB)**					
Type of Course	Core					
Course Code	**DA M 21 C13**					
Course Summary & Justification	It is meant for enhancing their skills in data analysis, data science and data analytics dealing with original data sets and big data. Case Studies, Presentations, Viva Voce Practical Records are to be maintained countersigned by the concerned faculty member and submitted at the end of the semester.					
Total StudentLearning Time (SLT)	Learning Approach	Lecture	Tutorial	Practical	Others	Total Learning Hours
	Authentic learning Collaborative learning Independent learning	20	20	60	20	120
Pre-requisites						

COURSE OUTCOMES (CO)

CO No.	Expected Course Outcome	Learning Domains	PSO No.
1	Enhancing their skills in data analysis, data science and data analytics dealing with original data sets and big data, case studies and presentations.	U,A,E,C,S	1,2,3,4,5

Remember (R), Understand (U), Apply (A), Analyse (An), Evaluate (E), Create (C), Skill (S), Interest (I) and Appreciation (Ap)

Teachingand Learning Approach	**Laboratory Procedure (Mode of transaction)** Direct Instruction: Lecture, Explicit Teaching, Demonstration, Hands on experimental sessions with original data sets and big data, Skill acquisition by laboratory training, case studies, presentations and discussions, viva voce etc.
Assessment Types	**Mode of Assessment: It is fully based on Continuous Internal Assessment** **Continuous Internal Assessment (CIA)-100** 1. Internal Laboratory Skill Tests -40 marks 2. Records, Accuracy, Lab discipline- 20 marks 3. Presentations, Viva-Voce -20 marks 4. Regularity, Punctuality, Attendance-20 marks

Approval Date	
Version	
Approval by	
Implementation Date	

SEMESTER III COURSES TOTAL CREDITS: 22

Course Code	Course Title	Teaching L+T+P	Credits
DA M 21 C14	MACHINE LEARNING	3+0+2	3
DA M 21 C15	STOCHASTIC MODELING & TIME SERIES FORECASTING	3+0+2	3
DA M 21 C16	ARTIFICIAL INTELLIGENCE	3 +0+2	3
	ELECTIVE 3	3 +0+2	3
	ELECTIVE 4	3 +0+2	3
DA M1 21 C17	PRACTICAL – DATA SCIENCE / ANALYTICS COMPUTER LAB	2 +0+3	3
	OPEN COURSE	4 +0+0	4
	Total Credits		22

MAHATMA GANDHI UNIVERSITY

DA M 21 C14: MACHINE LEARNING

School Name	**School of Data Analytics**					
Programme	**M.Sc. Data Science & Analytics**					
Course Name	**DA M 21 C14: MACHINE LEARNING**					
Type of Course	Core					
Course Code	**DA M 21 C14**					
Course Summary & Justification	To familiarize the students with various aspects of machine learning techniques and their applications.					
Total StudentLearningTime (SLT)	Learning Approach	Lecture	Tutorial	Practical	Others	Total LearningHours
	Authentic learning Collaborative learning Independent learning	60	10	30	20	120
Pre-requisites	Basic concepts in Linear Algebra and Statistics					

COURSE OUTCOMES (CO)

CO No.	Expected Course Outcome	Learning Domains	PSO No.
1	On completing this course, the student will be able to understood different techniques such as unsupervised learning,	U,A	1,2,3
2	Understand dimensionality reduction,	U,A,E	1,3,4,5
3	Familiarize PCA, SVM,	U,S,C	1,3,4,5
4.	Understand Discriminant function, multilayer preceptors, cluster analysis etc	U,A,An	1,3,4,5
Remember (R), Understand (U), Apply (A), Analyse (An), Evaluate (E), Create (C), Skill (S), Interest (I) and Appreciation (Ap)			

COURSE CONTENT

Module No	Module Content	CO	Hrs
1	Machine Learning- Examples of Machine Applications – Learning Associations – Classification- Regression- Unsupervised Learning- Reinforcement Learning. Supervised Learning: Learning class from examples- Probably Approximately Correct (PAC) Learning- Noise-Learning Multiple classes. Regression – Model Selection and Generalization. Introduction to Parametric methods- Maximum Likelihood Estimation: Bernoulli Density- Multinomial Density-Gaussian Density, Nonparametric Density Estimation: Histogram Estimator-Kernel Estimator-K-Nearest Neighbor Estimator .	1,2	15
2	Dimensionality Reduction: Introduction- Subset Selection- Principal Component Analysis, Feature Embedding-Factor Analysis- Singular Value Decomposition- Multidimensional Scaling- Linear Discriminant Analysis- Bayesian Decision Theory. Linear Discrimination: Introduction- Generalizing the Linear Model- Geometry of the Linear Discriminant- Pairwise Separation-Gradient Descent-Logistic Discrimination. Optical separating hyper plane – v-SVM, kernel tricks – vertical kernel- vertical kernel- defining kernel- multiclass kernel machines- one-class kernel machines	2,3	15
3.	Multilayer perceptron, Introduction, training a perceptron-learning Boolean functions- multilayer perceptron- back propagation algorithm- training procedures. Combining Multiple Learners, Rationale-Generating diverse learners- Model combination schemes- voting, Bagging- Boosting- fine tuning an Ensemble.	1,2,3	15

4	Cluster Analysis, Introduction-Mixture Densities, K-Means Clustering- Expectation-Maximization algorithm- Mixtures of Latent Variable Models-Supervised Learning after Clustering- Spectral Clustering- Hierarchical Clustering- Divisive Clustering- Choosing the number of Clusters.	2,3	15
	Total Credits of the Course	3	60

Teaching and Learning Approach	**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
Assessment Types	**Mode of Assessment** **A. Continuous Internal Assessment (CIA)-40 marks** 1. Internal Tests of maximum 20 marks 2. Seminar Presentation – a theme is to be discussed and identified to prepare a paper and present in the seminar Maximum marks 10 3. Write a detailed report on a given topic based on research findings and literature search – 10 marks **B. Semester End examination – 60 marks**

REFERENCES

1. E. Alpaydin (2014) Introduction to Machine Learning, 3rd Edition, MIT Press.
2. Frank Kane (2012) Data Science and Machine Learning, Manning Publications.
3. C.M.Bishop, Pattern Recognition and Machine Learning, Springer.
4. T. Hastie, R. Tibshirani and J. Friedman (2016) The Elements of Statistical Learning: Data Mining, Inference and Prediction, Springer, 2nd Edition,2009.

Approval Date	
Version	
Approval by	
Implementation Date	

MAHATMA GANDHI UNIVERSITY

DA M 21 C15: STOCHASTIC MODELLING AND TIME SERIES FORECASTING

School Name	**School of Data Analytics**					
Programme	**M.Sc. Data Science & Analytics**					
Course Name	**DA M 21 C15: STOCHASTIC MODELLING AND TIME SERIES FORECASTING**					
Type of Course	Core					
Course Code	**DA M 21 C15**					
Course Summary & Justification	To introduce the basics of stochastic processes and modeling as well as enable the students to analyze time series data and apply suitable techniques to model them and forecast future values.					
Total StudentLearning Time (SLT)	Learning Approach	Lecture	Tutorial	Practical	Others	Total Learning Hours
	Authentic learning Collaborative learning Independent learning	80	10	20	10	120
Pre-requisites	Basics of Probability Theory and Distribution Theory					

COURSE OUTCOMES (CO)

CO No.	Expected Course Outcome	Learning Domains	PSO No.
1	On completing this course, the student will be aware of various stochastic models.	U,A,E,	1,4,5
2	Understand time series models and can apply these to model data for predicting future values to make appropriate planning and decision making	U,A,S	1,2,3
Remember (R), Understand (U), Apply (A), Analyse (An), Evaluate (E), Create (C), Skill (S), Interest (I) and Appreciation (Ap)			

COURSE CONTENT

Module No	Module Content	CO	Hrs

1	Introduction to stochastic processes:- classification of stochastic processes, wide sense and strict sense stationary processes, processes with stationary independent increments, Markov process, Markov chains- transition probability matrices, Chapman-Kolmogorov equation, first passage probabilities, recurrent and transient states, mean recurrence time, stationary distributions, limiting probabilities, Random walk, random walk approximation of Brownian motion and diffusion process, Galton-Watson branching process, generating function relations, mean and variance, extinction probabilities.	1,2	20
2	Continuous time Markov chains, Poisson processes, properties, inter-arrival time distribution pure birth processes and the Yule processes, birth and death processes, Kolmogorov differential equations, linear growth process with immigration, steady-state solutions of Markovian queues - M/M/1, M/M/s, M/MM/∞ models, Renewal processes – basic concepts, examples, Poisson process viewed as a renewal process.	2,3	20
3.	Time series data, examples, Time series as stochastic process, Additive and multiplicative models, stationary time series- covariance stationarity, Modeling Time Series Data, Exponential Smoothing Methods - First-Order Exponential Smoothing, Second and higher Order Exponential Smoothing, Forecasting, Exponential Smoothing for Seasonal Data, Exponential Smoothers.	1,2,3	20
4	Time series modeling, Autocorrelation function (ACF), partial auto correlation function(PACF), correlogram, AR, MA, ARMA, ARIMA Models, Yule- Walker equations, Box-Jenkins Model fitting and diagnostics.. Forecasting future values, Non-linear time series models, ARCH and GARCH Process, GARCH process for modeling volatility, Non-Gaussian Time series modeling, Multivariate Time series analysis.	3	20
	Total Credits of the Course	3	80

TeachingandLearningApproach	**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

Assessment Types	Mode of Assessment **A. Continuous Internal Assessment (CIA)-40 marks** 1. Internal Tests of maximum 20 marks 2. Seminar Presentation – a theme is to be discussed and identified to prepare a paper and present in the seminar Maximum marks 10 3. Write a detailed report on a given topic based on research findings and literature search – 10 marks **B. Semester End examination – 60 marks**

REFERENCES

1. Karlin S. and Taylor H.M. (1975) A First Course in Stochastic Processes, Second Edition, Academic Press, New-York.
2. Montgomery, Douglas C., Cheryl L. Jennings, and Murat Kulahci. *Introduction to Time Series Analysis and Forecasting*. John Wiley & Sons, 2015.
3. Medhi J. (2017) Stochastic Processes, Second Edition , Wiley Eastern, New Delhi
4. Ross S.M. (2007) Stochastic Processes. Second Edition, Wiley Eastern, New Delhi
5. Brockwell P.J and Davis R.A. (2002) Introduction to Time Series and Forecasting Second edition, Springer-Verlag.
6. Ruey S. Tsay (2005). Analysis of Financial Time Series, Second Ed. Wiley & Sons

Approval Date	
Version	
Approval by	
Implementation Date	

	MAHATMA GANDHI UNIVERSITY
	DA M 21 C16: ARTFICIAL INTELLIGENCE

School Name	**School of Data Analytics**					
Programme	**M.Sc. Data Science & Analytics**					
Course Name	**DA M 21 C16: ARTFICIAL INTELLIGENCE**					
Type of Course	**Core**					
Course Code	**DA M 21 C16**					
Course Summary & Justification	To understand thinking and intelligence in ways that enable the construction of computer systems that are able to reason in uncertain environments					
Total StudentLearning Time (SLT)	Learning Approach	Lecture	Tutorial	Practical	Others	Total Learning Hours
	Authentic learning Collaborative learning Independent learning	60	10	30	20	120
Pre-requisites	Basics of Statistics					

COURSE OUTCOMES (CO)

CO No.	Expected Course Outcome	Learning Domains	PSO No.
1	**On completing this course, the student will be able t**o articulate and exemplify the basic knowledge artificial intelligence,.	**U/E,A**	1,2,3,4,5
2.	Understand the basics of knowledge representation	**U,A,C**	1,2,3,4,5
3.	can use AI programming languages and the methods of AI implementation	**U,A,S**	1,2,3
4.	can recommend AI strategies based on applications	**U,A,C**	1,2,3,4
Remember (R), Understand (U), Apply (A), Analyse (An), Evaluate (E), Create (C), Skill (S), Interest (I) and Appreciation (Ap)			

COURSE CONTENT

Module No	Module Content	CO	Hrs
1	Artificial Intelligence - Introduction, AI Problems, AI Techniques, The Level of the Model, Criteria For Success. Defining the Problem as a State Space Search, Problem Characteristics, Production Systems, Search: Issues in The Design of Search Programs, Un-Informed Search, BFS, DFS; Heuristic Search Techniques: Generate-And-Test, Hill Climbing, Best-First Search, A*Algorithm, Problem Reduction, AO*Algorithm, Constraint Satisfaction, Means-Ends Analysis. Knowledge Representation: Procedural Vs Declarative Knowledge, Representations & Approaches to Knowledge Representation, Forward Vs Backward Reasoning, Matching Techniques, Partial Matching, Fuzzy Matching Algorithms and RETE Matching Algorithms;	**1,2**	**15**
2.	Logic Based Programming-AI Programming languages: Overview of LISP, Search Strategies in LISP, Pattern matching in LISP , An Expert system Shell in LISP, Over view of Prolog, Production System using Prolog; Symbolic Logic: Propositional Logic, First Order Predicate Logic: Representing Instance and Relationships, Computable Functions and Predicates, Syntax & Semantics of FOPL, Normal Forms, Unification &Resolution, Representation Using Rules, Natural Deduction; Structured Representations of Knowledge: Semantic Nets, Partitioned Semantic Nets, Frames, Conceptual Dependency, Conceptual Graphs, Scripts, CYC;.	**2,3**	**15**
3.	Reasoning under Uncertainty: Introduction to Non-Monotonic Reasoning, Truth Maintenance Systems, Logics for non-monotonic Reasoning, Model and Temporal Logics; Statistical Reasoning: Bayes' Theorem, Certainty Factors and Rule-Based Systems, Bayesian Probabilistic Inference, Bayesian Networks, Dempster - Shafer Theory, Fuzzy Logic: Crisp Sets ,Fuzzy Sets, Fuzzy Logic Control, Fuzzy Inferences & Fuzzy Systems.	**2,3**	**15**
4	Experts Systems: Overview of an Expert System, Structure of an Expert Systems, Different Types of Expert Systems-Rule Based, Model Based, Case Based and Hybrid Expert Systems, Knowledge	**3**	**15**

	Ac - quisition and Validation Techniques, Black Board Architecture, Knowledge Building System Tools, Expert System Shells, Fuzzy Expert systems.		
	Total Credits of the Course	**3**	**60**

