(P.G. DEPARTMENT OF COMPUTER SCIENCE)

OUTLINES OF TESTS,
SYLLABI AND COURSES OF READING

FOR

M.Sc. ARTIFICIAL INTELLIGENCE AND DATA SCIENCE

(SEMESTER SYSTEM)

SECOND YEAR (Semester III & IV)

(2021-22 and 2022-23 Sessions)

FACULTY OF COMPUTING SCIENCES

SRI GURU TEG BAHADUR KHALSA COLLEGE

Sri Anandpur Sahib

An Autonomous College

Affiliated to Punjabi University, Patiala

APPROVED
Board of Studies Meeting held on 23rd June 2021

PROGRAMME OF STUDY OF M.Sc.(AI & DS) PART-II

(SEMESTER-III & IV)

FOR SESSIONS 2021-22, 2022-23

Semester-III

Paper Code	Name of Subject	Contact Hours per week				Examination scheme marks				Credit
		L	T	P	Total	Internal	External	Practical	Total	
MSAIDS-211	Data Analysis Using Python	5			5	30	70		100	5
MSAIDS-212	Soft Computing	4	1		5	30	70		100	5
MSAIDS-213	Digital Image Processing	5			5	30	70		100	5
MSAIDS-214	Computer Network Technologies	4	1		5	30	70		100	5
MSAIDS-215	Software Lab-III (Based on MSAIDS-211)			4	4	30		70	100	2
MSAIDS-216	Software Lab-IV (Based on MSAIDS-213)			4	4	30		70	100	2
	Total	**18**	**2**	**8**	**28**	**180**	**280**	**140**	**600**	**24**

Semester-IV

Paper Code	Name of Subject	Contact Hours per				Examination scheme marks				Credit
		L	T	P	Total	Internal	External	Practical	Total	
MSAIDS-221	Research Methodology	4	1		5	30	70		100	5
MSAIDS-222	Natural Language Processing Using Python	5			5	30	70		100	5
MSAIDS-223	Big Data Analytics	5			5	30	70		100	5
MSAIDS-224	Software Lab-V (Based on MSAIDS-222)			4	4	30		70	100	2
MSAIDS-225	Software Lab-VI (Based on MSAIDS-223)			4	4	30		70	100	2
MSAIDS-226	Project (2 Weeks In House Industrial Training)					30		70	100	5
	Total	**14**	**1**	**8**	**23**	**180**	**210**	**210**	**600**	**24**

APPROVED

NOTE:

The Break-up of Marks for Practical exams (External) will be as under:

1. Viva-Voce(External examination)	20Marks
2. Program Development and Execution	20Marks
3. File Record	30Marks

Internal Assessment will be based on Continuous Comprehensive Assessment (CCA) pattern and the breakup of Internal Assessment will be asunder:

a) Average of Two mid Semester Tests: **60%**

b) Assignment/Seminar/Class Test/Tutorial/Quiz etc.: **20%**

c) Attendance: **10%**

d) Class participation and behavior: **10%**

MSAIDS-211 Data Analysis using Python

Maximum Marks: 100 **Time Allowed: 3 Hours**
External Examination: 70 Marks **Pass Percentage: 35%**
Internal Assessment: 30 Marks **Teaching Hours per Week: 5**
Credits: 5 (5L)

Instruction for the Paper Setter: The question paper will consist of three sections: A, B & C. Sections A & B will have four questions each from the respective sections of the syllabus carrying 10 marks for each question. Section C will have 10 short-answer type questions carrying 30 marks for each question which will cover the entire syllabus uniformly.
Instruction for the candidates: Candidates are required to attempt two questions each from the Sections A & B of the question paper and the entire section C.

Course Objectives: The main objectives of this course are:

1. To install and use Jupyter notebooks.
2. To demonstrate the basic and advanced concepts of Numpy and its various functions.
3. To introduce Pandas Series and Data Frames.
4. To introduce data manipulation and cleaning techniques using Pandas.
5. To demonstrate the loading of various formats and various data manipulation operations like data wrangling, group operations etc.
6. To introduce data visualization and plotting tools.

