IPS Academy, Institute of Engineering & Science

(A UGC Autonomous Institute, Affiliated to RGPV, Bhopal)
Scheme Based on AICTE Flexible Curriculum

Department of Computer Science & Engineering

Master of Engineering (M. E.)

Computer Science & Engineering (Data Science)

I Semester

S.No.	Subject Code	Category	Subject Name	Maximum Marks Allotted					Total Marks	Contact Hours per week			Total Credits
				Theory			Practical						
				End Sem	Mid Sem. Exam	Quiz/ Assignment	End Sem	Term work Lab Work &Sessional		L	T	P	
1.	**PSCC-MCS101**	PSCC	**Introduction to Data Science**	60	25	15	–	–	**100**	3	1	–	**4**
2.	**PSCC-MCS102**	PSCC	**Advanced Data Structures and Algorithm**	60	25	15	–	–	**100**	3	1	–	**4**
3.	**PSMC-MCS101**	PSMC	**Mathematical Foundations for Data Science**	60	25	15	–	–	**100**	3	–	–	**3**
4.	**PSEC-MCS101**	PSEC	**Program Specific Elective Course-I**	60	25	15	–	–	**100**	3	–	–	**3**
5.	**LC-MCS101**	LC	**Data Science Lab**	–	–	–	60	40	**100**	–	–	4	**2**
6.	**LC-MCS102**	LC	**Advanced Data Structures Lab**	–	–	–	60	40	**100**	–	–	4	**2**
7.	**MLC-MCS101**	MLC	**Advanced DBMS**	60	25	15	–	–	**100**	2	–	–	**2**
8.	**AUD-MCS101**	AUD	**Audit Course-I**	–	–	–	–	–	–	2	–	–	**0**
			Total	**300**	**125**	**75**	**120**	**80**	**700**	**16**	**2**	**8**	**20**

Program Specific Elective-I	**Audit Course-I**
PSEC-MCS101(A) Computer Vision	AUD-MCS101(A) Disaster Management
PSEC-MCS101(B) Information Retrieval	AUD-MCS101(B) Stress Management
PSEC-MCS101(C) Big Data Analytics	AUD-MCS101(C) Economics Policies in India
PSEC-MCS101(D) Natural Language Processing	AUD-MCS101(D) ICT for Development

1 Hr Lecture	1 Hr Tutorial	2 Hr Practical
1 Credit	1 Credit	1 Credit

IPS Academy
Institute of Engineering & Science
Department of Computer Science & Engineering
I-Semester

PSCC-MCS101	**Introduction to Data Science**	**3L : 1T (4 hrs.)**	**4 credits**

Prerequisite: Computer programming, Mathematics, Analysis and Design of Algorithms

Course Objective:

The objective of this course is to provide a strong foundation of fundamental concepts in Data Science.

Course Contents: (40 hrs.)

Module 1: Introduction **(08 hrs.)**
Introduction to Data Science – Evolution of Data Science – Data Science Roles – Stages in a Data Science Project – Applications of Data Science in various fields – Data Security Issues.

Module 2: Data Collection and Data Pre-Processing **(08 hrs.)**
Data Collection Strategies – Data Pre-Processing Overview – Data Cleaning – Data Integration and Transformation – Data Reduction – Data Discretization.

Module 3: Exploratory Data Analytics **(08 hrs.)**
Descriptive Statistics – Mean, Standard Deviation, Skewness and Kurtosis – Box Plots – Pivot Table – Heat Map – Correlation Statistics – ANOVA.

Module 4: Model Development **(08 hrs.)**
Simple and Multiple Regression – Model Evaluation using Visualization – Residual Plot – distribution Plot – Polynomial Regression and Pipelines – Measures for In-sample Evaluation – Prediction and Decision Making.

Module 5: Model Evaluation **(08 hrs.)**
Generalization Error – Out-of-Sample Evaluation Metrics – Cross Validation – Overfitting – Under Fitting and Model Selection – Prediction by using Ridge Regression – Testing Multiple Parameters by using Grid Search.

