Semester: VII		
BIGDATA AND HADOOP &LAB		
Course Code: MVJ21IS71		**CIE Marks:50+50**
Credits: L:T:P: 3:0:2		**SEE Marks: 50 +50**
Hours:40 L+ 26 P		**SEE Duration: 03+03 Hours**

Course Learning Objectives: The students will be able to	
1	Understand Hadoop Distributed File System and examine MapReduce Programming
2	Explore Hadoop Tools and Manage Hadoop with Ambari
3	Appraise the role of Business Intelligence
4	Assess core data mining techniques for data analytics
5	Identify various Text Mining techniques

UNIT-I	
Introduction to big data and Hadoop Types of Digital Data, Introduction to Big Data, Big Data Analytics, History of Hadoop, Apache Hadoop, Analysing Data with Unix tools, Analysing Data with Hadoop, Hadoop Streaming, Hadoop Echo System, IBM Big Data Strategy,	**8 Hrs**
UNIT-II	
Introduction to Infosphere BigInsights and Big Sheets. HDFS(Hadoop Distributed File System)The Design of HDFS, HDFS Concepts, Command Line Interface, Hadoop file system interfaces, Data flow, Data Ingest with Flume and Scoop and Hadoop archives	**8 Hrs**
UNIT-III	
Hadoop I/O: Compression, Serialization, Avro and File-Based Data structures Map Reduce , Anatomy of a Map Reduce Job Run, Failures, Job Scheduling, Shuffle and Sort, Task Execution, Map Reduce Types and Formats, Map Reduce Features.	**8 Hrs**
UNIT-IV	
Hadoop Eco System Pig : Introduction to PIG, Execution Modes of Pig, Comparison of Pig with Databases, Grunt, Pig Latin, User DefinedFunctions, Data Processing operators. Hive : Hive Shell, Hive Services, Hive Metastore, Comparison with Traditional Databases, HiveQL, Tables, QueryingData and User Defined Functions.	**8 Hrs**
UNIT-V	
Hbase :HBasics, Concepts, Clients, Example, Hbase Versus RDBMS.Big SQL : Introduction , Data Analytics with RMachine Learning : Introduction, Supervised Learning, Unsupervised Learning, Collaborative Filtering.	**8 Hrs**

Course Outcomes: After completing the course, the students will be able to	
CO1	Master the concepts of HDFS and MapReduce framework

CO2	Investigate Hadoop related tools for Big Data Analytics and perform basic Hadoop Administration
CO3	Recognize the role of Business Intelligence, Data warehousing and Visualization in decision making
CO4	Infer the importance of core data mining techniques for data analytics
CO5	Compare and contrast different Text Mining Techniques

Reference Books	
1.	Tom White," Hadoop: The Definitive Guide",O'reily Media,Third Edition, 2012
2.	Seema Acharya, SubhasiniChellappan," Big Data Analytics",Wiley,2015

Continuous Internal Evaluation (CIE):

Theory for 50 Marks

CIE is executed by way of quizzes (Q), tests (T) and assignments. A minimum of three quizzes are conducted along with tests. Test portion is evaluated for 50 marks and quiz is evaluated for 10 marks. Faculty may adopt innovative methods for conducting quizzes effectively. The number of quizzes may be more than three (conduct additional quizzes and take best three). The three tests are conducted for 50 marks each and the average of all the tests are calculated for 50. The marks for the assignments are 20 (2 assignments for 10 marks each). The marks obtained in test, quiz and assignment are added to get marks out of 100 and report CIE for 50 marks.

Laboratory- 50 Marks

The laboratory session is held every week as per the time table and the performance of the student is evaluated in every session. The average of the marks over number of weeks is considered for 30 marks. At the end of the semester a test is conducted for 10 marks. The students are encouraged to implement additional innovative experiments in the lab and are awarded 10 marks. Total marks for the laboratory is 50.

Semester End Examination (SEE):

Total marks: 50+50=100

SEE for 50 marks are executed by means of an examination.

The Question paper for each course contains two parts, Part – A and Part – B. Part – A consists of objective type questions for 20 marks covering the entire syllabus. Part – B Students have to answer five questions, one from each unit for 16 marks adding up to 80 marks. Each main question may have a maximum of three sub divisions. Each unit will have internal choice in which both questions cover entire unit having same complexity in terms of COs and Bloom's taxonomy level.

CO-PO Mapping												
CO/PO	**PO1**	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PO8**	**PO9**	**PO10**	**PO11**	**PO12**
CO1	3	3	0	3	0	0	0	0	0	0	1	1
CO2	2	3	0	3	0	0	0	0	0	0	1	1
CO3	2	3	0	3	0	0	0	0	0	0	1	1
CO4	3	3	0	3	0	0	0	0	0	0	1	1
CO5	3	3	0	3	0	0	0	0	0	0	1	1

High-3, Medium-2, Low-1

Semester: IV		
BLOCK CHAIN TECHNOLOGY		
Course Code: MVJ21IS721		**CIE Marks:100**
Credits: L:T:P:S: 3:0:0		**SEE Marks: 100**
Hours: 40L+26T		**SEE Duration: 3 Hrs**
Course Learning Objectives: The students will be able to		
1	Understand how blockchain systems (mainly Bitcoin and Ethereum) work	
2	To securely interact with them,	
3	Design, build, and deploy smart contracts and distributed applications	

UNIT-I	
Distributed Database, Two General Problem, Byzantine General problem and Fault Tolerance, Hadoop Distributed File System, Distributed Hash Table, ASIC resistance, Turing Complete. Cryptography: Hash function, Digital Signature - ECDSA, Memory Hard Algorithm, Zero Knowledge Proof.	**8 Hrs**
UNIT-II	
Introduction, Advantage over conventional distributed database, Blockchain Network, Mining Mechanism, Distributed Consensus, Merkle Patricia Tree, Gas Limit, Transactions and Fee, Anonymity, Reward, Chain Policy, Life of Blockchain application, Soft & Hard Fork, Private and Public blockchain.	**8 Hrs**
UNIT-III	
Nakamoto consensus, Proof of Work, Proof of Stake, Proof of Burn, Difficulty Level, Sybil Attack, Energy utilization and alternate., Introducing modeling language for business resources and transactions, Introduction to key concepts related to smart contracts, accounts, transaction events, patterns and examples	**8 Hrs**
UNIT-IV	
History, Distributed Ledger, Bitcoin protocols - Mining strategy and rewards, Ethereum - Construction, DAO, Smart Contract, GHOST, Vulnerability, Attacks, Sidechain, Namecoin	**8 Hrs**
UNIT-V	
Stakeholders, Roots of Bit coin, Legal Aspects-Crypto currency Exchange, Black Market and Global Economy. Applications: Internet of Things, Medical Record	**8 Hrs**

Management System, Domain Name Service and future of Blockchain., Overview of how IoT can benefit from Blockchain implementation	

Course Outcomes: After completing the course, the students will be able to	
CO1	Learn design principles of Bitcoin and Ethereum and Nakamoto consensus.
CO2	. Explain the Simplified Payment Verification protocol.
CO3	Interact with a blockchain system by sending and reading transactions.
CO4	Design, build, and deploy a distributed application.
CO5	Evaluate security, privacy, and efficiency of a given blockchain system.

