Faculty of Science and Technology

Savitribai Phule Pune University

Maharashtra, India

Syllabus Structure

Master of Computer Engineering

(Data Science)

(Course 2020)

(With effect from 2020-21)

Savitribai Phule Pune University, Pune
Master of Computer Engineering (Data Science) (2020 Course)
(with effect from A.Y. 2020-21)

Semester I

Course Code	Course	Teaching Scheme Hours / Week		Examination Scheme and Marks					Credit		
		Theory	Practical	In-Sem	End- Sem	TW	OR/ PR	Total	TH	TW	PR
510301	Mathematical Foundations for Data Science	03	--	50	50	--	--	100	03	--	--
510302	Basics of Data Science	03	--	50	50	--	--	100	03	--	--
510303	Big Data Analytics	03	--	50	50	--	--	100	03	--	--
510304	Research Methodology	03	--	50	50	--	--	100	03	--	--
510305	Elective I	04	--	50	50	--	--	100	04	--	--
510306	Laboratory Proficiency I	--	08	--	--	50	50	100	--	02	02
Total		**16**	**08**	**250**	**250**	**50**	**50**	**600**	**16**	**02**	**02**
Total Credit									**20**		
510307	Non-Credit Course I								Grade		

Elective I

510305A	Data Storage Technologies and Networks	510305B	Information Systems Management
510305C	Data Preparation and Analysis	510305D	Artificial Intelligence for Data Science
510305E	Open Elective		

Semester II

Course Code	Course	Teaching Scheme Hours / Week		Examination Scheme and Marks					Credit		
		Theory	Practical	In-Sem	End- Sem	TW	OR/ PR	Total	TH	TW	PR
510308	Data Warehousing and Mining	03	--	50	50	--	--	100	03	--	--
510309	Machine Learning	03	--	50	50	--	--	100	03	--	--
510310	Soft Computing	03	--	50	50	--	--	100	03	--	--
510311	Elective II	04	--	50	50	--	--	100	04	--	--
510312	Mini Project with Seminar I	--	03	--	--	50	50	100	--	01	02
510313	Laboratory Proficiency II	--	08	--	--	50	50	100	--	02	02
Total		**13**	**12**	**200**	**200**	**100**	**100**	**600**	**13**	**03**	**04**
Total Credit									**20**		
510314	Non-Credit Course II								Grade		

Elective II

510311A	Distributed Databases	510311B	Recommender Systems
510311C	GPU Computing	510311D	Web Intelligence
510311E	Open Elective		

Savitribai Phule Pune University, Pune
Master of Computer Engineering (Data Science) (2020 Course)
(with effect from A.Y. 2021-22)

Semester III

Course Code	Course	Teaching Scheme Hours / Week		Examination Scheme and Marks					Credit		
		Theory	Practical	In-Sem	End- Sem	TW	OR/ PRE	Total	TH	TW	PR
610301	Deep Learning	03	--	50	50	--	--	100	03	--	--
610302	Data Modeling and Visualization	03	--	50	50	--	--	100	03	--	--
610303	Elective III	04	--	50	50	--	--	100	04	--	--
610304	Seminar on Industry Internship-I/ In-house Research Project-I	--	03	--	--	50	50	100	--	01	02
610305	Dissertation Stage I	--	08	--	--	50	50	100	--	04	04
Total		**10**	**11**	**150**	**150**	**100**	**100**	**500**	**10**	**05**	**06**
Total Credit									**21**		
610306	Non-Credit Course III								Grade		

Elective III

610303A	Real Time Analytics	610303B	Business Analytics
610303C	Computational Linguistic Analytics	610303D	Video Analytics
	610303E Open Elective		

Semester IV

Course Code	Course	Teaching Scheme Hours / Week	Examination Scheme and Marks			Credit	
		Practical	TW	OR/PRE	Total	TW	PR
610307	Seminar on Industry Internship-II / In-house Research Project-II	03	50	50	100	01	02
610308	Dissertation Stage II	18	150	50	200	08	10
Total		**21**	**200**	**100**	**300**	**09**	**12**
Total Credit						**21**	

Savitribai Phule Pune University, Pune

ME Data Sciences (2020 Course)

510301: Mathematical foundation for Data Science

Prerequisites: Basic Mathematics

Companion Course: Basics of Data Science

Course Objectives:

1. To understand role of discrete mathematics in data science.
2. To learn probability and apply it for real life problems in Data Science.
3. To understand basis of descriptive statistics measures and hypothesis.
4. To learn linear algebra and calculus concepts and applicability in Data Science.
5. To learn different linear regression methods used in machine learning

Course Outcomes:

On completion of the course, learner will be able to –

CO1: Apply measures of central tendency to analyze a payroll dataset.
CO2: Apply probabilistic model for credit card fraud detection.
CO3: Evaluate covariance and correlation of between two variables.
CO4: Demonstrate use eigenvalues and eigenvectors for a reducing dimension of a healthcare dataset
CO5: Apply simple regression model to predict the near future sales based on a time series data.

Course Contents

Unit I	**Discrete mathematics for Data Science**	**(07 Hours)**
Concept of set, cardinality of set, finite, infinite and uncountably infinite sets, Basic set operations, Principal of inclusion Exclusion, **Graph:** Basic terminologies, representation of graph, path and circuit, graph traversal, travelling salesperson problem, **Trees:** Basic terminologies, search tree: Binary & M-ary tree.		
# Exemplar / Case Studies	Discuss algorithm / program for Salesman problem	
*Mapping of Course Outcomes	CO1	
Unit II	**Linear Algebra and Calculus**	**(07 Hours)**
Linear Algebra: Matrix and vector algebra, systems of linear equations using matrices, linear independence, Matrix factorization concept/LU decomposition, Eigen values and eigenvectors, **Understanding of calculus:** concept of function and derivative, Multivariate calculus: concept, Partial Derivatives, chain rule, the Jacobian and the Hessian		
#Exemplar/Case Studies	1. Demonstration of dimensionality reduction using eigenvalues and eigenvector (PCA) 2. Discussion of Page rank algorithm using eigenvalues and eigenvector	
*Mapping of Course Outcomes	CO4	
Unit III	**Data Analysis & Probability Theory**	**(07 Hours)**
Data Representation, Average, Spread, Experiments, Outcomes, Events, Probability, Permutations and Combinations, Random Variables, Probability Distributions, Mean and Variance of a Distribution, Binomial, Poisson, and Hyper geometric Distributions, Normal Distribution, Distributions of Several Random Variables.		
#Exemplar/Case Studies	Discuss probabilistic model for predicting relations in social websites system	
*Mapping of Course Outcomes	**CO3**	
Unit IV	**Statistical Inference I**	**(07 Hours)**

Types of Statistical Inference, Descriptive Statistics, Inferential Statistics, Importance of Statistical Inference in Machine Learning, Descriptive Statistics, Measures of Central Tendency: Mean, Median, Mode, Mid-range, Measures of Dispersion: Range, Variance, Mean Deviation, Standard Deviation.

Coefficient of variation: Moments, Skewness, Kurtosis, One sample hypothesis testing, hypothesis, Testing of Hypothesis, Binomial distribution and normal distribution, Chi-Square Tests, t-test, ANOVA. Pearson Correlation.

#Exemplar/Case Studies	For a payroll dataset create Measure of central tendency and its measure of dispersion for statistical analysis of given data.
*Mapping of Course Outcomes	CO2

Unit V	Statistical Inference II	(07 Hours)

Measure of Relationship: Covariance, Karl Pearson's Coefficient of Correlation, Measures of Position: Percentile, Z-score, Quartiles, Bayes' Theorem, Bayes Classifier, Bayesian network, **Probabilistic models with hidden variables**

Exemplar/Case Studies	Create a probabilistic model for credit card fraud detection
Mapping of Course Outcomes	CO3

Unit VI	Regression Model	(07 Hours)

Introduction, types of regression. Simple regression- Types, Making predictions, Cost function, Gradient descent, Training, Model evaluation.

Multivariable regression: Growing complexity, Normalization, making predictions, initialize weights, Cost function, Simplifying with matrices, Bias term, Model evaluation

Exemplar/Case Studies	Create a probabilistic model for credit card fraud detection
Mapping of Course Outcomes	CO5

Textbooks:

1. Practical Statistics for Data Scientists: 50+ Essential Concepts Using R and Python, Bruce, Peter, Andrew Bruce, and Peter Gedeck, O'Reilly Media, 2020.
2. Liu, Chung Laung. Elements of discrete mathematics. Tata McGraw-Hill Education, 1987.
3. Introduction to Statistics and Data Analysis With Exercises, Solutions and Applications in R Authors: Heumann, Christian, Schomaker, Michael, Shalabh, Publisher" Springer 2016

Reference Books:

1. Applied Statistics and Probability for Engineers, Douglas C. Montgomery, George C. Runger, 2018, Wiley (Low price edition available)
2. Introduction to. Mathematics. Statistics. Robert V. Hogg. Allen T. Craig, Low price Indian edition by Pearson Education
3. Probability and Statistics for Engineers. Richard A. Johnson, Irwin Miller, John Freund
4. Mathematical Statistics with Applications. Irwin Miller, Marylees Miller, Pearson Education
5. The R Software-Fundamentals of Programming and Statistical Analysis -Pierre Lafaye de Micheaux, Rémy Drouilhet, Benoit Liquet, Springer 2013

MOOC Courses:

- Essentials of Data Science With R Software - Probability and Statistical Inference by Prof. Shalabh, IIT Kanpur.
-

E-books:

Important links:

Savitribai Phule Pune University, Pune
ME Data Sciences (2020 Course)
510302: Basics of Data Science

Teaching Scheme:	Credit	Examination Scheme:
TH: 03 hr/week	03	Mid Semester: 50 Marks End Semester: 50 Marks

Prerequisite Courses: Database System

Companion Course: Mathematics for Data Science

Course Objectives:

1. To understand the recommendation system and two basic architectures for a recommendation system.
2. To develop the fundamental knowledge and understand concepts to become a data science professional.
3. To learn statistical methods and machine learning algorithms required for Data Science.
4. To visualize data and use for communicating stories from data.
5. To study different types of recommendation systems.
6. To learn algorithms for analyzing and mining the structure of network graphs.

Course Outcomes:

On completion of the course, learner will be able to -

CO1: Apply data science processes to an e-commerce data and demonstrate the use of estimation methods for analyzing this data.
CO2: Compare and apply appropriate machine learning algorithms for classification.
CO3: Compare and choose one data visualization method for effective visualization of data.
CO4: Design a model of recommendation system based on the content of the data.
CO5: Apply standard clustering methods to analyze social network graph.

Course Contents

Unit I	Introduction to Data Science	

What is Data Science, importance of data science, Big data and data Science, The current Scenario, Industry Perspective Types of Data: Structured vs. Unstructured Data, Quantitative vs. Categorical Data, Big Data vs. Little Data, Data science process, Role Data Scientist.

Case Studies (if any)	Ecommerce Marketplace
Mapping of Course Outcomes	**CO1**

Unit II	Statistical Interference and Exploratory Data Analysis	

Introduction-Population and samples, Data Preparation, Exploratory Data Analysis-Summarizing Data, Data Distribution, Outlier Treatment, Measuring Symmetry, Continuous Distribution, Kernel Density, Estimation: Sample and Estimated Mean, Variance and Standard Scores, Covariance, and Pearson's and Spearman's Rank Correlation.

Case Studies (if any)	Demonstrate the case study of real direct online real estate using R language
Mapping of Course Outcomes	**CO1**

Unit III	Machine Learning Algorithms	

Machine Learning Algorithms: Linear Regression, K-nearest Neighbors(k-NN), K-mean, Spam Filters, Naive Bayes, and Wrangling: Naive Bayes, Comparing Naive Bayes to k-NN,Scraping the Web: APIs and Other Tools

Case Studies (if any)	Article Classification using naïve bayes
Mapping of Course Outcomes	**CO2**

Unit IV	Data Visualization	

Data visualisation: Introduction, Types of data visualisation, Data for visualisation: Data types, Data encodings, Retinal variables, Mapping variables to encodings, Visual encodings

Case Studies(if any)	Data Visualization on any problem
Mapping of Course Outcomes	**CO3**

Unit V	Recommendation Systems	

A Model for Recommendation Systems: The Utility Matrix, The Long Tail, Applications of Recommendation Systems, Populating the Utility Matrix, Content-Based Recommendations: Item Profiles, Discovering Features of Documents, Obtaining Item Features From Tags, Representing Item Profiles, User Profiles, Recommending Items to Users Based on Content, Collaborative Filtering: Measuring Similarity, The Duality of Similarity, Clustering Users and Items, Evaluation of Recommendation System		
Case Studies(if any)	MovieLens Case Study	
Mapping of Course Outcomes	**CO4**	
Unit VI	**Social Network Analysis**	
Social Networks as Graphs, Varieties of Social Networks, Graphs With Several Node Types, Clustering of Social-Network Graphs: Distance Measures for Social-Network Graphs, Applying Standard Clustering Methods, Betweenness, The Girvan-Newman Algorithm, Using Betweenness to Find Communities		
Case Studies(if any)	Community detection in social network	
Mapping of Course Outcomes	**CO5**	

Books & Other Resources:

Text Books:

1. Cathy O'Neil and Rachel Schutt. Doing Data Science, Straight Talk From The Frontline. O'Reilly.
2. Jure Leskovek, Anand Rajaraman and Jeffrey Ullman. Mining of Massive Datasets. v2.1,Cambridge University Press.

Reference Books:

1. Laura Igual and Santi Segui, Introduction to Data Science: A Python Approach to Concepts, Techniques and Applications,Springer; 1st ed. 2017 edition

MOOC Courses

- Data science for engineers Course philosophy and expectation https://nptel.ac.in/courses/106/106/106106179/
- Introduction to Data Analytics

https://nptel.ac.in/courses/110/106/110106064/

E-books: http://infolab.stanford.edu/~ullman/mmds/book0n.pdf

Important links:

Savitribai Phule Pune University, Pune
ME Data Sciences (2020 Course)
510303: Big Data Analytics

Teaching Scheme:	Credit	Examination Scheme:
TH: 03 hr/week	03	Mid Semester: 50 Marks End Semester: 50 Marks

Prerequisite Courses: Introduction to Probability theory, Statistics, Python/R

Companion Course: Data Preparation and Analysis

Course Objectives:

1. To understand the big data concepts and big data analytics lifecycle
2. To understand the big data analytics algorithms and tools
3. To understand the importance of big data visualization tools and techniques
4. To get acquainted with advancements in tools and techniques used for big data analytics

Course Outcomes:

On completion of the course, learner will be able to–

CO1: Design the data analytics life cycle for selected problem statement

CO2: Develop insights into the big data and present results for selected problem statement through visualization techniques

CO3: Demonstrate the use of Hadoop and its ecosystem elements to analyze big data.

CO4: Demonstrate use of advanced FOSS computing environments for big health care data.

Course Contents

Unit I	Basics of Big Data	07

Big data: characteristics, types, sources, architectures, Data analysis process, Data analytics lifecycle, Pre-processing data, Market and Business Drivers for Big Data Analytics, Business Problems Suited to Big Data Analytics

Case Studies (if any)	Case study on data analytics lifecycle
Mapping of Course Outcomes	**CO1, CO2**

Unit II	Technologies for big data analytics	07

Distributed and Parallel Computing for Big Data, Cloud Computing and Big Data, In-Memory Computing Technology for Big Data, Introduction to Hadoop, HDFS, MapReduce, YARN, HBase, Combining HDFS and HBase

#Case Studies (if any)	Using MapReduce to scale algorithms for Big Data analytics
***Mapping of Course Outcomes**	**CO3, CO4**

Unit III	Hadoop ecosystem for big data analytics	07

Hadoop ecosystem: Sqoop, Impala, Apache Flume, Pig, Hive, Data transformation and analysis using Pig, Data analysis using Hive and Impala, Mahout, Oozie, Zookeeper etc.