SECTION-A

Introduction to Data Analysis: Kinds of Data, Essential Python Libraries: NumPy, Pandas, Matplotlib, IPython and Jupyter, SciPy, scikit-learn, statsmodels
NumPy Basics: The NumPy ndarray: Creating ndarrays, Data types for ndarraays, Arithmetic with NumPy Arrays, Basic indexing, slicing, Boolean indexing, Fancy indexing, transposing arrays and swapping axes, Universal Functions: Fast Element-Wise Array Functions, Array-Oriented Programming with Arrays.
Advanced NumPy: ndarray Object Internals, Advanced Array Manipulation, Broadcasting.
Getting Started With Pandas: Introduction to pandas Data Structures: Series, Data frame, Index Objects
Essential Functionality: Reindexing, Dropping Entries from an Axis, Indexing, Selection, Filtering, Integer Indexes, Arithmetic and Data Alignment, Function Application and Mapping, Sorting and Ranking, Axis Indexes with Duplicate Labels, Summarizing and Computing Descriptive Statistics.
Data Loading, Storage, and File Formats: Reading and Writing Data in Text Format: csv files, json files.

SECTION-B

Data Cleaning and Preparation: Handling Missing Data: filtering out missing data, filling in missing data; Data Transformation: Removing duplicates, transforming data using a function or mapping, replacing values, renaming axis indexes, String Manipulation: string object methods.
Data Wrangling: Join, Combine, and Reshape, Hierarchical Indexing, Combining and Merging Datasets, Reshaping with hierarchical indexing.

Data Aggregation and Group Operations: GroupBy Mechanics, Iterating over groups, selecting a column or subset of columns, Data Aggregation
Advanced pandas: Categorical Data, Advanced GroupBy Use,
Plotting and Visualization: A Brief matplotlib API Primer: Figures, Subplots, colors, markers, Line styles, saving plots to file
Case Studies: 2012 Federal Election Commission Database.

Course Learning Outcomes: At the end of this course, students will be:

1. Be able to use Jupyter notebooks efficiently.
2. Learn how to work with NumPy datatypes.
3. Be proficient in pandas Series.
4. Be proficient in pandas DataFrames.
5. Understand how to use data visualization.
6. Know how to import and clean data.
7. Learn Data Wrangling and Aggregation operations.

Text Book:

1. Wes McKinney, Python for Data Analysis, Shroff Publications and Distributors

Reference Book:

1. Michael Milton, A Brain Friendly Guide: Head First Data Analysis, Shroff Publications and Distributors.
2. Bharti Motwani, Data Analytics using Python, Wiley publisher.
3. Anil Maheshwari, Data Analytics, McGraw Hill Education.

MSAIDS-212 Soft Computing

Maximum Marks: 100 **Time Allowed: 3 Hours**
External Examination: 70 Marks **Pass Percentage: 35%**
Internal Assessment: 30 Marks **Teaching Hours per Week: 5**
Credits: 5 (4L+1T)

Instruction for the Paper Setter: The question paper will consist of three sections: A, B & C. Sections A & B will have four questions each from the respective sections of the syllabus carrying 10 marks for each question. Section C will have 10 short-answer type questions carrying 30 marks for each question which will cover the entire syllabus uniformly.
Instruction for the candidates: Candidates are required to attempt two questions each from the Sections A & B of the question paper and the entire section C.

Course Objectives: The main objectives of this course are:

1. To introduce students to Soft Computing concepts and techniques.
2. To make the students familiar with architecture and various types of Neural Networks.
3. To introduce Fuzzy Logic, Architecture of Fuzzy Logic Systems and their applications.
4. To briefly introduce Genetic algorithms and its application areas.
5. To briefly explain Applications of Soft Computing.