Teachingand LearningApproach	**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
Assessment Types	**Mode of Assessment** **A. Continuous Internal Assessment (CIA)-40 marks** 1. Internal Tests of maximum 20 marks 2. Seminar Presentation – a theme is to be discussed and identified to prepare a paper and present in the seminar Maximum marks 10 3. Write a detailed report on a given topic based on research findings and literature search – 10 marks B. **Semester End examination – 60 marks**

REFERENCES

1. George F Luger (2016) Artificial Intelligence, Pearson EducationPublications
2. Elaine Rich and Knight (2017) Artificial Intelligence, Mc Graw-HillPublications
3. Patterson, D.W.(2005) Introduction to Artificial Intelligence &Expert Systems, Prentice Hall of India
4. Weiss.G, (2000) Multi Agent Systems- A Modern Approach to Distributed Artificial Intelligence, MIT Press.
5. Russell S. and Norvig, P.(2010) Artificial Intelligence : A Modern Approach, Prentice Hall of India

Approval Date	
Version	
Approval by	
Implementation Date	

ELECTIVE -3

ELECTIVE-4

	MAHATMA GANDHI UNIVERSITY
	DA M 21 C17 : PRACTICAL (DATA SCIENCE / ANALYTICS LAB)

SchoolName	**School of Data Analytics**					
Programme	**M.Sc. Data Science & Analytics**					
Course Name	**PRACTICAL (DATA SCIENCE / ANALYTICS LAB)**					
Type of Course	Core					
Course Code	**DA M 21 C17**					
Course Summary & Justification	It is meant for enhancing their skills in data analysis, data science and data analytics dealing with original data sets and big data. Case Studies, Presentations, Viva Voce etc with respect to the courses they have studied during this semester. Practical Records are to be maintained countersigned by the concerned faculty member and submitted at the end of the semester.					
Total StudentLearning Time (SLT)	Learning Approach	Lecture	Tutorial	Practical	Others	Total Learning Hours
	Authentic learning Collaborative learning Independent learning	20	20	60	20	120
Pre-requisites	Basics of Python and R Programing					

COURSE OUTCOMES (CO)

CO No.	Expected Course Outcome	Learning Domains	PSO No.
1	Enhancing their skills in data analysis, data science and data analytics dealing with original data sets and big data, case studies and presentations utilizing the knowledge they have obtained by studying the courses during this semester.	U,A,E,C,S	1,2,3,4,5
Remember (R), Understand (U), Apply (A), Analyze (An), Evaluate (E), Create (C), Skill (S), Interest (I) and Appreciation (Ap)			

Teachingand Learning Approach	**Laboratory Procedure (Mode of transaction)** Direct Instruction: Lecture, Explicit Teaching, Demonstration, Hands on experimental sessions with original data sets and big data, Skill acquisition by laboratory training, case studies, presentations and discussions, viva voce etc.

Assessment Types	**Mode of Assessment:** The Evaluation is fully based on Continuous Internal Assessment **Continuous Internal Assessment (CIA)-100** 1. Internal Laboratory Skill Tests -40 marks 2. Records, Accuracy, Lab discipline- 20 marks 3. Presentations, Viva-Voce -20 marks 4. Regularity, Punctuality, Attendance-20 marks

Approval Date	
Version	
Approval by	
Implementation Date	

OPEN COURSE

SEMESTER 1V COURSES TOTAL CREDITS : 21

Course Code	Course Title	Teaching L+T+P	Credits
	ELECTIVE 5	**3 +0+2**	**3**
	ELECTIVE 6	**3 +0+2**	**3**
DA M1 21 C18	**PROJECT / DISSERTATION**	**24**	**15**
	Total Credits		**21**

ELECTIVE COURSES

MAHATMA GANDHI UNIVERSITY

DA M 21 E01: NEURAL NETWORKS AND DEEP LEARNING

School Name	**School of Data Analytics**					
Programme	**M. Sc. Data Science & Analytics**					
Course Name	**NEURAL NETWORKS AND DEEP LEARNING**					
Type of Course	**Elective**					
Course Code	**DA M 21 E01**					
Course Summary & Justification	To introduce various concepts like ANN, CNN,SNN, RBN, ELU,RFN,RBM etc for deep learning.					
Total Student Learning Time (SLT)	Learning Approach	Lecture	Tutorial	Practical	Others	Total Learning Hours
	Authentic learning Collaborative learning Independent learning	60	20	20	20	120
Pre-requisites	Basics of Neural Networks					

COURSE OUTCOMES (CO)

CO No.	Expected Course Outcome	Learning Domains	PSO No.
1	Aware of different types of neural networks	**R,U,A**	1,2
2.	Understand the principles of soft computing.	U,A,An	**1,2,3**
3.	Able to carry out deep learning using Python.	**R,A,C**	**1,2,3**

Remember (R), Understand (U), Apply (A), Analyse (An), Evaluate (E), Create (C), Skill (S), Interest (I) and Appreciation (Ap)

COURSE CONTENT

Module No	Module Content	CO	Hrs
1	Neural Networks-Application Scope of Neural Networks-Fundamental Concept of ANN: The Artificial Neural Network-Biological Neural Network-Comparison between Biological Neuron and Artificial Neuron-Evolution of Neural Network. Basic models of ANN- Learning Methods- Activation Functions- Importance Terminologies of ANN.	**1**	**10**

2.	Shallow neural networks- Perceptron Networks-Theory- Perceptron Learning Rule Architecture-Flowchart for training Process- Perceptron Training Algorithm for Single and Multiple Output Classes. Back Propagation Network- Theory- Architecture-Flowchart for training process- Training Algorithm-Learning Factors for Back-Propagation Network. Radial Basis Function Network RBFN: Theory, Architecture, Flowchart and Algorithm.	**2,3**	**20**
3.	Conventional Neural Networks (CNN) – Introduction – Components of CNN Architecture – Rectified Linear Unit (ReLU) Layer – Exponential Linear Unit (ELU, or SELU) – Unique Properties of CNN –Architectures of CNN – Applications of CNN. Recurrent Neural Network-introduction- The Architecture of Recurrent Neural Network- The Challenges of Training Recurrent Networks- Echo-State Networks- Long Short-Term Memory (LSTM) – Applications of RNN.	**2,3**	**20**
4	Auto encoder- Introduction – Features of Auto encoder Types of Auto encoder Restricted Boltzmann Machine- Boltzmann Machine – RBM Architecture –Example – Types of RBM.	**2,3**	**10**
	Total Credits of the Course	**3**	**60**

Teachingand LearningApproach	**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
Assessment Types	**Mode of Assessment** **A. Continuous Internal Assessment (CIA)-40 marks** 1. Internal Tests of maximum 20 marks 2. Seminar Presentation – a theme is to be discussed and identified to prepare a paper and present in the seminar Maximum marks 10 3. Write a detailed report on a given topic based on research findings and literature search – 10 marks B. **Semester End examination – 60 marks**

REFERENCES

1.S.N.Sivanandam, S. N. Deepa, Principles of Soft Computing, Wiley-India, 3rd Edition, 2018.

2. S Lovelyn Rose, L Ashok Kumar, D Karthika Renuka, Deep Learning Using Python, Wiley-India, 1st Edition, 2019.

3. Frank Kane (2018) Machine Learning, Data Science and Deep Learning with Python, Manning Publications

Essential Reading / Recommended Reading

4. Charu C. Aggarwal, Neural Networks and Deep Learning, Springer, September 2018.

5. Francois Chollet, Deep Learning with Python, Manning Publications; 1st edition, 2017

6. John D. Kelleher, Deep Learning (MIT Press Essential Knowledge series), The MIT Press, 2018

Approval Date	
Version	
Approval by	
Implementation Date	

MAHATMA GANDHI UNIVERSITY

DA M 21 E 02: BAYESIAN INFERENCE & COMPUTING

School Name	**School of Data Analytics**					
Programme	**M.Sc. Data Science & Analytics**					
Course Name	**BAYESIAN INFERENCE & COMPUTING**					
Type of Course	Elective					
Course Code	**DA M 21 E 02**					
Course Summary & Justification	To introduce basic concepts and tools like prior information, posterior information in Bayesian Inference and Computing					
Total Student Learning Time (SLT)	Learning Approach	Lecture	Tutorial	Practical	Others	Total Learning Hours
	Authentic learning Collaborative learning Independent learning	60	20	20	20	120
Pre-requisite	**Knowledge in Basic Probability Theory and Distribution Theory**					

COURSE OUTCOMES

No.	Expected Course Outcome	Learning Domains	PSO No.

1	Students are aware of difference between subjective and frequentist probability, prior and posterior distributions	R/U/A	1,4,5
2	Students have understood the concepts prior information, posterior information, likelihood, loss function, minimax decision theory	U/A	1,4.5
3	Students can carry out Bayesian Estimation of parameters, construct credible intervals, HPD confidence intervals etc using different loss functions	U/An/Ap	1,4,5
4	Students can carry out Bayesian Testing using different approaches.	An/Ap	1,4,5
5	Students are able to apply the Bayesian computational tools like MCMC, Gibbs Sampler, Monte Carlo Methods, EM Algorithm etc for practical purposes in data science and analytics using prior information and sample data and R Program.	R/U/A/An /Ap	1,4,5

Remember (R), Understand (U), Apply (A), Analyse (An), Evaluate (E), Create (C), Skill (S), Interest (I) and Appreciation (Ap)

Module No	**Module Content**	**CO**	**Hours**
1	Basics of minimaxity: subjective and frequentist probability, Bayesian inference, Bayesian estimation, prior distributions, posterior distribution, loss function, principle of minimum expected posterior loss, SELF, AELF, 0-1 loss function, LINEX, quadratic and other common loss functions, Advantages of being a Bayesian.	1,2	15 hrs
2	Robustness and sensitivity, classes of priors, conjugate class, neighborhood class, One parameter exponential family, density ratio class, different methods of objective priors: Jeffrey's prior, probability matching prior, conjugate priors and mixtures, posterior robustness: measures and techniques, , HPD confidence intervals, credible intervals, prediction of a future observation.	1,2,4	15 hrs
3.	Basics of decision theory, multi-parameter models, Multivariate models, linear regression, asymptotic approximation to posterior distributions; Selection criteria and testing of hypothesis based on objective probabilities and Bayes' factors, Jeffrey's Approach and large sample methods: limit of posterior distribution, consistency of posterior distribution, asymptotic normality of posterior distribution.	1,2	15 hrs
4	Analytic approximation, E- M Algorithm, Monte Carlo sampling, Importance Sampling, Markov Chain Monte Carlo Methods, Metropolis – Hastings Algorithm, Gibbs sampling, Bootstrap and Jackknife, examples, convergence issues, applications	3	15 hrs
	Total Credits of the Course	**3**	**60 hrs**

Teachingand Learning Approach	**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
Assessment Types	**Mode of Assessment** **A. Continuous Internal Assessment (CIA)-40 marks** 1. Internal Tests of maximum -20 marks 2. Seminar Presentation – a theme is to be discussed and identified to prepare a paper and present in the seminar Maximum marks 10 3. Write a detailed report on a given topic based on research findings and literature search – 10 marks B. **Semester End examination – 60 marks**

Books for Reference

1. Albert Jim (2009) Bayesian Computation with R, second edition, Springer, New York

2. Bolstad W. M. and Curran, J.M. (2016) Introduction to Bayesian Statistics 3rd Ed. Wiley, New York

3. Christensen R. Johnson, W. Branscum A. and Hanson T.E. (2011) Bayesian Ideas and Data Analysis : An Introduction for Scientists and Statisticians, Chapman and Hall, London

4. A. Gelman, J.B. Carlin, H.S. Stern and D.B. Rubin (2004). Bayesian Data Analysis, 2nd Ed. Chapman & Hall

Essential Reading / Recommended Reading

1. Congdon P. (2006) Bayesian Statistical Modeling, Wiley, New York.

2. Ghosh, J.K. Delampady M. and T. Samantha (2006). An Introduction to Bayesian Analysis: Theory and Methods, Springer, New York.

3. Lee P.M. (2012) Bayesian Statistics: An Introduction-4th Ed. Hodder Arnold, New York.

4. Rao C.R. , Day D. (2006) Bayesian Thinking, Modeling and Computation, Handbook of Statistics, Vol.25.

Approval Date	
Version	
Approval by	
Implementation Date	

MAHATMA GANDHI UNIVERSITY

DAM21E03: INTRODUCTION TO BIG DATA AND HADOOP

School Name	**School of Data Analytics**					
Programme	**M.Sc. Data Science & analytics**					
Course Name	**INTRODUCTION TO BIG DATA AND HADOOP**					
Type of Course	Elective					
Course Code	**DAM21E03**					
Course Summary & Justification	To make the students aware about the different techniques for big data analytics.					
Total StudentLearning Time (SLT)	Learning Approach	Lecture	Tutorial	Practical	Others	Total Learning Hours
	Authentic learning Collaborative learning Independent learning	60	20	20	20	120
Pre-requisite	**Basics Knowledge of Bigdata**					

COURSE OUTCOMES (CO)

CO No.	Expected Course Outcome	Learning Domains	PSO No.
1	Enabled to use Hadoop	U,A	1,2,3
2	Understand RDBMS	E,U,A	1,2,4,5
3	Familiarized Mapreduce, HDFS	E,U,A	1,2,4,5
4	Undersatnd HIVE & PIG etc for big data analytics	E,U,A	1,2,4,5
Remember (R), Understand (U), Apply (A), Analyse (An), Evaluate (E), Create (C), Skill (S), Interest (I) and Appreciation (Ap)			