Case Studies(if any)	Sentiment analysis
Mapping of Course Outcomes	**CO3, CO4**

Unit IV	Big data analytics with Apache Spark	07

Apache Spark, Spark core, Interactive data analysis with spark shell, Writing a spark application, Spark RDD Optimization Techniques, Spark Algorithm, Spark SQL

Case Studies(if any)	Big data for maintaining HER: healthcare records
Mapping of Course Outcomes	**CO3, CO4**

Unit V	Programming languages for Big data analytics	07

Big data analytics with PySpark: Python and Apache Spark, Big data analytics with RHadoop: R and Hadoop, Text mining in R, Hadoop and Hive, Scala and Spark for Big data analytics, Hadoop on cloud

Case Studies(if any)	Cloudera platform: western union bank
Mapping of Course Outcomes	**CO4**

Unit VI	Visualization techniques and tools for big	07

	data	
Visualizing Big Data, Importance of data visualization, Challenges, Need for advanced visualization techniques, Tools used in data visualization, Big Data Visualization with R/Python/Tableau/other tools		
Case Studies(if any)	Industrial Big Data Visualization: A Case Study Using Flight Data Recordings	
Mapping of Course Outcomes	**CO2**	

Books & Other Resources:

Text Books:

1. DT Editorial Services, "Big Data, Black Book: Covers Hadoop 2, MapReduce, Hive, YARN, Pig, R and Data Visualization"
2. David Dietrich, Barry Hiller, "Data Science and Big Data Analytics", EMC education services, Wiley publications, 2012, ISBN0-07-120413-X
3. Mohammed Guller, "Big Data Analytics with Spark: A Practitioner's Guide to Using Spark for Large Scale Data Analysis", ISBN-13:978-1484209653
4. David Loshin, "Big Data Analytics From Strategic Planning to Enterprise Integration with Tools, Techniques, NoSQL, and Graph", Morgan Kaufmann

Reference Books:

1. Venkat Ankam, "Big Data Analytics", Packt Publishing
2. Jenny Kim, Benjamin Bengfort, "Data Analytics with Hadoop", O'Reilly Media, Inc., ISBN: 9781491913734
3. Glenn J. Myatt, "Making Sense of Data: A Practical Guide to Exploratory Data Analysis and Data Mining"

MOOC Courses

- **https://nptel.ac.in/courses/111/104/111104120/**
- **https://nptel.ac.in/courses/110/106/110106064/**
- **https://nptel.ac.in/courses/106/104/106104189/**

E-books:

- Hector Cuesta and Dr Sampath Kumar, "Practical Data Analysis", 2nd Edition

Important links:

- **https://hadoop.apache.org/**
- **http://spark.apache.org/**

Savitribai Phule Pune University, Pune
ME Data Sciences (2020 Course)
510304: Research Methodology

Teaching Scheme:	Credit	Examination Scheme:
TH: 03 hr/week	03	Mid Semester: 50 Marks End Semester: 50 Marks

Prerequisite Courses: -

Companion Course:
Mathematical Foundation of Data Science,Laboratory Proficiency I

Course Objectives:
1. To inculcate the research thinking
2. To understand various facets of research
3. To develop oral and written presentation skills
4. To acquaint learner with the IPR and plagiarism issues

Course Outcomes:
On completion of the course, learner will be able to–
CO1: Understand the importance of research, and critically review a research work
CO2: Design methods for data collection and analysis of data
CO3: Design and test the hypothesis
CO4: Write a survey paper and original research article with required analysis reports
CO5: Identify and differentiate intellectual properties
CO6: Write a patent draft

Course Contents

Unit I	Introduction to Research Methodology	7 hrs

Meaning, objectives, types, eight step model of research process: formulation of research problem, research design, data collection, sampling, preparing a research proposal, data processing and report writing.

Case Studies (if any)	Write a research proposal for Online flight booking system.
Mapping of Course	CO1

Unit II	Data and Sampling	7 hrs

Methods for primary and secondary data collection, data processing operations, measures for analyzing data, basics of sampling theory, concept of standard error, estimations on populations, determining sample size.

Case Studies (if any)	Collect data for overbooking decision for demand and revenue management of flights.
Mapping of Course Outcomes	CO2

Unit III	Tests in research - I	7 hrs

Concept and procedure of hypothesis testing, Hypothesis Testing: - of Means, for Differences between Means, for Comparing Two Related Samples, of Proportions, for Difference between Proportions, for Comparing a Variance to Some Hypothesized Population Variance, Testing the Equality of Variances of Two Normal Populations, of Correlation Coefficients.

Case Studies (if any)	
Mapping of Course Outcomes	CO2

Unit IV	Tests in research-II	7 hrs

Analysis of variance and covariance, Important Nonparametric or Distribution-free Test, Important Multivariate Techniques, Important Methods of Factor Analysis

Case Studies(if any)	
Mapping of Course Outcomes	CO4

Unit V	Introduction to Intellectual Property Rights	7 hrs

Intellectual property, types of IP, IPR in India, IP India and CIPAM; Paris and other international Conventions

related to IP, TRIPS, WIPO and WIPO's Hague system, Madrid system, Budapest system, PCT, Lisbon system.		
Case Studies(if any)	IP related laws in India	
Mapping of Course	**CO3, CO5**	
Unit VI	**Technical Content writing**	**7 hrs**
Types of technical content publishing: Patent, copyright, white paper, journal, conference, poster, short paper etc. Draft patent rules, drafting a patent and copyright, patentability, steps in patenting: searching prior art, preparing the patent application, claims, filing, prosecution, objections, appeal, issuance or rejection of patent. Writing various technical papers: survey paper, journal paper, indexing agencies, COPE, plagiarism		
Case Studies (if any)	Patent act, 1970 and Patent Rules 1972 (with amendments)	
Mapping of Course	**CO6**	

Books & Other Resources:

Text Books:

1. "Research Methodology- a step-by-step guide for beginners", Ranjit Kumar, 3rd edition, SAGE Publication
2. "Research Methodology- Methods and Techniques", C. R. Kothari, 3rd edition, New Age International Publishers.

Reference Books:

1. "Research Methods for Engineers", David V. Thiel, Cambridge University Press.

MOOC Courses:

E-books:

Important links:

- WIPO : https://www.wipo.int/portal/en/index.html
- IP India: http://www.ipindia.nic.in/
- Cell For IPR Promotion and Management : http://cipam.gov.in/
- Draft patent rules: http://cipam.gov.in/wp-content/uploads/2018/12/Draft-Patent-Rules-2018.pdf
- Manual of Patent Office Practice and Procedure: http://www.ipindia.nic.in/writereaddata/Portal/Images/pdf/Manual_for_Patent_Office_Practice_and_Procedure_.pdf
- WIPO IPR Resources: https://www.wipo.int/reference/en/

Savitribai Phule Pune University, Pune
ME Data Sciences (2020 Course)
510305A: Data Storage Technologies and Networks

Teaching Scheme:	Credit	Examination Scheme:
TH: 04 hr/week	**04**	**Mid Semester: 50 Marks** **End Semester: 50 Marks**

Prerequisite Courses: - Basic knowledge of Computer Architecture, Operating Systems, and Computer Networking is required

Companion Course: -

Course Objectives:

1. To understand storage systems
2. To learn data storage technologies
3. To understand storage networking fundamentals
4. To learn storage networking technologies
5. To acquaint learner with knowledge of how to secure storage infrastructure

Course Outcomes:

On completion of the course, learner will be able to –
CO1: Describe storage system architecture, its elements, and characteristics.
CO2: Compare the intelligent storage systems and select one for a storage application.
CO3: Demonstrate storage virtualization using Xen or KVM
CO4: Demonstrate the functioning of SAN and NAS using open-source simulators.
CO5: Describe the mechanisms to secure storage infrastructure.

Course Contents

Unit I	Introduction to storage system	7 hrs

Introduction to Information Storage - Information Storage, Data, Types of Data, Big Data, Information, Storage, Evolution of Storage Architecture, Data Center Infrastructure- Core Elements of a Data Center, Key Characteristics of a Data Center, Managing a Data Center,Data Center Environment – Application, Database Management System (DBMS), Host (Compute), Connectivity, Storage, Disk Drive Components, Disk Drive Performance, Host Access to Data, Direct-Attached Storage, Storage Design Based on Application, Disk Native Command Queuing, Introduction to Flash Drives

Case Studies (if any)	
Mapping of Course	**CO1**

Unit II	Intelligent Storage Systems	7 hrs

RAID Implementation Methods, RAID Array Components, RAID Techniques, RAID Levels, RAID Impact on Disk Performance, RAID Comparison, Components of an Intelligent Storage System- Front end, Cache, Back End, Physical disk, Storage Provisioning- Traditional Storage Provisioning, Comparison between Virtual and Traditional Storage Provisioning, Types of Intelligent Storage Systems- High-End Storage Systems, Midrange Storage Systems

Case Studies(if any)	
Mapping of Course	**CO2**

Unit III	Virtualization	7 hrs

Server and Storage I/O Fundamentals- Server and I/O Architectures, Storage Hierarchy, From Bits to Bytes, Disk Storage Fundamentals, Initiators and Targets, How write and read from a Storage Device, Storage Sharing vs. Data Sharing
Different Types of Storage: Not All Data Storage, I/O Connectivity and Networking Fundamentals, IT

Clouds, Virtualization: Servers, Storage, and Networking, Virtualization and Storage Services, Data and Storage Access		
Case Studies(if any)		
Mapping of Course	**CO3**	
Unit IV	**Storage Networking Technologies – SAN, iSCSI**	**7 hrs**
Fibre Channel Storage Area Networks - Fibre Channel: Overview, The SAN and Its Evolution, Components of FC SAN, FC Connectivity, Switched Fabric Ports, Fibre Channel Architecture- Fibre Channel Protocol Stack, Fibre Channel Addressing, World Wide Names, FC Frame, Structure and Organization of FC Data, Flow Control, Classes of Service, Zoning, FC SAN Topologies IP SAN – iSCSI- Components of iSCSI, iSCSI Host Connectivity, iSCSI Topologies, iSCSI Protocol Stack, FCIP - FCIP Protocol Stack, FCoE - I/O Consolidation Using FCoE, Components of an FCoE Network		
Case Studies(if any)		
Mapping of Course	**CO4**	
Unit V	**Storage Networking Technologies - NAS, Object-Based and Unified storage**	**7 hrs**
Introduction to NAS, Benefits, File Systems and Network File Sharing- Accessing a File System, Network File Sharing, Components of NAS- , NAS I/O Operation, NAS Implementations- Unified NAS, Unified NAS Connectivity, Gateway NAS, Gateway NAS Connectivity, Scale-Out NAS, Scale-Out NAS Connectivity, NAS File-Sharing Protocols – NFS, CIFS Object-Based and Unified Storage – Object-Based Storage Devices - Object-Based Storage Architecture, Components of OSD, Object Storage and Retrieval in OSD, Benefits of Object-Based Storage, Common Use Cases for Object-Based Storage, Content-Addressed Storage		
Case Studies(if any)		
Mapping of Course	**CO4**	
Unit VI	**Securing the Storage Infrastructure**	**7 hrs**
Information Security Framework, Risk Triad, Storage Security Domains – Securing application access domain, securing management access domain, Security Implementations in Storage Networking- FC SAN, NAS, IP Scan, Securing Storage Infrastructure in Virtualized and Cloud Environments – Security concerns, Security measures		
Case Studies (if any)	**-**	
Mapping of Course	**CO5**	
Books & Other Resources:		
Textbooks: 1. "Information storage and management", EMC Education Services, 2nd edition, SAGE Publication **2.** "Cloud and Virtual Data Storage Networking", Greg Schulz, CRC Press		
Reference Books: 1. "Storage Networks: The Complete Reference, Robert Spalding", Publisher: McGraw-Hill Osborne Media ISBN: 0072224762, 9780072224764 **2.** "Storage area network essentials", Richard Barker, Paul Massiglia, Wiley		
MOOC Courses • **https://swayam.gov.in/nd1_noc19_cs64/preview, Cloud Computing By Prof. Soumya Kanti Ghosh | IIT Kharagpur**		
E-books : • **Data Storage Technologies Kindle Edition by K.L. JAMES**		

Savitribai Phule Pune University, Pune
ME Data Sciences (2020 Course)
510305B: Information Systems Management

Teaching Scheme:	Credit	Examination Scheme:
TH: 04 hr /week	04	Mid Semester: 50 Marks End Semester: 50 Marks

Prerequisite Courses: Information Systems and Engineering Economics

Companion Course: NIL

Course Objectives:

1. To prepare the students for various forms of the Information Systems and its application in organizations.
2. To Prepare engineering students to do economic analyses in the decision making process to justify or reject alternatives / projects on an economic basis for an organization.
3. To learn the skills to make the best use of Business Intelligence
4. To learn the skills in building advanced Information Systems

Course Outcomes:

On completion of the course, learner will be able to–

CO1: Understand the activities that are undertaken while managing, designing, planning, implementation, and deployment of computerized information systems in an organization.
CO2: Perform and evaluate present worth, future worth and annual worth analyses on one of more economic alternatives.
CO3: Evaluate the decisions using What-If Analysis, Sensitivity analysis, Goal-seeking analysis, Optimization analysis techniques of DSS
CO4: Plan to implement a Business Intelligence Solution

Course Contents

Module I	Management Information System (MIS)	06 Hours

Managing Information Systems, Ethical and Social Issues, Information Technology Infrastructure and Choices, Information Systems Security and Control, Managing Data Resources, Business Process Integration and Enterprise Systems, ICT for Development and E-Governance.