SECTION-A

Introduction: Introduction to soft computing, Concept of computing systems, Elements of Soft Computing: Fuzzy Logic, Neural Networks, Genetic Algorithms, Adaptive Resonance Theory, "Soft" computing versus "Hard" computing, Characteristics of Soft computing.
Neural Networks: Introduction to Artificial Neural Networks, Functioning of Biological Brain & Neuron, Concept of Learning/Training, Model of an Artificial Neuron, Building Blocks: Single Layer Feed-forward, Multilayer Feed-forward, Feedback Network, Adjustments of Weights or Learning, Activation Functions, Learning and Adaption: Neural Network Rules; Back Propagation Network: Background, Back Propagation Learning, Back Propagation Algorithm.

SECTION-B

Fuzzy Logic: What is Fuzzy Logic, Characteristics of Fuzzy Logic System, Architecture of Fuzzy Logic System: Rule Base, Fuzzification, Inference Engine, De-Fuzzification; Membership Function, Classical Fuzzy Set Theory, Fuzzy Sets, Operation on Fuzzy Sets, Implication of Propositional Logic, Reasoning with Proposition, Predicate Logic, Fuzzy Reasoning Methods, Fuzzy Composition, Fuzzy Quantifiers, Fuzzy Inference, Applications of Fuzzy Logic.
Genetic Algorithms: Concepts, Creation of offspring, Working Principle, Encoding, Fitness Functions, Reproduction, Genetic Modeling; Generation Cycle & Convergence of Genetic Algorithm, Application areas of Genetic Algorithms.
Applications of Soft Computing: Image Registration, Object Recognition, Automated feature extraction, Navigation, Integration of soft computing and GIS for flood forecasting and monitoring, Landslide Susceptibility, Highway Alignment, Smart City Planning, Agriculture, Solid Waste Disposal.

Course Learning Outcomes: At the end of this course, students will be able to:

1. Conceptualize and parameterize various problems to be solved through basic soft computing techniques.
2. Identify and describe soft computing techniques and their roles in building intelligent machines.
3. To understand the fundamental theory and concepts of neural networks, Identify different neural network architectures and Back Propagation Algorithm.
4. Comprehend the fuzzy logic and the concept of fuzziness involved in various systems and fuzzy set theory.
5. Understand soft computing techniques and their role in problem solving.

Text Books:

1. S. N.Sivanandam and S.N.Deepa, "Principles of Soft Computing", Wiley India Pvt Ltd.
2. S, Rajasekaran & G.A. Vijayalakshmi Pai, Neural Networks, Fuzzy Logic & Genetic Algorithms, Synthesis & applications, PHI Publication.

Suggested Reading:

1. George J. Klir, Ute St. Clair, Bo Yuan, Fuzzy Set Theory: Foundations and Applications Prentice Hall.
2. David E. Goldberg, Genetic Algorithm in Search Optimization and Machine Learning Pearson Education India.

MSAIDS-213 Digital Image Processing

Maximum Marks: 100 **Time Allowed: 3 Hours**
External Examination: 70 Marks **Pass Percentage: 35%**
Internal Assessment: 30 Marks **Teaching Hours per Week: 5**
Credits: 5 (5L)

Instruction for the Paper Setter: The question paper will consist of three sections: A, B & C. Sections A & B will have four questions each from the respective sections of the syllabus carrying 10 marks for each question. Section C will have 10 short-answer type questions carrying 30 marks for each question which will cover the entire syllabus uniformly.
Instruction for the candidates: Candidates are required to attempt two questions each from the Sections A & B of the question paper and the entire section C.

Course Objectives: The main objectives of this course are:

1. To become familiar with digital image fundamentals.
2. To get exposed to simple image enhancement techniques in Spatial and Frequency domain.
3. To learn concepts of degradation function and restoration techniques.
4. To study the image segmentation and representation techniques.
5. To become familiar with image compression and recognition methods.