Course Outcome:

1. State the core concepts of Data Science, It's foundation and principles.
2. Examine the useful search techniques; learn their advantages, disadvantages and be able to develop intelligent systems.
3. Learn the practical applicability of intelligent systems, specifically its applications.
4. Understand AI in different areas like NLP, Pattern Recognition, game planning etc.
5. Understand important concepts like Expert Systems, DS applications.

List of Text / Reference Books:

1. Jojo Moolayil, “Smarter Decisions : The Intersection of IoT and Data Science”, PACKT, 2016.
2. Cathy O’Neil and Rachel Schutt , “Doing Data Science”, O'Reilly, 2015.
3. David Dietrich, Barry Heller, Beibei Yang, “Data Science and Big data Analytics”, EMC 2013
4. Raj, Pethuru, “Handbook of Research on Cloud Infrastructures for Big Data Analytics”, IGI Global

IPS Academy
Institute of Engineering & Science
Department of Computer Science & Engineering
I-Semester

PSCC-MCS102	**Advanced Data Structures and Algorithm**	**3L:1T (4 hrs.)**	**4 credits**

Prerequisite: Data Structure, Basic knowledge of C/C++,

Course Objective:

The objective of this course is intended to provide the foundations of the practical implementation and usage of Algorithms and Data Structures.

Course Contents: (40 hrs.)

Module 1: **(08 hrs.)**
Introduction to Data Structure: Concepts of Data and Information, Classification of Data Structures, Abstract Data Types, Implementation Aspects: Memory Representation, Data Structures Operations and its Cost Estimation, Introduction to linear data structures- Arrays, Linked List: Representation of linked list in memory, different implementation of linked list, Circular linked list, doubly linked list, Application of linked list: polynomial manipulation using linked list.

Module 2: **(08 hrs.)**
Stacks: Stacks as ADT, Different implementation of stack, multiple stacks, Application of Stack: Conversion of infix to postfix notation using stack, evaluation of postfix expression, Recursion, Queues: Queues as ADT, Different implementation of queue, Circular queue, Concept of Dqueue and Priority Queue, Queue simulation, Application of queues.

Module 3: **(10 hrs.)**
Algorithms, Designing algorithms, analyzing algorithms, asymptotic notations, heap and heap sort, Introduction to divide and conquer technique, analysis, design and comparison of various algorithms based on this technique, example binary search, merge sort, quick sort, strassen's matrix multiplication,

Module 4: **(08 hrs.)**

Study of Greedy Strategy, examples of greedy method like Optimal Merge Patterns, Huffman Coding, minimum spanning trees, knapsack problem, job sequencing with deadlines, Shortest Path Algorithm

Module 5: (06 hrs.)

Graphs: Introduction, Classification of graph: Directed and Undirected graphs, Representation, Graph algorithm: Minimum Spanning Tree (MST)- Kruskal, Prim's algorithms, Dijkstra's shortest path algorithm; Comparison between different graph algorithms, Application of graphs.

Course Outcome:

1. Design and analyze programming problem statements.
2. Choose appropriate data structures and algorithms, understand the ADT/libraries, and use it to design algorithms for a specific problem.
3. Learn basic concepts of algorithms and sorting & searching techniques.
4. Understand the necessary mathematical abstraction to solve problems and Come up with analysis of efficiency and proofs of correctness
5. Comprehend and select algorithm design approaches in a problem specific manner for tree and graph.

List of Text / Reference Books:

1. AM Tanenbaum, Y Langsam& MJ Augustein, "Data structure using C and C++", Prentice Hall India, Second Addition, 2000.
2. Robert Kruse, Bruse Leung, "Data structures & Program Design in C", Pearson Education, Second Addition, 2000.
3. E. Horowitz, Sahni and D. Mehta, "Analysis & Design of Algorithm", 2002, Galgotia Publication 2002
4. Aho, Hopcroft, Ullman, "Data Structures and Algorithms", Pearson Education.
5. N. Wirth, "Algorithms + Data Structure = Programs", Prentice Hall.
6. Richard, Gilberg Behrouz, Forouzan, "Data structure – A Pseudocode Approach with C", Thomson press.
7. Coremen Thomas, Leiserson CE, Rivest RL, "Introduction to Algorithms" PHI.
8. Ullmann, "Analysis & Design of Algorithm".
9. Michael T Goodrich, Robarto Tamassia, "Algorithm Design", Wiely Publication, India
10. Rajesh K Shukla, "Analysis and Design of Algorithms: A Beginner's Approach", Wiley publications, 2015.