Case Studies (if any)	In-house or cloud based ERP implementation, UIDAI Unique Identification Authority of India.
Mapping of Course Outcomes	CO1

Module II	Business Intelligence	07 Hours

Business Intelligence an Introduction: Introduction, Definition, History and Evolution, Difference between Information and Intelligence, Factors of Business Intelligence System - Business Intelligence Architecture, Real time Business Intelligence, Business Intelligence Applications, Business Intelligence Essentials: Introduction, Creating Business Intelligence Environment, Business Intelligence Landscape, Types of Business Intelligence, Business Intelligence Platform, Dynamic roles in Business Intelligence, Roles of Business Intelligence in Modern Business- Challenges of BI

Business Intelligence User Model: Introduction, Evolution of Business Intelligence, Business Intelligence Opportunity Analysis Overview, Content Management System, End User Segmentation, Basic Reporting and Querying, Online Analytical Processing, OLAP Techniques, OLAP Applications, Applying the OLAP to Data Warehousing, Benefits of using OLAP, Dashboard, Advanced/Emerging BI Technologies, Future

of Business Intelligence		
Case Studies(if any)	NIL	
Mapping of Course Outcomes	CO4	
Module III	**Building Advanced Information Systems**	**07 Hours**
Decision Support in Business, Decision Support Trends, Decision Support Systems, Management, Information Systems, Online Analytical Processing, Using Decision Support Systems, Executive Information Systems, Enterprise Portals and Decision Support, Knowledge Management Systems		
Case Studies(if any)	Real World Case: Hillman Group, Avnet, and Quaker Chemical: Process Transformation through Business Intelligence Deployments	
Mapping of Course Outcomes	CO3	
Module IV	**Economics and Management**	**07 Hours**
Engineering Economic Decisions, Time Value of Money, Understanding Money Management, Equivalence Calculations under Inflation, Present-Worth Analysis, Annual-Equivalence Analysis.		
Case Studies(if any)	Economic decisions done in Multi-national companies and comparative analysis of software enterprises from similar domains.	
Mapping of Course Outcomes	CO2	
Module V	**Applications of Business Intelligence**	**07 Hours**
Business Intelligence Strategy and Road Map: Introduction, Planning to implement a Business Intelligence Solution, Understand Limitations of Business Intelligence, Business Intelligence Usage, How to make the best use of Business Intelligence?, Implementing Business Intelligence: Implementation Strategy , Fundamental decisions,Business Intelligence Case Studies: Improving Operational Efficiency –Audi AG, Maximizing Profitability- The Frank Russell Company		
Case Studies(if any)	BI and Data mining Applications: ERP and BI, BI applications in CRM,BI in Marketing, Logistics and Productions Finance, Banking ,Telecommunications and fraud detection.	
Mapping of Course Outcomes	CO4	
Module VI	**Managing Information Systems Projects**	**06 Hours**
The importance of project management, Selecting projects, Establishing the business value of Information Systems, Managing project risk		
Case Studies(if any)	Hands on mini projects : Management Decision Problems, Improving Decision Making: Using Spreadsheet Software for Capital Budgeting for a New CAD System, Improving Decision Making: Using Web Tools for Buying and Financing a Home	
Mapping of Course Outcomes	CO1	

Books & Other Resources:

Text Books:

1. Rahul De, ―MIS: Management Information Systems in Business, Government and Society‖, Wiley India, ISBN: 13: 978-81-265-2019-0.
2. Chan S. Park , "Fundamentals of Engineering Economics‖, 3rd Edition, Pearson Education, ISBN 13: 978-02-737-7291-0
3. Kenneth C. Laudon, Jane P. Laudon, “Management Information Systems
4. MANAGING THE DIGITAL FIRM”, 12th Edition, Prentice Hall
5. James A. O’Brien, George M. Marakas, “INTRODUCTION TO INFORMATION SYSTEMS”, 15th Edition, McGraw-Hill

Reference Books:

1. William G. Sullivan, Elin M. Wicks, C. Patrick Koelling, Engineering Economy, Pearson Education, ISBN13: 978-01-334-3927-4

MOOC Couses: "Information Systems Specialization", offered by University of Minnesota

- https://www.coursera.org/specializations/information-systems
  "Enterprise Systems**"** by Jason Chan, Associate Professor, affiliated to University of Minnesota
- https://www.coursera.org/learn/enterprise-systems

  "It Infrastructure and Emerging Trends" by Soumya Sen, Associate Professor, affiliated to ***University of Minnesota***
- https://www.coursera.org/learn/it-infrastructure-and-emerging-trends

  "Analysis for business systems" by Ken Reily, Associate Professor, affiliated to ***University of Minnesota***
- https://www.coursera.org/learn/analysis-for-business-systems

"IS/IT Governance" by Gautam Ray, Associate Professor, affiliated to ***University of Minnesota***

- https://www.coursera.org/learn/is-it-governance

Books:

1. Business Intelligence Roadmap: The Complete Project Lifecycle For Decision-Support Applications by Larissa T. Moss & Shaku Atre
2. Data Strategy: How To Profit From A World Of Big Data, Analytics And The Internet Of Things by Bernard Marr
3. Business-Intelligence-by-Michael-Luckevich-Elizabeth-Vitt-Stacia-Misner- Elizabeth-Vitt - Michael-Luc
4. Definitive Guide to DAX, The: Business intelligence for Microsoft Power BI, SQL Server Analysis Services, and Excel, 2nd Edition
5. Oracle Business Intelligence with Machine Learning : Artificial Intelligence Techniques in OBIEE for Actionable BIBy Rosendo Abelleraand Lakshman Bulusu
6. Business Intelligence Guidebook by Rick Sherman Released November 2014 Publisher(s): Morgan Kaufmann ISBN: 9780124115286
7. Business Intelligence Strategy and Big Data Analytics by Steve Williams Released April 2016 Publisher(s): Morgan Kaufmann ISBN: 9780128094891

Important links:

- www.managementstudyguide.com
- https://www.coursera.org/specializations/information-systems

Savitribai Phule Pune University, Pune
ME Data Sciences (2020 Course)
510305C: Data Preparation and Analysis

Teaching Scheme:	Credit	Examination Scheme:
TH: 04 hr/week	04	Mid Semester: 50 Marks End Semester: 50 Marks

Prerequisite Courses:
Introduction to Probability theory, statistics, Python/R

Companion Course: Lab Practice I

Course Objectives:

1. To understand the importance of data and data preprocessing
2. To understand data cleaning and conditioning
3. To understand an ETL – Extract, Transform and Load – process and ETL tools
4. To get acquainted with data visualization techniques for exploratory analysis

Course Outcomes:

On completion of the course, learner will be able to–
CO1: Apply ETL process with ETL tools to datasets for data processing.
CO2: Prepare conditioned and preprocessed datasets using normalization method for data
CO3: Draw insights into the datasets using exploratory mechanism.
CO4: Demonstrate use of visualization tools for data preparation and analysis

Selection of Modules: Modules 1 to 3 are compulsory and select any one from modules 4, 5 and 6.

Course Contents

Module I	Data Gathering and Data Discovery	08
Identifying potential data sources, Gathering data, Data discovery- understanding the data, assessing data, data formats, Parsing, Selecting features, Transformation, Scalability and real-time issues		
Case Studies (if any)		
Mapping of Course Outcomes	**CO1**	
Module II	**Cleaning and Conditioning Data**	**08**
Data Preparation Basic Models: Data Integration, Data Cleaning, Data Normalization, Min-Max Normalization, Z-score Normalization, Decimal Scaling Normalization, Consistency checking, Heterogeneous and missing data, Dealing with missing values, Duplicate values, Noise, Inconsistent data, Outliers		
Case Studies(if any)		
Mapping of Course Outcomes	**CO2**	
Module III	**ETLT**	**08**
Transform and enrich data: Data Transformation, Linear Transformations, Quadratic Transformations, Non-polynomial Approximations of Transformations, Polynomial Approximations of Transformations, Rank Transformations, Box-Cox Transformations, Spreading the Histogram, Nominal to Binary Transformation, Transformations via Data Reduction, ETL tools		
Case Studies(if any)		
Mapping of Course Outcomes	**CO1**	
Module IV	**Exploratory Analysis**	**08**

Formulating Hypothesis, Data Terminology, Data Exploration, Data Exploration through Summary Statistics, Data Exploration through Plots, Feature Engineering, Feature selection, Feature transformation, Dimensionality reduction		
Case Studies(if any)		
Mapping of Course Outcomes	**CO3**	
Module V	**Data Visualization**	**08**
Visualization techniques, Different types of plots, Designing visualizations, Time series, Geolocated data, Correlations and connections, Hierarchies and networks, Interactivity		
Case Studies(if any)		
Mapping of Course Outcomes	**CO3, CO4**	
Module VI	**Advanced Tools for Data Preparation**	**08**
Web scraping, Data from social networks, Open-source tools for data preparation: Open Refine, R/Python libraries for data preparation and visualization		
Case Studies(if any)		
Mapping of Course Outcomes	**CO3, CO4**	

Books & Other Resources:

Text Books:

1. Glenn J. Myatt, "Making Sense of Data I: A Practical Guide to Exploratory Data Analysis and Data Mining"
2. Salvador García, Julián Luengo, Francisco Herrera, "Data Preprocessing in Data Mining"

Reference Books:

1. Mark Gardner, "Beginning R: The Statistical Programming Language", Wrox Publication, ISBN: 978-1-118-16430-3
2. David Dietrich, Barry Hiller, "Data Science and Big Data Analytics", EMC education services, Wiley publications, 2012, ISBN0-07-120413-X
3. Ruben Verborgh; Max De Wilde, "Using OpenRefine : the essential OpenRefine guide that takes you from data analysis and error fixing to linking your dataset to the Web"

MOOC Courses:

- **NPTEL course: Python for Data Science : https://swayam.gov.in/nd1_noc19_cs59/preview**

E-books:

- Jacqueline Kazil, Katharine Jarmu, "Data Wrangling with Python: Tips and Tools to Make Your Life Easier"
- Hector Cuesta and Dr Sampath Kumar, "Practical Data Analysis", 2nd Edition

Important links:

- https://openrefine.org/
- https://www.youtube.com/playlist?list=PLh2mXjKcTPSACrQxPM2_1Ojus5HX88ht7

Savitribai Phule Pune University, Pune
ME Data Sciences (2020 Course)

510305D: Artificial Intelligence for Data Science

Teaching Scheme:	Credit	Examination Scheme:
TH: 03 hr/week	03	Mid Semester: 50 Marks End Semester: 50 Marks

Prerequisite Courses: Data structure, Algorithms

Companion Course: Machine Learning, Soft Computing, Deep learning

Course Objectives:

Aim of the course is to introduce to the field of Artificial Intelligence (AI) with emphasis on its use to solve real world problems for which solutions are difficult to express using the traditional algorithmic approach. It explores the essential theory behind methodologies for developing systems that demonstrate intelligent behavior including dealing with uncertainty, learning from experience and following problem solving strategies found in nature.

1. To introduce the concepts of Artificial intelligence and methods
2. To provide the knowledge representation and Learning techniques to problem solving strategy
3. To design and solve real world problems using AI approaches
4. To implement AI techniques in different fields

Course Outcomes:

completion of the course, the learner will be able to–Able to Demonstrate knowledge of the fundamental principles of Artificial intelligent systems and would be able to analyze and compare the relative merits of a variety of AI problem solving techniques.

CO1: Identify the need of Intelligent agents in problem solving
CO2: Compare and analyze different search techniques applied for problem solving
CO3: Apply the knowledge representation method and reasoning for given decision problem
CO4: Design and analyze a learning technique for a given system in different AI application domains like marketing, healthcare, banking, finance, education.

Selection of Modules: Modules 1 to 3 are compulsory and select any one from modules 4, 5 and 6.

Unit I	**Introduction and Intelligent Agents**	**8 hrs**
Introduction: What is AI? Foundations History of Artificial Intelligence, The State of the Art Intelligent Agents: Agents and Environments, Good Behavior: The Concept of Rationality, The Nature of Environments, and The Structure of Agents.		
Case Studies (if any)	Intelligent agents in autonomous systems	
Mapping of Course Outcomes	**CO1**	
Unit II	**Problem-solving**	**8 hrs**
Solving Problems by Searching: Problem-Solving Agents, Uninformed Search Strategies, Informed (Heuristic) Search Strategies, Heuristic Functions, Beyond Classical Search Local Search Algorithms and Optimization Problems, Local Search in Continuous Spaces, Searching with Nondeterministic Actions, Searching with Partial Observations, Online Search Agents and Unknown Environments		
Case Studies(if any)	Search techniques for a sliding tile problem	
Mapping of Course Outcomes	**CO2**	
Unit III	**Knowledge, reasoning, and planning**	**9 hrs**

Knowledge based Agents, First-Order Logic and Its Inference, Classical Planning, Planning and Acting in the Real World, Knowledge Representation		
Case Studies(if any)		
Mapping of Course Outcomes	**CO3**	
Unit IV	**Uncertain knowledge and reasoning**	**9 hrs**
Quantifying Uncertainty, Probabilistic Reasoning, Probabilistic Reasoning over Time, Making Simple Decisions, Making Complex Decisions		
Case Studies(if any)	Application of planning to a production system	
Mapping of Course Outcomes for Unit IV	**CO3**	
Unit V	**Learning**	**9 hrs**
Learning from Examples, Knowledge in Learning, Learning Probabilistic Models, Reinforcement Learning		
Case Studies(if any)	E mail filtering with learning method	
Mapping of Course Outcomes	**CO4**	
Unit VI	**Applications with case studies**	**9 hrs**
AI Applications in various fields in marketing, healthcare, banking, finance, etc. Case Studies: Credit card Fraud Analysis, Sentiment Analysis, Recommendation Systems and Collaborative filtering, Uber Alternative Routing		
Case Studies (if any)	Application of AI and Machine Learning in e commerce	
Mapping of Course Outcomes	**CO4**	

Books & Other Resources:

Text Books:

1. Russell S. and Norvig P. (2009). Artificial Intelligence: A Modern Approach. Prentice-Hall, 3rd edition.
2. Elaine Rich, Kevin Knight and Nair, “Artificial Intelligence”, TMH,ISBN-978-0-07-008770-5

Reference Books:

1. Luger G.F. and Stubblefield W.A. (2008). Artificial Intelligence: Structures and strategies for Complex Problem Solving. Addison Wesley, 6th edition
2. Nilsson Nils J , “Artificial Intelligence: A new Synthesis, Morgan Kaufmann Publishers Inc. San Francisco, CA, ISBN: 978-1-55-860467-4
3. Patrick Henry Winston, “Artificial Intelligence”, Addison-Wesley Publishing Company, ISBN: 0-201-53377-4

MOOC Courses

- https://nptel.ac.in/courses/106/102/106102220/
- https://nptel.ac.in/courses/106/105/106105077/
- https://nptel.ac.in/courses/106/106/106106140/

E-books

- https://www.amazon.in/Artificial-Intelligence-As-AICTE-Intelligent/dp/8126579943

Important links:

- https://ieeexplore.ieee.org/
- https://en.wikipedia.org/wiki/Artificial_intelligence
- https://indiaai.in/

Savitribai Phule Pune University, Pune
ME Computer Engineering Data Science (2020 Course)
510306: Laboratory Proficiency-I

Teaching Scheme:	Credit	Examination Scheme:
PR: 08 hr/week	04	Term Work: 50 Marks Practical/Oral: 50 Marks

Prerequisite Courses: Knowledge of programming languages, Basics of Python/R

Companion Courses:
510301-Mathematical Foundations for Data Science
510302 – Basics of Data Science
510303 – Big Data Analytics
510304 – Research Methodology
510305 – Elective – I

All assignments are compulsory. Each student should implement the assignment individually. Laboratory teachers should make sure that the dataset/code/writeup is not the same.

510301-Mathematical Foundations for Data Science

Assignment 1

Choose a dataset from UCI Machine Learning repository (e.g. Cleveland).

a)Compute and display summary statistics for each feature available in the dataset. (eg. minimum, maximum, mean, range, standard deviation, variance and percentiles). Use a bar-graph to demonstrate your results.

b)Data Visualization-Create a histogram for each feature in the dataset to illustrate the feature distributions. Plot each histogram.

c)Create a boxplot for each feature in the dataset. All of the boxplots should be combined into a single plot. Compare distributions and identify outliers.

Assignment 2

a)Take any dataset from UCI repository (like air quality dataset) and perform regression analysis on it. Demonstrate your results using appropriate visualization techniques for numerical and categorical features (e.g. histogram, scatter plot, heat map, box plot).

b)Compute Eigen values and Eigen vectors for dataset in part a.

Useful links:

1. https://archive.ics.uci.edu/ml/datasets/heart+disease
2.https://archive.ics.uci.edu/ml/datasets/breast+cancer+wisconsin+(original)
3. https://archive.ics.uci.edu/ml/datasets/Air+Quality

510302 – Basics of DataScience

Assignment 1.

Implement Naive Bayes algorithm, using Java/Python/R to classify a dataset from UCI repository. (Do not use built-in functions for naive bayes). Compare the performance of your implementation with the Naive Bayes classifier from the Weka tool/R/Python. Present the Confusion matrix for each classifier. For measuring performance use at least five metrics such as accuracy, precision, recall, F-measure etc.

Assignment 2.

Take a sample dataset (The lab teacher may provide it). Plot the data using appropriate graphs (e.g. scatter diagram). Perform normality and symmetry tests on it using at least one graph method and at least one statistical test. Analyse the results. Then evaluate Spearman's Rank Correlation for this data.

510303 – Big Data Analytics

Assignment 1.