SECTION-A

Digital Image Fundamentals: Steps in Digital Image Processing, Components, Elements of Visual Perception, Image Sensing and Acquisition, Image Sampling and Quantization, Relationships between pixels, Color image fundamentals, RGB, HSI models, Two-dimensional mathematical preliminaries, 2D transforms, DFT, DCT.
Image Enhancement: Spatial Domain: Gray level transformations, Histogram processing, Basics of Spatial Filtering, Smoothing and Sharpening Spatial Filtering, Frequency Domain: Introduction to Fourier Transform, Smoothing and Sharpening frequency domain filters, Ideal, Butterworth and Gaussian filters, Homo-morphic filtering, Color image enhancement.

SECTION-B

Image Restoration: Image Restoration, degradation model, Properties, Noise models, Mean Filters, Order Statistics, Adaptive filters, Band reject Filters, Band pass Filters, Notch Filters, Optimum Notch Filtering, Inverse Filtering, Wiener filtering
Image Segmentation: Edge detection, Edge linking via Hough transform, Thresholding Region based segmentation, Region growing, Region splitting and merging, Morphological processing erosion and dilation, Segmentation by morphological watersheds, basic concepts, Dam construction, Watershed segmentation algorithm.
Image Compression and Recognition: Need for data compression, Huffman, Run Length Encoding, Shift codes, Arithmetic coding, JPEG standard, MPEG. Boundary representation, Boundary description, Fourier Descriptor, Regional Descriptors, Topological feature, Texture, Patterns and Pattern classes, Recognition based on matching.

Course Learning Outcomes: At the end of this course, students will be able to:

1. Know and understand the basics and fundamentals of digital image processing, such as digitization, sampling, quantization, and 2D-transforms.
2. Operate on images using the techniques of smoothing, sharpening and enhancement.
3. Understand the restoration concepts and filtering techniques.
4. Learn the basics of segmentation, features extraction, compression and recognition methods for color models.

Text Books:

1. Rafael C. Gonzalez, Richard E. Woods,Digital Image Processing Pearson, Third Edition.
2. Anil K. Jain,Fundamentals of Digital Image Processing Pearson.

References:

1. Kenneth R. Castleman, Digital Image Processing Pearson.
2. Rafael C. Gonzalez, Richard E. Woods, Steven Eddins, Digital Image Processing using MATLAB Pearson Education, Inc..
3. D,E. Dudgeon and RM. Mersereau, Multidimensional Digital Signal Processing Prentice Hall Professional Technical Reference.
4. William K. Pratt, Digital Image Processing John Wiley, New York.
5. Milan Sonka et al Image processing, analysis and machine vision Brookes/Cole, Vikas Publishing House, 2nd edition.

MSAIDS-214 Computer Network Technologies

Maximum Marks: 100 **Time Allowed: 3 Hours**
External Examination: 70 Marks **Pass Percentage: 35%**
Internal Assessment: 30 Marks **Teaching Hours per Week: 5**
Credits: 5 (4L+1T)

Instruction for the Paper Setter: The question paper will consist of three sections: A, B & C. Sections A & B will have four questions each from the respective sections of the syllabus carrying 10 marks for each question. Section C will have 10 short-answer type questions carrying 30 marks for each question which will cover the entire syllabus uniformly.
Instruction for the candidates: Candidates are required to attempt two questions each from the Sections A & B of the question paper and the entire section C.

Course Objectives: Thorough understanding of Network Technologies

SECTION-A

Data communication Techniques, Synchronous-Asynchronous Transmission, Digital Transmission, Transmission Media, Impairments, Data encoding Techniques
Communication Networks: Circuit switching, Message switching, Packet Switching. X.25, LAN Technologies, Virtual Circuits
Network Reference models – OSI and TCP/IP, Layered architecture
Data Link Layer: Design issue, framing, error control, flow control, HDLC, SDLC, data link layer in the Internet (SLIP, PPP)
Network Layer: Routing Algorithms, shortest path, distance vector routing, Link state routing, and multicast routing. Congestion control, traffic shaping, leaky bucket, token bucket, choke packets, load shedding, internetworking- connection oriented and connectionless, fragmentation, internet architecture and addressing, IP protocol, ICMP, APR, RARP, OSPF, BGP, CIDR, IPv6.