IPS Academy
Institute of Engineering & Science
Department of Computer Science & Engineering
I-Semester

PSMC-MCS101	**Mathematical Foundations for Data Science**	**3L (3 hrs.)**	**3 credits**

Prerequisite: Mathematics

Course Objective: The course will introduce students to the fundamental mathematical concepts required for a program in data science.

Course Contents: (40 hrs.)

Module 1: **(06 hrs.)**
Basics of Data Science: Introduction; Typology of problems; Importance of linear algebra, statistics and optimization from a data science perspective; Structured thinking for solving data science problems.

Module 2: **(12 hrs.)**
Linear Algebra: Matrices and their properties (determinants, traces, rank, nullity, etc.); Eigenvalues and eigenvectors; Matrix factorizations; Inner products; Distance measures; Projections; Notion of hyperplanes; half-planes.

Module 3: **(10 hrs.)**
Probability, Statistics and Random Processes: Probability theory and axioms; Random variables; Probability distributions and density functions (univariate and multivariate); Expectations and moments; Covariance and correlation; Statistics and sampling distributions; Hypothesis testing of means, proportions, variances and correlations; Confidence (statistical) intervals; Correlation functions; White-noise process.

Module 4: **(08 hrs.)**
Optimization: Unconstrained optimization; Necessary and sufficiency conditions for optima; Gradient descent methods; Constrained optimization, KKT conditions; Introduction to non-gradient techniques; Introduction to least squares optimization; Optimization view of machine learning.

Module 5: **(04 hrs.)**
Introduction to Data Science Methods: Linear regression as an exemplar function approximation problem; Linear classification problems.

Course Outcome:

1. To identify and solve problems using the concepts of linear algebra.
2. To understand and use the concepts of solution of partial differential equations and transform calculus in practical problems appear in Computer Science and Engineering.
3. To identify and solve problems regarding probability and hypothesis.
4. To explain and apply the concept of stochastic process and queuing theory in engineering problems.

List of Text / Reference Books:

1. M. K. Jain, Numerical solution of differential equations, John Wiley & Sons, 2010.
2. R. N.Bracewell, Fourier Transform & It's Applications, Tata McGraw Hill, 2014.
3. S. C. Chapra, Applied Numerical Methods with MATLAB, Tata McGraw Hill, 2007.
4. T. J. Ross, Fuzzy Logic with Engineering Application, Wiley publisher, 2011.
5. S. Ross, A first course in probability, Pearson education India, 6th edition, 2002.
6. Erwin kreyszig, Advanced Engineering Mathematics, John Wiley & Sons, 9th Edition, 2006.
7. B.S. Grewal, Higher Engineering Mathematics, Khanna Publishers, 36th Edition, 2010.
8. R. J. Beerwends, Fourier and Laplace transform, Cambridge university press, 2003.
9. E. J. Watson, Laplace transform and applications, Publisher: Van Nostrand Reinhold, 1980.
10. P. G. Hoel, S. C. Port, C. J. Stone, Introduction to probability theory, Universal Book Stall.
11. H. J. Zimmersoms , Fuzzy Sets Theory & its Applications.

IPS Academy
Institute of Engineering & Science
Department of Computer Science & Engineering
I-Semester

LC-MCS101	**Data Science Lab**	**4P (4 hrs.)**	**2 credits**

Prerequisite: Computer programming (C/C++), Analysis and Design of Algorithms, Mathematics

Course Objective:
To provide students with and practical base in Data Science.

List of Experiments:

Week 1
R AS CALCULATOR APPLICATION
a. Using with and without R objects on console
b. Using mathematical functions on console
c. Write an R script, to create R objects for calculator application and save in a specified location in disk

Week 2
DESCRIPTIVE STATISTICS IN R
a. Write an R script to find basic descriptive statistics using summary, str, quartile function on mtcars & cars datasets.
b. Write an R script to find subset of dataset by using subset (), aggregate () functions on iris dataset.

