MANONMANIAM SUNDARANAR UNIVERSITY
TIRUNELVELI-627 012, TAMILNADU, INDIA

CENTRE FOR INFORMATION TECHNOLOGY AND ENGINEERING

M.Sc. Data Analytics

(CBCS-University Department)

Regulations, Scheme and Syllabus

For those who joined from the academic year 2018-2019 onwards

Submitted by
Chairman, BOS and Head,
Centre for Information Technology and Engineering,

to

The Registrar
Manonmaniam Sundaranar University
Tirunelveli - 12

MANONMANIAM SUNDARANAR UNIVERSITY
TIRUNELVELI-627 012, TAMILNADU, INDIA

Centre for Information Technology and Engineering
M.Sc. Data Analytics
(CBCS-University Department)

Regulations, Scheme and Syllabus
For those who joined from the academic year 2018-2019 onwards

Preamble:

Objective of the Course: M.Sc. degree programme in Data Analytics is designed to provide the students a long term career scope in the IT industry of present and forthcoming decades. Through the program, the students shall be exposed to the wide application possibilities of Data Analytics, the Data storage and processing platforms, the most popular tools and techniques for Data Analytics so that they acquire skills to develop real-time Data Analytics and Big Data Analytics projects to pursue career in contemporary IT industry and carry out research as time, environment and society dictates.

Curriculum Design strategy: Curriculum Highlights The M.Sc. degree programme in Data Analytics (M.Sc.DA) is a fine curriculum aimed squarely at producing graduates with the multi-faceted skills needed to draw insights from complex data sets, and to be able to communicate those insights effectively. The Super-Child of five major fields, Mathematics, Statistics, Computer Science, Information Technology and Business Management; it is the product of the interdisciplinary group of faculty members and practicing professionals involving mathematicians, statisticians, computer scientists, IT professionals, economists, operations researchers, and experts in various fields of business and management.

Uniqueness of M.Sc.DA programme: The M.Sc.DA is designed as a single, fully-integrated course of study—not a menu of core and elective courses—taught exclusively to students in the program. It is highly interactive. Students work together in teams and receive personalized coaching to improve their productivity. It is an intensive 20-month learning experience designed to immerse students into the acquisition of practical

knowledge and application of methods and techniques. The curriculum is carefully calibrated that would get updated continuously to meet the evolving challenges facing data scientists. The CITE department possesses classrooms, team rooms, study spaces, and other amenities under one roof, as well as the faculty and staff who are available to interact with students throughout the day.

Students Centric Curriculum Design: M.Sc.DA students can hone their skills working on challenging problems with actual data shared from business organizations. The Practical Project work spans eight months and culminates with an executive-level report and presentation to the industry sponsor. Students will work with leading industry-standard programming tools. M.Sc.DA students will work with every industry that include some of the world's leading organizations and best known brands.

Learning Freedom to Students through MOOCs: Courses can be done by a student on Massive Open Online Courses - MOOC platform SWAYAM, edX, etc that can be credit transferred to the course basket as equivalent to classroom based courses based on the BoS recommendations approved from time-to-time.

With two decades of experienced faculty and industry expertise, the curriculum has contents capable of producing superior student outcomes.

A. REGULATIONS/ PROGRAM SPECIFIC REQUIREMENTS

A1: Duration of the Course:

The M.Sc. DA programme is a 2 years full time programme spread over four semesters.

A2: Eligibility for Admission:

The minimum eligibility conditions for admission to the M.Sc. programme in Data Analytics are given below.

The candidates who seek admission into the first semester of the M.Sc. programme in Data Analytics course will be required to have passed the Bachelor's / Master's degree or equivalent in Science (B.Sc. /M.Sc.) or Engineering (B.E. /M.E.) or Management (BBA /MBA) or Commerce or Finance or Economics or Econometrics

from Manonmaniam Sundaranar University or any other Indian University or equivalent in any one of the following disciplines:

1.Statistics

2.Mathematics

3. Business Administration

4.Information Technology

5.Information Technology and E-Commerce

6.Computer Science

7.Computer Applications

8. Any other discipline with Statistics/Mathematics or Commerce /Finance with Applications of Statistics/Mathematics/Computer/IT as a subject.

A3: Credit Requirement for the Degree:

The general Regulations of the Choice Based Credit System programme of Manonmaniam Sundaranar University are applicable to this programme. The University requirement for the M.Sc. programme is completion of 90 credits of course work, out of which 5 credits should be through the mini project, 18 credits should be through the 4th semester main project work, remaining 64 credits should be through Core, Elective and Supportive Course papers. A Core course has 4 credits, elective, Supportive subjects weigh 3 credits and Practical subjects weigh 2 credits. No candidate will be eligible for the Degree of M.Sc.(Master of Science) in Data Analytics unless the candidate has undergone the prescribed courses of study for a period not less than 4 semesters and has acquired 90 credits and other passing requirements in all subjects of study. The marks, M_i obtained by the student in each subject, i shall be multiplied by the credit of that subject, C_i; such marks of all 'n' subjects are added up and divided by the total credit (90) to obtain the Consolidated Percentage of Marks.

$$\text{Consolidated Percentage of Marks} = \frac{\sum_{i=0}^{n} C_i \times M_i}{\sum_{i=0}^{n} C_i}$$

A5: Attendance Requirement:

A candidate will be permitted to appear for the semester examination only if the candidate keeps not less than 75 percent attendance. The University condonation rules are applicable for those who lack minimum of 75% attendance. The candidates with less than 60% attendance will have to repeat the concerned entire semester.