SECTION - B

Transport Layer: Transport Service, quality of service, connection management, addressing, flow control and buffering, multiplexing, Internet transport protocols- TCP and UDP
Session layer: Dialogue management, synchronization and remote procedure call.
Presentation layer: date representation, data compression, network security and cryptography
Application layer: SMTP and World Wide Web

Course Outcomes

1. Recognize the technological trends of Computer Networking.
2. Discuss the key technological components of the Network.
3. Evaluate the challenges in building networks and solutions to those.

Reference Books:

1. A. S. Tanenbaum, "Computer Networks", Pearson Education
2. W. Stallings," Data and Computer Communications", PHI
3. J.F. Kurose, K.W. Ross, "Computer Networking: A Top-Down Approach featuring the

Internet", Pearson Education
4. L.L. Peterson, B.S. Davie, “Computer Networks: A Systems Approach”, Pearson Education

MSAIDS-215 Software Lab-III
(Based on MSAIDS-211)

Maximum Marks: 100 **Time Allowed: 3 Hours**
External Examination: 70 Marks **Pass Percentage: 35%**
Internal Assessment: 30 Marks **Teaching Hours per Week: 4**
Credits: 2 (2P)

This laboratory course will mainly comprise of exercises based on subject MSAIDS-211: Data Analysis using Python. Students are required to develop programs based upon:

1. NumPy basics
2. Advanced NumPy
3. Pandas- Series, Dataframes
4. Essential functionalities of Pandas
5. Data loading, storage of various file formats
6. Data wrangling
7. Group operations
8. Data visualization using matplotlib.

The Break-up of Marks for Practical exams (External) will be as under:

1. Viva Voce 20Marks
2. Program Development and Execution 30Marks
3. File Record 20Marks

MSAIDS-216 Software Lab-IV
(Based on MSAIDS-213)

Maximum Marks: 100 **Time Allowed: 3 Hours**
External Examination: 70 Marks **Pass Percentage: 35%**
Internal Assessment: 30 Marks **Teaching Hours per Week: 4**
Credits: 2 (2P)

This laboratory course will mainly comprise of exercises based on subject MSAIDS-213: Digital Image Processing. Students are required to develop the given programs with internal documentation:

1. Program to extract different Attributes of an Image.
2. Program for Image Negation.
3. To fill the region of interest for the image.
4. Program for Histogram Mapping and Equalization.
5. Program for Image Smoothening and Sharpening.
6. Program for Edge Detection.
7. Program for Morphological Operations: erosion and dilation.
8. Program for Pseudo Coloring.
9. Program for Chain Coding.
10. Program for DCT/IDCT Computation.

The Break-up of Marks for Practical exams (External) will be as under:

1. Viva Voce 20Marks
2. Program Development and Execution 30Marks
3. File Record 20Marks

MSAIDS-221 Research Methodology

Maximum Marks: 100 **Time Allowed: 3 Hours**
External Examination: 70 Marks **Pass Percentage: 35%**
Internal Assessment: 30 Marks **Teaching Hours per Week: 5**
Credits: 5 (4L+1T)

Instruction for the Paper Setter: The question paper will consist of three sections: A, B & C. Sections A & B will have four questions each from the respective sections of the syllabus carrying 10 marks for each question. Section C will have 10 short-answer type questions carrying 30 marks for each question which will cover the entire syllabus uniformly.
Instruction for the candidates: Candidates are required to attempt two questions each from the Sections A & B of the question paper and the entire section C.

Course Objectives: The main objectives of this course are:

1. To familiarize participants with basic of research and the research process.
2. To make the students identify appropriate research topics, select and define appropriate research problem and parameters.
3. To enable the participants in conducting research work and formulating research synopsis and report.
4. To enable the students to identify and discuss the complex issues inherent in selecting a research problem, selecting an appropriate research design and implementing a research project.
5. To develop an understanding of various research designs and techniques.
6. To identify various sources of information for literature review and data collection.