COURSE CONTENT

Module No	Module Content	CO	Hrs

1	Distributed file system – Big Data and its importance, Four Vs, Drivers for Big data, Big data analytics, Big data applications, Algorithms using map reduce, Matrix-Vector Multiplication by Map Reduce. Apache Hadoop– Moving Data in and out of Hadoop – Understanding inputs and outputs ofMapReduce – Data Serialization, Problems with traditional large-scale systems-Requirements for a new approach- Hadoop – Scaling-Distributed Framework- Hadoop v/s RDBMS-Brief history of Hadoop. Installing and Configuring Hadoop; Configurations of Hadoop: Hadoop Processes (NN, SNN, JT, DN, TT)-Temporary directory – UI-Common errors when running Hadoop cluster, solutions. Setting up Hadoop on a local Ubuntu host: Prerequisites, downloading Hadoop, setting up SSH, configuring the pseudo-distributed mode, HDFS directory, NameNode, Examples of MapReduce, Using Elastic MapReduce, Comparison of local versus EMR Hadoop. **Lab Exercise** 1. Word count application in Hadoop. 2. Sorting the data using MapReduce. 3. Finding max and min value in Hadoop.	1,2	15
2	Understanding MapReduce: Key/value pairs, Hadoop Java API for MapReduce, Writing MapReduce programs, Hadoop-specific data types, Input/output. Developing MapReduce Programs: Using languages other than Java with Hadoop, Analyzing a large dataset. Advanced MapReduce Techniques: Simple, advanced, and in-between Joins, Graph algorithms, using language-independent data structures. Hadoop configuration properties – Setting up a cluster, Cluster access control, managing the NameNode, Managing HDFS, MapReduce management, Scaling. **Lab Exercise:** 1. Implementation of decision tree algorithms using MapReduce. 2. Implementation of K-means Clustering using MapReduce. 3. Generation of Frequent Item set using MapReduce.	2,3	15
3.	Hadoop Streaming - Streaming Command Options – Specifying a Java Class as the Mapper/Reducer – Packaging Files With Job Submissions – Specifying Other Plug-ins for Jobs. HIVE & PIG ; Architecture, Installation, Configuration, Hive vs RDBMS, Tables, DDL & DML, Partitioning & Bucketing, Hive Web Interface, Pig, Use case of Pig, Pig Components, Data Model, Pig Latin. **Lab Exercise:** 1. Count the number of missing and invalid values through joining two large given datasets. 2. Using hadoop's map-reduce, Evaluating Number of Products Sold in Each Country in the online shopping portal. Dataset	1,4	15

	is given. 3. Analyze the sentiment for product reviews, this work proposes a MapReduce technique provided by Apache Hadoop. 4. Trend Analysis based on Access Pattern over Web Logs using Hadoop. 5. Service Rating Prediction by Exploring Social Mobile Users Geographical Locations.		
4	**Module 4** Hbase RDBMS VsNoSQL, Hbasics, Installation, Building an online query application – Schema design, Loading Data, Online Queries, Successful service. Hands On: Single Node Hadoop Cluster Set up in any cloud service provider- How to create instance. How to connect that Instance Using putty.Installing Hadoop framework on this instance. Run sample programs which come with Hadoop framework. Lab Exercise: 1. Big Data Analytics Framework Based Simulated Performance and Operational Efficiencies Through Patient Records in Hospital System.	2,3,4	15
	Total Credits of the Course	3	60

Teachingand LearningApproach	**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
Assessment Types	**Mode of Assessment** **C. Continuous Internal Assessment (CIA)-40 marks** 1. Internal Tests of maximum -20 marks 2. Seminar Presentation – a theme is to be discussed and identified to prepare a paper and present in the seminar Maximum marks 10 3. Write a detailed report on a given topic based on research findings and literature search – 10 marks D. **Semester End examination – 60 marks**

REFERENCES

1. Boris lublinsky, Kevin t. Smith, Alexey Yakubovich, Professional Hadoop Solutions, Wiley, 2015.
2. Tom White, Hadoop: The Definitive Guide, O'Reilly Media Inc., 2015.
3. Garry Turkington, Hadoop Beginner's Guide, Packt Publishing, 2013.
4. Pethuru Raj, Anupama Raman, DhivyaNagaraj and Siddhartha Duggirala, High-Performance Big-Data Analytics: Computing Systems and Approaches, Springer, 2015.
5. Jonathan R. Owens, Jon Lentz and Brian Femiano, Hadoop Real-World Solutions Cookbook, Packt Publishing, 2013.

Approval Date	
Version	
Approval by	
Implementation Date	

MAHATMA GANDHI UNIVERSITY

DAM21 E04: INFORMATION RETRIEVAL TECHNIQUES

SchoolName	**School of Data Analytics**					
Programme	**M. Sc. Data Science & Analytics**					
Course Name	**INFORMATION RETRIEVAL TECHNIQUES**					
Type of Course	Elective					
Course Code	**DAM21 E04**					
Course Summary & Justification	To introduce various information retrieval techniques					
Total StudentLearning Time (SLT)	Learning Approach	Lecture	Tutorial	Practical	Others	Total LearningHours
	Authentic learning Collaborative learning Independent learning	60	20	20	20	120
Pre-requisite	**Basics of information retrieval**					

COURSE OUTCOMES (CO)

CO No.	Expected Course Outcome	Learning Domains	PSO No.
1	Aware of various techniques for information retrieval	U,E	1,2,3,4
2	Understand search engines and web navigation	U,R,A	1,2,3
3	Understand , link analysis, XML retrieval	U,R,A	1,2,3
4	Familiarized information filtering and categorization.	U,R,A	1,3
5	Apply information retrieval techniques to real contexts	U,R,A,C,S	1,2,3,4,5

Remember (R), Understand (U), Apply (A), Analyse (An), Evaluate (E), Create (C), Skill (S), Interest (I) and Appreciation (Ap)

COURSE CONTENT

Module No	Module Content	Credits	Hrs

1	Introduction- History of Information Retrieval (IR) –Components of IR-Issues– Open source Search engine Frameworks- The impact of the web on IR-The role of artificial intelligence (AI) in IR–IR Versus Web Search- Components of a Search engine – Characterizing the web. Boolean and vector-space retrieval models – Term weighting- TF-IDF weighting – cosine similarity – Preprocessing- Inverted indices – efficient processing with sparse vectors– Language Model based IR– Probabilistic IR– Latent Semantic Indexing- Relevance feedback and query expansion.	1,2,3	15
2	Web search overview, web structure, the user, paid placement, search engine optimization/spam. Web size measurement-search engine optimization/spam –Web Search Architectures-crawling-meta crawlers –Focused Crawling-web indexes –-Near-duplicate detection- Index Compression– XML retrieval	2,3	15
3.	Link Analysis – hubs and authorities–Page Rank and HITS algorithms- Searching and Ranking– Relevance Scoring and ranking for Web–Similarity- Hadoop & MapReduce –Evaluation- Personalized search-Collaborative filtering and content-based recommendation of documents and products–handling "invisible" Web- Snippet generation, Summarization, Question Answering, Cross- Lingual Retrieval	3,4	15
4	Information filtering; organization and relevance feedback – Text Mining- Text classification and clustering- Categorization algorithms: naïve Bayes; decision trees; and nearest neighbor – Clustering algorithms: agglomerative clustering; k-means; expectation maximization (EM) algorithm. Presentation of IR, Case Studies	5	15
	Total Credits of the Course	3	60

Teachingand LearningApproach	**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
Assessment Types	**Mode of Assessment** C. Continuous Internal Assessment (CIA) 1. Internal Tests of maximum 20 marks 2. Seminar Presentation – a theme is to be discussed and identified to prepare a paper and present in the seminar Maximum marks 10 3. Write a detailed report on a given topic based on research findings and literature search – 10 marks D. **Semester End examination – 60 marks**

REFERENCES

1. Manning, C. D., Schütze, H. &Raghavan, P. (2008). Introduction to Information Retrieval (Vol.39). Cambridge University Press.
2. Ricardo Baeza-Yates &Berthier Ribeiro-Neto (2011). Modern Information Retrieval: The Concepts and Technology behind Search 2nd Edition, ACM Press Books.
3. Bruce Croft, Donald Metzler and Trevor Strohman (2009). Search Engines: Information Retrieval in Practice, 1st Edition Addison Wesley.
4. Mark Levene (2010) An Introduction to Search Engines and Web Navigation, 2nd Edition Wiley.

Additional References

1. Stefan Buettcher, Charles L. A. Clarke, Gordon V. Cormack(2010). Information Retrieval: Implementing and Evaluating Search Engines, The MIT Press.
2. Grossman, D. A., & Frieder, O. (2012). Information retrieval: Algorithms and Heuristics
Springer Science & Business Media.
3. Manu Konchady (2008) "Building Search Applications: Lucene, Ling Pipe", First Edition, Gate Mustru Publishing.

Approval Date	
Version	
Approval by	
Implementation Date	

MAHATMA GANDHI UNIVERSITY

DAM21E05: DATA WAREHOUSING & DATA MINING

SchoolName	School of Data Analytics
Programme	M.Sc. Microbiology/Biotechnology/Biochemistry/Biophysics
Course Name	**DATA WAREHOUSING & DATA MINING**
Type of Course	Elective
Course Code	**DAM21 E05**

Course Summary & Justification	To introduce basic ideas and advanced techniques in data warehousing and data mining.					
Semester	Second					
Total Student Learning Time (SLT)	Learning Approach	Lecture	Tutorial	Practical	Others	Total Learning Hours
	Authentic learning Collaborative learning Independent learning	60	20	20	20	120
Pre-requisite	Basic idea about Data Mining					

COURSE OUTCOMES (CO)

CO No.	Expected Course Outcome	Learning Domains	PSO No.
1.	Understood the important tools and techniques such as OLAP, OLTP,	U,A,R	1,2,3
2.	Understand Cluster Analysis for data warehousing	U,A,C,S	1,2,3
3.	Familiarized data mining	U,A,S	1,2,3
4.	Equipped to apply data warehousing and data mining techniques in real applications	U,A,C,S,An	1,2,3,4
**Remember (R), Understand (U), Apply (A), Analyse (An), Evaluate (E), Create (C), Skill (S), Interest (I) and Appreciation (Ap)*			

COURSE CONTENT

Module No	Module Content	CO	Hrs
1	Introduction to Data Warehousing: Introduction, Advantages and Disadvantages of Data Warehousing, Data Warehouse, Data Mart, Aspects of Data Mart, Online Analytical Processing, Characteristics of OLAP, OLAP Tools, OLAP Data Modeling, OLAP Tools and the Internet, Difference between OLAP and OLTP, Multidimensional Data Model	1,2	15
2	Data Mining: Introduction, Techniques, Issues and challenges, applications, Data preprocessing, Knowledge representation, Various risks in Data Mining, Advantages and disadvantages of Data Mining, Ethical issues in Data Mining, Analysis of Ethical issues. Association Rule Mining: Introduction, Methods to discover association rules, Association rules with item constraints, Decision Trees: Introduction, Tree construction principle, Decision tree construction algorithm, Pruning techniques, Integration of pruning and construction.	2,3	15

3.	Cluster analysis: Introduction, clustering paradigms, Similarity and distance, Density, Characteristics of clustering algorithms, Center based clustering techniques, Hierarchical clustering, Density based clustering, Other clustering techniques, Scalable clustering algorithms, Cluster evaluation Rough set theory, use of rough set theory for classification & feature selection. ROC Curves: Introduction, ROC Space, Curves, Efficient generation of Curves, Areaunder ROC Curve, Averaging ROC curves, Applications	3,4	15
4	Advanced techniques: Web mining – Introduction, Web content mining, Web structure mining, Web usage mining; Text mining- Unstructured text, Episode rule discovery from text, Text clustering; Temporal data mining – Temporal association rules, Sequence mining, Episode discovery, time series analysis; Spatial data mining – Spatial mining tasks, Spatial clustering, Spatial trends.	4	15
	Total Credits of the Course	3	60

Teaching and Learning Approach	**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
Assessment Types	**Mode of Assessment** **A. Continuous Internal Assessment (CIA)-40 Marks** 1. Internal Tests of maximum 20 marks 2. Seminar Presentation – a theme is to be discussed and identified to prepare a paper and present in the seminar - Maximum marks 10 3. Write a detailed report on a given topic based on research findings and literature search – 10 marks **B. Semester End examination – 60 marks**

REFERENCES

1. Alex Berson & Stephen J. Smith (1997) Data Warehousing, Data Mining & OLAP Computing Mc Graw Hill Publishers
2. Data Mining Techniques: A.K. Pujari, Universities Press, 2001
3. Mastering Data Mining: M. Berry and G. Linoff, John Wiley & Sons., 2000

Approval Date	
Version	
Approval by	

MAHATMA GANDHI UNIVERSITY

DAM21 E06: WEB SCRAPING AND TEXT MINING

School Name	**School of Data Analytics**					
Programme	**M.Sc. Data Science & Analytics**					
Course Name	**WEB SCRAPING AND TEXT MINING**					
Type of Course	Elective					
Course Code	**DAM21 E06**					
Course Summary & Justification	To introduce web and data technologies like HTML, XML, JSON, Xpath, AJAX etc for web scrapping and text mining.					
Total Student Learning Time (SLT)	Learning Approach	Lecture	Tutorial	Practical	Others	Total Learning Hours
	E.g., Authentic learning Collaborative learning Independent learning	60	20	20	20	120
Pre-requisite	**Basic idea about HTML,XML etc**					

No.	Expected Course Outcome	Learning Domains	PSO No.
1	equipped with tools for web scrapping	R/U/A	1,2,4,5
2	Understand text mining	U/A	12,,4,5
3	Apply Web scrapping and text mining in real contexts.	U/An/Ap	1,2,4,5

Remember (R), Understand (U), Apply (A), Analyse (An), Evaluate (E), Create (C), Skill (S), Interest (I) and Appreciation (Ap)

COURSE CONTENT

Module No	Module Content	CO	Hours
1	Introduction to Web and Data Technologies: HTML,Browser presentation and source code, Syntax rules, Tags and attributes, Parsing,; XML and JSON :A short example XML document, XML syntax rules, When is an XML document well formed or valid?, XML extensions and technologies, XML and R in practice, A short example JSON document, JSON syntax rules, JSON and R in practice,	!,2	15 hrs

2	Xpath - a query language for web documents, Identifying node sets with Xpath, Extracting node elements, HTTP: HTTP fundamentals, Advanced features of HTTP, Protocols beyond HTTP, HTTP in action,	1,2,	15 hrs
3.	AJAX: JavaScript, XHR, Exploring AJAX with Web Developer Tools, SQL and Relational Databases: Overview and terminology, Relational Databases, SQL: a language to communicate with Databases, R packages to manage databases	1	10 hrs
4	Web Scraping and Text Mining: Scraping the Web**:** Retrieval scenarios, Scraping data from AJAX-enriched web pages with Selenium / R web driver, Retrieving data from APIs, Authentication with Oauth, Extraction strategies, Web scraping: Good practice, Statistical Text Processing: The running example: Classifying press releases of the British government, Processing textual data, Supervised learning techniques, Unsupervised learning techniques, Managing Data Projects: Interacting with the file system, Organizing scraping procedures, Executing R scripts on a regular basis	2,3	20 hrs
	Total Credits of the Course	**3**	**60 hrs**
	Books for Reference		
	1. Munzert, S., Rubba, C., Mei Bner, P., Nyhuis, D. (2014) *Automated Data Collection with R: A Practical Guide to Web Scraping and Text Mining*. John Wiley & Sons, 2. Ryan Mitchell (2018) Web Scrapping with Python : Collecting More Data from the Modern Web ; Second Edition , O'Reily Publishers, USA		