Demonstrate application of Apache spark to analyse streaming data from social media. (Installation of multi-node Hadoop as well as Spark is to be done by students.)

Assignment 2.

Take any text or image dataset (e.g. Stanford Sentiment Treebank, Sentiment140, Amazon Product data) and perform analysis on it.
Useful links:
1. https://nlp.stanford.edu
2. http://cs.stanford.edu/people/alecmgo/trainingandtestdata.zip
3. https://www.kaggle.com/lakshmi25npathi/imdb-dataset-of-50k-movie-reviews

4. https://snap.stanford.edu/data/amazon/productGraph/

510304 – Research Methodology

Assignment 1.

Select appropriate research topics in consultation with the lab teacher. Prepare a research proposal for the same. Follow the standard format for preparation of research proposals.

Assignment 2.

Prepare a patent application for the system mentioned above.

510305 – Elective – I

510305A: Data Storage Technologies and Networks

Mini-project: Build Cloud storage service system based on open source tools. Design and develop applications to upload and download the data of different types (block, object, file).

510305B: Information Systems Management

Mini-project: Design an ERP system for college using appropriate Information Systems Management concepts.

510305C: Data Preparation and Analysis

Mini-project:

a) Use ETL tools/R/Python for applying various transformations on free datasets available
b) Use Open Refine to preprocess raw data from websites

510305D: Artificial Intelligence for Data Science

Mini-project: Develop an application with prediction using learning techniques in AI.

Savitribai Phule Pune University, Pune
ME Data Sciences (2020 Course)
510308: Data Warehousing and Mining

Teaching Scheme:	Credit	Examination Scheme:
TH: 03 hr/week	**03**	**Mid Semester: 50 Marks** **End Semester: 50 Marks**

Prerequisite Courses: Database Management Systems, Data Mining

Companion Course:

Course Objectives:

1. Understand the necessity of Data Warehousing and its continuous growth.
2. Understand Planning and Management of Data Warehouse.
3. Understand issues in various Architectural types of Data warehouse.
4. Understand the application of various models of Data Warehouse.
5. Understand the web-enabled data warehouse and role of data mining

Course Outcomes:
On completion of the course, learner will be able to
CO1: Decide the type of Data warehouse to build.
CO2: Perform Requirement gathering and Design suitable architecture for Data warehouse project.
CO3: Design and prepare data for Data warehouse using ETL tools
CO4: Build web-enabled data warehouse
CO5: Analyze and Apply Data Mining techniques on real life applications
CO6: Demonstrate phases in data warehouse development life cycle with Data warehouse project.

Unit I	Overview and Concept	

Need for Data warehousing: Escalating need for strategic information, failure of past system, operational versus decision support systems, data warehouse defined, data warehouse movement, Evolution of Business Intelligence.
Data ware housing building blocks: Data ware house and Data Marts, architecture, components, metadata
Trends in warehousing: Continue growth in data warehousing, significant trends

Case Studies (if any)	Discuss design of data warehouse/ Data Mart for suitable system
Mapping of Course Outcomes	**CO1**

Unit II	Architecture and Infrastructure	

Principles of dimension modelling: Dimensional modelling basics, star schema, advantages, examples. Advanced dimension modelling: updates to dimensional tables, miscellaneous dimensions, snowflake schema, Aggregate Fact tables, families of Stars.
Data Extraction, Transformation, and Loading: ETL overview, requirement, Data extraction, Data Transformation, Data Loading, other integration approaches

Case Studies(if any)	Discuss different architectural details using suitable application
Mapping of Course Outcomes	**CO2**

Unit III	Design and Data Preparation	

Principles of dimension modeling: Dimensional modeling basics, star schema, advantages, examples.
Advanced dimension modeling: updates to dimensional tables, miscellaneous dimensions, snowflake schema, Aggregate Fact tables, families of Stars.

Data Extraction, Transformation, and Loading: ETL overview, requirement, Data extraction, Data Transformation, Data Loading, other integration approaches		
Case Studies(if any)	Demonstration of Pentaho , Apache Kafka	
Mapping of Course Outcomes	**CO3**	
Unit IV	**Information Access and delivery**	
Users of Information: Information from the data ware house, who will use information, information delivery, delivery tools OnLine Analytical Processing in the data warehouse: Demand for OLAP, major features and functions, OLAP models, OLAP implementation considerations Data ware house and the web: Web enables data warehouse, web-based information delivery, OLAP and the WEB, building a web-enabled data warehouse		
Case Studies(if any)	Discovering web access patterns and trends by applying OLAP	
Mapping of Course Outcomes	**CO4**	
Unit V	**Data Mining**	
Why data mining, what is Data Mining, Data mining as a process of Knowledge Discovery ,Major issues in Data Mining, Mining Frequent Pattern, Classification : basic concept and methods, Clustering : basic concept and methods, Data Mining Trends and Research Frontiers : Mining Complex Data Types, other Methodologies of Data Mining, Data Mining Applications, Data Mining trends.		
Case Studies(if any)	Discovering web access patterns and trends by Data Mining Technology on web logs	
Mapping of Course Outcomes	**CO5**	
Unit VI	**Implementation and Maintenance**	
Physical design process: Physical design steps, considerations, physical storage, indexing, performance enhancement techniques. Data warehouse development: Data warehouse testing, major deployment activity, security, backup and recovery. Growth and Maintenance: Monitoring the data warehouse, user training and support, managing the data warehouse.		
Case Studies(if any)	Discuss test cases for data warehouse applications	
Mapping of Course Outcomes	**CO6**	

Books & Other Resources:

Text Books:

1. Data Warehousing Fundamentals – PaulrajPonnaiah Wiley student Edition
2. Data Mining – Concepts and Techniques - Jiawei Han & Micheline Kamber, Morgan Kaufmann Publishers, Elsevier,2nd Edition, 2006.
3. Data Warehousing in the Real World – Sam Aanhory& Dennis Murray Pearson Edition Asia.

Reference Books:

1. Matthew A. Russell, "Mining the Social Web: Data Mining Facebook, Twitter, LinkedIn, Google+, GitHub, and More" , Shroff Publishers, 2nd Edition, ISBN: 9780596006068
2. G.K. Gupta, “Introduction to Data Miing with Case Studies,” PHI Learning Private Limited, 2nd Edition, ISBN 978-81-203-4326-9.

MOOC Courses :

1. Coursera 4 weeks course on “ Data Warehouse Concepts, Design, and Data Integration” by University of Colorado System Instructor Name : Michael Manninohttps://www.coursera.org/learn/dwdesign
2. Coursera 4 weeks course on “ Relational Database Support for Data Warehouses”, offered by

University of Colorado System Instructor Name : Michael Manninohttps://www.coursera.org/learn/dwrelational

3. Coursera 4 weeks course on " Business Intelligence Concepts, Tools, and Applications" offered by University of Colorado System Instructor Name : Prof. Jahangir Kairimi https://www.coursera.org/learn/business-intelligence-tools
4. NPTEL course on "Data Mining"offered by IIT KharagpurInstructor Name: Prof.PabitraMitrahttps://nptel.ac.in/courses/106105174/

E-books

Important links:

Savitribai Phule Pune University, Pune
ME Data Sciences (2020 Course)
510309: Machine Learning

Teaching Scheme:	Credit	Examination Scheme:
TH: 03 hr/week	**03**	**Mid Semester: 50 Marks** **End Semester: 50 Marks**

Prerequisite Courses: Big Data Analytics

Companion Course: Deep Learning

Course Objectives:

1. To understand Human learning aspects
2. To learn the primitives in learning process by computer
3. To Understand nature of problems solved with Machine Learning
4. To acquaint with the basic concepts and techniques of Machine Learning.
5. To learn the means for categorization of the information

Course Outcomes:

On completion of the course, learner will be able to

CO1:Acquire fundamental knowledge of learning theory

CO2:Design and evaluate various machine learning algorithms

CO3:Use machine learning methods for multivariate data analysis in various scientific fields

CO4: Choose and apply appropriate Machine Learning Techniques for analysis, forecasting, categorization and clustering of the data

Unit I	Machine Learning Concepts	09 Hours

Introduction to Machine Learning, Machine Learning applications, Types of learning: Supervised, Unsupervised and semi-supervised, reinforcement learning techniques, Models of Machine learning: Geometric model, Probabilistic Models, Logical Models, Grouping and grading models, Parametric and non-parametric models, Predictive and descriptive learning, Classification concepts, Binary and multi-class classification.

Mapping of Course Outcomes	CO1-Acquire fundamental knowledge of learning theory

Unit II	Learning Theory	09 Hours

Features: Feature Extraction, Feature Construction and Transformation, Feature Selection, Dimensionality Reduction: Subset selection, the Curse of dimensionality, Principle Components analysis, Independent Component analysis, Factor analysis, Multidimensional scaling, Linear discriminant analysis, Bias/Variance tradeoff, Union and chernoff/ Hoeffding bounds, VC dimension, Probably Approximately Correct (PAC) learning, Concept learning, the hypothesis space, Least general generalization, Internal disjunction, Paths through the hypothesis space, model Evaluation and selection

Mapping of Course Outcomes	CO1-Acquire fundamental knowledge of learning theory

Unit III	Geometric Models	09 Hours

Regression, Logistic regression , Assessing performance of regression - Error measures, Overfitting, Least square method, Multivariate Linear regression, Regression for Classification, Perceptron, Muli-layer perceptron, Simple neural network, Kernel based methods, Support vector machines(SVM), Soft margin SVM, Support Vector Machines as a linear and non-linear classifier, Limitations of SVM, Concept of Relevance Vector, K-nearest neighbor algorithm

Mapping of Course Outcomes	CO2- Design and evaluate various machine learning algorithms	
Unit IV	**Logical, Grouping And Grading Models**	**09 Hours**
Decision Tree Representation, Alternative measures for selecting attributes, Decision tree algorithm: ID3, Minimum Description length decision trees, Ranking and probability estimation trees, Regression trees, Clustering trees, Rule learning for subgroup discovery, Association rule mining, Distance based clustering-K-means algorithm, Choosing number of clusters, Clustering around medoids – silhouettes, Hierarchical clustering, Ensemble methods: Bagging and Boosting		
Mapping of Course Outcomes	CO2-Design and evaluate various machine learning algorithms	
Unit V	**Probabilistic Models**	**09 Hours**
Uncertainty, Normal distribution and its geometric interpretations, Baye's theorem, Naïve Bayes Classifier, Bayesian network, Discriminative learning with maximum likelihood, Probabilistic models with hidden variables, Hidden Markov model, Expectation Maximization methods, Gaussian Mixtures and compression based models		
Mapping of Course Outcomes	CO2-Design and evaluate various machine learning algorithms	
Unit VI	**Case Studies on Advanced Machine Learning Techniques**	**09 Hours**
Diagnosis of human disease, Diagnosis of crop disease, Text mining tasks like semantic analysis, author profiling, author identification, language identification, summarization etc., Prediction & forecasting, Fraud detection, Learning to rate vulnerabilities and predict exploits		
Mapping of Course Outcomes	1) CO3-Use machine learning methods for multivariate data analysis in various scientific fields 2) CO4-Choose and apply appropriate Machine Learning Techniques for analysis, forecasting, categorization and clustering of the data	

Books & Other Resources:

Text Books:

1. Peter Flach, Machine Learning: The Art and Science of Algorithms that make sense of data, Cambridge University Press, 1st Edition, 2012, ISBN No.: 978-1-316-50611-0
2. Ethem Alpaydin, Introduction to Machine Learning, PHI, 2nd edition, 2013, 978-0-262-01243-0
3. Kevin Murphy, Machine Learning: a Probabilistic Approach, MIT Press, 1st Edition, 2012, ISBN No.: 978-0262-30616-4

Reference Books:

1. C.M. Bishop, Pattern Recognition and Machine learning, Springer, 1st Edition, 2013, ISBN No.: 978-81-322-0906-5
2. Hastie, Tibshirani, Friedman, Introduction to statistical machine learning with applications in R, Springer, 2nd Edition, 2013, ISBN No.: 978-1-4614-7138-7
3. Tom Mitchell, Machine Learning, McGraw Hill, 1997, 0-07-042807-7
4. Parag Kulkarni, Reinforcement and Systemic Machine learning for Decision Making, Wiley-IEEE Press, 2012, 978-0-470-91999-6
5. M. F. Der, L. K. Saul, S. Savage, and G. M. Voelker (2014). Knock it off: profiling the online storefronts of counterfeit merchandise. In Proceedings of the Twentieth ACM Conference on Knowledge Discovery and Data Mining (KDD-14), pages 1759-1768. New York, NY.

6. J. T. Ma, L. K. Saul, S. Savage, and G. M. Voelker (2011). Learning to detect malicious URLs. ACM Transactions on Intelligent Systems and Technology 2(3), pages 30:1-24.
7. D.-K. Kim, G. M. Voelker, and L. K. Saul (2013). A variational approximation for topic modeling of hierarchical corpora. To appear in Proceedings of the 30th International Conference on Machine Learning (ICML-13). Atlanta, GA.
8. M. Bozorgi, L. K. Saul, S. Savage, and G. M. Voelker (2010). Beyond heuristics: learning to classify vulnerabilities and predict exploits. In Proceedings of the Sixteenth ACM Conference on Knowledge Discovery and Data Mining (KDD-10), pages 105-113. Washington, DC

MOOC Courses :

- Introduction to Machine Learning, By prof. Balaraman Ravindran

Savitribai Phule Pune University, Pune
ME Data Sciences (2020 Course)
510310: Soft Computing

Teaching Scheme:	Credit	Examination Scheme:
TH: 03 hrs/week	03	Mid Semester: 50 Marks End Semester: 50 Marks

Prerequisite Courses: Mathematics, Algorithms and analysis, Programming Language Python/ Java/C++

Companion Course: Machine Learning, Data Analytics, AI

Course Objectives:

1. To provide effective and efficient problem solving with soft computing methodologies
2. To develop Intelligent systems with soft computing
3. To give data analysis solutions with soft computing techniques
4. To implement soft computing solutions for real world problems

Course Outcomes:

On completion of the course, learner will be able to

CO1: Identify the components of soft computing and compare soft computing techniques.
CO2:Design a fuzzy inference system for a given system with set of fuzzy rules
CO3:Apply genetic algorithm for solution of an optimization function
CO4:Design a neural network solution for a classification problem
CO5:Identify , analyze and evaluate a hybrid soft computing technique for a given application

Unit I	Introduction to Soft Computing and Fuzzy logic	8 hrs
Introduction to soft computing: , Paradigms soft computing, Features, Components, Techniques, Applications, Neural Networks, Fuzzy logic, Genetic Algorithms, Hybrid systems, Introduction to Fuzzy logic: Classical and Fuzzy sets, operations, properties, Fuzzy Relations.		
Case Studies (if any)	Selection of Fuzzy membership to temperature sensor control system	
Mapping of Course Outcomes	**CO1**	

Unit II	Fuzzy Systems	8 hrs
Membership Functions, Fuzzification and Methods, Defuzzification and Methods, Fuzzy Logic, Fuzzy Rules and Fuzzy Reasoning, Fuzzy Inference Systems, Fuzzy Expert Systems, Fuzzy Decision Making. Fuzzy Control Systems, Fuzzy Classification.		
Case Studies(if any)	Fuzzy inference system for air conditioner	
Mapping of Course Outcomes	**CO2**	

Unit III	Genetic Algorithms	8 hrs
Introduction to Genetic Algorithms (GA), Search space, Working Principle, Simple GA, Operators, Fitness function, Multi-level Optimization.		
Case Studies(if any)	Application of GA for resource planning problem	
Mapping of Course Outcomes	**CO3**	

Unit IV	Introduction to Neural Networks	8 hrs
Neural Network, Models, Terminologies, Supervised Learning, Learning rules and various activation functions, Single layer Perceptrons , BackPropagation networks, Architecture of Backpropagation(BP)		

Case Studies(if any)	Backpropogation algorithm for disease detection	
Mapping of Course Outcomes	**CO4**	
Unit V	**Advance Neural Networks**	**8 hrs**
Neural Networks as Associative Memories - Hopfield Networks, Bidirectional Associative Memory Networks Unsupervised Learning, Kohonen Self Organizing Maps and Counter Propogation Networks Neural Network Classification, Deep learning		
Case Studies(if any)	Neural Network system for weather forecasting	
Mapping of Course Outcomes	**CO4**	
Unit VI	**Hybrid Systems and Soft Computing Applications**	**8 hrs**
Neuro-Fuzzy Hybrid Systems, Genetic Neuro Hybrid Systems, Fuzzy-Genetic Hybrid Systems. GA based Back propagation Networks, Optimization of TSP using GA, Fuzzy controllers for Robot.		
Case Studies(if any)	**Hybrid system for customer segmentation**	
Mapping of Course Outcomes	**CO5**	

Books & Other Resources:

Text Books:

1. S.N. Sivanandam & S.N.Deepa "Principles of Soft computing", John Wiley & Sons,
2. S.Rajasekaran, G. A. Vijayalakshami, Neural Networks, Fuzzy Logic and Genetic Algorithms: Synthesis & Applications, PHI.
3. David E. Goldberg., Genetic Algorithms: in Search and Optimization,PHI
4. Jyh:Shing Roger Jang, Chuen:Tsai Sun, EijiMizutani, Neuro:Fuzzy and Soft Computing, Prentice:Hall of India, 2003

Reference Books:

1. Timothy J. Ross, Fuzzy Logic with Engineering Applications (Wiley)
2. George J. Klir and Bo Yuan, Fuzzy Sets and Fuzzy Logic: Theory and Applications, Prentice Hall,
3. An Introduction to Genetic Algorithm Melanic Mitchell (MIT Press)
4. Evolutionary Algorithm for Solving Multi-objective, Optimization Problems (2^{nd} Edition), Collelo, Lament, Veldhnizer (Springer)
5. Neural Networks and Learning Machines Simon Haykin (PHI).
6. Neural Networks, Fuzzy logic, and Genetic Algorithms, S. Rajasekaran & G. A. V. Pai, PHI.