Reference Books	
1.	"Bitcoin and Cryptocurrency Technologies: A Comprehensive Introduction,",Arvind Narayanan, Joseph Bonneau, Edward Felten, Andrew Miller and Steven Goldfeder,Princeton University Press,July 19, 2016
2.	"Mastering Bitcoin: Unlocking Digital Cryptocurrencies" ,Antonopoulos

Continuous Internal Evaluation (CIE):

Theory for 50 Marks

CIE is executed by way of quizzes (Q), tests (T) and assignments. A minimum of three quizzes are conducted along with tests. Test portion is evaluated for 50 marks and quiz is evaluated for 10 marks. Faculty may adopt innovative methods for conducting quizzes effectively. The number of quizzes may be more than three (conduct additional quizzes and take best three). The three tests are conducted for 50 marks each and the average of all the tests are calculated for 50. The marks for the assignments are 20 (2 assignments for 10 marks each). The marks obtained in test, quiz and assignment are added to get marks out of 100 and report CIE for 50 marks.

Semester End Examination (SEE):

Total marks: 50+50=100

SEE for 50 marks is executed by means of an examination. The Question paper for each course contains two parts, Part – A and Part – B. Part – A consists of objective type questions for 20 marks covering the entire syllabus. Part – B Students have to answer five questions, one from each unit for 16 marks adding up to 80 marks. Each main question may have a maximum of three sub divisions. Each unit will have internal choice in which both questions cover entire unit having same complexity in terms of COs and Bloom's taxonomy level.

CO-PO Mapping

CO/PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	3	3	0	2	0	0	0	0	0	0	1	1
CO2	3	3	0	2	0	0	0	0	0	0	1	1
CO3	3	3	0	3	0	0	0	0	0	0	0	1
CO4	2	2	0	3	0	0	0	0	0	0	1	1
CO5	2	2	0	2	0	0	0	0	0	0	0	1

High-3, Medium-2, Low-1

Semester: VII		
DEEP LEARNING		
Course Code: MVJ21IS722		**CIE Marks:100**
Credits: L:T:P:S: 3:0:0		**SEE Marks: 100**
Hours: 40L+26T		**SEE Duration: 3 Hrs**
Course Learning Objectives: The students will be able to		
1	Explain the fundamentals of Deep Learning	
2	Familiarize with Tensor Flow, Installation of software module	
3	Design and build support vector machine	

UNIT-I	
Feedforward Neural networks. Gradient descent and the backpropagation algorithm. Unit saturation, aka the vanishing gradient problem, and ways to mitigate it. RelU Heuristics for avoiding bad local minima. Heuristics for faster training. Nestors accelerated gradient descent. Regularization. Dropout. Convolutional Neural Networks Architectures, convolution / pooling layers	**8 Hrs**
UNIT-II	
Recurrent Neural Networks , LSTM, GRU, Encoder Decoder architectures, Deep Unsupervised Learning , Autoencoders (standard, sparse, denoising, contractive, etc), Variational Autoencoders, Adversarial Generative Networks, Autoencoder and DBM	**8 Hrs**
UNIT-III	
Applications of Deep Learning to Computer Vision , Image segmentation, object detection, automatic image captioning, Image generation with Generative adversarial networks, video to text with LSTM models. Attention models for computer vision tasks. Applications of Deep Learning to NLP: Introduction to NLP and Vector Space Model of Semantics	**8 Hrs**
UNIT-IV	
Word Vector Representations: Continuous Skip-Gram Model, Continuous Bag-ofWords model (CBOW), Glove, Evaluations and Applications in word similarity, analogy reasoning Named Entity Recognition, Opinion Mining using Recurrent	**8 Hrs**

Neural Networks , Parsing and Sentiment Analysis using Recursive Neural Networks ,	
UNIT-V	
Sentence Classification using Convolutional Neural Networks , Dialogue Generation with LSTMs , Applications of Dynamic Memory Networks in NLP , Recent Research in NLP using Deep Learning: Factoid Question Asnwering, similar question detection, Dialogue topic tracking, Neural Summarization, Smart Reply	**8 Hrs**

Course Outcomes: After completing the course, the students will be able to	
CO1	Basics of Deep Learning
CO2	Understand TensorFlow and Reinforcement Learning
CO3	Explain RNN and Unsupervised Feature Learning
CO4	Explain RNN and Unsupervised Feature Learning
CO5	Explain Architecture of CNNs

Reference Books	
1.	Bengio, Yoshua, Ian J. Goodfellow, and Aaron Courville,"Deep learning",An MIT Press book in preparation,2015
2.	Bengio, Yoshua," Learning deep architectures for AI " . Foundations and trends in Machine Learning 2.1,2009: 1127

Continuous Internal Evaluation (CIE):

Theory for 50 Marks

CIE is executed by way of quizzes (Q), tests (T) and assignments. A minimum of three quizzes are conducted along with tests. Test portion is evaluated for 50 marks and quiz is evaluated for 10 marks. Faculty may adopt innovative methods for conducting quizzes effectively. The number of quizzes may be more than three (conduct additional quizzes and take best three). The three tests are conducted for 50 marks each and the average of all the tests are calculated for 50. The marks for the assignments are 20 (2 assignments for 10 marks each). The marks obtained in test, quiz and assignment are added to get marks out of 100 and report CIE for 50 marks.

Semester End Examination (SEE):

Total marks: 50+50=100

SEE for 50 marks is executed by means of an examination. The Question paper for each course contains two parts, Part – A and Part – B. Part – A consists of objective type questions for 20 marks covering the entire syllabus. Part – B Students have to answer five questions, one from each unit for 16 marks adding up to 80 marks. Each main question may have a maximum of three

sub divisions. Each unit will have internal choice in which both questions cover entire unit having same complexity in terms of COs and Bloom's taxonomy level.

CO-PO Mapping												
CO/PO	**PO1**	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PO8**	**PO9**	**PO10**	**PO11**	**PO12**
CO1	3	3	0	3	0	0	0	0	0	0	1	1
CO2	2	3	0	3	0	0	0	0	0	0	1	1
CO3	2	3	0	3	0	0	0	0	0	0	1	1
CO4	3	3	0	3	0	0	0	0	0	0	1	1
CO5	3	3	0	3	0	0	0	0	0	0	1	1

Semester: VII		
FILE STRUCTURES		
Course Code: MVJ21IS723		**CIE Marks:100**
Credits: L:T:P:S: 3:0:0		**SEE Marks: 100**
Hours: 40L+26T		**SEE Duration: 3 Hrs**
Course Learning Objectives: The students will be able to		
1	Explain the fundamentals of file structures and their management.	
2	Measure the performance of different file structures	
3	Organize different file structures in the memory.	
4	Demonstrate hashing and indexing techniques.	

UNIT-I	
Introduction: File Structures: The Heart of the file structure Design, A Short History of File Structure Design, A Conceptual Toolkit; Fundamental File Operations: Physical Files and Logical Files, Opening Files, Closing Files, Reading and Writing, Seeking, Special Characters, The Unix Directory Structure, Physical devices and Logical Files, File-related Header Files, UNIX file System Commands; Secondary Storage and System Software: Disks, Magnetic Tape, Disk versus Tape; CD-ROM: Introduction, Physical Organization, Strengths and Weaknesses; Storage as Hierarchy, A journey of a Byte, Buffer Management, Input /Output in UNIX. Fundamental File Structure Concepts, Managing Files of Records : Field and Record Organization, Using Classes to Manipulate Buffers, Using Inheritance for Record Buffer Classes, Managing Fixed Length, Fixed Field Buffers, An Object-Oriented Class for Record Files, Record Access, More about Record Structures, Encapsulating Record Operations in a Single Class, File Access and File Organization.	**8 Hrs**
UNIT-II	
Organization of Files for Performance, Indexing: Data Compression, Reclaiming Space in files, Internal Sorting and Binary Searching, Keysorting; What is an Index? A Simple Index for Entry-Sequenced File, Using Template Classes in C++ for Object	**8 Hrs**