Teaching and Learning Approach	**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
Assessment Types	**Mode of Assessment** **A. Continuous Internal Assessment (CIA)-40 Marks** 1. Internal Tests of maximum 20 marks 2. Seminar Presentation – a theme is to be discussed and identified to prepare a paper and present in the seminar - Maximum marks 10 3. Write a detailed report on a given topic based on

	research findings and literature search – 10 marks **B. Semester End examination – 60 marks**

Approval Date	
Version	
Approval by	
Implementation Date	

MAHATMA GANDHI UNIVERSITY

DAM21E07: NATURAL LANGUAGE PROCESSING WITH PYTHON

School Name	**School of Data Analytics**					
Programme	**M.Sc. Data Science & Analytics**					
Course Name	**NATURAL LANGUAGE PROCESSING WITH PYTHON**					
Type of Course	**Elective**					
Course Code	**DAM21E07**					
Course Summary & Justification	To introduce various techniques for natural language processing using Python.					
Total StudentLearning Time (SLT)	Learning Approach	Lecture	Tutorial	Practical	Others	Total Learning Hours
	Authentic learning Collaborative learning Independent learning	60	20	20	20	120

Pre-requisites	Basics of Python, Classification Methods etc

COURSE OUTCOMES (CO)

CO No.	Expected Course Outcome	Learning Domains	PSO No.
1	Aware of Text classification	**U,A,E**	1,2, 3,4,5
2.	**Understand** Text summarization	**U,E,A**	**2,3,4,**
3.	**Familiarized** Semantic and sentiment analysis,	**U,A,R**	**1,2,3,4,5**
4.	Understand classification algorithms and can apply these in practical real life problems.	**U,C,A,S**	**2,3,5**
Remember (R), Understand (U), Apply (A), Analyse (An), Evaluate (E), Create (C), Skill (S), Interest (I) and Appreciation (Ap)			

COURSE CONTENT

Module No	Module Content	CO	Hrs
1	What is Natural Language Processing? Language Processing and Python, Natural Language Basics, Natural Language, Linguistics, Language Syntax and Structure, Language Semantics, Text Corpora, Natural Language Processing, Text Analytics; Processing and Understanding Text: Text Tokenization, Text Normalization, Understanding Text Syntax and Structure.	**1,2**	**10**
2.	Text Classification: What Is Text Classification? Automated Text Classification, Text Classification Blueprint, Text Normalization, Feature Extraction, Classification Algorithms, Evaluating Classification Models, Building a Multi-Class Classification System	**2**	**20**
3.	Text Summarization: Text Summarization and Information Extraction, important concepts, Text Normalization Feature Extraction, Key phrase Extraction, Topic Modeling, Automated Document Summarization; Text Similarity and Clustering: Important Concepts, Text Normalization, Feature Extraction, Text Similarity, Analyzing Term Similarity, Analyzing Document Similarity, Document Clustering	**2,3**	**20**
4	Semantic and Sentiment Analysis: Semantic Analysis, Exploring WordNet, Word Sense Disambiguation, Named Entity Recognition, Analyzing Semantic Representations; Stemming and Lemmatization, Synsets and Hypernyms.	**3,4**	**10**
	Total Credits of the Course	**3**	**60**

Teachingand LearningApproach	**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.
Assessment Types	**Mode of Assessment** A. **Continuous Internal Assessment (CIA)-40 marks** 1. Internal Tests of maximum 20 marks 2. Seminar Presentation – a theme is to be discussed and identified to prepare a paper and present in the seminar Maximum marks 10 3. Write a detailed report on a given topic based on research findings and literature search – 10 marks **B. Semester End examination – 60 marks**

REFERENCES

1. Dipanjan Sarkar, Text Analytics with Python, Apress/Springer, 2016
2. Bird, Steven, Ewan Klein, and Edward Loper. (2009). Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit. O'Reilly Media, Inc.,

Approval Date	
Version	
Approval by	
Implementation Date	

MAHATMA GANDHI UNIVERSITY

DAM21E08: FRAUD ANALYTICS

School Name	**School of Data Analytics**					
Programme	**M. Sc. Data Science & Analytics**					
Course Name	**FRAUD ANALYTICS**					
Type of Course	**Elective**					
Course Code	**DAM21E08**					
Course Summary & Justification	To introduce how one can treat the Internet as a source of data and analyze web-scale data using distributed computing					
Total StudentLearning Time (SLT)	Learning Approach	Lecture	Tutorial	Practical	Others	Total Learning Hours
	Authentic learning Collaborative learning Independent learning	60	20	20	20	120
Pre-requisites	Basics of coding theory, logistic regression etc					

COURSE OUTCOMES (CO)

CO No.	Expected Course Outcome	Learning Domains	PSO No.
1	Gained the basic knowledge on analysis of fraud and fraud detection models;	**U,A,C**	2,4,5
2.	compare different models, develop automation process of fraud detention	**U,An,E**	**2,3,4,5**
3.	Equipped to formulate and evaluate fraud detection	**A,U,S**	**2,5**

Remember (R), Understand (U), Apply (A), Analyse (An), Evaluate (E), Create (C), Skill (S), Interest (I) and Appreciation (Ap)

COURSE CONTENT

Module No	Module Content	CO	Hrs
1	Formulation and evaluation of fraud detection - Fraud detection using data analysis Obtain and cleanse the data for fraud detection-preprocess data for fraud detection - sampling, missing values, outliers, categorization etc.- Explain characteristics and components of the data and	**1**	**15**

	assess its completeness		
2.	Identify known fraud symptoms -and use digital analysis to identify unknown fraud symptoms- Fraud detection models using supervised analytics (logistic regression, decision trees, neural networks, ensemble models, etc.	**1**	**15**
3.	Automating fraud detection process -Fraud detection models using unsupervised analytics hierarchical clustering, non-hierarchical clustering, agglomerative and divisive techniques k-means clustering, self organizing maps, etc.	**2**	**15**
4	Fraud detection models using social network analytics homophily, featurization, egonets, PageRank, bigraphs etc. - Verification of results and understand how to prosecute fraud. Fraud detection and prevention-case studies.	**1,2**	**15**
	Total Credits of the Course	**4**	**60**

Teachingand LearningApproach	**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
Assessment Types	**Mode of Assessment** **A. Continuous Internal Assessment (CIA)-40 marks** 1. Internal Tests of maximum 20 marks 2. Seminar Presentation – a theme is to be discussed and identified to prepare a paper and present in the seminar Maximum marks 10 3. Write a detailed report on a given topic based on research findings and literature search – 10 marks **B.Semester End examination – 60 marks**

REFERENCES

1. Baesens, B., Van Vlasselaer, V., & Verbeke, W. (2015). Fraud analytic using descriptive, predictive, and social network techniques: a guide to data science for fraud detection. John Wiley & Sons.

2. Nigrini, M. J. (2011). Forensic analytics: methods and techniques for forensic accounting investigations (Vol. 558). John Wiley & Sons.

Approval Date	
Version	
Approval by	
Implementation Date	

MAHATMA GANDHI UNIVERSITY

DAM21 E09: INTERNET OF THINGS IN THE CLOUD

School Name	**School of Data Analytics**					
Programme	**M.Sc. Data Science & Analytics**					
Course Name	**INTERNET OF THINGS IN THE CLOUD**					
Type of Course	Elective					
Course Code	**DAM21 E09**					
Course Summary & Justification	To introduce how one can treat the Internet as a source of data and analyze web-scale data using distributed computing					
Total StudentLearning Time (SLT)	Learning Approach	Lecture	Tutorial	Practical	Others	Total LearningHours
	Authentic learning Collaborative learning Independent learning	60	20	20	20	120
Pre-requisites	Basics of Python and Java					

COURSE OUTCOMES (CO)

CO No.	Expected Course Outcome	Learning Domains	PSO No.
1	Become proficient in Data cleaning and Text mining	U,A	1, 2.3.4.5
2	Know how to use Python to scrape the web II	U,A,E,An	1,2,3,4,5
3	Can retrieve data from web sites and APIs over the Internet using Python	U,A,E	1,2,3,4,5
4	Aware of Web Services using the Java Script Object Notation (JSON) data format	U,A,AnE	1,2,3,4,5
5	Able to apply the techniques to execute many thing using internet in the cloud	S,U,E,An/C	1,2,3,4,5
Remember (R), Understand (U), Apply (A), Analyse (An), Evaluate (E), Create (C), Skill (S), Interest (I) and Appreciation (Ap)			

COURSE CONTENT

Module No	Module Content	CO	Hrs
1	Introduction to Python II (parsing & using libraries) Using Python to scrape the web I (regex & other libraries); Using Python to scrape the web II (data cleaning), Text Mining with Python (nltk).	1,2	15
2	Regular Expressions- Extracting Data With Regular Expressions, Networks and Sockets- protocol - web browsers used to retrieve documents, web applications used to interact with Application Program Interfaces (APIs)	2,3	15
3.	Use Python to retrieve data from web sites and APIs over the Internet; Reading Web Data From Python - Scraping HTML Data How to retrieve and parse XML (eXtensible Markup Language) data- extensible Markup Language - Extracting Data from XML.	3,4	15
4	API's, Web Services using the Java Script Object Notation (JSON) data format; Practical case studies of web scrapping.	3,4,5	15
	Total Credits of the Course	3	60

Teachingand LearningApproach	**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
Assessment Types	**Mode of Assessment** **A. Continuous Internal Assessment (CIA)-40 marks** 1. Internal Tests of maximum 20 marks 2. Seminar Presentation – a theme is to be discussed and identified to prepare a paper and present in the seminar Maximum marks 10 3. Write a detailed report on a given topic based on research findings and literature search – 10 marks B. **Semester End examination – 60 marks**

REFERENCES

1. Severance,C.R., Blumenberg,S.,& Hauser, E.(2016).Python for Everybody: Exploring Data in Python 3. Create Space Independent Publishing Platform.

2. Munzert, S., Rubba, C., Meißner, P., & Nyhuis, D. (2014). Automated Data Collection with R: A Practical Guide to Web Scraping and Text Mining. John Wiley &Sons.

3. Hersent, O., Boswarthick, D., & Elloumi, O. (2012). The Internet of Things: Applications to the Smart Grid and Building. JohnWiley&Sons.

Approval Date	
Version	
Approval by	
Implementation Date	

MAHATMA GANDHI UNIVERSITY

DAM21E10: OPERATIONS RESEARCH

School Name	**School of Data Analytics**					
Programme	**M.Sc. Data Science & Analytics**					
Course Name	**OPERATIONS RESEARCH**					
Type of Course	Elective					
Course Code	**DAM21E10**					
Course Summary & Justification	To provide sound knowledge of Operations Research with applications to data science and analytics					
Total StudentLearning Time (SLT)	Learning Approach	Lecture	Tutorial	Practical	Others	Total LearningHours
	Authentic learning Collaborative learning Independent learning	80	10	20	10	120

Pre-requisites	Basic knowledge in Statistical Distribution, Stochastic Processes

COURSE OUTCOMES (CO)

CO No.	Expected Course Outcome	Learning Domains	PSO No.
1.	Students have understood linear programming problem, graphical and Simplex methods to solve it.		2
2	Students have understood transportation problem and different methods to solve it.		3,4
3	Students have understood integer programming and dynamic programming		2,5
4	Students have understood queuing theory and are capable to solve different types of queues and inventory problems		2,4,5
Remember (R), Understand (U), Apply (A), Analyse (An), Evaluate (E), Create (C), Skill (S), Interest (I) and Appreciation (Ap)2,5			

COURSE CONTENT

Module No	Module Content	CO	Hrs
1	Linear programming: Mathematical Model, assumptions of linear programming, Solutions of linear programming problems – Graphical Method, Simplex method, Artificial Variable Method, Two phase Method, Big M Method, Applications, Duality, Dual simplex method, Introduction to sensitivity analysis.	1	20
2	Special types of Linear programming problems- Transportation Problem (TP) – Mathematical formulation of Transportation Problem, Basic feasible solution in TP, Degeneracy in TP, Initial basic feasible solutions to TP, Matrix Minima Method, Row Minima Method, Column Minima Method, Vogel's Approximation Method, Optimal Solution to TP, MODI Method, Stepping Stone Method, Assignment problems – Definition, Hungarian Method	1	20
3.	Integer Programming: Pure Integer Programming, Mixed Integer Programming, Solution Methods – Cutting plane method, branch and bound method. Binary Integer Linear programming- Travelling salesman problems – Iterative method, Branch and bound method; Dynamic programming: Deterministic and	2,3	20

	Probabilistic Dynamic programming. Linear programming by dynamic programming approach.		
4	Queuing Model: Elements and Characteristics of queuing systems., Classification of queuing systems–Structures of Basic Queuing System, Stationary solutions to Poisson Queuing System – M/M/1: ∞/FIFO, M/M/1: ∞/SIRO, M/M/1: N/FIFO ,M/M/C: N/FIFO, M/M/C: Inventory Models: Deterministic inventory models, economic order quantity and its extensions - Stochastic inventory models, setting safety stocks - Reorder point order quantity models.	4	20
	Total Credits of the Course	3	80

Teachingand LearningApproach	**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
Assessment Types	**Mode of Assessment** **A. Continuous Internal Assessment (CIA)-40 marks** 1. Internal Tests of maximum 20 marks 2. Seminar Presentation – a theme is to be discussed and identified to prepare a paper and present in the seminar Maximum marks 10 3. Write a detailed report on a given topic based on research findings and literature search – 10 marks B. **Semester End examination – 60 marks**

REFERENCES

1. Ravindran, A., Philips, D.T Solberg, J. J. (2007) Operations Research: Principles and Practice, Wiley

2. J. K. Sharma (2009), Operations Research – Theory and Applications, 4th Ed, Mc Millan Publishing.

3. Taha, H.A. (2007), Operations Research, 8th Ed., Mc Millan Publishing Company

4. Kantiswaroop, P. K. Guptha, & Manmohan (2007) Operations Research, 13th Ed, Sulthan Chand & Sons.

5. Beightler C. S, & Philips D T (2009), 'Foundations of Optimization', 2nd Ed., PHI

6. Mc Millan Claude Jr (1979), 'Mathematical Programming', 2nd Ed. Wiley Series.

Approval Date	
Version	
Approval by	
Implementation Date	

MAHATMA GANDHI UNIVERSITY

DAM21E11: CLOUD COMPUTING

School Name	**School of Data Analytics**					
Programme	**M.Sc. Data Science & Analytics**					
Course Name	**CLOUD COMPUTING**					
Type of Course	Elective					
Course Code	**DAM21 E11**					
Course Summary & Justification	To introduce recent trends in cloud computing, architecture and service models like XaaS, IaaS, SaaS, PaaS and illustrate through case studies.					
Total StudentLearning Time (SLT)	Learning Approach	Lecture	Tutorial	Practical	Others	Total Learning Hours
	Authentic learning Collaborative learning Independent learning	60	10	30	20	120
Pre-requisites	Basics of Cloud computing					

COURSE OUTCOMES (CO)

CO No.	Expected Course Outcome	Learning Domains	PSO No.
1	Gained knowledge in recent trends in cloud computing, architecture,.	U,A,R	1,2,3.4.