MOOC Courses

- https://swayam.gov.in/nd1_noc20_cs17/preview

E-books:

- **https://www.amazon.com/Soft-Computing-Neuro-Fuzzy-Genetic-Algorithms-ebook/dp/B00LOBIAPG**
- **https://bookboon.com/en/introduction-to-soft-computing-ebook**
- **http://freecomputerbooks.com/Introduction-to-Soft-Computing.html**

Important links:

- **https://www.journals.elsevier.com/applied-soft-computing**
- **http://www.soft-computing.de/linkC.html**

Savitribai Phule Pune University, Pune
ME Data Sciences (2020 Course)
510311A:Distributed Databases

Teaching Scheme:	Credit	Examination Scheme:
TH: 04 hr/week	04	Mid semester: 50 Marks End Semester: 50 Marks

Prerequisite Courses: Database Management Systems

Companion Course:

Course Objectives:

1. Understand the various aspects in Distributed Data.
2. Understand query processing and optimization in Distributed Database.
3. Management of distributed data with different levels of transparency.
4. Understand how to use database management tools in resolving deadlock situations.

Course Outcomes:

On completion of the course, learner will be able to

CO1:Design distributed database for any real world application.
CO2:Write query for data manipulation on Distributed Database.
CO3:Manage Transaction using fragmentation.
CO4:Handle deadlock situation in Distributed Database.
CO5:Apply security policies on Distributed Databases.
CO6:Manage data from Heterogeneous databases.

Selection of Modules: Modules 1 to 3 are compulsory and select any one from modules 4, 5 and 6.

Module I	Overview of Distributed Database Design	9 Hours

What is Distributed Database System (DDBS), Features of DDBS, promises of DDBS, Design issue in DDBS, Distributed DBMS architecture:- Client/server System, Peer-to-Peer, Multi-Database system,Levels of distribution transparency : Reference Architecture for Distributed Databases, Types of Data Fragmentation, Integrity Constraints in Distributed Databases
Framework of Distributed Databases Design, Design of Database Fragmentation, Allocation of fragments, Transparencies in Distributed Database Design.

Case Studies (if any)	RAID Distributed Database Management System
Mapping of Course Outcomes	CO1

Module II	Distributed Query Processing And Optimization	8 Hours

Concept, objective, and phases of distributed query processing, Translation of global queries to fragment queries, Query optimization in centralized databases, framework for query optimization in Distributed databases, join queries, general queries.

Case Studies(if any)	
Mapping of Course Outcomes	CO2

Module III	Transactions Management	9 Hours

TRANSLATION OF GLOBAL QUERIES TO FRAGMENT QUERIES: Equivalence Transformations For Queries, Transforming Global Queries into Fragment Queries, Distributed Grouping and Aggregate Function Evaluation, Parametric Queries.
THE MANAGEMENT OF DISTRIBUTED TRANSACTIONS: A Framework for Transaction Management, Supporting Atomicity of Distributed Transactions, Concurrency Control for Distributed Transactions, Architectural Aspects of

Distributed Transactions, Transaction Schedules in Distributed databases		
Case Studies(if any)		
Mapping of Course Outcomes	**CO3**	
Module IV	**Concurrency Control and Reliability.**	**9 Hours**
CONCURRENCY CONTROL: Foundations of Distributed Concurrency Control, Distributed Deadlocks, Concurrency Control Based on Timestamps, Optimistic Methods for Distributed Concurrency Control. Introduction to Deadlock, Distributed Deadlock prevention, avoidance, detection and recovery, Two-Phase and Three-Phase Commit Protocol. RELIABILITY: Basic Concepts, Non Blocking Commitment Protocols, Reliability and Concurrency Control, Determining a Consistent View of the Network, Detection and Resolution of Inconsistency, Checkpoints And Cold Restart. Catalog Management in Distributed Databases, Authorization and Protection.		
Case Studies(if any)		
Mapping of Course Outcomes	**CO4**	
Module V	**Security Aspects in DDBMS**	**9 Hours**
Study of a variety of attacks on the components of system (such as on routing protocols in ad hoc networks), privacy issues in Peer to Peer systems, trusted collaboration and dissemination of data among cooperative entities, Security problems, security policies, DAC methods, MAC methods, security models for DDBMS		
Case Studies(if any)		
Mapping of Course Outcomes	**CO5**	
Module VI	**Heterogeneous Database**	**9 Hours**
Architecture of Heterogeneous Database, Interface Standards for Relational Database :ODBC ODBC architecture, functionality and usage of ODBC Database Integration:- Schema Translation and schema Integration, Query processing issues in Heterogeneous database.		
Case Studies(if any)	Design heterogeneous distributed database	
Mapping of Course Outcomes	**CO6**	

Books & Other Resources:

Text Books:

1. Distributed Databases principles & systems by Stefano Ceri, Giuseppe Pelagatti, 2nd edition, McGraw-Hill, New York, 1985, ISBN 0-07-010829-3.
2. N.TamerOzsu, Patrick Valduriez, “Principles of Distributed Database Systems”, 2nd , Illustrated Edition, Prentice Hall International Inc., 1999, ISBN 0136597076, 9780136597070.
3. Database system Concept by Silberschatz And Korth 6th Edition,Tata Mcgraw Hill Education Private Limited, ISBN - 9789332901384

Reference Books:

1. Database Systems: A Practical Approach to Design, Implementation and Management- Thomas Connolly, Carolyn Begg, Pearson Publisher, 4th Edition.
2. Database Management Systems - Raghu Ramakrishnan and Johannes Gehrke, McGraw-Hill Education publisher, illustrated Edition,2003,ISBN0072465638, 9780072465631
3. Carlo Zaniolo, Stefano Ceri, Christos Faloustsos, R.T.Snodgrass, V.S.Subrahmanian, “Advanced Database Systems”, Morgan Kaufman, 1997.

MOOC Courses- Database Management Systems : 12 weeks course (20% coverage of syllabus)https://nptel.ac.in/courses/106/106/106106220/

E-books-

- Distributed Database Management Systems: A Practical Approach Kindle Edition by Saeed K. Rahimi, Frank S. Haug, 1st Edition, Wiley-IEEE Computer Society, ASIN: B005CDYQSC https://books.google.co.in/books?id=VryuBgAAQBAJ&printsec=frontcover&dq=Distributed+Database+Management+Systems:&hl=en&sa=X&ved=2ahUKEwiw8urYzKzqAhViyjgGHdv_Cc4Q6AEwAHoECAYQAg#v=onepage&q=Distributed%20Database%20Management%20Systems%3A&f=false

Important links:

Savitribai Phule Pune University, Pune ME Data Sciences (2020 Course) 510311B: Recommender Systems		
Teaching Scheme:	**Credit**	**Examination Scheme:**
TH: 04 hr/week	**04**	**Mid Semester: 50 Marks** **End Semester: 50 Marks**

Prerequisite Courses: Completed Machine Learning Crash Course or you have equivalent knowledge.

Companion Course: Machine Learning with TensorFlow.

Course Objectives:

1. Describe the purpose of recommendation systems.
2. Understand the components of a recommendation system including candidate generation, scoring,and ranking.
3. Familiarity with linear algebra

Course Outcomes:

On completion of the course, learner will be able to

CO1:Predict the "rating" or "preference" a user would give to an item

CO2:Product recommenders for services such as Amazon, or content recommenders for social media platforms such as Facebook and Twitter

CO3:Explain a variety of approaches for building recommender systems

CO4:Describe system evaluation methods from both algorithmic and users' perspectives

Selection of Modules: Modules 1 to 3 are compulsory and select any one from modules 4, 5 and 6.

Module I	**Introduction**	**8 Hours**
Recommender system functions, Linear Algebra notation: Matrix addition, Multiplication, transposition, and inverses; covariance matrices, Understanding ratings, Applications of recommendation systems, Issues with recommender system.		
Case Studies (if any)		
Mapping of Course Outcomes	**CO1**	
Module II	**Collaborative Filtering**	**8 Hours**
User-based nearest neighbor recommendation, Item-based nearest neighbor recommendation, Model based and pre-processing based approaches, Attacks on collaborative recommender systems.		
Case Studies(if any)		
Mapping of Course Outcomes	**CO2**	
Module III	**Content & knowledge based recommendation**	**8 Hours**
High level architecture of content-based systems, Advantages and drawbacks of content based filtering, Item profiles, Discovering features of documents, Obtaining item features from tags, Representing item profiles, Methods for learning user profiles, Similarity based retrieval, Classification algorithms. Knowledge based recommendation: Knowledge representation and reasoning, Constraint based recommenders, Case based recommenders.		
Case Studies(if any)		
Mapping of Course Outcomes	**CO3**	
Module IV	**Hybrid approaches**	**8 Hours**

Opportunities for hybridization, Monolithic hybridization design: Feature combination, Feature augmentation, Parallelized hybridization design: Weighted, Switching, Mixed, Pipelined hybridization design: Cascade Meta-level, Limitations of hybridization strategies		
Case Studies(if any)		
Mapping of Course Outcomes	**CO3**	
Unit V	**Evaluating Recommender System**	**8 Hours**
Introduction, General properties of evaluation research, Evaluation designs, Evaluation on historical datasets, Error metrics, Decision-Support metrics, User-Centred metrics.		
Case Studies(if any)		
Mapping of Course Outcomes	**CO4**	
Module VI	**Recommender Systems and communities**	**8 Hours**
Communities, collaboration and recommender systems in personalized web search, Social tagging recommender systems, Trust and recommendations, Group recommender systems.		
Case Studies(if any)	Social Networks	
Mapping of Course Outcomes		

Books & Other Resources:

Text Books:

1. Francesco Ricci, Lior Rokach, and Bracha Shapira, eds. Recommender Systems Handbook, 2nd edition.Spring US, 2015.
2. Jannach D., Zanker M., Fel FeringA., Recommender Systems: An Introduction, Cambridge University Press, 2011.

Reference Books:

1. Charu C. Aggarwal, Recommender Systems: The Textbook, 1/e, Springer, 2016.
2. Ricci F., Rokach L., Shapira D., Kantor B.P., Recommender Systems Handbook,Springer, 2011.
3. Manouselis N., Drachsler H., Verbert K., Duval E., Recommender Systems for Learning, Springer, 2013

MOOC Courses: Courses available on Coursera and Udemy only.

E-books:

- Recommender systems**:** The Textbook Kindle edition by Charu C. Aggarwal.
- Hands-On Recommendation Systems with Python: Start building powerful and personalized, recommendation engines with Python by Rounak Banik.
- Building Recommendation Systems using Python by MehreenTahir.

Important links:

Savitribai Phule Pune University, Pune
ME Data Sciences (2020 Course)
510311C: GPU Computing

Teaching Scheme:	Credit	Examination Scheme:
TH: 04 hr/week	04	Mid Semester: 50 Marks End Semester: 50 Marks

Prerequisite Courses: Parallel programming concepts, languages, and Platforms

Companion Course: Soft Computing

Course Objectives:

1. To understand the different approaches of parallel programming.
2. To study massively parallel computing hardware and programming models.
3. To be conversant with GPGPU programming with CUDA.
4. To develop parallel programs in heterogeneous environments with OpenCL.
5. To understand machine learning using GPU.

Course Outcomes:

On completion of the course, learner will be able to

CO1:Analyze and measure performance of modern parallel computing systems.
CO2:Design and Implement parallel programs on GPUs.
CO3:Develop a high-performance parallel application in CUDA.
CO4:Build parallel programming logic on current system architectures using OpenCL.
CO5:Implement machine learning using GPU.

Selection of Modules: Modules 1 to 3 are compulsory and select any one from modules 4, 5 and 6.

Module I	Understanding Parallelism with GPUs.	9 Hours

Review of traditional computer architecture – basic five stage RISC pipeline, cache memory, register file, SIMD instructions, and GPU architectures - streaming multi processors, cache hierarchy, the graphics pipeline, parallel programming languages and models. Understanding Parallelism with GPUs.

Case Studies (if any)	
Mapping of Course Outcomes	CO1

Module II	Grids, Blocks, and Threads	9 Hours

Grids, Blocks, and Threads Introduction to Data Parallelism and CUDA C, Data-Parallel Execution Model, CUDA Memories-Memory types and memory Access Efficiency, Performance Considerations-Warps ,Thread Execution, Global Memory Bandwidth, Dynamic Partitioning of Execution Resources, Instruction Mix and Thread Granularity, the CUDA extensions to the C language, and the basic programming/debugging tools.

Case Studies(if any)	Prefix Sum
Mapping of Course Outcomes	CO1 , CO2

Module III	Memory Handling and Synchronization	8 Hours

Memory Handling with CUDA- The basic CUDA memory/threading model, floating-point considerations in parallel computing and common data-parallel programming patterns needed to develop a high-performance parallel application.Programs for concurrent Data Structure such as Worklists, Linked-lists. Synchronization across CPU and GPU.