I/O, Object-Oriented support for Indexed, Entry-Sequenced Files of Data Objects, Indexes that are too large to hold in Memory, Indexing to provide access by Multiple keys, Retrieval Using Combinations of Secondary Keys, Improving the Secondary Index structure: Inverted Lists, Selective indexes, Binding.	
UNIT-III	
Consequential Processing and the Sorting of Large Files: A Model for Implementing Cosequential Processes, Application of the Model to a General Ledger Program, Extension of the Model to include Mutiway Merging, A Second Look at Sorting in Memory, Merging as a Way of Sorting Large Files on Disk. Multi-Level Indexing and B-Trees: The invention of B-Tree, Statement of the problem, Indexing with Binary Search Trees; Multi-Level Indexing, B-Trees, Example of Creating a B-Tree, An Object-Oriented Representation of B-Trees, B-Tree Methods; Nomenclature, Formal Definition of B-Tree Properties, Worst-case Search Depth, Deletion, Merging and Redistribution, Redistribution during insertion; B* Trees, Buffering of pages; Virtual BTrees; Variable-length Records and keys.	**8 Hrs**
UNIT-IV	
Indexed Sequential File Access and Prefix B + Trees: Indexed Sequential Access, Maintaining a Sequence Set, Adding a Simple Index to the Sequence Set, The Content of the Index: Separators Instead of Keys, The Simple Prefix B+ Tree and its maintenance, Index Set Block Size, Internal Structure of Index Set Blocks: A Variable-order B- Tree, Loading a Simple Prefix B+ Trees, B-Trees, B+ Trees and Simple Prefix B+ Trees in Perspective	**8 Hrs**
UNIT-V	
Hashing: Introduction, A Simple Hashing Algorithm, Hashing Functions and Record Distribution, How much Extra Memory should be used?, Collision resolution by progressive overflow, Buckets, Making deletions, Other collision resolution techniques, Patterns of record access. Extendible Hashing: How Extendible Hashing Works, Implementation, Deletion, Extendible Hashing Performance, Alternative Approaches.	**8 Hrs**

Course Outcomes: After completing the course, the students will be able to	
CO1	Choose appropriate file structure for storage representation.
CO2	Identify a suitable sorting technique to arrange the data
CO3	Select suitable indexing for better performance to a given problem.
CO4	Select suitable hashing techniques for better performance to a given problem.
CO5	Identify the sorting techniques

Reference Books	
1.	Michael J. Folk, Bill Zoellick, Greg Riccardi: File Structures-An Object Oriented Approach with C++, 3rd Edition, Pearson Education, 1998.
2.	K.R. Venugopal, K.G. Srinivas, P.M. Krishnaraj: File Structures Using C++, Tata McGraw-Hill, 2008.

3.	Scot Robert Ladd: C++ Components and Algorithms, BPB Publications, 1993

Continuous Internal Evaluation (CIE):

Theory for 50 Marks

CIE is executed by way of quizzes (Q), tests (T) and assignments. A minimum of three quizzes are conducted along with tests. Test portion is evaluated for 50 marks and quiz is evaluated for 10 marks. Faculty may adopt innovative methods for conducting quizzes effectively. The number of quizzes may be more than three (conduct additional quizzes and take best three). The three tests are conducted for 50 marks each and the average of all the tests are calculated for 50. The marks for the assignments are 20 (2 assignments for 10 marks each). The marks obtained in test, quiz and assignment are added to get marks out of 100 and report CIE for 50 marks.

Semester End Examination (SEE):

Total marks: 50+50=100

SEE for 50 marks is executed by means of an examination. The Question paper for each course contains two parts, Part – A and Part – B. Part – A consists of objective type questions for 20 marks covering the entire syllabus. Part – B Students have to answer five questions, one from each unit for 16 marks adding up to 80 marks. Each main question may have a maximum of three subdivisions. Each unit will have internal choice in which both questions cover entire unit having same complexity in terms of COs and Bloom's taxonomy level.

CO-PO Mapping												
CO/PO	**PO1**	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PO8**	**PO9**	**PO10**	**PO11**	**PO12**
CO1	3	3	0	3	0	0	0	0	0	0	1	1
CO2	2	3	0	3	0	0	0	0	0	0	1	1
CO3	2	3	0	3	0	0	0	0	0	0	1	1
CO4	3	3	0	3	0	0	0	0	0	0	1	1
CO5	3	3	0	3	0	0	0	0	0	0	1	1

Semester: VII		
INTERNET OF THINGS		
Course Code: MVJ21IS731		**CIE Marks:100**
Credits: L:T:P:S: 3:0:0		**SEE Marks: 100**
Hours: 40L+26T		**SEE Duration: 3 Hrs**
Course Learning Objectives: The students will be able to		
1	Assess the genesis and impact of IoT applications, architectures in real world	
2	Illustrate diverse methods of deploying smart objects and connect them to network.	
3	Compare different Application protocols for IoT.	

UNIT-I	
What is IoT, Genesis of IoT, IoT and Digitization, IoT Impact, Convergence of IT and IoT, IoT Challenges, IoT Network Architecture and Design, Drivers Behind New Network Architectures, Comparing IoT Architectures, A Simplified IoT Architecture, The Core IoT Functional Stack, IoT Data Management and Compute Stack.	**8 Hrs**
UNIT-II	
Smart Objects: The "Things" in IoT, Sensors, Actuators, and Smart Objects, Sensor Networks, Connecting Smart Objects, Communications Criteria, IoT Access Technologies, IP as the IoT Network Layer, The Business Case for IP, The need for Optimization, Optimizing IP for IoT, Profiles and Compliances	**8 Hrs**
UNIT-III	
Application Protocols for IoT, The Transport Layer, IoT Application Transport Methods, Data and Analytics for IoT, An Introduction to Data Analytics for IoT, Machine Learning, Big Data Analytics Tools and Technology, Edge Streaming Analytics, Network Analytics, Securing IoT, A Brief History of IOT Security,	**8 Hrs**
UNIT-IV	
Common Challenges in OT Security, How IT and OT Security Practices and Systems Vary, Formal Risk Analysis Structures: OCTAVE and FAIR, The Phased Application of Security in an Operational Environment, IoT Physical Devices and Endpoints - Arduino UNO: Introduction to Arduino, Arduino UNO, Installing the Software, Fundamentals of Arduino Programming. IoT Physical Devices and Endpoints	**8 Hrs**
UNIT-V	
RaspberryPi: Introduction to RaspberryPi, About the RaspberryPi Board: Hardware Layout, Operating Systems on RaspberryPi, Configuring RaspberryPi, Programming RaspberryPi with Python, Wireless Temperature Monitoring System Using Pi, DS18B20 Temperature Sensor, Connecting Raspberry Pi via SSH, Accessing Temperature from DS18B20 sensors, Remote access to RaspberryPi, Smart and Connected Cities, An IoT Strategy for Smarter Cities, Smart City IoT Architecture, Smart City Security Architecture, 08 Smart City Use-Case Examples.	**8 Hrs**

Course Outcomes: After completing the course, the students will be able to	
CO1	Describe the characteristics and key technologies for IoT system
CO2	. Interfacing Sensor and Actuator with Arduino development board.
CO3	Implementing IoT device by interfacing communication module and cloud
CO4	Describe protocols of resource constraint network
CO5	Elaborate the need for Data Analytics and Security in IoT

Reference Books	
1.	"IoTFundamentals: Networking Technologies, Protocols, and Use Cases for the Internet

	ofThings", David Hanes, Gonzalo Salgueiro, Patrick Grossetete, Robert Barton, Jerome Henry, 1stEdition, Pearson Education (Cisco Press Indian Reprint). (ISBN: 978-9386873743
2.	"Internet of Things",Srinivasa K G,CENGAGE Leaning India,2017

Continuous Internal Evaluation (CIE):

Theory for 50 Marks

CIE is executed by way of quizzes (Q), tests (T) and assignments. A minimum of three quizzes are conducted along with tests. Test portion is evaluated for 50 marks and quiz is evaluated for 10 marks. Faculty may adopt innovative methods for conducting quizzes effectively. The number of quizzes may be more than three (conduct additional quizzes and take best three). The three tests are conducted for 50 marks each and the average of all the tests are calculated for 50. The marks for the assignments are 20 (2 assignments for 10 marks each). The marks obtained in test, quiz and assignment are added to get marks out of 100 and report CIE for 50 marks.