2	Understand service models, platform and storage and web services.	R,U	1,2,3,4,
3	Enabled to apply these techniques to real problems	R,U,A,S	1,2,3,4,
Remember (R), Understand (U), Apply (A), Analyse (An), Evaluate (E), Create (C), Skill (S), Interest (I) and Appreciation (Ap)			

COURSE CONTENT

Module No	Module Content	CO	Hrs
1	Overview of Computing Paradigm, Recent trends in Computing Grid Computing, Cluster Computing, Distributed Computing, Utility Computing, Cloud Computing Evolution of cloud computing Business driver for adopting cloud computing. Introduction to Cloud Computing. Cloud Computing (NIST Model) Introduction to Cloud Computing, History of Cloud Computing, Cloud service providers. Properties, Characteristics & Disadvantages Pros and Cons of Cloud Computing.	1,2	15
2	Cloud Computing Architecture, Comparison with traditional computing architecture (client/server), Services provided at various levels, How Cloud Computing Works, Role of Networks in Cloud computing, protocols used, Role of Web services.	2,3	15
3.	Service Models (XaaS), Infrastructure as a Service(IaaS),Platform as a Service(PaaS), Software as a Service(SaaS), Deployment Models, Public cloud, Private cloud, Hybrid cloud, Community cloud. Introduction to virtualization, Different approaches to virtualization, Hypervisors, Machine Image Virtual Machine (VM). Examples, Amazon EC2, Renting, EC2 Compute Unit	2,3	15
4	Platform and Storage, pricing, customers Eucalyptus, Platform as a Service(PaaS),Introduction to PaaS, What is PaaS, Service Oriented Architecture (SOA). Examples: Google App Engine, Microsoft Azure, Sales Force.com's Force.com platform; Software as a Service (SaaS): Introduction to SaaS, Web services, Web 2.0, Web OS, Case; Study on SaaS Cloud Security. Case Study on Open Source & Commercial Clouds	3	15
	Total Credits of the Course	3	60

Teachingand Learning Approach	**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
Assessment Types	**Mode of Assessment** **A. Continuous Internal Assessment (CIA)-40 marks** 1. Internal Tests of maximum 20 marks 2. Seminar Presentation – a theme is to be discussed and identified to prepare a paper and present in the seminar Maximum marks 10 3. Write a detailed report on a given topic based on research findings and

	literature search – 10 marks **B, Semester End Examination – 60 marks**

REFERENCES

1. Barrie Sosinsky (2010) Cloud Computing Bible, *Wiley-India*,
2. Kirsh D. and Hurwitz J.(2018) Cloud Computing for Dummies 2nd Edition
3. Ruparelia, N.B.(2016) Cloud Computing, MIT Press
4. Orban S.(2019) Ahead in the Cloud: Best Practices for Navigating, AWS

Approval Date	
Version	
Approval by	
Implementation Date	

MAHATMA GANDHI UNIVERSITY

DAM21 E12: BUSINESS INTELLIGENCE & ANALYTICS

School Name	**School of Data Analytics**					
Programme	**M. Sc. Data Science & Analytics**					
Course Name	**BUSINESS INTELLIGENCE & ANALYTICS**					
Type of Course	**Elective**					
E12:	**DAM21E12:**					
Course Summary & Justification	To introduce various concepts in business intelligence, data warehousing, data mining, business models, knowledge management, data extraction, data life cycle.					
Total StudentLearning Time (SLT)	Learning Approach	Lecture	Tutorial	Practical	Others	Total Learning Hours
	Authentic learning Collaborative learning Independent learning	60	20	20	20	120
Pre-requisites	Basics of Plant tissue culture					

COURSE OUTCOMES (CO)

CO No.	Expected Course Outcome	Learning Domains	PSO No.

1	Understood the tools for business intelligence and total quality management.	**U,A,C**	1,2,3
2.	Familiar with basic methodology in data warehousing and data mining as well as knowledge management.	U,A,E	**1,2,3,4**
3.	capable of applying these for data extraction, strategy development, business intelligence etc	**U,A,C,S**	**1,2,3,4**
Remember (R), Understand (U), Apply (A), Analyse (An), Evaluate (E), Create (C), Skill (S), Interest (I) and Appreciation (Ap)			

COURSE CONTENT

Module No	Module Content	CO	Hrs
1	Business Intelligence: Introduction, Definition, Business Intelligence Segments, Difference between Information and Intelligence, Defining Business Intelligence Value Chain, Factors of Business Intelligence System, Real time Business Intelligence, Business Intelligence Applications. Creating Business Intelligence Environment, Business Intelligence Landscape, Types of Business Intelligence, Business Intelligence Platform, Dynamic roles in Business Intelligence, Roles of Business Intelligence in Modern Business- Challenges of BI . Business Intelligence Types: Introduction, Multiplicity of Business Intelligence Tools, Types of Business Intelligence Tools, Modern Business Intelligence, the Enterprise Business Intelligence, Information Workers. Architecting the Data: Types of Data, Enterprise Data Model, Enterprise Subject Area Model, Enterprise Conceptual Model, Enterprise Conceptual Entity Model, Granularity of the Data, Data Reporting and Query Tools, Data Partitioning, Meta data, Total Data Quality Management (TDQM)	**1,2**	**15**
2.	Introduction to Data Mining: Definition of Data Mining, Architecture of Data Mining, Kinds of Data which can be mined, Functionalities of Data Mining, Classification on Data Mining system, Various risks in Data Mining, Advantages and disadvantages of Data Mining, Ethical issues in Data Mining, Analysis of Ethical issues. Introduction to Data Warehousing: Introduction, Advantages and Disadvantages of Data Warehousing, Data Warehouse, Data Mart, Aspects of Data Mart, Online Analytical Processing, Characteristics of OLAP, OLAP Tools, OLAP Data Modeling, OLAP Tools and the Internet, Difference between OLAP and OLTP, Multidimensional Data Model	**2,3**	**15**
3.	Types of Business Models, B2B Business Intelligence Model, Electronic Data Interchange & E-Commerce Models, Advantages of E-Commerce for B2B Businesses, Systems for Improving B2B E-Commerce, B2C Business	**3**	**15**

	Intelligence Model, Need of B2C model in Data warehousing, Different types of B2B intelligence Models. Knowledge Management: Introduction, Characteristics of Knowledge Management, Knowledge assets, Generic Knowledge Management Process, Knowledge Management Technologies, Essentials of Knowledge Management Process		
4	Data Extraction: Introduction, Data Extraction, Role of ETL process, Importance of source identification, Various data extraction techniques, Logical extraction methods, Physical extraction methods, Change data capture Business Intelligence Life Cycle: Introduction, Business Intelligence Lifecycle, Enterprise Performance Life Cycle (EPLC)Framework Elements, Life Cycle Phases, BI Strategy, Objectives and Deliverables, Transformation Roadmap, Building a transformation roadmap, BI Development Stages and Steps, Parallel Development Tracks, BI Framework.	**3**	**10**
	Total Credits of the Course	**3**	**60**

Teachingand LearningApproach	**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
Assessment Types	**Mode of Assessment** **A. Continuous Internal Assessment (CIA)-40 marks** 1. Internal Tests of maximum 20 marks 2. Seminar Presentation – a theme is to be discussed and identified to prepare a paper and present in the seminar Maximum marks 10 3. Write a detailed report on a given topic based on research findings and literature search – 10 marks **B, Semester End examination – 60 marks**

REFERENCES

1. Business Intelligence Guidebook: From Data Integration to Analytics by Rick Sherman
2. Business Intelligence Roadmap: The Complete Project Lifecycle for Decision-Support Applications by Larissa T. Moss and Shaku Atre
3. The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling by Ralph Kimball and Margy Ross
4. Successful Business Intelligence, Second Edition: Unlock the Value of BI & Big

Data by Cindi Howson 5. Business Intelligence for Dummies by Swain Scheps

Approval Date	
Version	
Approval by	
Implementation Date	

MAHATMA GANDHI UNIVERSITY

DAM21 E13: DATA ANALYTICS COMPUTING

School Name	**School of Data Analytics**					
Programme	**M.Sc. Data Science & Analytics**					
Course Name	**DATA ANALYTICS COMPUTING**					
Type of Course	Elective					
Course Code	**DAM21E13**					
Course Summary & Justification	to train students to use R and Python for all types of computational works in Data Science and Data Analytics.					
Semester	Third					
Total StudentLearning Time (SLT)	Learning Approach	Lecture	Tutorial	Practical	Others	Total Learning Hours
	Authentic learning Collaborative learning Independent learning	40	20	40	20	120
Pre-requisites						

COURSE OUTCOMES (CO)

CO	Expected Course Outcome	Learning	PSO No.

No.		Domains	
1	Students are thorough with the concepts of computation: algorithms, convergence	U,A,C	1,2,3
2	Understand Complexity with illustrations, sorting & searching algorithms and numerical methods.	U,A,E	1,2,3,4
3	Proficient in computing methodologies like Monte-Carlo simulations of random numbers and various statistical methods,	A,S,U	1,3,4,5
4	Understand memory handling strategies for big data	U,A,An	1,2
Remember (R), Understand (U), Apply (A), Analyse (An), Evaluate (E), Create (C), Skill (S), Interest (I) and Appreciation (Ap)			

COURSE CONTENT

Module No	Module Content	CO	Hours
1	**Computer Package** L0 -- Installation of R Studio and understanding the basic framework L1 – Basic computational structures – Iterations and Recursions L2 -- Sequences and Arrays in R – Search and Sort Algorithms L2 -- Vectors and Matrices in R – Solving systems of linear equations L3 -- Functions in R – Plotting (2D, Contour, 3D), Differentiation, Root finding L4 – Linear Models in R – Gradient descent, Linear regression L5 – Eigen values and vectors computation and Singular Value Decomposition in R L6 – Handling sparse matrices in R – Basic operations on sparse matrices	1,2,3	15
2	**Concept of Computations** L1 -- Algorithms – Search and Sort, Divide and Conquer, Greedy Algorithms – motivating example from set cover for large data sets. L2 – Computational Complexity – Growth of functions, Order notation L3 – Computational Complexity -- Convergence, Error Estimation L6 – Binary Trees and Graphs as Computational Models	4,5,6	15
3.	**Numerical Methods** L2 -- Solving system of linear equations – Gauss - Jordan (concept of pivoting) L3 -- Solving non-linear equations – Newton-Raphson, Steepest Descent L4 – Optimizing cost functions – Gradient descent, Least square regression L5 – Iterative methods in Linear Algebra – Power iteration, Eigenvalues, SVD	1,2,3	15

4	**Computing Methodologies** L7 – Probability Distributions and Random Sampling from Probability Distributions, Simulated Annealing L8 – Monte-Carlo Simulation, Importance Sampling, MCMC Methods, EM Agorithm, Jackknife, Bootstrapping – Case studies **Memory Handling** L6 – Sparse Matrix – Store, Search and Basic operations L9 – Pruning and Sampling algorithms, Streaming data, External sorting	4,5,6	15
	Total Credits of the Course	3	60

Teaching and Learning Approach	**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
Assessment Types	**Mode of Assessment** **A. Continuous Internal Assessment (CIA)-40 marks** 1. Internal Computational Skill Tests- maximum 20 marks 2. Seminar Presentation –Case study reports for big data sets as illustration- 10 marks 3. Write a detailed report on various computation techniques – 10 marks B. **Semester End Practical examination – 60 marks**

REFERENCES

1. John M. Chambers (2008) Software for Data Analysis – Programming with R : Springer
2. Samuel Conte and Carl de Boor (1980) Elementary Numerical Analysis – An Algorithmic Approach : McGraw-Hill Education
3. Rivest and Stein(2009), Introduction to Algorithms : Cormen, Leiserson, The MT Press 3rd Edition
4.Heuman, C., Schomaker, M. and Shalab (2016) Introduction to Statistics and Data Analysis with Applications in R, Springer.
5.Purohit, S.G, Gore, S., and Deshmukh, S.R.(2018) Statistics Using R, Narosa Publishers

Approval Date	
Version	
Approval by	
Implementation Date	

MAHATMA GANDHI UNIVERSITY

DAM21 E14: COMPLEX NETWORK ANALYSIS

SchoolName	**School of Data Analytics**					
Programme	**M.Sc. Data Science & Analytics**					
Course Name	**COMPLEX NETWORK ANALYSIS**					
Type of Course	Elective`					
Course Code	**DAM21E14**					
Course Summary & Justification	to introduce various concepts to model and visualize network structure and understand its dynamics.					
Total StudentLearning Time (SLT)	Learning Approach	Lecture	Tutorial	Practical	Others	Total LearningHours
	Authentic learning Collaborative learning Independent learning	60	20	20	20	120
Pre-requisite						

COURSE OUTCOMES (CO)

CO No.	Expected Course Outcome	Learning Domains	PSO No.
1	Understand network data and representations; Execute graph algorithms;.	A,U,E	1,2,4,5
2	Carryout basic transformation and visualization;.	A,U,S	1,2,3,4,5
3	Analyse graph visualization algorithms; Analyse real-world networks	A,U,R	1,2,3,4,5