Week 3
READING AND WRITING DIFFERENT TYPES OF DATASETS
a. Reading different types of data sets (.txt, .csv) from Web and disk and writing in file in specific disk location.
b. Reading Excel data sheet in R.
c. Reading XML dataset in R.

Week 4
VISUALIZATIONS
a. Find the data distributions using box and scatter plot.
b. Find the outliers using plot.
c. Plot the histogram, bar chart and pie chart on sample data.

Week 5
CORRELATION AND COVARIANCE
a. Find the correlation matrix.
b. Plot the correlation plot on dataset and visualize giving an overview of

relationships among data on iris data.
c. Analysis of covariance: variance (ANOVA), if data have categorical variables on iris data.

Week 6
REGRESSION MODEL
Import a data from web storage. Name the dataset and now do Logistic Regression to find out relation between variables that are affecting the admission of a student in a institute based on his or her GRE score, GPA obtained and rank of the student. Also check the model is fit or not. Require (foreign), equire (MASS).

Week 7
MULTIPLE REGRESSION MODEL
Apply multiple regressions, if data have a continuous
Independent variable. Apply on above dataset.

Week 8
REGRESSION MODEL FOR PREDICTION
Apply regression Model techniques to predict the data on above dataset.

Week 9
CLASSIFICATION MODEL
a. Install relevant package for classification.
b. Choose classifier for classification problem.
c. Evaluate the performance of classifie

Week 10
CLUSTERING MODEL
a. Clustering algorithms for unsupervised classification.
b. Plot the cluster data using R visualizations.

Course Outcomes:

1. Able to understanding of the major areas and challenges of DS.
2. Ability to apply basic DS algorithms to solve problems.
3. Able to describe search strategies and solve problems by applying a suitable search method.
4. Able to describe and apply knowledge representation.

IPS Academy
Institute of Engineering & Science
Department of Computer Science & Engineering
I-Semester

LC-MCS102	**Advanced Data Structures Lab**	**4P (4 hrs.)**	**2 credits**

Prerequisite: Data Structure, Basic knowledge of C/C++,

Course Objectives:

1. Understand the foundations of the practical implementation and usage of Algorithms and Data Structures.

List of Experiments:

Week 1

1. Write a program for Iterative and Recursive Binary Search.

Week 2

1. Write a program for Merge Sort.
2. Write a program for Quick Sort.

Week 3

1. Write a program for Strassen's Matrix Multiplication.
2. Write a program for optimal merge patterns.
3.

Week 4

1. Write a program for Huffman coding.

Week 5

1. Write a program for minimum spanning trees using Kruskal's a nd P r im s algorithm.
2. Write a program for single sources shortest path algorithm.

Week 6

1. Write a program for Floye-Warshal algorithm.
2. Write a program for traveling salesman problem and Hamiltonian cycle problem.

Week 7

1. Min/Max Heap
2. Leftist Heap

Week 8

1. AVL Trees
2. Red-Black Trees

3. B-Trees
4. Segment Trees

Week 9

1. Line segment intersection

Week 10

1. Convex Hull
2. Voronoi Diagram

Course Outcomes:

1. Design and analyze programming problem statements.
2. Choose appropriate data structures and algorithms, understand the ADT/libraries, and use it to design algorithms for a specific problem.
3. Learn basic concepts of algorithms and sorting & searching techniques.
4. Implement heap and various tree structure like AVL, Red-black, B and Segment trees.
5. Solve the problems such as line segment intersection, convex shell and Voronoi diagram.

IPS Academy
Institute of Engineering & Science
Department of Computer Science & Engineering
I-Semester

MLC-MCS101	Advanced DBMS	2L (2 hrs.)	2 credits

Prerequisite: Database Management System
.

Course Objective:

The objective of the course is to present an introduction to advance concepts in database management systems, with an emphasis on how to organize, maintain and retrieve efficiently and effectively information from a DBMS.

Course Contents: (40 hrs.)