Semester End Examination (SEE):

Total marks: 50+50=100

SEE for 50 marks is executed by means of an examination. The Question paper for each course contains two parts, Part – A and Part – B. Part – A consists of objective type questions for 20 marks covering the entire syllabus. Part – B Students have to answer five questions, one from each unit for 16 marks adding up to 80 marks. Each main question may have a maximum of three sub divisions. Each unit will have internal choice in which both questions cover entire unit having same complexity in terms of COs and Bloom's taxonomy level.

CO-PO Mapping												
CO/PO	**PO1**	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PO8**	**PO9**	**PO10**	**PO11**	**PO12**
CO1	3	3	0	3	0	0	0	0	0	0	1	1
CO2	2	3	0	3	0	0	0	0	0	0	1	1
CO3	2	3	0	3	0	0	0	0	0	0	1	1
CO4	3	3	0	3	0	0	0	0	0	0	1	1

High-3, Medium-2, Low-1

Semester: VII		
NATURAL LANGUAGE PROCESSING		
Course Code: MVJ21IS732		**CIE Marks:100**
Credits: L:T:P:S: 3:0:0		**SEE Marks: 100**
Hours: 40L+26T		**SEE Duration: 3 Hrs**
Course Learning Objectives: The students will be able to		

1	Expose students to the concepts of n-grams and Language Modelling with n-gram
2	Expose students to the Natural Language Processing pipeline
3	Expose students to the Information Extraction problems and end to end Natural Language Generation problems as applications of Natural Language Processing

UNIT-I	
Text Normalization, Morphology and Finite State Transducer: Concept/ Types of Ambiguity in Natural Language Processing, Empirical Laws: Zipf's Law, Heap's Law. Text Normalization: Content and Function Words, Type vs. Token, Unix Tools for Crude Tokenization and Normalization, Word Tokenization and Normalization, Lemmatization and Stemming, Sentence Segmentation. Morphology and Finite State Transducers: Survey of English Morphology, Finite State Morphological Parsing, Combining FST Lexicon and Rules, Lexicon - Free FST - The Porter Stemmer, Human Morphological Parsing	**8 Hrs**
UNIT-II	
N-Grams, Edit Distance and Language Modelling: n-grams, Evaluating Language Models - Perplexity, Generalization and Zeros, Smoothing - Kneser-Ney Smoothing, Web and Stupid Back Off, Perplexity's Relation to Entropy. Spelling Correction and Noisy Channel: Noisy Channel Model, Real World Spelling Error, Minimum Edit Distance Algorithm, Improved Edit Models. Word Classes and Part-of-Speech (POS) Tagging: English Word Classes, Penn Tagsets for English, Rule-Based Part-of-Speech Tagging, Transformation-Based Tagging, POS Tagging using Hidden Markov Model, Maximum Entropy Model and Conditional Random Fields, Neural Language Models with Deep Artificial Neural Network	**8 Hrs**
UNIT-III	
Parsing: Context Free Grammar. Syntactic Parsing: Ambiguity Presented By Parse Trees, CKY Parsing, Chart Parsing and Earley Parser. Partial Parsing: Chunking. Statistical Parsing: Probabilistic Context Free Grammar, Probabilistic CKY Parsing of PCFG, Problems with PCFG, Probabilistic Lexicalized PCFG. Introduction to Dependency Parsing: Dependency Relations, Dependency Formalisms, Dependency Tree Banks, Evaluating Parsers.	**8 Hrs**
UNIT-IV	
Semantics - Lexical semantics: Word Senses and Relations Between Word Senses, WordNet: A Database of Lexical Relations, Word Sense Disambiguation - Overview, Supervised Word Sense Disambiguation, WSD - Dictionary and Thesaurus Methods, Semi- Supervised WSD, Unsupervised Word Sense Induction. Word Similarity or Semantic Relatedness Based On Thesaurus: Resnik Similarity, Lin Similarity, Jiang-Conrath Distance, Extended Gloss Overlap And Extended Lesk Method. Lexicons For Sentiment and Affect Extraction: Available Sentiment Lexicons, Using Wordnet Synonyms And Antonyms - Sentiwordnet, Supervised	**8 Hrs**

Learning of Word Sentiments, Using Lexicon For Sentiment Recognition, Lexicons For Emotions And Other Affective States.	
UNIT-V	
Information Retrieval, Natural Language Generation and Neural Network Methods for Natural Language Processing - Information retrieval: Information Extraction vs. Retrieval, Information Extraction Sub-Problems, Named Entity Recognition - Practical NER Architectures. Natural Language Generation: An Architecture, Question Answering System - IR Based Factoid Question Answering, Knowledge Based Question Answering, IBM's Watson, Dialogue System And Chatbot - Rule Based And Corpus Based Chatbots.	**8 Hrs**

Course Outcomes: After completing the course, the students will be able to	
CO1	Implement meaningful course or research projects using current Natural Language Processing technology
CO2	Analyze the natural language text.
CO3	Define the importance of natural language.
CO4	Understand the concepts Text mining.
CO5	Illustrate information retrieval techniques

Reference Books	
1.	Daniel Jurafsky and James H Martin, "Speech and Natural Language Processing"http://web.stanford. edu/~jurafsky/slp3/,3rd Edition Draft
2.	Yoav Goldberg "Neural Network Methods for Natural Language Processing",Morgan and Claypool Publishers

Continuous Internal Evaluation (CIE):

Theory for 50 Marks

CIE is executed by way of quizzes (Q), tests (T) and assignments. A minimum of three quizzes are conducted along with tests. Test portion is evaluated for 50 marks and quiz is evaluated for 10 marks. Faculty may adopt innovative methods for conducting quizzes effectively. The number of quizzes may be more than three (conduct additional quizzes and take best three). The three tests are conducted for 50 marks each and the average of all the tests are calculated for 50. The marks for the assignments are 20 (2 assignments for 10 marks each). The marks obtained in test, quiz and assignment are added to get marks out of 100 and report CIE for 50 marks.

Semester End Examination (SEE):

Total marks: 50+50=100

SEE for 50 marks is executed by means of an examination. The Question paper for each course contains two parts, Part – A and Part – B. Part – A consists of objective type questions for 20 marks covering the entire syllabus. Part – B Students have to answer five questions, one from each unit for 16 marks adding up to 80 marks. Each main question may have a maximum of three sub divisions. Each unit will have internal choice in which both questions cover entire unit having same complexity in terms of COs and Bloom's taxonomy level.