Case Studies(if any)	Graph algorithms, Simulations,
Mapping of Course Outcomes	CO2 , CO3

Module IV	Designing GPU-Based Systems.	8 Hours
Parallel Programming and Computational Thinking, MPI-CUDA programming in a heterogeneous computing cluster. Dynamic parallelism, Unified Virtual Memory, CPU vs GPU, GPU hardware overview, GPU memory architecture, GPU properties, compute capability of GPU, multi- GPU solution. Multi-GPU processing, Peer access, Heterogeneous processing		
Case Studies(if any)	Molecular Visualization and Analysis	
Mapping of Course Outcomes	**CO3**	
Module V	**Introduction to OpenCL**	**8 Hours**
Introduction to OpenCL-The OpenCL Platform Model , The OpenCL Execution Model, Kernels and the OpenCL Programming Model, The OpenCL Memory Model, OpenCL basics with Examples. OpenCL for Heterogeneous Computing-Memory performance considerations in OpenCL. OpenCL runtime and concurrency model-Commands and the Queuing Model, Multiple Command- Queues,The Kernel Execution Domain-Work Items, Work-Groups, NDRanges ,Naive and Built-In Kernels		
Case Studies(if any)	Dissecting OpenCL on a Heterogeneous System on AMD FX-8350 CPU, AMD Radeon R9 290X GPU or as per available latest configuration of CPU.	
Mapping of Course Outcomes	**CO4**	
Module VI	**Machine learning applications with CUDA**	**8 Hours**
Containerization on GPU-Enabled Platforms, concept of Containerization, working of open and closed environments as local and cloud containers **Accelerated Machine learning on GPUS**, Exploring the Pytorch and Neural networks.		
Case Studies(if any)	GPU Enabled Machine Learning	
Mapping of Course Outcomes	**CO5**	

Books & Other Resources:

http://www.cs.columbia.edu/~m-reed/gpu.html

https://developer.nvidia.com/udacity-cs344-intro-parallel-programming

Text Books:

1. "Programming Massively Parallel Processors" - David Kirk and Wen-meiHwu
2. " Heterogeneous Computing with OpenCL" -- Benedict Gaster,Lee Howes, David R. Kaeli
3. Hands-On GPU Computing with Python: (Kindle Edition) by Bandyopadhyay, Avimanyu

Reference Books:

1) Shane Cook, "CUDA Programming: A Developer's Guide to Parallel Computing with GPUs", Morgan Kaufmann Publishers Inc. San Francisco, CA, USA 2013 ISBN: 9780124159884

2) CUDA BY EXAMPLE by Jason Sanders,Edvard Kandrot

MOOC Courses- https://swayam.gov.in/nd1_noc20_cs41/preview

E-books -Hands-On GPU Computing with Python: Explore the capabilities of GPUs for solving high performance computational problems Kindle Edition

Important links:

https://developer.nvidia.com/

https://www.khronos.org/opencl/

Savitribai Phule Pune University, Pune
ME Data Sciences (2020 Course)

510311D: Web Intelligence

Teaching Scheme	Credit Scheme	Examination Scheme and Marks
Lecture: 04 Hours/Week	**04**	**In_Semester(TH): 50 Marks** **End_Semester(TH): 50 Marks**

Prerequisites: Basic Mathematics, algorithms and data Structures

Companion Course :

Course Objectives:

1. To learn web intelligence basics to build website using intelligent technologies
2. To learn the semantic web technologies.
3. To learn web spiders to create specialized search engines.
4. To study web structure mining methods.
5. To learn Algorithmic Aspects of Web Intelligent Systems
6. To study social network intelligence as link analysis of the web.

Course Outcomes:

On completion of the course, learner will be able to–

CO1:Build website using intelligent technologies

CO2:Apply various semantic web technologies for building layered language models.

CO3:Learn and apply web spiders for specialized search engines.

CO4:Apply web structure mining methods for mining the data.

CO5:Use algorithmic aspects of web intelligent systems for web document pre-fetching on the Internet

CO6: Apply Social Network Intelligence for identifying fractal nature of the web and knowledge management.

Selection of Modules: Modules 1 to 3 are compulsory and select any one from modules 4, 5 and 6.

Module I	**Web Intelligence and Information Retrieval**	**(07 Hours)**
What is web intelligence? Towards an Intelligent web, knowledge, Web mining, Building better websites using intelligent technologies, benefits of intelligent web. Information Retrieval: Introduction, document representation, retrieval models, evaluation of retrieval performance, public domain information retrieval systems.		
Case Studies(if any)		
Mapping of Course Outcomes	CO1	

Module II	**Semantic Web**	**(07 Hours)**
Semantic Web technologies, Introduction, layered-language model, metadata and ontologies, ontology language for web, ontologies for knowledge management-ontology usage scenario, ontologies as RDBMS schema, Topic ontology versus schema ontologies, proton ontology, Semantic web services- WSMO aproach, OWL-s approach, SWSF approach, WSDLS approach, The link between SWS and existing Web services standards.		
Case Studies(if any)		
Mapping of Course Outcomes	CO2	

Module III	**Web Content Mining**	**(07 Hours)**
Introduction, Web Spiders for Personal Search- Personal Web Spiders Using Web Spiders to Create Specialized Search Engines- Specialized Search Engines, Focused Spidering Algorithms for Specialized Search Engines, Web Content Mining- opinion mining, structure mining. social Networks- Finding communities, usage mining, example: query log analysis, advanced example- web spam detection		

Case Studies(if any)		
Mapping of Course Outcomes	CO3	
Module IV	**Web Structure Mining**	**(07 Hours)**
Introduction, Hyper link structure, Web search and hyper link, Modeling web topology: Page rank algorithm, Hyperlink induced topic search (HITS), comparison of Page rank and HITS, Random walks on the web, Other approaches to study web link structure—Social Networks, Reference and index pages.		
Case Studies(if any)		
Mapping of Course Outcomes	CO4	
Module V	**Algorithmic Aspects of Web Intelligent Systems**	**(07 Hours)**
An Overview of the System- User Interface, Performance, Users and Authentication Techniques, Agent's Inference Engine Algorithms- Data Characteristics and Generic Handling Techniques, Choosing the Next Document, Finding Interesting Object Collections and Predicting Votes by Matching Users, Finding an Interesting Documents Collection and Predicting Votes Using Na'ive Bayes Analysis, Matching Related Documents Web Document Prefetching on the Internet- Introduction: Prefetching at Different Stages, Conditions of Content Prefetching, Classifying Prefetching Methods, Prefetching Structure and Optimization, Performance Evaluations on Prefetching, Other Variants of Prefetching, Related Applications .		
Case Studies(if any)		
Mapping of Course Outcomes	CO5	
Module VI	**Social Network Intelligence**	**(07 Hours)**
Social Networks: From the Web to Knowledge Management - Link Analysis of the Web, Communities on the Web, Connectivity and the Diameter of the Web, Fractal Nature of the Web, Social Networks for Knowledge Management, A Ranking Algorithm Based on Graph Topology to Generate Reputation or Relevance- Social Networks, Ranking Algorithm, Experiments About Ranking, Reputation, and Relevance		
Case Studies(if any)		
Mapping of Course Outcomes	CO6	

Books

Text Books:

1) Ning Zhong, JimingLiu, Yiyu Yao, " Web Intelligence", Springer, ISBN: 978-3-642-07936-8
2) John Davies, Rudy Studer, Paul Warren, Semantic Web Technologies: Trends and Research in Ontology-based Systems, ISBN: 978-0-470-02596-3
3) Pawan Lingras, Rajendra Akerkar, "Building an Intelligent Web: Theory and Practice", *ISBN*-10: 076374137X

Reference Books:

1. Dallas Marks, Heather Sinkwitz, Jim Brogden, Gabriel Orthous , "SAP Business Objects Web Intelligence: The Comprehensive Guide", 3rd edition , Galileo Press
2. Akerkar, R. & Lingras, "Building an Intelligent Web: Theory and Practice, Jones and Bartlett Publishers, Sudbury, Massachusetts. ISBN-13: 978-0-7637-4137-2
3. Ian H. & Frank, E," Data Mining: Practical Machine Learning Tools and Techniques", 2nd Edition, Morgan Kaufman. ISBN 0120884070, 9780120884070

Savitribai Phule Pune University, Pune
ME Computer Engineering Data Science (2020 Course)
5103013: Laboratory Proficiency-II

Teaching Scheme:	Credit	Examination Scheme:
PR: 8 hr/week	04	Term Work: 50 Marks Practical/Oral: 50 Marks

Companion Courses:
510308 – Data warehousing and mining
510309 – Machine Learning
510310 – Soft Computing
510311 – Elective – II

All assignments are compulsory. Each student should implement the assignment individually. Laboratory teachers should make sure that the example/dataset/code is not the same. Apart from the Weka tool, SAS University edition is also suggested for use.

510308- Data warehousing and mining

Assignment 1.

For an organization of your choice, choose a set of business processes. Design star / snow flake schemas for analysing these processes. Create a fact constellation schema by combining them.

Assignment 2.

Extract data from different data sources, apply suitable transformations and load into destination tables using an ETL tool. For Example: Business Origination: Sales, Order, and Marketing Process.

510309 – Machine Learning

Assignment 1.
a) Using appropriate dataset from UCI machine learning repository design a decision tree. Implement two different decision tree algorithms. Find the root node of the decision tree.
b) Extract confusion matrix from the test results. Compare the performance of the two decision tree algorithms in terms of at least six relevant measures.
c) Now, classify the data using three decision tree algorithms from the Weka tool and compare the performance of your implementations with the results from the Weka tool.

Assignment 2.

Implement k-NN classifier to classify a standard dataset (from UCI machine learning repository). Use Java/Python/R for implementation. Test the performance for various values of k. Now, classify the same dataset using distance-weighted k-NN and Locally weighted averaging methods. Compare the performance on at least six standard performance measures.

510310 – Soft Computing

Assignment 1.

Implement genetic algorithms for finding the optimal allocation of resources for certain applications. (e.g. computer laboratory allocation, radio resource allocation, human resource allocation, irrigation (water resource) planning etc.)

Assignment 2.

Implement back-propagation algorithm and GA based back-propagation neural network for detection of a disease. (You may choose heart disease, cancer disease, plant diseases, skin disease) Dataset could be structured or unstructured. Compare the performance of these two algorithms using appropriate performance measures.

510311 – Elective – II

510311A: Distributed Databases

Mini-project: Design and implement the distributed architecture for the Hadoop having Name node, Tracker node and data nodes. Implement Digital Library Infrastructure using Hadoop or Similar recent technology for distributed database storage. To develop front end GUI and algorithm for searching the multimedia resource files, presentations in the selected domain, author, book title, ISBN. Use different search exploration techniques.

510311B: Recommender Systems

Mini-project: Design and Implement recommender algorithms using an open source toolkit. Use appropriate standard dataset.

510311C: GPU Computing

Mini-project: Image clustering using OpenCL.

510311D: Web Intelligence

Mini-project: Implement web-scraping for required text/image data for recommender system design.

Savitribai Phule Pune University, Pune
ME Data Sciences (2020 Course)
610301: Deep Learning

Teaching Scheme:	Credit	Examination Scheme:
TH: 03 hr/week	03	Mid Semester: 50 Marks End Semester: 50 Marks

Prerequisite Courses: Mathematical Foundation of Data Science, Machine Learning

Companion Course: --

Course Objectives:

1. To introduce major deep learning algorithms
2. To introduce optimization techniques to training deep neural networks
3. Learn regularization techniques to training deep neural networks
4. To introduce Convolutional Neural Networks and its applications
5. Learn autoencoders and its applications
6. Learn deep recurrent and memory networks

Course Outcomes:

On completion of the course, learner will be able to

CO1:Identify the deep learning algorithms which are more appropriate for various types of learning tasks in various domains
CO2:To apply optimization techniques to training deep neural networks
CO3:Apply regularization techniques to improve the performance of deep learning algorithms.
CO3:Implement deep learning algorithms and solve real-world problems in computer vision.
CO5:Apply autoencoders to solve real world problems.
CO6:Implement deep learning algorithms and solve real-world problems in Natural Language Processing

Unit I	Introduction to deep learning	8 Hours

Perceptrons, Perceptron Learning Algorithm, Sigmoid Neuron, Shallow neural networks, Deep neural networks, Feedforward Neural networks, Gradient descent and the backpropagation algorithm

Case Studies (if any)	Implementation of neural networks from scratch
Mapping of Course Outcomes	**CO1**

Unit II	Optimization and Deep Learning	8 Hours

Learning Parameters of a feedforward neural network, the vanishing gradient problem, and ways to mitigate it, RelU Heuristics for avoiding bad local minima, Heuristics for faster training, Nesterov Accelerated GD, Stochastic GD, AdaGrad, RMSProp, Adam, Momentum.Adagrad, Principal Component Analysis and its interpretations, Singular Value Decomposition.

Case Studies(if any)	Application of PCA for Dimentionality reduction,Speeding ML algorithm
Mapping of Course Outcomes	**CO2**

Unit III	Regularization Techniques	6 Hours

Bias Variance Tradeoff, L2 regularization, Early stopping, Dataset augmentation, Parameter sharing and tying, Injecting noise at input, Ensemble methods, Dropout		
Case Studies(if any)	Case study of dataset augmentation, dropout on image classification dataset	
Mapping of Course Outcomes	**CO3**	
Unit IV	**Convolutional Neural Networks**	**8 Hours**
Convolutional Neural Networks, LeNet, AlexNet, ZF-Net, VGGNet, GoogLeNet, ResNet, Visualizing Convolutional Neural Networks, Guided Backpropagation, Deep Dream, Deep Art, Fooling Convolutional Neural Networks		
Case Studies(if any)	Image classifier for identifying cat vs dogs using CNN	
Mapping of Course Outcomes	**CO4**	
Unit V	**Deep Unsupervised Learning**	**7 Hours**
Autoencoders: standard, sparse, denoising, contractive, Variational Autoencoders, Adversarial Generative Networks, Autoencoder and DBM		
Case Studies(if any)	Feature Extraction using autoencoders	
Mapping of Course Outcomes	**CO5**	
Unit VI	**Sequence Models:**	**7 Hours**
RNN, LSTM, GRU models, Application to NLP, language models, machine translation, image captioning, video processing, visual question answering, video processing, learning from descriptions, Attention Mechanism, Attention over images		
Case Studies(if any)	Sentiment Analysis using Recurrent Neural Networks	
Mapping of Course Outcomes	**CO6**	

Books & Other Resources:

Text Books:

1. Goodfellow, Y. Bengio, A. Courville, Deep Learning, MIT Press, 2016.

Reference Books:

1. Neural Networks: A Systematic Introduction, Raúl Rojas, 1996
2. Pattern Recognition and Machine Learning, Christopher Bishop, 2007

MOOC Courses:

1. Deep Learning Part-I, Swayam Prof.Mitesh M. Khapra

2. Neural Networks and Deep Learning, Coursera, Andrew Ng

3. Deep Learning for Computer Vision, Prof. Vineeth N Balasubramanian

E-books: **http://www.deeplearningbook.org**
Important links: **1. https://towardsdatascience.com** **2. https://www.kaggle.com** **3. http://deeplearning.net/**

Savitribai Phule Pune University, Pune ME Data Sciences (2020 Course) 610302: Data Modeling and Visualization		
Teaching Scheme:	**Credit**	**Examination Scheme:**
TH: 03hr/week	**03**	**Mid Semester: 50 Marks** **End Semester: 50 Marks**

Prerequisite Courses: Computer Graphics, Data mining, Image processing, Statistical methods

Companion Course:

Course Objectives:

1. To map element of visualization well to perceive information well
2. To learn different types of data and its visualization
3. To study quantitative and non quantitative data visualization.
4. To study the pattern for static and moving data

Course Outcomes:

On completion of the course, learner will be able to

CO1:Understand types of data and data visualization methods
CO2:Understand the need of data visualization.
CO3:Apply visualization technique well for quantitative data
CO4:Understand patterns in motion
CO5:Evaluate the performance of visualization technique
CO6:Apply data visualization using open source tool Tableau

Unit I	**Introduction to Data Visualization**	**8Hours**

Need for data visualization. Types of Data, Stages of Data visualization, Fitts Law, Human visual perception and cognition