Module 1: **(10 hrs.)**
An overview of database, The Extended Entity Relationship Model and Object Model: The ER model revisited, Motivation for complex data types, User defined abstract data types and structured types, Subclasses, Super classes, Inheritance, Specialization and Generalization,

Module 2: **(8 hrs.)**
Query Processing, Optimization & Database Tuning: Algorithms for Executing Query Operations, Heuristics for Query Optimizations, Estimations of Query Processing Cost, Join Strategies for Parallel Processors, Database Workloads, Tuning Decisions, DBMS Benchmarks, Clustering & Indexing, Multiple Attribute Search Keys, Query Evaluation Plans, Pipelined Evaluations

Module 3: **(8 hrs.)**
Distributed Database System: Structure of Distributed Database, Data Fragmentation, Data Model, Query Processing, Semi Join, Parallel & Pipeline Join, Distributed Query Processing,

Module 4: **(08 hrs.)**
Enhanced Data Model for Advanced Applications: Database Operating System, Introduction to Temporal Database Concepts, Spatial And Multimedia Databases, Active Database System, Deductive Databases, Web Databases, Issues in Real Time Database Design.

Module 5: **(06 hrs.)**
Accessing databases from Web, database technology to Web related areas such as semi-structured databases and data integration, XML, XQuery, XPath, XML Schemas, distributed database design.

Course Outcome:

1. Explain and evaluate the fundamental theories and requirements that influence the design of modern database systems
2. Assess and apply database functions and packages suitable for enterprise database development and database management
3. Critically evaluate alternative designs and architectures for databases and discuss and evaluate methods of storing, managing and interrogating complex data.
4. Explain and critically evaluate different database solutions for data exchange.
5. Analyze the background processes involved in queries and transactions, and explain how these impact on database operation and design

List of Text / Reference Books:

1. R. Ramakrishnan, J. Gehrke, “Database Management Systems”, McGraw Hill, 2004
2. A. Silberschatz, H. Korth, S. Sudarshan, “Database system concepts”, 5 Edition, McGraw Hill, 2008.
3. Martin Kleppmann, “Designing Data-Intensive Applications: The Big Ideas Behind Reliable, Scalable, and Maintainable Systems”, 1st Ed, O'Reilly 2017.
4. Date C. J., “An Introduction to Database Systems”, Addison Wesley Longman (8th Ed), 2003.
5. Majumdar & Bhattacharya, “Database Management System”, TMH.
6. Elmasri, Navathe, “Fundamentals of Database Systems”, Addison Wesley.

IPS Academy
Institute of Engineering & Science
Department of Computer Science & Engineering
I-Semester

PSEC-MCS101	Computer Vision	3L (3 hrs.)	3 credits

Prerequisite: Data structures, Computer Graphics, Mathematics

Course Objective:

The main objective of this course is to introduce students the fundamentals of image formation, methods and techniques of computer vision and pattern recognition and to develop an appreciation for various issues in the design of computer vision and object recognition systems.

Course Contents: (40 hrs.)

Module 1: **(10 hrs.)**
Introduction: Image Processing, Computer Vision and Computer Graphics, What is Computer Vision-Low-level, Mid-level, High-level, Overview of Diverse Computer Vision Applications: Document Image Analysis, Biometrics, Object Recognition, Tracking, Medical Image Analysis, Content-Based Image Retrieval, Video Data Processing, Multimedia, Virtual Reality and Augmented Reality

Module 2: **(10 hrs.)**
Image Formation Models: Monocular imaging system, Radiosity: The 'Physics' of Image Formation, Radiance, Irradiance, BRDF, color etc, Orthographic & Perspective Projection, Camera model and Camera calibration, Binocular imaging systems, Multiple views geometry, Structure determination, shape from shading, Photometric Stereo, Depth from Defocus, Construction of 3D model from images.

Module 3: **(8 hrs.)**
Image Processing and Feature Extraction: Image preprocessing, Image representations (continuous and discrete), Edge detection, Motion Estimation: Regularization theory, Optical computation, Stereo Vision, Motion estimation, Structure from motion.