CO-PO Mapping												
CO/PO	**PO1**	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PO8**	**PO9**	**PO10**	**PO11**	**PO12**
CO1	3	3	0	2	0	0	0	0	0	0	1	1
CO2	3	3	0	2	0	0	0	0	0	0	1	1
CO3	3	3	0	3	0	0	0	0	0	0	0	1
CO4	2	2	0	3	0	0	0	0	0	0	1	1
CO5	2	2	0	2	0	0	0	0	0	0	0	1

High-3, Medium-2, Low-1

Semester: VII		
DATA SECURITY AND PRIVACY		
Course Code: MVJ21IS733		**CIE Marks:100**
Credits: L:T:P:S: 3:0:0		**SEE Marks: 100**
Hours: 40L+26T		**SEE Duration: 3 Hrs**
Course Learning Objectives: The students will be able to		
1	Identify standard algorithms used to provide confidentiality, integrity and authenticity for data	
2	Distinguish key distribution and management schemes.	
3	Deploy encryption techniques to secure data in transit across data networks	
4	Implement security applications in the field of Information technology	
5	Demonstrate data privacy	

UNIT-I	
Symmetric Cipher Model, Cryptography, Cryptanalysis and Brute-Force Attack, Substitution Techniques, Caesar Cipher, Mono-alphabetic Cipher, Playfair Cipher, Hill Cipher, Poly alphabetic Cipher, One Time Pad. Block Ciphers and the data encryption standard: Traditional block Cipher structure, stream Ciphers and block Ciphers, Motivation for the feistel Cipher structure, the feistel Cipher, The data encryption standard, DES encryption, DES decryption, A DES example, results, the avalanche effect, the strength of DES, the use of 56-Bit Keys, the nature of the DES	**8 Hrs**

algorithm, timing attacks, Block cipher design principles, number of rounds, design of function F, key schedule algorithm.	
UNIT-II	
Principles of public-key cryptosystems. Public-key cryptosystems. Applications for public-key cryptosystems, requirements for public-key cryptosystems. Public-key cryptanalysis. The RSA algorithm, description of the algorithm, computational aspects, the security of RSA. Other Public-Key Cryptosystems: Diffiehellman key exchange, The algorithm, key exchange protocols, man in the middle attack, Elgamal Cryptographic systems, Elliptic curve arithmetic, abelian groups, elliptic curves 27.09.2022 over real numbers, elliptic curves over Zp, elliptic curves overGF(2m), Elliptic curve cryptography, Analog of Diffie-hellman key exchange, Elliptic curve encryption/ decryption, security of Elliptic curve cryptography, Pseudorandom number generation based on a asymmetric cipher	**8 Hrs**
UNIT-III	
Symmetric key distribution using Symmetric encryption, A key distribution scenario, Hierarchical key control, session key lifetime, a transparent key control scheme, Decentralized key control, controlling key usage, Symmetric key distribution using asymmetric encryption, simple secret key distribution, secret key distribution with confidentiality and authentication, A hybrid scheme, distribution of public keys, public announcement of public keys, publicly available directory, public key authority, public keys certificates, X509 certificates. Certificates, X-509 version 3, Public Key infrastructure	**8 Hrs**
UNIT-IV	
Privacy-Preserving Data Mining Algorithms, The Randomization Method, Group Based Anonymization.	**8 Hrs**
UNIT-V	
Distributed Privacy-Preserving Data Mining, Privacy-Preservation of Application Results, Limitations of Privacy: The Curse of Dimensionality, Applications of Privacy-Preserving Data Mining	**8 Hrs**

Course Outcomes: After completing the course, the students will be able to	
CO1	Identify the vulnerabilities in any computing system and hence to choose security solution.
CO2	Plan to resolve the identified security issues.
CO3	Analyse security mechanisms using theoretical approache
CO4	Recognize the importance of data privacy, limitations and applications
CO5	Organize the privacy preserving algorithms

Reference Books	
1.	Cryptography and Network Security, William Stallings

2.	., Pearson 7th edition. 4. Privacy Preserving Data Mining: Models and Algorithms, Charu C. Aggarwal, Philip S Yu, Kluwer Academic Publishers, 2008, ISBN 978-0-387-70991-8, DOI 10.1007/978- 0-387-70992-5
3.	Cryptography and Network Security, Atul Kahate, McGraw Hill Education, 4th Edition.
4.	Cryptography and Information Security, V K Pachghare, 2nd edition, PHI.

Continuous Internal Evaluation (CIE):

Theory for 50 Marks

CIE is executed by way of quizzes (Q), tests (T) and assignments. A minimum of three quizzes are conducted along with tests. Test portion is evaluated for 50 marks and quiz is evaluated for 10 marks. Faculty may adopt innovative methods for conducting quizzes effectively. The number of quizzes may be more than three (conduct additional quizzes and take best three). The three tests are conducted for 50 marks each and the average of all the tests are calculated for 50. The marks for the assignments are 20 (2 assignments for 10 marks each). The marks obtained in test, quiz and assignment are added to get marks out of 100 and report CIE for 50 marks.

Semester End Examination (SEE):

Total marks: 50+50=100

SEE for 50 marks is executed by means of an examination. The Question paper for each course contains two parts, Part – A and Part – B. Part – A consists of objective type questions for 20 marks covering the entire syllabus. Part – B Students have to answer five questions, one from each unit for 16 marks adding up to 80 marks. Each main question may have a maximum of three sub divisions. Each unit will have internal choice in which both questions cover entire unit having same complexity in terms of COs and Bloom's taxonomy level.

CO-PO Mapping												
CO/PO	**PO1**	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PO8**	**PO9**	**PO10**	**PO11**	**PO12**
CO1	3	3	0	2	0	0	0	0	0	0	1	1
CO2	3	3	0	2	0	0	0	0	0	0	1	1
CO3	3	3	0	3	0	0	0	0	0	0	0	1
CO4	2	2	0	3	0	0	0	0	0	0	1	1
CO5	2	2	0	2	0	0	0	0	0	0	0	1

High-3, Medium-2, Low-1

Semester: VII		
CLOUD COMPUTING		
Course Code: MVJ21IS741		**CIE Marks:100**
Credits: L:T:P:S: 3:0:0		**SEE Marks: 100**
Hours: 40L+26T		**SEE Duration: 3 Hrs**
Course Learning Objectives: The students will be able to		
1	To provide students with the fundamentals and essentials of Cloud Computing.	
2	To provide students a sound foundation of the Cloud Computing so that they are able to start using and adopting Cloud Computing services and tools in their real life scenarios.	
3	To enable students exploring some important cloud computing driven commercial systems and applications	

UNIT-I	
Introduction to Networking, Data communication, Cloud Computing, Origin of Cloud Computing, Basic Concepts and Terminology. Goals and Benefits, Risks and Challenges, Roles and Boundaries, Cloud Characteristics .Cloud Delivery Models, Cloud Deployment Models	**8 Hrs**
UNIT-II	
Broadband Networks and Internet Architecture, Data Center Technology, Virtualization Technology.Web Technology, Multitenant Technology, Service Technology .Applications, Cloud computing for Healthcare, Energy Systems, Transportation Systems, Manufacturing Industry	**8 Hrs**
UNIT-III	
Cloud Infrastructure Mechanisms: Logical Network Perimeter, Virtual Server: Cloud Storage Device, Cloud Usage Monitor, Resource Replication, Ready-Made Environment .Specialized Cloud Mechanisms: Automated Scaling Listener, Load Balancer, SLA Monitor, Pay Per Use Monitor: Audit Monitor, Failover System, Hypervisor, Resource Cluster, Multi:Device Broker	**8 Hrs**
UNIT-IV	
Cloud Management Mechanisms: Remote Administration System, Resource Management System, SLA Management System, Billing Management System .Cost Metrics and Pricing Models: Business Cost Metrics, Cloud Usage Cost Metrics, Cost Management Considerations . Service Quality Metrics and SLAs: Service Quality Metrics, SLA Guidelines.	**8 Hrs**
UNIT-V	
Fundamental Cloud Architectures: Illustration with Case Study Fundamental Cloud Security: Basic Terms and Concepts, Threat Agents, Cloud Security Threats .Cloud Security Mechanisms: Encryption, Hashing: Digital Signature, Public Key Infrastructure, Identity and Access Management	**8 Hrs**

Course Outcomes: After completing the course, the students will be able to	
CO1	Use the concepts of classes and objects in Object Oriented Programming. Use UML to model a complex system by defining actors and use cases.
CO2	Construct Class Models and analyze the dynamics of a system using Activity, Sequence, State and Process models.
CO3	Depict the architecture of a software system by using component and deployment models and design a database based on a class model.
CO4	Use GRASP and SOLID principles in the design of software.
CO5	Apply software design patterns in a variety of situations.