Remember (R), Understand (U), Apply (A), Analyse (An), Evaluate (E), Create (C), Skill (S), Interest (I) and Appreciation (Ap)

COURSE CONTENT

Module No	Module Content	CO	Hours
1	What is Network Science? What is (not) network science? The main premise of network science, History and relation to graph theory, physics, sociology, and other disciplines, Examples of	1	15

	networks from different application domains. Networks in epidemiology, biology, Neuro Science, Social Science, Economics, Network Contagion and Epidemics, Epidemics on networks, Epidemic modeling (SI, SIS, SIR, etc) under homogeneous mixing, Epidemic modeling under arbitrary degree distributions, Basic reproductive number and super spreaders, Vaccination strategies. Relevant Concepts From Graph Theory, Undirected, directed, signed, weighted and spatial networks, Paths, connected components, random walks, etc, Directed Acyclic Graphs, Bipartite graphs, Max-flow/min-cut; Degree Distribution and ER Graphs; Degree distribution, Friendship paradox, ER graphs and their degree distribution, Giant component size in ER graphs, Assortative vs dis-assortative networks.; Random vs. Real Graphs and "Scale Free" Networks, The degree distribution of real-world networks, Power-law degree distributions, Preferential attachment model, How to detect a power-law and estimate the exponent, Configuration model and degree-preserving randomization		
2	Network Paths, Clustering and "Small World" Property, Clustering and transitivity in networks, Diameter and characteristic path length, Small-world networks and the Watts-Strogatz model, Network motifs; Centrality and Network-core Metrics and Algorithms, Link-based centrality metrics, Path-based centrality metrics, k-core decomposition, Core-periphery structure, Rich-club set of nodes Community Detection and Hierarchical Modularity, Hierarchical clustering in networks; Modularity metric, Algorithms for modularity maximization, Limitations of modularity.; Hiearchical modularity, Advanced Topics in Community Detection, Overlapping communities, Dynamic communities, Comparing community structures, The role of nodes within and between communities, Applications of community detection.	2	15
3.	Influence Phenomena on Networks, Influence vs contagion, The linear threshold model and the Independent cascades model, Empirical studies in information and behavior spreading, Seeding strategies on how to maximize influence, Cascades and community structure; Other Dynamic Processes Of/On Networks, Percolation, random failures, and targeted attacks on networks, Search on networks, Synchronization on networks, Controlling networks, Co-evolutionary networks.	2,3	15
4	Models of Static and Dynamic Networks, Stochastic network models that generate power-law degree distributions, Optimization-based network models, Stochastic block models, Models that generate hierarchical modularity; Statistical Analysis of Network Data, Network sampling methods, Estimation of network metrics, Association networks, Network tomography.	1,3,4	15
	Total Credits of the Course	3	60

Books for Reference

Compulsory Reading:

1. A-L. Barabási , Network Science , available online, 2015.
2. M.E.J. Newman, Networks - An Introduction , Oxford University Press, 2010.
3. D. Easley and J. Kleinberg, Networks, Crowds and Markets , Cambridge Univ Press, 2010 (also available online).

Additional References

1. R. Cohen and S. Havlin, Complex Networks - Structure, Robustness and Function , Cambridge Univ Press, 2010.
2. M.O. Jackson, Social and Economic Networks , Princeton Univ Press, 2008.
3. A. Barrat, M. Barthelemy and A. Vespignani, Dynamical Processes on Complex Networks, Cambridge Univ Press, 2008.
4. E. Kolaczyk, Statistical Analysis of Network Data, Springer, 2009.
5. S. Wasserman, K. Faust, Social Network Analysis: Methods and Applications, Cambridge Univ Press, 1994.
6. P. Van Mieghem, Graph Spectra for Complex Networks, Cambridge University Press, 2011.
7. R. Diestel, Graph Theory (4th edition), Springer, 2010.
8. R.K.Ahuja and T.L.Magnanti, Network Flows: Theory, Algorithms, and Application , Pearson, 1993.
9. Networks in Epidemiology: An Integrated Modeling Environment to Study the Co-evolution of Networks, Individual Behavior and Epidemics by Chris Barrett et al.
10. Networks in Biology: Network Inference, Analysis, and Modeling in Systems Biology by Reka Albert
11. Networks in Neuroscience: Complex brain networks: graph theoretical analysis of structural and functional systems by Ed Bullmore and Olaf Sporns
12. Networks in Social Science: Network Analysis in the Social Sciences by Stephen Borgatti et al.
13. Networks in Economics: Economic Networks: The New Challenges by Frank Schweitzer et al.
14. Networks in Ecology: Networks in Ecology by Jordi Bascompte
15. Networks and the Internet: Network Topologies: Inference, Modelling and Generation by Hamed Haddadi et al.

Teachingand LearningApproach	**Classroom Procedure (Mode of transaction)** Direct Instruction, 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
Assessment Types	**Mode of Assessment** **A. Continuous Internal Assessment (CIA)-40 marks** At least Two Internal Tests -20 marks Assignment – Every student needs to write an assignment on a given topic based on the available published literature – 10 marks Seminar Presentation – A case study needs to be presented and discussed with the class- 10 marks **B. Semester End examination – 60 marks**

Approval Date	
Version	
Approval by	
Implementation Date	

MAHATMA GANDHI UNIVERSITY

DAM21E15: BIOSTATISTICS & EPIDEMIOLOGY

School Name	**School of Data Analytics**
Programme	**M.Sc. Data Science & Analytics**
Course Name	**BIOSTATISTICS & EPIDEMIOLOGY**
Type of Course	Elective
Course Code	**DAM21E15**

Course Summary & Justification	To make aware of various biostatistical and epidemiological measures and methods for studying about incidence of diseases, disease tracking etc using data science and analytics.					
Total StudentLearning Time (SLT)	Learning Approach	Lecture	Tutorial	Practical	Others	Total LearningHours
	Authentic learning Collaborative learning Independent learning	80	10	20	10	120
Pre-requisites	**Basic knowledge in Statistical Distribution, Stochastic Processes**					

COURSE OUTCOMES (CO)

CO No.	Expected Course Outcome	Learning Domains	PSO No.
1	Students are able to plan and conduct epidemiological studies using appropriate methods in Biostatistics and Data Analytics	U/R/A/An/E	1,2,3,4,5, 6,7
2	Now they can conduct cohort studies, case control studies etc and can calculate incidence rate, case fatality rate etc. to assess and validate the situation.	U/R/A/An/E	1,2,3,4,5, 6,7
3	Now they can conduct disease tracking and learning using data analytics and data science	U/R/A/An/E	1,2,3,4,5, 6,7
4			
Remember (R), Understand (U), Apply (A), Analyse (An), Evaluate (E), Create (C), Skill (S), Interest (I) and Appreciation (Ap)2,5			

COURSE CONTENT

Module No	Module Content	CO	Hrs
1	Basic concepts in epidemiology; Measures of Exposures, Types of exposures, Sources of exposures, Measures of outcome; Disease registries, Classification of diseases, Measures of disease frequency: Prevalence, Incidence, Risk, Odds of disease, Incidence time, Incidence rate, Relationship between prevalence, rate and risk, Routine data to measure disease occurrence, cumulative rate, cumulative risk, proportional incidence, Case fatality; Standardization - direct method of Standardization, indirect method of standardization.	1	20
2	Type of study design- Intervention studies, Cohort studies, case-control studies, cross-sectional studies, ecological studies; Relative and absolute measures of effect, Confidence intervals and significance tests for measures of occurrence and effect; Validity and reliability of measures of exposure and outcome: Sensitivity, Specificity, Predictive	1,2	20

	value method for selecting a positivity criterion, Receiver Operator Characteristic (ROC) curve, Intra and Inter-observer reliability, Kappa measure of agreement.		
3.	Case-control studies: Definition of cases, and controls, methods of selecting cases and controls, matching, sample size, power calculations, Basic methods of analysis of grouped data, Basic methods of analysis of matched data. Logistic regression for case-control studies, estimation and interpretation of logistic parameters, matched analysis- estimation of logistic parameters, Categorical data analysis. Cohort studies: Prospective cohort studies: planning and execution, retrospective cohort, nested case-control, case-cohort studies: planning and execution, matching and efficiency in cohort studies, cohort studies –statistical analysis. Longitudinal studies: Design, execution and analysis of longitudinal studies, repeated measurement analysis.	1,2	20
4	Sources of bias, Selection bias, measurement bias, misclassification of exposure and outcome, Differential and non-differential exposure and outcome classification, Confounding, Assessment of confounding, Mantel-Haenszel summary measures of effect, Interaction, Mantel-Haenszel method to adjust for several confounders, Confidence intervals and statistical tests for adjusted relative measures of effect; Excess risk and Attributable risk; Causation and Hill's criteria.	1,2	20
	Total Credits of the Course	3	80

Teachingand LearningApproach	**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
Assessment Types	**Mode of Assessment** **A. Continuous Internal Assessment (CIA)-40 marks** 1. Internal Tests of maximum 20 marks 2. Seminar Presentation – a theme is to be discussed and identified to prepare a paper and present in the seminar Maximum marks 10 3. Write a detailed report on a given topic based on research findings and literature search – 10 marks B. **Semester End examination – 60 marks**

REFERENCES

Compulsory Reading:

1. Leon Gordis (2008) Epidemiology, Elsevier Publishers
2. Rothman K.J and Greenland S (1998). Modem Epidemiology, Third edition, Lippincott

3. Isabel dos Santos Silva (1999) Cancer Epidemiology: Principles and Methods, International Agency for Research on Cancer.

Further Reading:

- Penny Web ,Chiris Bain & Sandi Pirozzo (2005).Essential Epidemiology - An Introduction for students & Health Professionals, Cambridge University Press
- R Bonita, R Beaglehole. T Kjellström, (2006): Basic Epidemiology 2nd Edition.
- Clayton, D. and Hills, M. (2013). Statistical Methods in Epidemiology, OUP
- Ahrens W. and Pigcot I.(2005). Handbook of Epidemiology, Springer.

Approval Date	
Version	
Approval by	
Implementation Date	

MAHATMA GANDHI UNIVERSITY

DAM21E16: BIOINFORMATICS AND COMPUTATIONAL BIOLOGY

School Name	**School of Data Analytics**					
Programme	**M.Sc. Data Science & Analytics**					
Course Name	**BIOINFORMATICS AND COMPUTATIONAL BIOLOGY**					
Type of Course	Elective					
Course Code	**DAM21E16**					
Course Summary & Justification	To make students aware of basics of bioinformatics and computational biology and make them familiar with data bases like NCBI-GENBANK. It is also expected to give them training in RASMOL software and computer aided drug design.					
Total StudentLearning Time (SLT)	Learning Approach	Lecture	Tutorial	Practical	Others	Total LearningHours

	Authentic learning Collaborative learning Independent learning	70	10	30	10	120
Pre-requisites	**Basic knowledge in Biology**					

COURSE OUTCOMES (CO)

CO No.	Expected Course Outcome	Learning Domains	PSO No.
1.	Students have understood basics of bioinformatics and computational biology.	U/R/A/An/E	1,2,3,4,5,6,7
2	They are aware of different data bases like NCBI-GENBANK and can use RASMOL for drug development.	U/R/A/An/E/C/S/Ap	1,2,3,4,5,6,7
3	They can apply these for bioinformatics modelling and pharmaceuticals and genome analysis	U/R/A/An/E	1,2,3,4,5,6,7
4	Students are aware of Hidden Markov Models and can apply these for parameter estimation	U/R/A/An/E/C/S/Ap	1,2,3,4,5,6,7
Remember (R), Understand (U), Apply (A), Analyse (An), Evaluate (E), Create (C), Skill (S), Interest (I) and Appreciation (Ap)2,5			

COURSE CONTENT

Module No	Module Content	CO	Hrs
1	Basics of Bioinformatics, Definition, importance and role of bioinformatics in life sciences. Computational Biology Biological sequences, Biological databases: Nucleotide sequence databases (NCBI- GENBANK, DDBJ and EMBL). Protein databases - structure and sequence databases (PDB, SWISSPROT and UNIPROT).	1	15
2	Introduction to Sequence alignment: Local alignment and Global alignment, Pair wise alignment (BLAST and FASTA) and multiple sequence alignment. Phylogenetic Tree construction and Analysis, Molecular visualization software - RASMOL. Basic concepts of Drug discovery pipe line, computer aided drug discovery and its applications. Human Genome Project and genome analysis.	1	20
3.	Applications of HMM to biological sequence analysis, Markov chain	1	20

	as a classifier, use of Markov chain Model for demarcation of a region in Biological sequence analysis, Applications in genetic sequence analysis such as detection of CPG Island. Testing whether given stretch of sequence is coming from CPG Island (use of Markov model for discrimination) **,** Markov model based classification & cluster analysis, testing order of a Markov model, testing homogeneity of two Markov models, Use of these tests to design clustering algorithm.		
4	Hidden Markov/chains, Difference between these and simple Markov chains, Analysis of Hidden Markov Models/chains, Verterbi's algorithm, Forward and backward algorithm for hidden Markov model, Parameter estimation in hidden Markov model when path is known as well as unknown, Baum – Welch algorithm	1	15
	Total Credits of the Course	3	70

Teachingand LearningApproach	**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
Assessment Types	**Mode of Assessment** **A. Continuous Internal Assessment (CIA)-40 marks** 1. Internal Tests of maximum 20 marks 2. Seminar Presentation – a theme is to be discussed and identified to prepare a paper and present in the seminar Maximum marks 10 3. Write a detailed report on a given topic based on research findings and literature search – 10 marks B. **Semester End examination – 60 marks**

REFERENCES

Compulsory Reading:

1. Alexander Isaac: (2006). Introduction to Mathematical Methods in Bioinformatics. Springer.
2. Durbin R., Eddy S., Krogh A., Michelson G. (1998). Biological Sequence Analysis, Cambridge University Press.