Module 4: **(08 hrs.)**
Shape Representation and Segmentation: Contour based representation, Region based representation, Deformable curves and surfaces, Snakes and active contours, Level set representations, Fourier and wavelet descriptors, Medial representations, Multi resolution analysis

Module 5: **(04 hrs.)**
Object recognition: Hough transforms and other simple object recognition methods, Shape correspondence and shape matching

Course Outcome:

1. To implement fundamental image processing techniques required for computer vision.
2. Understand Image formation process.
3. Extract features form images and do analysis of images.
4. To perform shape analysis and generate 3D model from images.
5. Understand video processing, object recognition, motion computation and 3D vision and geometry

List of Text / Reference Books:

1. R. C. Gonzalez, R. E. Woods. "Digital Image Processing", Addison Wesley Longman, Inc., 1992.
2. D. H. Ballard, C. M. Brown., "Computer Vision", Prentice-Hall, Englewood Cliffs, 1982.
3. Richard Szeliski, "Computer Vision: Algorithms and Applications (CVAA)", Springer, 2010.
4. E. R. Davies, "Computer & Machine Vision", Fourth Edition, Academic Press, 2012.
5. Simon J. D. Prince, "Computer Vision: Models, Learning, and Inference", Cambridge University Press, 2012.
6. Mark Nixon and Alberto S. Aquado, "Feature Extraction & Image Processing for Computer Vision", Third Edition, Academic Press, 2012.
7. Sonka, Hlavac, and Boyle Thomson, "Image Processing, Analysis, and Machine Vision".
8. D. Forsyth and J. Ponce, "Computer Vision - A modern approach", Prentice Hall Robot Vision, by B. K. P. Horn, McGraw-Hill.
9. E. Trucco and A. Verri, "Introductory Techniques for 3D Computer Vision", by, Publisher: Prentice Hall.

IPS Academy
Institute of Engineering & Science
Department of Computer Science & Engineering
I-Semester

PSEC-MCS101	Information Retrieval	3L (3 hrs.)	3 credits

Prerequisite: Data Mining, Data Structure and Algorithms

Course Objective:

The main objective of this course is to present the scientific support in the field of information search and retrieval. This course explores the fundamental relationship between information retrieval, hypermedia architectures, and semantic models, thus deploying and testing several important retrieval models such as vector space, Boolean and query expansion.

Course Contents: (30 hrs.)

Module 1: **(8 hrs.)**
Introduction to Information Retrieval, Retrieval strategies: vector space model, Probabilistic retrieval strategies: Simple term weights, Non binary independence model, Language models.

Module 2: **(6 hrs.)**
Retrieval Utilities: Relevance feedback, clustering, N-grams, Regression analysis, Thesauri

Module 3: **(6 hrs.)**
Retrieval utilities: Semantic networks, parsing, Cross–Language: Information Retrieval: Introduction, Crossing the Language barrier.

Module 4: **(4 hrs.)**
Efficiency: Inverted Index, Query processing, Signature files, Duplicate document detection.

Module 5: **(06 hrs.)**
Integrating structured data and text: A historical progression, Information retrieval as relational application, Semi Structured search using a relational schema, Distributed Information Retrieval: A theoretical Model of Distributed retrieval, web search

Course Outcome:

1. Understanding information retrieval strategies with examples.
2. Knowledge gathering about Retrieval Utilities and relevance feedback.
3. Applying and examine case studies
4. Understanding concept of query processing.
5. Knowledge gathering about integrated structure data, text and models.

List of Text / Reference Books:

1. David A. Grossman, OphirFrieder, “Information Retrieval – Algorithms and Heuristics”, Springer, 2nd Edition(Distributed by Universal Press), 2004
2. Gerald J Kowalski, Mark T Maybury, “Information Storage and Retrieval Systems: Theory and Implementation”, Springer, 2004.
3. Soumen Chakrabarti, “Mining the Web: Discovering Knowledge from Hypertext Data, Morgan”, Kaufmann Publishers, 2002.
4. Christopher D Manning, Prabhakar Raghavan, HinrichSchutze, “An Introduction to Information Retrieval”, Cambridge University Press, England, 2009.