Reference Books	
1.	Thomas Erl, ZaighamMahmood,RichardoPuttini, "Cloud Computing:Concepts", Prentice Hall/PearsonPTR, ISBN: 9780133387520,Fourth Printing, 2014
2.	ArshdeepBahga, Vijay Madisetti:"Cloud Computing: A Hands-On Approach", University Press, ISBN: 9780996025508,2016
3.	K.Chandrasekaran ,"Essentials of Cloud Computing ",Chapman and Hall/CRC Press, ISBN 9781482205435,2014
4.	Thomas Erl, Robert Cope, Amin Naserpour,Cloud Computing Design Patterns, Prentice Hall/Service Tech Press, Pearson, ISBN: 978-0133858563,2015

Continuous Internal Evaluation (CIE):

Theory for 50 Marks

CIE is executed by way of quizzes (Q), tests (T) and assignments. A minimum of three quizzes are conducted along with tests. Test portion is evaluated for 50 marks and quiz is evaluated for 10 marks. Faculty may adopt innovative methods for conducting quizzes effectively. The number of quizzes may be more than three (conduct additional quizzes and take best three). The three tests are conducted for 50 marks each and the average of all the tests are calculated for 50. The marks for the assignments are 20 (2 assignments for 10 marks each). The marks obtained in test, quiz and assignment are added to get marks out of 100 and report CIE for 50 marks.

Semester End Examination (SEE):

Total marks: 50+50=100

SEE for 50 marks is executed by means of an examination. The Question paper for each course contains two parts, Part – A and Part – B. Part – A consists of objective type questions for 20 marks covering the entire syllabus. Part – B Students have to answer five questions, one from each unit for 16 marks adding up to 80 marks. Each main question may have a maximum of three

sub divisions. Each unit will have internal choice in which both questions cover entire unit having same complexity in terms of COs and Bloom's taxonomy level.

CO-PO Mapping												
CO/PO	**PO1**	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PO8**	**PO9**	**PO10**	**PO11**	**PO12**
CO1	3	3	3	3	3		2					
CO2	3	3	3	3	3		2					
CO3	3	3	3	3	3	3						3
CO4	3	3	3	3	3	3		3				3
CO5	3	3	3	3	3	3						3

High-3, Medium-2, Low-1

Semester: VII		
INTRODUCTION TO AI		
Course Code: MVJ21IS742		**CIE Marks:100**
Credits: L:T:P:S: 3:0:0		**SEE Marks: 100**
Hours: 40L+26T		**SEE Duration: 3 Hrs**
Course Learning Objectives: The students will be able to		
1	Identify the problems where AI is required and the different methods available.	
2	Compare and contrast different AI techniques available.	
3	Define and explain learning algorithms.	
4	Design different learning algorithms for improving the performance of AI systems.	
5	Implement projects using different AI learning techniques	

UNIT-I	
What is artificial intelligence, Problems, Problem Spaces and search, Heuristic search technique.Application: Solving various AI based problems.	**8 Hrs**
UNIT-II	
Knowledge Representation Issues, Using Predicate Logic, Representing knowledge using Rules.Application: Developing information about the objects	**8 Hrs**
UNIT-III	
Symbolic Reasoning under Uncertainty, Statistical reasoning, Weak Slot and Filter Structures. Application: Connecting one concept to another , combining ideas about data.	**8 Hrs**
UNIT-IV	
Strong slot-and-filler structures, Game Playing.Application: Designing Smart Games	**8 Hrs**
UNIT-V	

Natural Language Processing, Learning, Expert Systems.Application: Sentiment analysis	**8 Hrs**

Course Outcomes: After completing the course, the students will be able to	
CO1	Identify the AI based problems.
CO2	Apply techniques to solve problems
CO3	Define learning and explain various learning techniques.
CO4	Discuss expert systems
CO5	Implement projects using different AI learning techniques.

Reference Books	
1.	E. Rich , K. Knight & S. B. Nair - Artificial Intelligence, 3/e, McGraw Hill.
2.	Stuart Russel, Peter Norvig, "Artificial Intelligence: A Modern Approach" , 2nd Edition, Pearson Education, 2003.
3.	Dan W. Patterson, Introduction to Artificial Intelligence and Expert Systems – Prentice Hal of India.
4.	G. Luger, "Artificial Intelligence: Structures and Strategies for complex problem Solving", Fourth Edition, Pearson Education, 2002.

Continuous Internal Evaluation (CIE):

Theory for 50 Marks

CIE is executed by way of quizzes (Q), tests (T) and assignments. A minimum of three quizzes are conducted along with tests. Test portion is evaluated for 50 marks and quiz is evaluated for 10 marks. Faculty may adopt innovative methods for conducting quizzes effectively. The number of quizzes may be more than three (conduct additional quizzes and take best three). The three tests are conducted for 50 marks each and the average of all the tests are calculated for 50. The marks for the assignments are 20 (2 assignments for 10 marks each). The marks obtained in test, quiz and assignment are added to get marks out of 100 and report CIE for 50 marks.

Semester End Examination (SEE):

Total marks: 50+50=100

SEE for 50 marks is executed by means of an examination. The Question paper for each course contains two parts, Part – A and Part – B. Part – A consists of objective type questions for 20 marks covering the entire syllabus. Part – B Students have to answer five questions, one from each unit for 16 marks adding up to 80 marks. Each main question may have a maximum of three

sub divisions. Each unit will have internal choice in which both questions cover entire unit having same complexity in terms of COs and Bloom's taxonomy level.

CO-PO Mapping												
CO/PO	**PO1**	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PO8**	**PO9**	**PO10**	**PO11**	**PO12**
CO1	2	2	3		2							
CO2	2	3	3	3	2							
CO3		2	2	2								
CO4		2	2	3								
CO5	3	3	3	3	3							

High-3, Medium-2, Low-1

Semester: VII		
PYTHON PROGRAMMING		
Course Code: MVJ21IS743		**CIE Marks:100**
Credits: L:T:P:S: 3:0:0		**SEE Marks: 100**
Hours: 40L+26T		**SEE Duration: 3 Hrs**
Course Learning Objectives: The students will be able to		
1	Learn fundamental features of object-oriented language	
2	Design, write, debug, run Python Programs	
3	Develop console -based applications using Python	
4	Develop console & windows applications using Python	
5	Introduce event driven Graphical User Interface (GUI) programming using Python built in functions	

UNIT-I	
Why should you learn to write programs, Introduction to Python, Variables, expressions and statements, Conditional execution, Functions. Application: In learning and implementing small project process	**8 Hrs**
UNIT-II	
Iteration, Strings, Files.Application: Pattern recognition and Reading resultant column in supervised learning data set	**8 Hrs**
UNIT-III	
Lists, Dictionaries, Tuples, Regular Expressions.Application: Handling query languages and Managing Large set of data with respect to database	**8 Hrs**
UNIT-IV	
Classes and objects, Classes and functions, Classes and methods.Application:	**8 Hrs**

Designing games and puzzles	
UNIT-V	
Networked programs, Using Web Services, Using databases and SQL.Application: Music composition and movie development	**8 Hrs**

Course Outcomes: After completing the course, the students will be able to	
CO1	Understand Python syntax and semantics and be fluent in the use of Python flow control and functions.
CO2	Demonstrate proficiency in handling Strings and File Systems.
CO3	Implement Python Programs using core data structures like Lists, Dictionaries and use Regular Expressions.
CO4	Interpret the concepts of Object-Oriented Programming as used in Python.
CO5	Implement exemplary applications related to Network Programming, Web Services and Databases in Python.