Case Studies (if any)	**Installation of Tableau Public and analysing different types of data.**
Mapping of Course Outcomes	**CO1,CO2**

Unit II	**Visualization of numerical data**	**8 Hours**

Types of Data visualization: Basic charts, scatter plots, Histogram ,advanced visualization Techniques like streamline and statistical measures

Case Studies(if any)	**Perform constellation modelling of high dimensional data. And analyse the properties**
Mapping of Course Outcomes	**CO2**

Unit III	**Visualization of non-numeric data**	**6 Hours**

Plots , Graphs, networks, Hierarchies, symbol and shaded maps, treemap

Case Studies(if any)	A roadmap with symbols representing cities and colored lines representing roads between the cities. Provide node-link diagram. Perform search to find the node symbol and extract the alternate paths.
Mapping of Course Outcomes	**CO2**

Unit IV	**High dimensional data**	

Mapping of high dimensional data into suitable visualization method- Principal component analysis, multidimensional, clustering study of High dimensional data visualization in R, Python , Google chart

API		
Case Studies(if any)	Make use of IMDB movie dataset and apply classification and use suitable data visualization techniques.	
Mapping of Course Outcomes	**CO3,CO1**	
Unit V	**Static and moving data**	
Gestalt laws, texture theory and data mapping, perception of transparency/; overlapping data, perceiving patterns in multidimensional discrete data, patterns in motion		
Case Studies(if any)	**Take the example of traffic signal, analyse the pattern and use suitable method to visualize pattern in motion.**	
Mapping of Course Outcomes	**CO1,CO4**	
Unit VI	**Evaluation and visualization tools**	
Evaluation of visualization, Tableau , Desktop workspace in Tableau , visual control, data analytics		
Case Studies(if any)	Data analytics in Tableau	
Mapping of Course Outcomes	**CO5,CO6**	
Books & Other Resources:		
Text Books: 1. Information visualization perception for design, colin ware, MK publication		
Reference Books: 1. Big data black book, Dream tech publication 2. Handbook for visualizing : a handbook for data driven design by Andy krik		
MOOC Courses: Coursera course on data visualization		
E-books:		
Important links:		

Savitribai Phule Pune University, Pune ME Data Sciences (2020 Course) 610303A: Real Time Analytics		
Teaching Scheme:	**Credit**	**Examination Scheme:**
TH: 03 hr/week	**03**	**Mid Semester: 50 Marks** **End Semester: 50 Marks**

Prerequisite Courses: Time series and Forecasting

Companion Course: Real time Analytics with Apache storm

Course Objectives:

1. To teach the fundamental techniques and principles in achieving data analytics with scalability and streaming capability.
2. To provide an overview of an exciting growing field of data analytics.
3. To enable students to have skills that will help them to solve complex real-world problems in decision support.

Course Outcomes:

On completion of the course, learner will be able to--

CO1: Understand & apply appropriate analytical techniques.
CO2:Apply analytics for decision making in healthcare services.
CO3:Learn and understand open source tools like Google Analytics

Selection of Modules: Modules 1 to 3 are compulsory and select any one from modules 4, 5 and 6.

Module I	**Fundamentals of Data Analytics**	**8 Hours**
Data Analytics Basics, Data Types, Analytics Types, Data Analytics Steps: Data Pre-Processing, Data Imputation, Data Cleaning, Data Transformation, Data Visualization, and Data Engineering. Descriptive, Predictive, and Prescriptive Analytics.		
Case Studies (if any)	Any Exploratory Data Analysis (EDA) can be done here Ex: https://towardsdatascience.com/exploratory-data-analysis-in-python-c9a77dfa39ce	
Mapping of Course Outcomes		
Module II	**Data Analytics with Python**	**8 Hours**
Data Analytics using Python, Statistical Procedures, Web Scraping in Python, Advanced analytics, NumPy, Pandas, SciPy, Matplotlib.		
Case Studies(if any)	Web Scrapping must be emphasized.	
Mapping of Course Outcomes		
Module III	**Time Series Analysis**	**8 Hours**
Box-Jenkins Methodology for ARIMA models: Examining correlation and stationarity of time series data, ARIMA models for time series data (An Auto-regressive model of order one and a Moving Average Model of order one).		
Case Studies(if any)	ARIMA is used for time series analysis to get moving avg,share market analysis can be done here Ex:https://towardsdatascience.com/stock-market-analysis-using-arima-8731ded2447a	
Mapping of Course Outcomes		

Module IV	Streaming Data	6 Hours
Streaming Analytics Architecture: Designing Real-Time Streaming Architectures, Service Configuration and Coordination.		
Case Studies(if any)	Real-Time Analytics with Network Data: This section explains Apache Storm based real-time analytics solution, using an example of a telecom service provider. In the network of a telecom service provider, there can be different sources of incoming data, like: 1. Stream of data generated due to use of services by subscribers 2. Performance data of access network, as reported by network probes Data related with new subscription orders, activation and terminate orders.	
Mapping of Course Outcomes		
Unit V	**Streaming Data Analysis**	**7 Hours**
Data-Flow Management in Streaming Analysis, Processing Streaming Data, Storing Streaming Data		
Case Studies(if any)	Case study can be done on any social media site Ex: https://www.dataquest.io/blog/streaming-data-python/	
Mapping of Course Outcomes		
Module VI	**Market Basket Analysis, Recommender system**	**7 Hours**
Todays ecommerce system, apriori algorithm. YOLO: real time object Detection		
Case Studies(if any)		
Mapping of Course Outcomes		

Books & Other Resources:

Text Books:

1. Anil Maheshwari, “Data Analytics made accessible,” Amazon Digital Publication, 2014.
2. Byron Ellis,”Real–Time Analytics:Techniques to Analyze and Visualize Streaming Data”, WILEY Publication.
3. Michael Minelli, Michelle Chambers, and Ambiga Dhiraj, "Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Businesses", Wiley, 2013.

Reference Books:

1. Thomas H. Davenport, Jeanne G. Harris and Robert Morison, “Analytics at
   a) Work: Smarter Decisions, Better Results”, Harvard Business Press, 2010 .
2. Spyros Makridakis, Steven C.Wheelwright and Rob J. Hyndman. Forecasting
   b) methods and Applications, Third Edition”, John Wiley & Sons Inc., New York (Chapters 1, 4 and 7), 2005.

MOOC Courses:

E-books:

- Anil Maheshwari, “Data Analytics made accessible,” Amazon Digital Publication, 2014
- Real-Time Analytics, Techniques to analyze and visualize streaming Data by Byron Ellis.

Important links:

Savitribai Phule Pune University, Pune ME Data Sciences (2020 Course) 610303B: Business Analytics		
Teaching Scheme:	**Credit**	**Examination Scheme:**
TH: 04 hr/week	**04**	**Mid Semester: 50 Marks** **End Semester: 50 Marks**

Prerequisite Courses: Basic Statistics, Basic Mathematics, Basic Management and Basics of Data Mining

Companion Course:

Course Objectives:

1. Understand the role of business analytics within an organization.
2. Analyze data using statistical and data mining techniques and understand relationships between the underlying business processes of an organization.
3. Understanding insights of managers to solve business problems and to support managerial decision making.
4. Survey the processes needed to develop, report, and analyze business data.
5. Use decision-making tools/Operations research techniques.
6. Mange business process using analytical and management tools.

Course Outcomes:

On completion of the course, learner will be able to

CO1:Analyze and visualize data in different industries such as manufacturing, service, retail, software, banking and finance, sports etc.
CO2:Use technical skills in descriptive modeling to support business decision-making.
CO3:Use technical skills in predictive modeling to support business decision-making.
CO4:Use technical skills in prescriptive modeling to support business decision-making.
CO5: Demonstrate decision making with and without Risk for solving problems in different industries.

Selection of Modules: Modules 1 to 3 are compulsory and select any one from modules 4, 5 and 6.

Module I	**Overview of Business analytics**	

Business analytics: Overview of Business analytics, Scope of Business analytics, Business Analytics Process, Relationship of Business Analytics Process and organization, competitive advantages of Business Analytics. Organization Structures of Business analytics, Team management, Management Issues, Designing Information Policy, Outsourcing, Ensuring Data Quality, Measuring contribution of Business analytics, Managing Changes. Database Analytics

Case Studies (if any)	
Mapping of Course Outcomes	**CO1**

Module II	**Descriptive Analytics**	

Descriptive Analytics :
Statistical Tools: Statistical Notation, Descriptive Statistical methods, Review of probability distribution and data modeling, sampling and estimation methods overview. Important Resources, Business Analytics Personnel, Data and models for Business analytics, problem solving, Visualizing and Exploring Data, Business Analytics Technology.

Case Studies(if any)	Tableau – Data visualization tool
Mapping of Course Outcomes	**CO2**

Module III	**Predictive Analytics**	

Trendlines and Regression Analysis Learning Objectives : Modeling Relationships and Trends in Data , Simple Linear Regression, Residual Analysis and Regression Assumptions , Multiple Linear Regression, Building Good Regression Models, Regression with Categorical Independent Variables, Regression Models with Nonlinear Terms. **Forecasting Methods:** Forecasting Techniques: Qualitative and Judgmental Forecasting, Statistical Forecasting Models, Forecasting Models for Stationary Time Series, Forecasting Models for Time Series with a Linear Trend, Forecasting Time Series with Seasonality, Regression Forecasting with Casual Variables, Selecting Appropriate Forecasting Models.		
Case Studies(if any)	Healthcare data analysis	
Mapping of Course Outcomes	**CO3**	
Module IV	**Prescriptive Analytics**	
Linear Optimization : Optimization Models, Linear Optimization in Bank Financial Planning, Analytics in Practice: Using Optimization Models for Sales Planning at NBC, Developing Linear Optimization Models, Identifying Decision Variables, the Objective, and Constraints , Developing a Mathematical Model ,Implementing Linear Optimization Models, Solving Linear Optimization Models , Graphical Interpretation of Linear Optimization with Two Variables , Applications of Linear Optimization, Integer Linear Optimization Models , Models with Binary Variables, Nonlinear Optimization Models.		
Case Studies(if any)	Portfolio Analysis	
Mapping of Course Outcomes	**CO4**	
Module V	**Simulation and Risk Analysis**	
Model-Building Strategies: Building Models Using Logic and Business Principles ,Building Models Using Influence Diagrams, Building Models Using Historical Data,Model Assumptions, Complexity, and Realis. Analysing uncertainty. Introduction of simulation and Risk Analysis, Types of simulations, Risk Management, Risk Assessment, Impact Analysis, Monte Carlo Simulation Using Analytic Solver Platform, New-Product Development Model, Newsvendor Model, Overbooking Model, Cash Budget Model.		
Case Studies(if any)		
Mapping of Course Outcomes for Unit V	**CO5**	
Unit VI	**Decision Analysis**	
Decision Analysis: Formulating Decision Problems, Decision Strategies with the without Outcome Probabilities, Decision Trees, The Value of Information, Utility and Decision Making. Recent Trends in: Embedded and collaborative business intelligence, Visual data recovery, Data Storytelling and Data journalism, Decision Tree and Risk		
Case Studies(if any)		
Mapping of Course Outcomes	**CO5**	
Books & Other Resources:		
Text Books: 1) Business analytics Principles, Concepts, and Applications by Marc J. Schniederjans, Dara G. Schniederjans, Christopher M. Starkey, Pearson FT Press. 2) Business Analytics: Methods, Models and Decisions by James Evans, persons Education, 3rd Edition		
3) **Reference Books:** 4) R. Sharda, D. Delen, and E. Turban, Business Intelligence and Analytics. Systems for Decision Support,10th Edition. Pearson/Prentice Hall, 2015. ISBN-13: 978-0-13-305090-5, ISBN-10: 0-13-305090-4; 5) Carlo Vercellis, "Business Intelligence - Data Mining and Optimization for Decision Making", Wiley Publications, ISBN: 9780470753866		
MOOC Courses : Syllabus covered(90%) 1) Business Analytics for Management Decision : 12 weeks NPTEL course		

2) Series of Coursera course on Business Analytics
E-books: 1. Business Analytics A Practitioner's Guide by **Rahul Saxena**, **Anand SrinivasanInternational Series in Operations Research & Management Science**, Springer New York, December 5, 2012,ISBN: 9781461460800.
Important links:

Savitribai Phule Pune University, Pune
ME Data Sciences (2020 Course)
610303C: Computational Linguistic Analytics

Teaching Scheme:	Credit	Examination Scheme:
TH: 04hr/week	04	Mid Semester: 50 Marks End Semester: 50 Marks

Prerequisite Courses: Theory of Computation

Companion Course:

Course Objectives:

1. To understand grounded introduction to contemporary work in Computational Linguistics
2. To learn standard methods for processing words (morphology)
3. To learn standard methods for sentence processing (parsing and generation)
4. To acquaint methodologies for semantic analysis.

Course Outcomes:

On completion of the course, learner will be able to--

CO1: Understand key computational notions.

CO2:Developalgorithms and software for intelligently processing language data.

CO3:Translate one language into another using morphology and syntax of the given sentence

CO4:Be specialists in the application of computers to the processing of natural languages.

CO5:Become aware of the Legal, ethical and security issues concerning data, including aggregated data.

CO6:get opportunity to work in real life research project jobs in the field of computational linguistics, also known as text analytics, natural language processing and informatics.

Selection of Modules: Modules 1 to 3 are compulsory and select any one from modules 4, 5 and 6.

Module I	Computational Semantics	08 Hours

Identification of syntactic structure in natural language. Parsing algorithms for popular grammar formalisms. Application of statistical information to parsing. Parser evaluation. Extraction of parse features. Popular semantic resources. Techniques for building new resources from unstructured text data. Cross validation, ROC curves, feature engineering and regularization.

Case Studies (if any)	
Mapping of Course Outcomes	CO1

Module II	Advanced Machine Learning	12 Hours

Supervised machine learning with focus on classification. K-NN, Decision trees, SVM, combining models via ensembling: boosting, bagging, random forests. Basic machine learning concepts: generalization error and overfitting. Introduction to optimization, Gradient Descent and Stochastic Gradient Descent. Roundoff error and finite differences. Clustering, association rules, model fitting via EM algorithm. Finding groups and other structures in unlabeled and high dimensional data.

Case Studies(if any)	
Mapping of Course Outcomes	CO2

Module III	Computational Morphology and Machine Translation	08

Sub-word phenomenon approaches. Automatic morphological analysis of diverse languages. Parts of speech tagging. Word segmentation and character-level neural network models. Key methodologies for automatic translation between languages with focus on statistical and neural machine translation approaches. Applying machine translation (MT) architecture to analogous monolingual tasks. MT evaluation.