SECTION-A

Objectives and types of research: Definition and types of research (Descriptive and analytical research, applied and fundamental research, qualitative and quantitative research, conceptual and empirical research).
Research problem formulation: Defining and formulating research problem and its necessity, selecting the problem, literature review and its importance; Primary and secondary data sources-library (books, journals, periodicals, reference sources, abstracting and indexing sources, reviews, monographs), patents, web (search engines, online libraries, online journals, e-books, e-encyclopedia, institutional websites); Journals and books-standards of research journals (impact factor, ISSN, ISBN, online and print journals, indexed journals, peer reviewed journals), citation index, H-index; Identifying gaps areas from literature review.
Research design and methods: Developing the research hypothesis; Research design – basic principles and need, important concepts; Observations and facts, laws and theories, prediction and explanation, induction, deduction; Development of models, developing a research plan, exploration, description, diagnosis, experimentation.
Data collection: Execution of research, observation and collection of data, methods of data collection, primary data, secondary data.
Documentation: Techniques and importance of documentation; Role of internet, information technology and computers in research and documentation.

SECTION-B

Reporting and thesis writing: Structure and components of research report, types of report-monographs, review articles, research papers, thesis, books, technical reports and their significance; Different steps in preparation of a written scientific document- layout, structure and language of reports, illustrations and tables, bibliography, references, footnotes.

Presentation of research papers: Poster presentations-layout and format; Oral presentation-planning, preparation, use of visual art, importance of effective communication.

Application of intellectual property rights: Commercialization, copyright, royalty, intellectual property rights and patent law; Plagiarism-concept and authentication of originality of research; Citation and acknowledgement; Reproducibility and accountability.

Cost analysis of project: Cost incurred on raw materials, different testing procedures, cost of instrumentation, downstream processing cost (wherever required); Cost of clinical trials.

Research grants: National/International funding agencies; Government and private bodies.

Course Learning Outcomes: At the end of this course, students will be able to:

1. Develop understanding on various kinds of research, objectives of doing research, research process, research designs and sampling.
2. Read, comprehend and explain research articles in their academic discipline.
3. Effectively present a Research paper.
4. To make the students able to identify and discuss the concepts and procedures of sampling, data collection, analysis and reporting.

Recommended Reading:

1. Statistics – An Introductory Analysis by Taro Yamane, Harper International Edition.
2. An Introduction to Statistical Methods by C.B. Gupta, Vikas Publ. Co., Jallandhar.
3. Research Methodology: Methods and Techniques by CR Kothari and Gaurav Garg by NewAge International Publishers (Third Edition), ISBN-10:8122436235, ISBN-13:978-8122436235
4. Research Methodology: A step-by-step Guide for Beginners by Ranjit Kumar, SAGE Publications, ISBN-13: 978-1849203012

MSAIDS-222 Natural Language Processing using Python

Maximum Marks: 100 **Time Allowed: 3 Hours**
External Examination: 70 Marks **Pass Percentage: 35%**
Internal Assessment: 30 Marks **Teaching Hours per Week: 5**
Credits: 5 (5L)

Instruction for the Paper Setter: The question paper will consist of three sections: A, B & C. Sections A & B will have four questions each from the respective sections of the syllabus carrying 10 marks for each question. Section C will have 10 short-answer type questions carrying 30 marks for each question which will cover the entire syllabus uniformly.
Instruction for the candidates: Candidates are required to attempt two questions each from the Sections A & B of the question paper and the entire section C.

Course Objectives: The main objectives of this course are:

1. This course introduces the fundamental concepts and techniques of natural language processing (NLP).
2. Teach students the leading trends and systems in natural language processing.
3. Make them understand the concepts of morphology, syntax, semantics and pragmatics of the language and that they are able to give the appropriate examples that will illustrate the above mentioned concepts.
4. Enable students to be capable to describe the application based on natural language processing.
5. Enable students to implement natural language processing using NLTK.