Reference Books	
1.	Charles R. Severance, "Python for Everybody: Exploring Data Using Python 3", 1st Edition, CreateSpace Independent Publishing Platform, 2016. (http://do1.drchuck.com/pythonlearn/EN_us/pythonlearn.pdf)
2.	Allen B. Downey, "Think Python: How to Think Like a Computer Scientist", 2ndEdition, Green Tea Press, 2015. (http://greenteapress.com/thinkpython2/thinkpython2.pdf)
3.	Charles Dierbach, "Introduction to Computer Science Using Python", 1st Edition, Wiley India Pvt Ltd. ISBN-13: 978-8126556014
4.	Charles R. Severance, "Python for Everybody: Exploring Data Using Python 3", 1st Edition, CreateSpace Independent Publishing Platform, 2016. (http://do1.drchuck.com/pythonlearn/EN_us/pythonlearn.pdf)

Continuous Internal Evaluation (CIE):

Theory for 50 Marks

CIE is executed by way of quizzes (Q), tests (T) and assignments. A minimum of three quizzes are conducted along with tests. Test portion is evaluated for 50 marks and quiz is evaluated for 10 marks. Faculty may adopt innovative methods for conducting quizzes effectively. The number of quizzes may be more than three (conduct additional quizzes and take best three). The three tests are conducted for 50 marks each and the average of all the tests are calculated for 50. The marks for the assignments are 20 (2 assignments for 10 marks each). The marks obtained in test, quiz and assignment are added to get marks out of 100 and report CIE for 50 marks.

Semester End Examination (SEE):

Total marks: 50+50=100

SEE for 50 marks is executed by means of an examination. The Question paper for each course contains two parts, Part – A and Part – B. Part – A consists of objective type questions for 20 marks covering the entire syllabus. Part – B Students have to answer five questions, one from each unit for 16 marks adding up to 80 marks. Each main question may have a maximum of three sub divisions. Each unit will have internal choice in which both questions cover entire unit having same complexity in terms of COs and Bloom's taxonomy level.

CO-PO Mapping												
CO/PO	**PO1**	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PO8**	**PO9**	**PO10**	**PO11**	**PO12**
CO1	3	3	3		2							
CO2	2	2	3		2							
CO3	3	2	3		2							
CO4	2	1	3		2							
CO5	2	1	3		2							

High-3, Medium-2, Low-1

Semester: VII		
INTRODUCTION TO BIGDATA		
Course Code: MVJ21IS744		**CIE Marks:100**
Credits: L:T:P:S: 3:0:0		**SEE Marks: 100**
Hours: 40L+26T		**SEE Duration: 3 Hrs**
Course Learning Objectives: The students will be able to		
1	Understand Hadoop Distributed File system and examine MapReduce Programming	
2	Explore Hadoop tools and manage Hadoop with Sqoop	
3	Appraise the role of data mining and its applications across industries	
4	Identify various Text Mining techniques	

UNIT-I	
Hadoop Distributed file system:HDFS Design, Features, HDFS Components, HDFS user commands Hadoop MapReduce Framework: The MapReduce Model, Map-reduce Parallel Data Flow,Map Reduce Programming	**8 Hrs**
UNIT-II	
Essential Hadoop Tools:Using apache Pig, Using Apache Hive, Using Apache Sqoop, Using Apache Apache Flume, Apache H Base	**8 Hrs**
UNIT-III	
Data Warehousing: Introduction, Design Consideration, DW Development Approaches, DW Architectures Data Mining: Introduction, Gathering, and	**8 Hrs**

Selection, data cleaning and preparation, outputs ofData Mining, Data Mining Techniques	
UNIT-IV	
Decision Trees: Introduction, Decision Tree Problem, Decision Tree Constructions, Lessons from Construction Trees. Decision Tree Algorithm Regressions: Introduction, Correlations and Relationships, Non-Linear Regression, Logistic Regression, Advantages and disadvantages.	**8 Hrs**
UNIT-V	
Text Mining: Introduction, Text Mining Applications, Text Mining Process, Term Document Matrix, Mining the TDM, Comparison, Best Practices Web Mining: Introduction, Web Content Mining, Web Structured Mining, Web Usage Mining, Web Mining Algorithms.	**8 Hrs**

Course Outcomes: After completing the course, the students will be able to	
CO1	Master the concepts of HDFS and MapReduce framework
CO2	Investigate Hadoop related tools for Big Data Analytics and perform basic
CO3	Infer the importance of core data mining techniques for data analytics
CO4	Use Machine Learning algorithms for real world big data.
CO5	Use MapReduce Algorithms in real world big data.

Reference Books	
1.	Douglas Eadline,"Hadoop 2 Quick-Start Guide: Learn the Essentials of Big DataComputing in the Apache Hadoop 2 Ecosystem", 1stEdition, Pearson Education,2016. 2. Anil Maheshwari, “Data Analytics”, 1stEdition, McGraw Hill Education,2017

Continuous Internal Evaluation (CIE):

Theory for 50 Marks

CIE is executed by way of quizzes (Q), tests (T) and assignments. A minimum of three quizzes are conducted along with tests. Test portion is evaluated for 50 marks and quiz is evaluated for 10 marks. Faculty may adopt innovative methods for conducting quizzes effectively. The number of quizzes may be more than three (conduct additional quizzes and take best three). The three tests are conducted for 50 marks each and the average of all the tests are calculated for 50. The marks for the assignments are 20 (2 assignments for 10 marks each). The marks obtained in test, quiz and assignment are added to get marks out of 100 and report CIE for 50 marks.

Semester End Examination (SEE):

Total marks: 50+50=100

SEE for 50 marks is executed by means of an examination. The Question paper for each course contains two parts, Part – A and Part – B. Part – A consists of objective type questions for 20 marks covering the entire syllabus. Part – B Students have to answer five questions, one from each unit for 16 marks adding up to 80 marks. Each main question may have a maximum of three sub divisions. Each unit will have internal choice in which both questions cover entire unit having same complexity in terms of COs and Bloom's taxonomy level.

CO-PO Mapping												
CO/PO	**PO1**	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PO8**	**PO9**	**PO10**	**PO11**	**PO12**
CO1	3	3	0	2	0	0	0	0	0	0	1	1
CO2	3	3	0	2	0	0	0	0	0	0	1	1
CO3	3	3	0	3	0	0	0	0	0	0	0	1
CO4	2	2	0	3	0	0	0	0	0	0	1	1
CO5	2	2	0	2	0	0	0	0	0	0	0	1

High-3, Medium-2, Low-1

Semester: VII		
PROJECT PHASE I		
Course Code: MVJ21ISPR76		**CIE Marks:100**
Credits: L:T:P:S:0:0:4		**SEE Marks: 100**
Hours:		**SEE Duration: 3 Hrs**

Course objective :

To support independent learning and innovative attitude.

To guide to select and utilize adequate information from varied resources upholding ethics.