IPS Academy
Institute of Engineering & Science
Department of Computer Science & Engineering

I-Semester

PSEC-MCS101	Big Data Analytics	3L (3 hrs.)	3 credits

Prerequisite: Data structures and algorithms, Data Base Management System

Course Objective: The main goal of this course is to help students learn, understand and practice big data analytics and machine learning approaches, which include the study of modern computing big data technologies and scaling up machine learning techniques focusing on industry applications.

Course Contents: (40 hrs.)

Module 1: **(06 hrs.)**

Introduction to Big data, Big data characteristics, Types of big data, Traditional versus Big data, Evolution of Big data, challenges with Big Data, Technologies available for Big Data, Infrastructure for Big data, Use of Data Analytics, Desired properties of Big Data system.

Module 2: **(10 hrs.)**

Introduction to Hadoop, Core Hadoop components, Hadoop Eco system, Hive Physical Architecture, Hadoop limitations, RDBMS Versus Hadoop, Hadoop Distributed File system, Processing Data with Hadoop, Managing Resources and Application with Hadoop YARN, MapReduce programming.

Module 3: **(10 hrs.)**

Introduction to Hive: Hive Architecture, Hive Data types, Hive Query Language, Introduction to Pig, Anatomy of Pig, Pig on Hadoop, Use Case for Pig, ETL Processing, Data types in Pig, running Pig, Execution model of Pig, Operators, functions.

Module 4: **(08 hrs.)**

Introduction to NoSQL, NoSQL Business Drivers, NoSQL Data architectural patterns, Variations of NOSQL, Architectural patterns using NoSQL, Introduction to MangoDB

Module 5: **(06 hrs.)**

Mining social Network Graphs: Introduction Applications of social Network mining, Social Networks as a Graph, Types of social Networks, Clustering of social Graphs, Direct Discovery of communities in a social graph, Introduction to recommender system.

Course Outcome:

1. Students should be able to understand the concept and challenges of Big data.
2. Students should be able to demonstrate knowledge of big data analytics.
3. Students should be able to develop Big Data Solutions using Hadoop Eco System.
4. Students should be able to gain hands-on experience on large-scale analytics tools.
5. Students should be able to analyze the social network graphs.

List of Text / Reference Books:

1. Radha Shankarmani, M. Vijaylakshmi, " Big Data Analytics", Wiley, Second edition, Januray 2016.
2. Seema Acharya, Subhashini Chellappan, " Big Data and Analytics", Wiley, Second edition, 2015.
3. Michael Minelli, Michele Chambers, AmbigaDhiraj, “Big Data Big Analytics”,Wiley, First edition, February 2013.
4. Tom White, “HADOOP: The definitive Guide”, O Reilly 2012
5. Chris Eaton, Dirk deroos, “Understanding Big Data”, McGraw Hill, 2012.

IPS Academy
Institute of Engineering & Science
Department of Computer Science & Engineering
I-Semester

PSEC-MCS101	Natural Language Processing	3L (3 hrs.)	3 credits

Prerequisite: Artificial Intelligence, Compiler Design, Theory of Computation

Course Objective:

This course introduces the fundamental concepts and techniques of natural language processing (NLP). Students will gain an in-depth understanding of the computational properties of natural languages and the commonly used algorithms for processing linguistic information.

Course Contents: (40 hrs.)

Module 1: **(06 hrs.)**
Introduction to AI and NLP, History of NLP, Study of Human Languages, Ambiguity and Uncertainty in language, NLP Phases, Introduction to Linguistic Resources, Corpus, Elements of Corpus Design, Types of Treebank Corpus, Applications of Treebank Corpus

Module 2: **(08 hrs.)**
Parsing techniques, context free grammer, recursive transitions nets (RNT), augmented transition nets (ATN), case and logic grammers, semantic analysis, Regular Expression, Properties and Examples, Finite State Autometa, Concept of derivation, types of derivations

Module 3: **(10 hrs.)**
Game playing, Minimax search procedure, alpha-beta cutoffs, additional refinements, Planning, Overview an example domain the block word, component of planning systems, goal stack planning, non linear planning.