SECTION-A

Introduction and Overview: What is Natural Language Processing (NLP)? Applications of NLP, Understanding Natural Language Processing (NLP), Rule-based NLP vs Statistical NLP.
Components of Natural Language Processing (NLP): Natural Language Understanding (NLU), Natural Language Generation (NLG).
Phases of NLP: Lexical Analysis and Morphological, Syntactic Analysis (Parsing), Semantic Analysis, Discourse Integration, Pragmatic Analysis.
Basic Text Processing: Regular Expression, Word Tokenization, Sentence Segmentation.

SECTION-B

Easy to Use NLP Libraries: Various libraries for implementing NLP.
Overview of NLTK: Exploring Features of NLTK, Setting the NLTK Environment.
Extracting, Cleaning and Pre-processing Text using NLTK: Stemming, Lemmatization, Part-of-Speech (PoS) tagging, Chunking, Chinking, Named Entity Recognition (NER), Word Net, Bag of Words.

Course Learning Outcomes: At the end of this course, the students will learn various concepts such as fundamentals of NLP, its phases, Tokenization, Stemming, Lemmatization, POS tagging, Named Entity Recognition, Syntax Tree Parsing, and so on using Python's most famous NLTK package.

Text Books:

1. Speech and Language processing An introduction to Natural Language Processing, Computational Linguistics and speech Recognition by Daniel Jurafsky and James H. Martin (ISBN13: 978-131873216).
2. Natural Language Processing with Python by Steven Bird, Ewan Klein, Edward Lopper(ISBN13:978-0596516499)

Reference Book:

1. Handbook of Natural Language Processing, Second Edition—Nitin Indurkhya, Fred J. Damerau, Fred J. Damerau (ISBN13: 978-1420085921).

MSAIDS-223 Big Data Analytics

Maximum Marks: 100 **Time Allowed: 3 Hours**
External Examination: 70 Marks **Pass Percentage: 35%**
Internal Assessment: 30 Marks **Teaching Hours per Week: 5**
Credits: 5 (5L)

Instruction for the Paper Setter: The question paper will consist of three sections: A, B & C. Sections A & B will have four questions each from the respective sections of the syllabus carrying 10 marks for each question. Section C will have 10 short-answer type questions carrying 30 marks for each question which will cover the entire syllabus uniformly.
Instruction for the candidates: Candidates are required to attempt two questions each from the Sections A & B of the question paper and the entire section C.

Course Objectives: The main objectives of this course are:

1. To optimize business decisions and create competitive advantage with Big Data analytics
2. To explore the fundamental concepts of big data analytics.
3. To learn to analyze the big data using intelligent techniques.
4. To understand the various search methods and visualization techniques.
5. To learn to use various techniques for mining data stream.
6. To understand the applications using Map Reduce Concepts.
7. To introduce programming tools PIG & HIVE in Hadoop echo system.

SECTION-A

Introduction to big data : Introduction to Big Data Platform – Challenges of Conventional Systems - Intelligent data analysis – Nature of Data - Analytic Processes and Tools - Analysis vs Reporting.
Mining data streams : Introduction To Streams Concepts – Stream Data Model and Architecture - Stream Computing - Sampling Data in a Stream – Filtering Streams – Counting Distinct Elements in a Stream – Estimating Moments – Counting Oneness in a Window – Decaying Window - Real time Analytics Platform(RTAP) Applications – Case Studies - Real Time Sentiment Analysis- Stock Market Predictions.

SECTION-B

Hadoop: History of Hadoop- the Hadoop Distributed File System – Components of Hadoop Analysing the Data with Hadoop- Scaling Out- Hadoop Streaming- Design of HDFS-Java interfaces to HDFS Basics- Developing a Map Reduce Application-How Map Reduce Works-Anatomy of a Map Reduce Job run-Failures-Job Scheduling-Shuffle and Sort – Task execution - Map Reduce Types and Formats- Map Reduce Features- Hadoop environment.
Frameworks: Applications on Big Data Using Pig and Hive – Data processing operators in Pig – Hive services – HiveQL – Querying Data in Hive.