Further Reading:

3. Rajan S S and Balaji R, (2002), Introduction to Bioinformatics, Himalaya Publishing

House

4. Waterman, M.S. (2000). Introduction to Computational Biology, CRC Pres

Approval Date	
Version	
Approval by	
Implementation Date	

MAHATMA GANDHI UNIVERSITY

DAM21E17: ADVANCED RESAMPLING TECHNIQUES

School Name	**School of Data Analytics**					
Programme	**M.Sc. Data Science & Analytics**					
Course Name	**ADVANCED RESAMPLING TECHNIQUES**					
Type of Course	Elective					
Course Code	**DAM21E17**					
Course Summary & Justification	To enable the students to understand advanced resampling methodologies and techniques such as Jackknife and bootstrapping.					
Total StudentLearning Time (SLT)	Learning Approach	Lecture	Tutorial	Practical	Others	Total LearningHours
	Authentic learning Collaborative learning Independent learning	80	10	20	10	120
Pre-requisites	Understanding of basics knowledge of resampling techniques such as Jackknife and bootstrapping					

COURSE OUTCOMES (CO)

CO	Expected Course Outcome	Learning	PSO No.

No.		Domains	
1.	The students are now aware of the resampling techniques and are able to apply Jackknifing and bootstrapping techniques and compare their performance	A,An, Ap	1, 2
2	. Students are able to apply the above techniques in advanced research like micro-array analysis, gene sequencing, etc.	E,C,S	3,4
3	Students get exposure to Resampling in non-i.i.d. models, Resampling in linear models: with special emphasis on residual bootstrap and weighted bootstrap,	E,C,S	2,5
4	They have understood the concept of robust and efficient resampling schemes, estimating equation and generalized bootstrap and can apply these.	E,C,S	2,4,5
Remember (R), Understand (U), Apply (A), Analyse (An), Evaluate (E), Create (C), Skill (S), Interest (I) and Appreciation (Ap)2,5			

COURSE CONTENT

Module No	Module Content	CO	Hrs
1	Normal approximation and its limitations. Shortcomings of analytic derivations, with examples. Estimates of the distribution function: parametric and non-parametric MLE. U-Statistics, Mm-Estimators, Resampling - Purpose of resampling, Examples, estimating the estimating variance, the sampling distribution and other features of a statistic.	1,2	20
2	Jackknife. Bias reduction. Estimation of variance. Delete 1 and delete d jackknives. Examples. Bootstrap. Parametric and non-parametric bootstrap. Estimation of variance, estimation of distribution function. Examples. Comparison between bootstrap approximation and normal approximation. Examples.	2,3	20
3.	Notions of variance consistency and distributional consistency. Jackknife distributional inconsistency. Bootstrap distributional consistency. Comparisons between bootstrap and jackknife. Examples.	4,5	20
4	Resampling in non-i.i.d. models: need for other resampling schemes. Examples. Resampling in linear models: special emphasis on residual bootstrap and weighted bootstrap, Concept of robust and efficient resampling schemes. Introduction to estimating equation and generalized bootstrap. Examples.	3,4,5	20
	Total Credits of the Course	3	80

Teachingand LearningApproach	**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
Assessment Types	**Mode of Assessment** **A. Continuous Internal Assessment (CIA)-40 marks** 1. Internal Tests of maximum 20 marks 2. Seminar Presentation – a theme is to be discussed and identified to prepare a paper and present in the seminar Maximum marks 10 3. Write a detailed report on a given topic based on research findings and literature search – 10 marks B. **Semester End examination – 60 marks**

REFERENCES

Compulsory Reading:

1. Bose, Arup and Chatterjee, Snigdhansu (2018). U-Statistics, Mm-Estimators and Resampling, Hindusthan Book Agency, June 2018.
2. Davidson, A. C. and Hinkley D. V. (1997). Bootstrap methods and their applications.
3. Efron, B. (1982). The Jackknife, the Bootstrap and other Resampling Plans. CBMS-NSF Regional Conference Series in Applied Mathematics,No 38.
4. Efron, B. and Tibshirani, R. J. (1993). An Introduction to the Bootstrap. Chapman & Hall/CRC.
5. Shao, J. and Tu, D. (1995). The Jackknife and Bootstrap. Springer.

Further Reading:

- Barbe, P.. and Bertail, P. (1995). *The Weighted Bootstrap*. Lecture Notes in Statistics, Vol 98.
- Gine, E. (1997). *Lectures on Some Aspects of the Bootstrap*. Lecture Notes in Mathematics,Vol 1665. Springer
- Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*. Springer.
  .
- Politis, D. N., Romano J. P. and Wolf, M. (1999). *Subsampling*. Springer.

Approval Date	
Version	
Approval by	
Implementation Date	

MAHATMA GANDHI UNIVERSITY

DAM21E18: TIME SERIES ANALYSIS & FORECASTING

School Name	**School of Data Analytics**					
Programme	**M.Sc. Data Science & Analytics**					
Course Name	**TIME SERIES ANALYSIS & FORECASTING**					
Type of Course	Elective					
Course Code	**DAM21E18**					
Course Summary & Justification	By the end of this course the student will be able to analyse time series data and identify and interpret various types of behaviour of the time series. They will be to model a time series data and forecast future values from it.					
Total StudentLearning Time (SLT)	Learning Approach	Lecture	Tutorial	Practical	Others	Total LearningHours
	Authentic learning Collaborative learning Independent learning	80	10	20	10	120
Pre-requisites	**Basic knowledge in Statistical Distribution, Stochastic Processes**					

COURSE OUTCOMES (CO)

CO No.	Expected Course Outcome	Learning Domains	PSO No.
1.	Now students are aware of various aspects of time series analysis and modeling.	U/R/A/An/E/Ap	1,2,3,4,5,6,7
2	They are able to analyze a given time series data and fit suitable ARIMA model using Box-Jenkins method.	U/R/A/An/E/S/C	1,2,3,4,5,6,7
3	Students are aware of spectral density, periodogram etc as well as various non-linear models like ARCH, GARH, Bilinear, Multivariate and Vector Autoregression.	U/R/A/An/E/S/C	1,2,3,4,5,6,7
4	Now students are able to model a given time series data and forecast future values		2,4,5

Remember (R), Understand (U), Apply (A), Analyse (An), Evaluate (E), Create (C), Skill (S), Interest (I) and Appreciation (Ap)2,5

COURSE CONTENT

Module No	Module Content	CO	Hrs
1	Time series as a discrete parameter stochastic process, Auto-covariance and auto-correlation functions, Partial Auto-correlation function and their properties, Stationary processes, Wold representation of linear stationary processes, Detailed study of the Box - Jenkins linear time series models: Autoregressive, Moving Average, Autoregressive Moving Average and Autoregressive Integrated Moving Average models.		20
2	Estimation of ARMA models: Yule-Walker estimation for AR Processes, Maximum likelihood and least squares estimation for ARMA Processes. Choice of AR and MA periods, Forecasting using ARIMA models, Residual analysis and diagnostic checking. Forecasting of future values..		20
3.	Non-Gaussian Time Series Modeling- exponential, gamma and Laplace, Cauchy, Stable stationary marginal distributions; Diagnostic checking and Forecasting; Integer Valued time series models, Count data modeling, ACT and PACF and model identification, Minification models, Mini-max processes		20
4	Spectral density of a stationary time series and its elementary properties, Periodogram, Spectral density of AR, MA, ARMA processes. Seasonal ARIMA models Introduction to ARCH and GARCH models and their applications; non-linear time series models, bilinear models; Multivariate Time Series Analysis and Vector Auto-regression.		20
	Total Credits of the Course	3	80

Teachingand LearningApproach	**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
Assessment Types	**Mode of Assessment** **A. Continuous Internal Assessment (CIA)-40 marks** 1. Internal Tests of maximum 20 marks 2. Seminar Presentation – a theme is to be discussed and identified to prepare a paper and present in the seminar Maximum marks 10 3. Write a detailed report on a given topic based on research findings and literature search – 10 marks B. **Semester End examination – 60 marks**

REFERENCES

Compulsory Reading:

1. Brockwell P.J and Davis R.A. (2002) Introduction to Time Series and Forecasting Second Edition, Springer-Verlag.
2. Box G.E.P, Jenkins G.M. and Reinsel G.C. (2008) Time Series Analysis: Forecasting and Control, Fourth Edition, Wiley.
3. Abraham B. and Ledolter J.C. (2005) Statistical Methods for Forecasting, Second Edition Wiley.

Further Reading:

- Cryer, J. D. and Chan, K. (2008). Time Series Analysis with Applications in R, Second Edition, Springer-Verlag.
- Shumway, R. H. and Stoffer, D. S. (2011) Time Series Analysis and Its Applications with R Examples, Third Edition, Springer-Verlag.

Approval Date	
Version	
Approval by	
Implementation Date	

MAHATMA GANDHI UNIVERSITY

DAM21E19: DEMOGRAPHY & POPULATION STUDIES

School Name	**School of Data Analytics**					
Programme	**M.Sc. Data Science & Analytics**					
Course Name	**DEMOGRAPHY & POPULATION STUDIES**					
Type of Course	Elective					
Course Code	**DAM21E19**					
Course Summary & Justification	Students are now able to understand and use various mortality rates , to construct life tables, to calculate and use various characteristics of life time models and population growth models.					
Total StudentLearning Time (SLT)	Learning Approach	Lecture	Tutorial	Practical	Others	Total LearningHours
	Authentic learning Collaborative learning Independent learning	80	10	20	10	120
Pre-requisites	Understanding the basics of Indian population census and statistical models etc.					

COURSE OUTCOMES (CO)

CO No.	Expected Course Outcome	Learning Domains	PSO No.
1.	Now students have understood various mortality and morbidity measures	U/R/A/An/E	1,2,3,4,5, 6,7
2	They are equipped to develop life tables, fertility indices etc using different methods.	U/R/A/An/E	1,2,3,4,5, 6,7
3	Also they can construct population growth indices and develop population growth models.		2,5
4			2,4,5
***Remember (R), Understand (U), Apply (A), Analyse (An), Evaluate (E), Create (C), Skill**			

(S), Interest (I) and Appreciation (Ap)2,5

COURSE CONTENT

Module No	Module Content	CO	Hrs
1	Sources of mortality data-mortality measures-ratios and proportions, crude mortality rates, specific rates- measures Mortality trends, levels and determinants in India with special reference to infant mortality and maternal mortality, Cause of death statistics standardization of mortality rates, direct and indirect methods, gradation of mortality data, fitting Gompertz and Makeham curves, Concepts and definitions of health and morbidity, morbidity rates, Sources of data on mortality and morbidity.	1,2	20
2	Life tables-complete life table-relation between life table functions, abridged life table-relation between abridged life table functions,construction of life tables, Greville's formula, Reed and Merrell's formula- sampling distribution of life table functions, multivariate pgf –estimation of survival probability by method of MLE.	2	20
3.	Fertility models, fertility indices relation between CBR,GFR,TFR and NRR, stochastic models on fertility and human reproductive process, Dandekar's modified binomial and Poisson models, Brass, Singh models, models for waiting time distributions, Sheps and Perrin model.	2	20
4	Population growth indices, logistic model, fitting logistic, other growth models, Lotka's stable population, analysis, quasi stable population, effect of declining mortality and fertility on age structure, population projections, component method- Leslie matrix technique, properties of time independent Leslie matrix-models under random environment.	2	20
	Total Credits of the Course	3	80

Teachingand LearningApproach	**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
Assessment Types	**Mode of Assessment** **A. Continuous Internal Assessment (CIA)-40 marks** 1. Internal Tests of maximum 20 marks 2. Seminar Presentation – a theme is to be discussed and identified to prepare a paper and present in the seminar Maximum marks 10 3. Write a detailed report on a given topic based on research findings

	and literature search – 10 marks B. **Semester End examination – 60 marks**

REFERENCES

Compulsory Reading:

1. Biswas, S. (2007) Applied Stochastic Processes-A Biostatistical and Population Oriented Approach, Second Edition, New Central Book Agency.
2. Pollard, J. H. (1975) Mathematical Models for the Growth of Human Population, Cambridge University Press.
3. Shrivastava OS, 1998, Demography and Population Studies, Vikas Publishing House Pvt Ltd , 2nd Edition.

Further Reading:

- Biswas, S. (1988) Stochastic Processes in Demography and Applications, Wiley Eastern.
- Keyfitz, N. (1977) Applied Mathematical Demography A Wiley Interscience publication.
- Ramkumar, R. (1986) Technical Demography, Wiley Eastern.
- Srinivasan, K. (1970) Basic Demographic Techniques and Applications.
- Keyfitz, Nathan. (1977). Introduction to the Mathematics of Population, Addision-Wesley Publishing Company, Massachusetts.
- Pathak, K. B. and Ram, F. (2015). Techniques of Demographic Analysis, 2nd Revised ed. Himalaya Publishing House, New Delhi.

Approval Date	
Version	
Approval by	
Implementation Date	

MAHATMA GANDHI UNIVERSITY

DAM21E20: INDUSTRIAL STATISTICS & QUALITY CONTROL

School Name	**School of Data Analytics**					
Programme	**M.Sc. Data Science & Analytics**					
Course Name	**INDUSTRIAL STATISTICS & QUALITY CONTROL**					
Type of Course	Elective					
Course Code	**DAM21E20**					
Course Summary & Justification	To make the students aware of the modern quality assurance techniques and methods applied in industry and engineering					
Total StudentLearning Time (SLT)	Learning Approach	Lecture	Tutorial	Practical	Others	Total LearningHours
	Authentic learning Collaborative learning Independent learning	80	10	20	10	120
Pre-requisites	Elementary knowledge of statistical methods and probability. Basic knowledge of various Statistical Distributions					

COURSE OUTCOMES (CO)

CO No.	Expected Course Outcome	Learning Domains	PSO No.
1.	Students are able to carry out quality assessment in various contexts using control charts..	U, A, An	1,2
2	They are aware of process capability indices for efficient process control	A, An, App	3,4
3	Students are now aware of acceptance sampling for variables and attributes	C,S,I	2,5
4	They are aware of OC, ASN etc for comparison	A, An, App	2,4,5

Remember (R), Understand (U), Apply (A), Analyse (An), Evaluate (E), Create (C), Skill (S), Interest (I) and Appreciation (Ap)2,5

COURSE CONTENT

Module No	Module Content	CO	Hrs
1	Meaning of quality, and need for quality control. Meaning and scope of statistical process control, General theory of control charts, Shewhart control charts for variables- mean charts, R-charts, and S-charts, Moving-average control charts. Attribute control charts - p, np, c, u charts. OC and ARL curves of control charts.	1	20
2	Modified control charts. Control charts with memory - EWMA charts, CUSUM charts. Economic design of mean charts Process capability analysis, process capability indices – Cp Cpk, Cpm. Statistical aspect of six sigma philosophy, The Taguchi Method: The Taguchi philosophy of Quality, Quality in the service sector and TQM,	1	20
3.	Statistical product control- basic ideas. Acceptance sampling for attributes - single sampling, double sampling, multiple sampling and sequential sampling plans. ASN curves. Measuring performance of sampling plans through OC curves. Rectifying inspection plans.AOQ and ATI curves,	1	20
4	Acceptance sampling by variables. Sampling plan for a single specification limit with known and unknown variance. Performance evaluation through OC curves. Designing a variable sampling plan with a specified OC curve. Multivariate statistical process control, Hotelling's T^2control chart.	1	20
	Total Credits of the Course	3	80

Teachingand LearningApproach	**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

Assessment Types	**Mode of Assessment** **A. Continuous Internal Assessment (CIA)-40 marks** 1. Internal Tests of maximum 20 marks 2. Seminar Presentation – a theme is to be discussed and identified to prepare a paper and present in the seminar Maximum marks 10 3. Write a detailed report on a given topic based on research findings and literature search – 10 marks B. **Semester End examination – 60 marks**

REFERENCES

Compulsory Reading:

1. Montgomery, D.C. (2012). Introduction to Statistical Quality Control, Seventh edition, Wiley.
2. Duncan, A.J. (1986) Quality Control and Industrial Statistics, Irwin, Homewood
3. Grant E.L. and Leaven Worth, R.S. (1980) Statistical Quality Control, McGraw Hill.