To guide to organize the work in the appropriate manner and present information (acknowledging the sources) clearly.

To develop interactive, communication, organisation, time management, and presentation skills.

To impart flexibility and adaptability.

To inspire independent and team working.

To expand intellectual capacity, credibility, judgement, intuition.

To adhere to punctuality, setting and meeting deadlines.

To instil responsibilities to oneself and others.

To train students to present the topic of project work in a seminar without any fear, face audience confidently, enhance communication skill, involve in group discussion to present and exchange

ideas
Project Work Phase - II:Each student of the project batch shall involve in carrying out the project work jointly in constant consultation with internal guide, co-guide, and external guide and prepare the project report as per the norms avoiding plagiarism..
Course outcomes: Present the project and be able to defend it. Make links across different areas of knowledge and to generate, develop and evaluate ideas and information so as to apply these skills to the project task. Habituated to critical thinking and use problem solving skills. Communicate effectively and to present ideas clearly and coherently in both the written and oral forms. Work in a team to achieve common goal. Learn on their own, reflect on their learning and take appropriate actions to improve it
CIE procedure for Mini - Project: CIE procedure for Project Work Phase - 1: (i)Single discipline:The CIE marks shall be awarded by a committee consisting of the Head of the concerned Department and two senior faculty members of the Department, one of whom shall be the Guide. The CIE marks awarded for the project work phase -1, shall be based on the evaluation of project work phase -1 Report, project presentation skill and question and answer session in the ratio 50:25:25.The marks awarded for the project report shall be the same for all the batch mates. (ii) Interdisciplinary:Continuous Internal Evaluation shall be group wise at the college level with the participation of all guides of the college. Participation of external guide/s, if any, is desirable.

The CIE marks awarded for the project work phase -1, shall be based on the evaluation of project work phase -1 Report, project presentation skill and question and answer session in the ratio 50:25:25.The marks awarded for the project report shall be the same for all the batch mates

Semester: VII		
NOSQL DATABASE		
Course Code: MVJ21IS77		**CIE Marks:100**
Credits: L:T:P:S: 1:0:0		**SEE Marks: 100**
Hours: 40L+26T		**SEE Duration: 2 Hrs**
Course Learning Objectives: The students will be able to		
1	Recognize and Describe the four types of NoSQL Databases, the Document-oriented, KeyValue	
2	Pairs, Column-oriented and Graph databases useful for diverse applications.	
3	Apply performance tuning on Column-oriented NoSQL databases and Document-oriented NoSQL Databases.	
4	Differentiate the detailed architecture of column oriented NoSQL database, Document database and Graph Database and relate usage of processor, memory, storage and file system commands	
5	Evaluate several applications for location based service and recommendation services. Devise an application using the components of NoSQL.	

UNIT-I	
Why NoSQL? The Value of Relational Databases, Getting at Persistent Data, Concurrency, Integration, A (Mostly) Standard Model, Impedance Mismatch, Application and Integration Databases, Attack of the Clusters, The Emergence of NoSQL, Aggregate Data Models; Aggregates, Example of Relations and Aggregates, Consequences of Aggregate Orientation, Key-Value and Document Data Models, Column-Family Stores, Summarizing AggregateOriented Databases. More Details on Data Models; Relationships, Graph Databases, Schemaless Databases, Materialized Views, Modeling for Data Access,	**8 Hrs**
UNIT-II	
Distribution Models; Single Server, Sharding, Master-Slave Replication, Peer-to-Peer Replication, Combining Sharding and Replication. Consistency, Update Consistency, Read Consistency, Relaxing Consistency, The CAP Theorem, Relaxing Durability, Quorums. Version Stamps, Business and System Transactions, Version Stamps on Multiple Nodes	**8 Hrs**
UNIT-III	
Map-Reduce, Basic Map-Reduce, Partitioning and Combining, Composing Map-Reduce Calculations, A Two Stage Map-Reduce Example, Incremental Map-	**8 Hrs**

Reduce Key-Value Databases, What Is a Key-Value Store, Key-Value Store Features, Consistency, Transactions, Query Features, Structure of Data, Scaling, Suitable Use Cases, Storing Session Information, User Profiles, Preference, Shopping Cart Data, When Not to Use, Relationships among Data, Multioperation Transactions, Query by Data, Operations by Sets	
UNIT-IV	
Document Databases, What Is a Document Database?, Features, Consistency, Transactions, Availability, Query Features, Scaling, Suitable Use Cases, Event Logging, Content Management Systems, Blogging Platforms, Web Analytics or Real-Time Analytics, E- Commerce Applications, When Not to Use, Complex Transactions Spanning Dif erent Operations, Queries against Varying Aggregate Structure	**8 Hrs**
UNIT-V	
Graph Databases, What Is a Graph Database?, Features, Consistency, Transactions, Availability, Query Features, Scaling, Suitable Use Cases, Connected Data, Routing, Dispatch, and Location-Based Services, Recommendation Engines, When Not to Use.	**8 Hrs**

Course Outcomes: After completing the course, the students will be able to	
CO1	Demonstrate an understanding of the detailed architecture of Column Oriented NoSQL databases, Document databases, Graph databases.
CO2	Use the concepts pertaining to all the types of databases.
CO3	Analyze the structural Models of NoSQL.
CO4	Develop various applications using NoSQL databases

Reference Books	
1.	Sadalage, P. & Fowler, NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence, Pearson Addision Wesley, 2012
2.	Dan Sullivan, "NoSQL For Mere Mortals", 1st Edition, Pearson Education India, 2015. (ISBN13: 978-9332557338)
3.	Dan McCreary and Ann Kelly, "Making Sense of NoSQL: A guide for Managers and the Rest of us", 1st Edition, Manning Publication/Dreamtech Press, 2013. (ISBN-13: 978-9351192022)
4.	Kristina Chodorow, "Mongodb: The Definitive Guide- Powerful and Scalable Data Storage", 2nd Edition, O'Reilly Publications, 2013. (ISBN-13: 978-9351102694)

Continuous Internal Evaluation (CIE):

Theory for 50 Marks

CIE is executed by way of quizzes (Q), tests (T) and assignments. A minimum of three quizzes are conducted along with tests. Test portion is evaluated for 50 marks and quiz is evaluated for 10 marks. Faculty may adopt innovative methods for conducting quizzes effectively. The number of quizzes may be more than three (conduct additional quizzes and take best three). The three tests

are conducted for 50 marks each and the average of all the tests are calculated for 50. The marks for the assignments are 20 (2 assignments for 10 marks each). The marks obtained in test, quiz and assignment are added to get marks out of 100 and report CIE for 50 marks.

Semester End Examination (SEE):

Total marks: 50+50=100

SEE for 50 marks is executed by means of an examination. The Question paper for each course contains two parts, Part – A and Part – B. Part – A consists of objective type questions for 20 marks covering the entire syllabus. Part – B Students have to answer five questions, one from each unit for 16 marks adding up to 80 marks. Each main question may have a maximum of three sub divisions. Each unit will have internal choice in which both questions cover entire unit having same complexity in terms of COs and Bloom's taxonomy level.

CO-PO Mapping												
CO/PO	**PO1**	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PO8**	**PO9**	**PO10**	**PO11**	**PO12**
CO1	3	3	0	2	0	0	0	0	0	0	1	1
CO2	3	3	0	2	0	0	0	0	0	0	1	1
CO3	3	3	0	3	0	0	0	0	0	0	0	1
CO4	2	2	0	3	0	0	0	0	0	0	1	1
CO5	2	2	0	2	0	0	0	0	0	0	0	1

High-3, Medium-2, Low-1