Learning Outcomes:

Students will be able to:

1. Work with big data platform and explore the big data analytics techniques business applications.

2. Design efficient algorithms for mining the data from large volumes.
3. Analyze the HADOOP and Map Reduce technologies associated with big data analytics.
4. Explore on Big Data applications Using Pig and Hive.
5. Understand the fundamentals of various big data analytics techniques.
6. Build a complete business data analytics solution

References:

1. Michael Berthold, David J. Hand, "Intelligent Data Analysis", Springer.
2. Tom White "Hadoop: The Definitive Guide" Third Edition, O'reilly Media.
3. Chris Eaton, Dirk DeRoos, Tom Deutsch, George Lapis, Paul Zikopoulos,
4. "Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data", McGrawHill Publishing.
5. Anand Rajaraman and Jeffrey David Ullman, "Mining of Massive Datasets", CUP.
6. Bill Franks, "Taming the Big Data Tidal Wave: Finding Opportunities in Huge Data Streams with Advanced Analytics", John Wiley& sons.
7. Glenn J. Myatt, "Making Sense of Data", John Wiley & Sons.
8. Pete Warden, "Big Data Glossary", O'Reilly.
9. Jiawei Han, Micheline Kamber "Data Mining Concepts and Techniques", 2nd Edition, Elsevier, Reprinted .
10. Da Ruan, Guoquing Chen, Etienne E.Kerre, Geert Wets, "Intelligent Data Mining", Springer.
11. Paul Zikopoulos, Dirkde Roos, Krishnan Parasuraman, Thomas Deutsch, James Giles , David Corrigan, "Harness the Power of Big Data The IBM Big Data Platform", Tata McGraw Hill Publications.
12. Arshdeep Bahga, Vijay Madisetti, "Big Data Science & Analytics: A Hands-On Approach ",VPT.
13. Bart Baesens "Analytics in a Big Data World: The Essential Guide to Data
14. Science and its Applications (WILEY Big Data Series)", John Wiley & Sons.

MSAIDS-224 Software Lab-V
(Based on MSAIDS-222)

Maximum Marks: 100 **Time Allowed: 3 Hours**
External Examination: 70 Marks **Pass Percentage: 35%**
Internal Assessment: 30 Marks **Teaching Hours per Week: 4**
Credits: 2 (2P)

This laboratory course will mainly comprise of exercises based on subject MSAIDS-222: Natural Language Processing using Python.

The Break-up of Marks for Practical exams (External) will be as under:

1. Viva Voce 20Marks
2. Program Development and Execution 30Marks
3. File Record 20Marks

MSAIDS-225 Software Lab-VI
(Based on MSAIDS-223)

Maximum Marks: 100 **Time Allowed: 3 Hours**
External Examination: 70 Marks **Pass Percentage: 35%**
Internal Assessment: 30 Marks **Teaching Hours per Week: 4**
Credits: 2 (2P)

This laboratory course will mainly comprise of exercises based on subject MSAIDS-223: Big Data Analytics.

The Break-up of Marks for Practical exams (External) will be as under:

1. Viva Voce 20Marks
2. Program Development and Execution 30Marks
3. File Record 20Marks

MSAIDS-226 Project

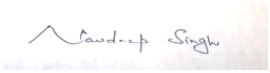
Maximum Marks: 100 **Pass Percentage: 35%**
External Examination: 70 Marks **Credits: 5**
Internal Assessment: 30 Marks

In this course, Students have to make a Project based on any of the technologies learnt so far. For this, a Two-Week In-House Industrial Training will be provided to the students. Students have to submit a Project Report to their Internal Supervisor. The marks distribution of the External Examination will be as follow:

Project Report & Presentation:	**45 Marks**
Viva-Voce:	**25 Marks**

Members of Board of Studies

1. Dr. Surender Kumar 2. Dr.Gurpreet Singh Lehal 3. Dr. Gurvinder Singh




4. Dr. Navdeep Singh 5. Mr. UpkarSingh 6. Mr. Sachin Kumar

7. Prof. Tajinder Kaur 8. Prof. Paramjit Kaur 9. Prof. Amandeep Kaur

APPROVED
Board of Studies Meeting held on 23rd June 2021