Further Reading:

- Amitava Mitra. (2016). Fundamentals of Quality Control and Improvement, 4th edition Pearson Education Asia.
- Mittag, H.J. and Rinne, H. (1993) Statistical Methods for Quality Assurance, Chapman & Hall.
- David Hoyle.(2017). ISO 9000 Quality Systems Handbook, Routledge; 6 Edition.
- Rabbit, J T and Bergle, P.A. The ISO 9000 Book, Quality Resources, Second Edition,
- Schilling, E.G. (1982) Acceptance Sampling in Quality Control, Marcel Dekker.
- Wetherill, G. B. and Brown, D.W (1991). Statistical Process Control: Theory and Practice. 3rd Ed. Chapman and Hall, USA.

Approval Date	
Version	
Approval by	
Implementation Date	

OPEN COURSE

MAHATMA GANDHI UNIVERSITY

DA M 21 O 01: ADVANCED STATISTICAL TECHNIQUES FOR DATA ANALYTICS

School Name	**School of Data Analytics**					
Programme	**M.Sc. Data Science & Analytics**					
Course Name	**Advanced Statistical Techniques for Data Analytics**					
Type of Course	**Open Course**					
Course Code	**DA M 21 O 01**					
Course Summary & Justification	To provide sound knowledge of Operations Research with applications to data science and analytics					
Total StudentLearning Time (SLT)	Learning Approach	Lecture	Tutorial	Practical	Others	Total LearningHours
	Authentic learning Collaborative learning Independent learning	75	10	25	10	120
Pre-requisites	**Basic knowledge in Statistical Distribution, Stochastic Processes**					

COURSE OUTCOMES (CO)

CO No.	Expected Course Outcome	Learning Domains	PSO No.
1.	Students become aware of various univariate and multivariate distributions and their important characteristics	A, App, C	1, 2
2	Students are able to carry out uni-variate and multivariate estimation and testing methods, diamension reduction techniques, classification and clustering techniques etc	An, App, S	3,4
3	Students are aware of various stochastic process models like Markov Chains, Poisson Processes, Birth Death Processes and their applications	E, App,I	2,5
4	Now students are familiar with different time series models and their applications	App, I	2,4,5

Remember (R), Understand (U), Apply (A), Analyse (An), Evaluate (E), Create (C), Skill (S), Interest (I) and Appreciation (Ap)2,5

COURSE CONTENT

Module No	Module Content	CO	Hrs
1	Basic concepts on univariate and multivariate random variables, Concept of random vector, Mean vector and Dispersion matrix, Binomial, Poisson and Multinomial distributions, Univariate and Multivariate normal distributions, Marginal and conditional distributions and their applications, univariate tests for mean and equality of means of two normal populations.		15
2	Sample mean vector and its distribution, Tests for mean, when population variance is known and unknown, Hotelling's T^2 and Mahalanobis' D^2 statistics and applications. Tests of hypotheses about the mean vectors, Wishart distribution, Discriminant analysis, Principal components, Canonical correlations, Factor analysis, Cluster analysis.		15
3.	Stochastic modelling, Stochastic processes, Markov processes, Markov chains, Poisson processes, Pure birth processes and the Yule processes, birth and death processes, Kolmogorov differential equations, linear growth process with immigration, steady-state solutions of Markovian queues - M/M/1, M/M/s, M/MM/∞ models, Renewal processes, Branching processes-basic concepts and applications.		15
4	Time series data, examples, Time series as stochastic process, Additive and multiplicative models, stationary time series- covariance stationarity, Modeling Time Series Data, Exponential Smoothing and Forecasting, AR, MA, ARMA, ARIMA Models. Diagnostic checking		15
5	Time series modeling, Autocorrelation function (ACF), partial auto correlation function (PACF), correlogram, Yule-Walker equations, Box-Jenkins Model fitting and diagnostics, Forecasting future values, Non-linear and non- Gaussian time series models, ARCH and GARCH Models. Integer Valued Time Series of Count data, Model identification.		15
	Total Credits of the Course	3	75

Teachingand LearningApproach	**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
Assessment Types	**Mode of Assessment** **A. Continuous Internal Assessment (CIA)-40 marks** 1. Internal Tests of maximum 20 marks 2. Seminar Presentation – a theme is to be discussed and identified to prepare a paper and present in the seminar Maximum marks 10 3. Write a detailed report on a given topic based on research findings

	and literature search – 10 marks B. **Semester End examination – 60 marks**

REFERENCES

References

1. Rencher, A. C. (2012) Methods of Multivariate Analysis. (3rd ed.) John Wiley.
2. Johnson R.A. and Wichern D.W. (2008) Applied Multivariate Statistical Analysis.sss 6th Edition, Pearson Education.
3. Medhi J. (2017) Stochastic Processes, Second Edition, Wiley Eastern, New Delhi
4. Karlin S. and Taylor H.M. (1975) A First Course in Stochastic Processes, Second Edition, Academic Press, New-York.
5. Brockwell P.J and Davis R.A. (2002) Introduction to Time Series and Forecasting Second edition, Springer-Verlag.

Approval Date	
Version	
Approval by	
Implementation Date	

Rubrics selected for OBE implementation

1. **Overall performance** in each course of the semester on a continuous basis
2. Response to **critical theoretical questions** in each course
3. Procedural approach adopted towards **lab oriented critical questions**
4. Response to **socially relevant issues and recent trends** in each course
5. **Aptitude to research** and **specific research problem** in each course

PART 1Task Description

1. Written Examinations
2. Assignments/Case Studies
3. Seminars / Case Study Presentations
4. Practical Examination, Skill Testing
5. Viva Voce

PART II Scale- Continuous mode

Excellent, Satisfactory, Needs improvement (Remedial practices recommended)

PART III Dimensions

Written Examination-Content, Communicating

Assignment -Content, level of Comprehension

Seminar, Case Study-Content, Performance, Execution, Application

Practical Exam- Conduct of Practical, Observation and recording

Viva Voce -Response to questions, Attitude

PART IV Description of the dimensions

Content-Brief and meaningful

Comprehension- Precise and effective

Communicating-Direct and orderly

Procedure adopted- Scientific Suitability and easiness

Conduct of practical-Accuracy and reproducibility

Observation and recording- Sharp and systematic

Response to questions- Analytical approach and level of accuracy

Attitude- Positive and self-inspiring

MODEL QUESTION PAPERS

Mahatma Gandhi University, Kottayam

School of Data Analytics

M. Sc. Data Science & Analytics Semester 1 Examination

DA M 21 C01 INTRODUCTION TO DATA SCIENCE& DATA ANALYTICS

Time:3 hours **Max Marks:60**

PART A (Answer all questions. Each question carries 1 mark)

1. Which is the preferred way to manage and query data that resides in database?
   a)sql b)html c)xml d)None of these
2. Which among the following involves in data engineering.
   a. Acquiring data b. Ingesting data
   c. Transforming data d. All of these
3. To find the percentage score and age of the students in a class. Score and age can be considered as two variables. What type of dataset is this?
   a. Numerical b. Bivariate c. Multivariate d. Categorical
4. Which among the following is not a data type in R
   a. Series b. Vector c. List d. Data Frame
5. Consider the following statements

PART B (Answer any 5 questions. Each question carries 2 marks)

6. Define Data set.
7. List the Applications of Data Science .
8. Explain the role of Mathematics in Data science.
9. Define the term ‘Meta data’.
10. List out the difference between structured and unstructured data?
11. How can we extract data from R Data Frame?
12. Briefly explain data types in Mathematics ?
13. What is list? Write the syntax for creating List in R Programming.

PART C (Answer any 5 questions.Each question carries 5 marks)

14. Who is a ‘Data Scientist’? What are the skills needed for a Data Scientist?

15. Define Data science. Why we need hacker mindset in Data science?

16. How can we create data sets using R Language?
17. What are Factors in R language?How it differ from List?
18. Explain 3 Vs Of Big data.
19. What is Exploratory Data Analysis? Explain
20. Write a note on 'Advanced Computing'?
21. What are unstructured data? Discuss the advantages of unstructured data.

PART D (Answer any 2 questions.Each question carries 10 marks)

22. Data Science is "Mash-up of Disciplines", Substantiate
23. What is Data set? Explain different data sets in detail.
24. What are data objects in R language? Explain DataFrame object.
25. Illustrate the steps in doing data science

Mahatma Gandhi University, Kottayam

School of Data Analytics

M. Sc. Data Science & Analytics: Semester 1 Examination

DA M 21C04 DATA BASE MANAGEMENT SYSTEMS

Time:3 hours **Max Marks:60**

PART A (Answer all questions. Each question carries 1 mark)

1. Which of the following is used to denote the selection operation in relational algebra?
   a) Pi (Greek) b) Sigma (Greek)
   c) Lambda (Greek) d) Omega (Greek)
2. In the _____ normal form, a composite attribute is converted to individual attributes.
   a) first b) Second c) Third d) Fourth
3. Which of the following refers to the level of data abstraction that describes exactly how the data actually stored?
   a)Conceptual Level b)Physical Level c)File Level d)Logical Level

4. Which of the following makes the transaction permanent in the database?.

a) View b) commit c) rollback d) flask back

5. OLAP stands for

a) Online Analytical Processing b) Online analysis processing

d) Online transaction processing d) Online aggregate processing

PART B (Answer any 5 questions. Each question carries 2 marks)

6. Define Functional dependency
7. What are **Views** in Database.
8. Describe NoSQL.
9. Draw any four symbols used in ER diagram and mention its purpose.
10. What are the advantages of DDBMS.
11. What is Replication in data storage?
12. Explain the data warehouse life cycle?
13. How to detect the presence of deadlock in a concurrent schedule?

PART C (Answer any 5 questions.Each question carries 5 marks)

14. Explain DML Commands with example.
15. What are Key attribute? Given relational schema R(P Q R S T) having following attributes P Q R S and T, also there is a set of functional dependency denoted by FD = { P->QR, RS->T, Q->S, T-> P }.Find the Key attribute.
16. Draw an E-R diagram of a college database with entities student, staff, course, teacher, clerk, department & hostel. Relationship names must be meaningful
17. What is multi valued dependency? Explain with an example.
18. What is Serializability? Differentiate between serial schedule and serializable schedule.
19. Explain OODBMS.
20. Explain components of data warehousing.
21. Write a note on OLAP tools.

PART D (Answer any 2 questions.Each question carries 10 marks)

22. What are data models? Illustrate different data models.
23. What is normalization? Explain 1NF,2NF and 3NF in detail.
24. a)What are transactions? Explain Properties of a transaction.

    b)Explain Two Phase Locking protocol.
25. What is data warehousing? List the advantage of Data warehouse. Explain how to build a data warehouse?

Mahatma Gandhi University, Kottayam

School of Data Analytics

M.Sc Data Science & Analytics: Semester 1 Semester Examination

DA M 21C05: PROGRAMMING IN PYTHON FOR DATA SCIENCE

Time: 3 hours **Max Marks:60**

PART A (Answer all questions. Each question carries 1 mark)

1. Which character is used in Python to make a single line comment?
   a)# b); c)// d)&
2. Which of the following functions is a built-in function in python?
   a. factorial() b) print()
   c) seed() d) sqrt()
3. NumPy stands for ____
   a. Number Python b. Numerical Python c. Numbers in Python d. None of the above
4. Which of the following statement is wrong
5. Recommanded way to load matplotlib library is

   a) import matplotlib.pyplot as plt b) import matplotlib.pyplot

   c) import matplotlib as plt d) import matplotlib

PART B (Answer any 5 questions; Each carries 2 marks)

6. What are variables?Mention the rules for creating legal variable names in Python.
7. What are Identifiers?
8. What is Numpy?
9. Describe any four Numpy array attributes.
10. How to create a Series data structure in Python?
11. Describe the features of data frame?
12. Develop a python program to draw a simple Scatter plot ?
13. Write a note on dictionary in Python.

PART C (Answer any 5 questions.Each question carries 5 marks)

14. Write a python program to check whether the given number is odd or even,If it is even Print the multiplication table of the number. Else Print sum of first three multiples of the number.The number should be entered by the user
15. Illustrate any two control flow statements in Python with sample code segment.
16. Explain different arithmetic,Mathematical functions on numpy with code segment?

17. Write a numpy program to create
    a. 3x3 Identity Matrix
    b. An array of all even numbers from 10 to 50
18. Illustrate different Arithmetic, Mathematical functions on Numpy with code segment
19. Write a Python Program to add,subtract,multiply and divide two pandas series.
20. Explain merge function in pandas.
21. With example illustrate how a scatter plot will draw using matplotlib ?

PART D (Answer any 2 questions.Each question carries 10 marks)

22. Explain looping statements in Python.
23. How to create a Numpy array? Develop a program to add,subtract and multiply two 2D arrays.
24. Explain different join operation in Pandas with sample data sets.
25. What is Matplotlib? Summarize the Features of Matplotlib. Describe how it differ from Altair.