Case Studies(if	

any)		
Mapping of Course Outcomes	**CO3**	
Module IV	**Advanced Computational Semantics**	**08 Hours**
Text corpora collection and curation. Methods to pull representative datasets from internet sources. Techniques for efficient and reliable annotation. Application of machine learning to various semantic tasks: Information extraction, semantic role labelling, semantic parsing, discourse parsing, question answering, summarization and natural language inference. Cutting edge techniques in natural languageprocessing.		
Case Studies(if any)	Latest innovations in neural network architectures.	
Mapping of Course Outcomes	**CO4, CO6**	
Module V	**Statistical NLP**	**08 Hours**
Text and document classification, Classification of selected words or phrases in sentential or broader contexts, Sequence labeling, Structure assignment to sentences, Sentence transduction, Knowledge transfer from other (related) languages		
Case Studies(if any)		
Mapping of Course Outcomes	**CO2**	
Module VI	**Sentiment Analysis**	**08 Hours**
Sentiment identification and Analysis, Text polarity and emotion classification. Fine-grained mining, Sentiment in social networks, Legal, ethical and security issues concerning data, including aggregated data. Proactive compliance with rules and, in their absence, principles for the responsible management of sensitive data.		
Case Studies(if any)	Identification and analysis of opinion, especially social media, aspectual mining	
Mapping of Course Outcomes	**CO4, CO5, CO6**	
Books & Other Resources:		
Text Books: 1. Dan Jurafsky and James H. Martin. Speech and Language Processing (3rd ed. draft), Prentice-Hall, 2000 2. Igor A. Bolshakov and Alexander Gelbukh, Computational Linguistics :Models, Resources, Applications 3. Patrick Blackburn and Kristina Striegnitz (BS) Natural Language Processing Techniques in Prolog 4. Patrick Blackburn and Johan Bos (BB1) Representation and Inference for Natural Language A First Course in Computational Semantics 5. Dipanjan SarkarText Analytics with Python: A Practical Real-World Approach to Gaining Actionable Insights from your Data		
Reference Books: 1. Jacob Eisenstein. Natural Language Processing 2. Yoav Goldberg. A Primer on Neural Network Models for Natural Language Processing		
MOOC Courses: Text Mining and Analytics offered by Illinois, Coursera		

E-books

- Computational Linguistics: Models, Resources, Applications
- Speech and Language Processing (Dan Jurafsky, et al)
- O'Reilly® Natural Language Processing with Python

Important links:

Savitribai Phule Pune University, Pune
ME Data Sciences (2020 Course)
610303D: Video Analytics

Teaching Scheme:	Credit	Examination Scheme:
TH: 04hr/week	04	Mid Semester: 50 Marks End Semester: 50 Marks

Prerequisite Courses: Linear Algebra/Probability Review/Matrix theory

Companion Course: Image processing, Machine Learning, Deep Learning

Course Objectives:

1. To understand image preprocessing, post processing methods
2. To learn various object recognition methods
3. Study various types of camera, camera models and understand 3D vision and its geometry
4. To use appropriate motion analysis methods for real world computer vision applications, understand different motion analysis methods.

Course Outcomes:

On completion of the course, learner will be able to:

CO1:Identify and analyze suitable methods of Image low level and high level processing such as, Image preprocessing, Image Enhancement, Image segmentation, feature extraction for given case

CO2:Study/computer vision applications

CO3:Model and apply various camera model to obtain 3D vision

CO4: Design and analyze various motion analysis methods for real world computer vision applications.

CO5:Apply various object recognition methods for computer vision real time applications

CO6:Identify and analyze various intelligent video analytics use cases

Selection of Modules: Modules 1 to 3 are compulsory and select any one from modules 4 and 5.

Module I	Introduction to Computer Vision	Credit 1

Motivation, Relationships to other fields, Image preprocessing, Image Enhancement, Image segmentation, Feature Extraction: Shape representation and description : Contour-based shape representation and description, region based shape representation and description, statistical and syntactic texture description methods

Camera Models :Cameras: Pinhole cameras, cameras with lenses, the Human eye, Sensing, 3D cameras. 3D vision tasks, Basic of projective geometry, A single perspective camera, two cameras, stereopsis, Use of 3D vision: Shape from X

Case Studies (if any)	4) An optic music recognition (OMR) system/ Automated image analysis in cardiology 5) Automated identification of airway tree
Mapping of Course Outcomes	**CO1,CO2**

Module II	Motion Analysis	Credit 1

Differential Motion Analysis methods, Change detection, Segmentation using motion, Image flow, segmentation using Moving camera, Optical flow, Analysis based on correspondence of interest points, detection of specific motion patterns, video tracking, motion models to aid tracking

Case Studies(if any)	Visual surveillance system/Crowd detection
Mapping of Course Outcomes	**CO3**

Module III	Object Recognition	Credit 1

Knowledge representation, Statistical Pattern Recognition, Neural Nets, Syntactic pattern recognition, Recognition as graph matching, Optimization techniques in recognition, fuzzy systems, texture recognition methods

Case Studies(if any)	Face Mask Detection

Mapping of Course Outcomes	**CO4**	
Module IV	**Intelligent Video Analytics**	**Credit 1**
Real-time video analytics and video mining, temporal and spatial event recognition,Vision-based activity recognition,Behaviour Analysis,Content-Based Analysis of Digital Video		
Case Studies(if any)	Traffic controlling systems at airport/Vehicle counting	
Mapping of Course Outcomes	**CO3,CO4,CO5**	
Unit V	**Video Analytics: State-of-the-art and the Future**	**Credit 1**
Video Analytics: state of the art applications with reference to computer vision applications, Deep learning in video analytics, Human motion recognition and its applications, Video Analytics for Business Intelligence, Virtual reality/Augmented reality applications, and Healthcare applications.		
Case Studies(if any)	Bicycle detection with the deep learning	
Mapping of Course Outcomes	**CO3,CO4,CO5**	

Books & Other Resources:
White paper: Video Analytics: Technologies and use cases
https://wso2.com/whitepapers/innovating-with-video-analytics-technologies-and-use-cases/#07

Text Books:

1. Sonka, Hlavac, Boyle, "Digital Image Processing and Computer Vision"- CENGAGE Learning, Indian Edition
2. Ramesh Jain, Kasturi, Schunck, "Machine Vision",McGraw-Hill

Reference Books:

1) Milan Sonka, Vaclav Hlavac, Roger Boyle, "Image Processing, Analysis, and Machine Vision"(2nd edition), Thomson Learning
2) David Forsyth, Jean Ponce, "Computer Vision", Pearson Education
3) Jan Eril Solem, "Programming Computer Vision with python", O'REILLY
4) Video Analytics for Business Intelligence, Editors: Caifeng Shan, Fatih Porikli, Tao Xiang, Shaogang Gong ((https://link.springer.com/book/10.1007/978-3-642-28598-1)

MOOC Courses:

1) Introduction to Computer Vision with Watson and OpenCV (Coursera)
2) Introduction to Intel® Distribution of OpenVINO™ toolkit for Computer Vision Applications (Coursera)
3) Computer Vision Basics (Coursera)
4) Deep Learning in Computer Vision (Coursera)
5) Fundamentals of Digital Image and Video Processing (Coursera)

E-books:

- http://vision.stanford.edu/teaching/cs131_fall1718/files/cs131-class-notes.pdf (90% syllabus coverage)
- https://www.researchgate.net/publication/337293786_Modern_Deep_Learning_and_Advanced_Computer_Vision_Book
- http://programmingcomputervision.com/downloads/ProgrammingComputerVision_CCdraft.pdf (for assignment)
- Content-Based Analysis of Digital Video (Kindle Edition)by Alan Hanjalic (Author) Publisher: Springer; 2004 edition (8 May 2007)

Important links:

Savitribai Phule Pune University, Pune
ME Data Sciences (2020 Course)
610304: Industry Internship-I/ In house Research Project - I

Teaching Scheme:	Credit	Examination Scheme:
PR: 03 Hr/Week	03	TW: 50 Marks OR/PRE : 50 Marks

Prerequisite Courses:

Companion Course:

Course Objectives:

- To identify the domain of research
- To learn to communicate in a scientific language through collaboration with a guide.
- To categorize the research material confined to the domain of choice

Course Outcomes:
On completion of the course, learner will be able to–

CO1:Conduct thorough literature survey confined to the domain of choice

CO2:Develop presentation skills to deliver the technical contents

CO3:Furnish the report of the technical research domain

CO4:Analyze the findings and work of various authors confined to the chosen domain

Conduction guidelines

The preferences/choices of the domain will be taken from the students. The guide needs to be allocated based on the preference/choices. The research project should be assigned to students. In case of Industry Internship-I, the assigned guide from college has to monitor and evaluate the progress of the student. The student has to exhibit the continuous progress through regular reporting and presentations and proper documentation. The continuous assessment of the progress needs to be documented unambiguously.

Savitribai Phule Pune University, Pune
ME Data Sciences (2020 Course)
610305- Dissertation Stage I

Teaching Scheme:	Credit	Examination Scheme:
TH: 08 hr/week	**08**	**Mid Semester: 50 Marks** **End Semester: 50 Marks**

Prerequisite Courses:

Companion Course:

Course Objectives:

1. To identify the domain of research
2. To learn to communicate in a scientific language through collaboration with a guide.
3. To understand the various means of technical publications and terminologies associated with publications
4. To categorize the research material confined to the domain of choice
5. To formulate research problems with the help of the guide/mentor elaborating the research.
6. To acquire information independently and assess its relevance for answering the research questions.

Course Outcomes:

On completion of the course, learner will be able to–

CO1: Conduct thorough literature survey confined to the domain of choice
CO2:Develop presentation skills to deliver the technical contents
CO3: Furnish the report of the technical research domain
CO4:Analyze the findings and work of various authors confined to the chosen domain

Dissertation Stage–I is an integral part of the Dissertation work. In this, the student shall complete the partial work of the Dissertation which will consist of problem statement, literature review, design, scheme of implementation (Mathematical Model/SRS/UML/ERD/block diagram/ PERT chart,) and Layout & Design of the Set-up.

The student is expected to complete the dissertation at least up to the design phase. As a part of the progress report of Dissertation work Stage-I, the candidate shall deliver a presentation on the advancement in Technology pertaining to the selected dissertation topic. The student shall submit the duly approved and certified progress report of Dissertation Stage-I in standard format for satisfactory completion of the work by the concerned guide and head of the Department/Institute.

The examiner will be assessed by a panel of examiners of which one is necessarily an external examiner. The assessment will be broadly based on literature study, work undergone, content delivery, presentation skills, documentation and report.

The students are expected to validate their study undertaken by publishing it at standard platforms.

The investigations and findings need to be validated appropriately at standard platforms – conference and/or peer reviewed journal.

The student has to exhibit the continuous progress through regular reporting and presentations and proper documentation of the frequency of the activities at the sole discretion of the PG coordination.

The continuous assessment of the progress needs to be documented unambiguously. For standardization and documentation, it is recommended to follow the formats and guidelines circulated / as in the dissertation workbook approved by the Board of Studies. Follow guidelines and formats as mentioned in Dissertation Workbook.

Savitribai Phule Pune University, Pune
ME Data Sciences (2020 Course)
610307: Industry Internship-I/ In house Research Project – II

Teaching Scheme:	Credit	Examination Scheme:
TH: 03 hr/week	03	TW: 50 Marks OR/PRE : 50 Marks

Companion Course:

Course Objectives:

1. To identify the domain of research
2. To learn to communicate in a scientific language through collaboration with a guide.
3. To categorize the research material confined to the domain of choice

Course Outcomes:

On completion of the course, learner will be able to–

CO1:Conduct thorough literature survey confined to the domain of choice
CO2:Develop presentation skills to deliver the technical contents
CO3:Furnish the report of the technical research domain
CO4:Analyze the findings and work of various authors confined to the chosen domain

Conduction guidelines

Industry or research internship should include partial/complete project implementation. Student should be allocated to the research guide in first semester itself and same guide should be continued for the **:** Industry Internship-I/ In house Research Project – II. Otherwise the preferences/choices of the domain should be taken from the students. The guide needs to be allocated based on the preference/choices. The research project should be assigned to students. In case of Industry Internship-I, the assigned guide from college has to monitor and evaluate the progress of the student. The student has to exhibit the continuous progress through regular reporting and presentations and proper documentation. The continuous assessment of the progress needs to be documented unambiguously.

Savitribai Phule Pune University, Pune
ME Data Sciences (2020 Course)
610308: Dissertation Stage II

Teaching Scheme:	Credit	Examination Scheme:
PR: 18hr/week	18	TW: 150 Marks OR/PRE: 50 Marks

Course Objectives:

1. To follow SDLC meticulously and meet the objectives of proposed work
2. To test rigorously before deployment of system
3. To validate the work undertaken
4. To consolidate the work as furnished report

Course Outcomes:

On completion of the course, learner will be able to–

CO1:Show evidence of independent investigation

CO2:Critically analyze the results and their interpretation ; infer findings

CO3:Report and present the original results in an orderly way and placing the open questions in the right perspective.

CO4:Link techniques and results from literature as well as actual research and future research lines with the research.

CO5:Appreciate practical implications and constraints of the specialist subject

Guidelines:

In Dissertation Work Stage–II, the student shall consolidate and complete the remaining part of the dissertation which will consist of Selection of Technology, Installations, UML implementations, testing, Results, measuring performance, discussions using data tables per parameter considered for the improvement with existing/known algorithms/systems, comparative analysis, validation of results and conclusions. The student shall prepare the duly certified final report of Dissertation in standard format for satisfactory completion of the work by the concerned guide and head of the Department/Institute.

The students are expected to validate their study undertaken by publishing it at standard platforms.

The investigations and findings need to be validated appropriately at standard platforms – conference and/or peer reviewed journal.

The student has to exhibit continuous progress through regular reporting and presentations and proper documentation of the frequency of the activities in the sole discretion of the PG coordination.
The continuous assessment of the progress needs to be documented unambiguously.

It is recommended to continue with guidelines and formats as mentioned in the DissertationWorkbook approved by the Board of Studies.

Task Force at curriculum Design

1. Advisors, the team of Board of Studies-

Dr.Varsha Patil (Chairman) , Dr. Shirish Sane, Dr. Sunil Bhirud, Dr. Manik Dhore, Dr. Rajesh Prasad, Dr. Parikshit Mahalle, Dr. Pramod Patil, Dr. Geetanjali Kale, Dr. Sachin Lodha, Dr. Venkatesharan, , Dr. Suhasini Itkar, Dr. R. V. Patil and Dr. P. M. Yawalkar, Dr. Girish Khilari

2. Team Leader – Dr. Geetanjali Kale

3. Course Design Teams:

Sr. No.	Name of the Subject	Name of the Staff
1	Mathematical Foundation of Data Science	Dr. Prof. G. V. Kale Prof. H. P. Channe Prof. V. J. Damle
2	Data Science	Prof. A. G. Phakatkar Prof. Pranjali Joshi
3	Big Data Analytics	Prof. R. A. Kulkarni Prof. M. S. Wakode
4	Research Methodology	Prof. P. S. Game Dr. Prof. A. R. Buchade
5	Laboratory Proficiency-I	Prof. P. S. Game Prof. H. P. Channe
6	Data Storage Technologies and Networks (Elective I)	Dr. Prof. A. R.Buchade Prof. R. S. Paswan
7	Information Systems Management	Dr. Prof. S. D. Kale Prof. A. A. Chandorkar Prof. Y. A. Handge
8	Data Preparation and Analysis (Elective I)	Prof. M. S. Wakode
9	Artificial Intelligence for Data Science (Elective I)	Dr. Prof. A. R. Deshpande Dr. Prof. B. A. Sonkamble
10	Data Warehousing & Mining	Prof. K. C. Waghmare
11	Machine Learning	Dr.Prof.S.D.Kale Prof.P.P.Joshi
12	Soft Computing	Dr. Prof. A.R.Deshpande
13	Laboratory Proficiency-II	Prof.P.S.Game Dr.Prof.S.D.Kale
14	Distributed Databases (Elective II)	Prof.P.P.Joshi

15	Recommender Systems (Elective II)	Prof. V.V.Bagade
16	GPU COMPUTING (Elective II)	Prof. R.A.Kulkarni Prof.P.P.Joshi
17	Web Intelligence (Elective II)	Dr. Swati Bhavsar
18	Deep Learning	Prof. H.P. Channe
19	Data Modeling and Visualization	Prof. R.S.Paswan Dr. Prof.S.S.Sonawane
20	Real Time Analytics (Elective III)	Prof. V.V.Bagade
21	Business Analytics (Elective III)	Prof. R.S.Paswan Prof. K.C.Waghmare
22	Computational Linguistic Analytics (Elective III)	Prof. M.S.Takalikar
23	Video Analytics (Elective III)	Dr.Prof.A.S.Ghotkar