Module 4: **(10 hrs.)**
Approaches and Methods to Word Sense Disambiguation (WSD), Applications of Word Sense Disambiguation (WSD), Difficulties in WSD, Concept of Coherence, Discourse structure, Algorithms for Discourse Segmentation, Text Coherence, Building Hierarchical Discourse Structure, Reference Resolution, Terminology Used in Reference Resolution, Types of Referring Expressions, Reference Resolution Tasks.

Module 5: **(06 hrs.)**
Natural Language Processing — Part of Speech (PoS) Tagging, Natural Language Processing: Natural Language Inception, Information Retrieval, Applications of NLP, Types of Machine Translation Systems, Approaches to Machine Translation (MT), Fighting Spam, Sentiment Analysis, Introduction to Python.

Course Outcome:

1. Understand approaches to syntax and semantics in NLP.
2. Understand approaches to discourse, generation, dialogue and summarization within NLP.
3. Understand current methods for statistical approaches to machine translation.
4. Acquiring knowledge about Word Sense Disambiguation and its applications.
5. Understand machine learning techniques in NLP and various algorithms applied within NLP.

List of Text / Reference Books:

1. Jurafsky and Martin, "Speech and Language Processing", Prentice Hall, 2000.
2. Lutz and Ascher, "Learning Python", O'Reilly, ISBN: 0596002815.
3. Manning and Schutze, "Statistical Natural Language Processing", MIT Press; 1st edition (June 18, 1999).
4. Tom Mitchell, “Machine Learning”, McGraw Hill, 1997, ISBN 0070428077.
5. Cover, T. M. and J. A. Thomas, “ Elements of Information Theory” Wiley Publications, 1991.
6. Charniak, E, “Statistical Language Learning”, The MIT Press. 1996.
7. Jelinek F, “Statistical Methods for Speech Recognition”, The MIT Press, 1998.

IPS Academy
Institute of Engineering & Science
Department of Computer Science & Engineering
I-Semester

AUD-MCS101	ICT for Development	2L (2 hrs.)	

Prerequisite: Basic Computer Engineering

Course Objective:

This course introduces the fundamental concepts and techniques of ICT and Development.

Course Contents: (24 hrs.)

Module 1: **(04 hrs.)**
Learning Presentation Skills: What is Presentation, Key elements of Presentation, Presentation Delivery Rules of a professional Presentation, Positive and Nervous making Factors, Tips for dynamic Deliveries, Dealing with questions in Presentation.

Module 2: **(04 hrs.)**

Mock Presentation:

Delivering Presentation, Mock Presentation, handle difficult Presentation tactics, Presentation Skills

Module 3: **(06 Hrs)**

Learning Group Discussion Skills: What is Group Discussion, Personality Qualities in Group Discussion, Need of Group Discussion in Industry/Organization, Organization's Group Discussion, Perspective Aspects of Group Discussion, Types of Topics Managing Score Cards in Group Discussion,

Module 4: Profile Development, Personality Skills and Assessments **(06 Hrs)**

Learning CV Building Skills, Planning Profile Building Factors, Execution Strategies for Profile Building, Learning through Professional, Role Plays Learning Situation, Reactions Picture & Word Perceptions, Career Advancement Assessment, Attention Disorder Assessment, Comm. Skill Assessment, Employability Skills Assessment

Module 5: **(04 hrs.)**
Mock Group Discussions:

Group Discussion on Factual & Abstract Topics, Techniques to Imitate Group Discussion, Do's and Don'ts in Group Discussion, Positive and Negative factors in Group Discussion

Course Outcome:

- Develop and Present Professional Presentations
- Develop discussion skills and further participate in discussions
- Develop Reasoning Skills in Discussion
- Plan primary CV with a view to develop it further in upcoming semesters
- Asses himself/herself in various career related assessments

List of Text / Reference Books:

1. The "Art of Winning" by M Ashraf Rizvi.
2. "How to Succeed in Group Discussions" by Dr. S K Mandal
3. "Presentation Skills: The Essential Guide for Students" by Patsy McCarthy&Caroline Hatcher