A6: Assessment

The assessment will comprise Continuous Internal Assessment (CIA) carrying a maximum of 25% marks and end-semester Examination carrying a maximum of 75% marks in each theory subject (Core/Elective/Supportive Course). For practical subjects, Internship and Major Project, the CIA is carried out for 25% marks and the External Assessment (Final Lab Exam, Lab Report, Viva-Voce for Practical Subjects and Final Project Presentation, Project Report, Viva-Voce for Internship and Major Project) is for 75% marks.

Semester examination will be conducted for all subjects of study, at the end of each Semester.

If a Student wants to carry out the final Major project work in 4^{th} semester in an IT company, the student can get permission from the concerned Project Supervisor and Head of the Department after submitting the Acceptance Letter from the IT Company.

Assessing and Transferring credits earned from a MOOC Course

To assess and transfer the number of credits to be given to a MOOC course done by a candidate, a Viva Voce has to be conducted for the candidate by a department committee chaired by the Head of the department with two other faculty members from the department. On successful completion of the Viva Voce, the department committee will recommend the credit transfer of the course to the office of the Controller of Examinations.

A7: Passing Requirements

A candidate who secures not less than 50 percent marks in end-semester examination and not less than 50 percent of the total marks (Continuous Internal Assessment + end-semester examination) in any subject of study will be declared to have passed the subject.

A Candidate who successfully completes the course and satisfies the passing requirements in all the subjects of study and curricular requirements will be declared to have qualified for the award of the Degree.

A8: Classification of successful candidates

The candidates who passed written papers, practical papers and Projects shall be classified as follows. Total Marks secured in written papers, practical papers and Project work altogether put as overall percentage along with the credits.

The classification is as follows,

Marks Overall %	Classification
1. 75% and above with a First attempt Pass in all subjects	I Class with Distinction
2. i) 75% above from multiple attempts	I Class
ii) 60% to below 75%	I Class
3. 50% to below 60%	II Class

A9: Academic Session

The academic year normally begins in July every year and ends in April.

These regulations will come into effect from the academic year 2018-2019 onwards.

PROGRAMME EDUCATIONAL OBJECTIVES (PEO)

The objectives of the program are to:

PEO1. To Establish Data Analytics as the need of the hour for IT industries.

PEO2. To acquire knowledge in the basics of Data Analytics,

PEO3. To acquire knowledge in the statistical and mathematical concepts, tools and techniques for Data Analytics

PEO4. To empower the student to gain specialist expertise in at least one of the domain areas of Data Analytics like Social Media Data Analytics, Marketing Data Analytics, Financial data Analytics etc. using Descriptive, Diagnostic, Predictive and Prescriptive approaches in Data Analytics.

PEO5. To acquire knowledge in the necessary ethical skills as a Data Analytics professional or Scientist.

PEO6. To understand the need for life-long learning as an IT professional.

PROGRAMME SPECIFIC OUTCOMES (PSO):

PSO1. Analytics Knowledge: Able to identify the scope of Data Analytics in a particular domain.

PSO2. Problem Analysis: Able to identify, plan, design and complete Data Analytics Projects.

PSO3. Modern Tool usage: Able to take up Data Analytics as time, environment and society dictates using modern tools and techniques.

PSO4. Ethics: Able to understand his/her ethical responsibility.

PSO5. Research: Competent to solve complex problems using research based knowledge, methods and techniques.

PSO6. Specific and Team work solving: Ability to understand each phase in the data analytics projects and work as an individual or a group to deliver a best analysis.

PEO vs. PSO Mapping

	PSO1	PSO2	PSO3	PSO4	PSO5	PSO6
PE01	M	S	S	M	M	S
PE02	S	S	S	M	M	M
PE03	M	S	S	M	M	M
PE04	S	S	S	S	S	S
PE05	M	S	M	S	M	L
PE06	L	M	M	M	S	S

S- Strong, M – Middle, L – Low

GRADUATE ATTRIBUTES of M.Sc. DATA ANALYTICS

GA1. Deep Discipline Knowledge: Informed and trained in Data Analytics by cutting edge research, frame-structured throughout their program of studies acquired from personal interaction with research active educators and Industry Practitioner.

GA2. Critical Thinking and Problem Solving: Steeped in Data Analytics research methods and rigour based on empirical evidence and the scientific approach to knowledge development demonstrated through appropriate and relevant assessment. Acquire knowledge in the statistical and mathematical concepts, tools and techniques for Data Analytics

GA3. Teamwork and Communication Skills: Developed from, with, and via the honed through assessment and practice throughout the program of studies encouraged and valued in all aspects of learning.

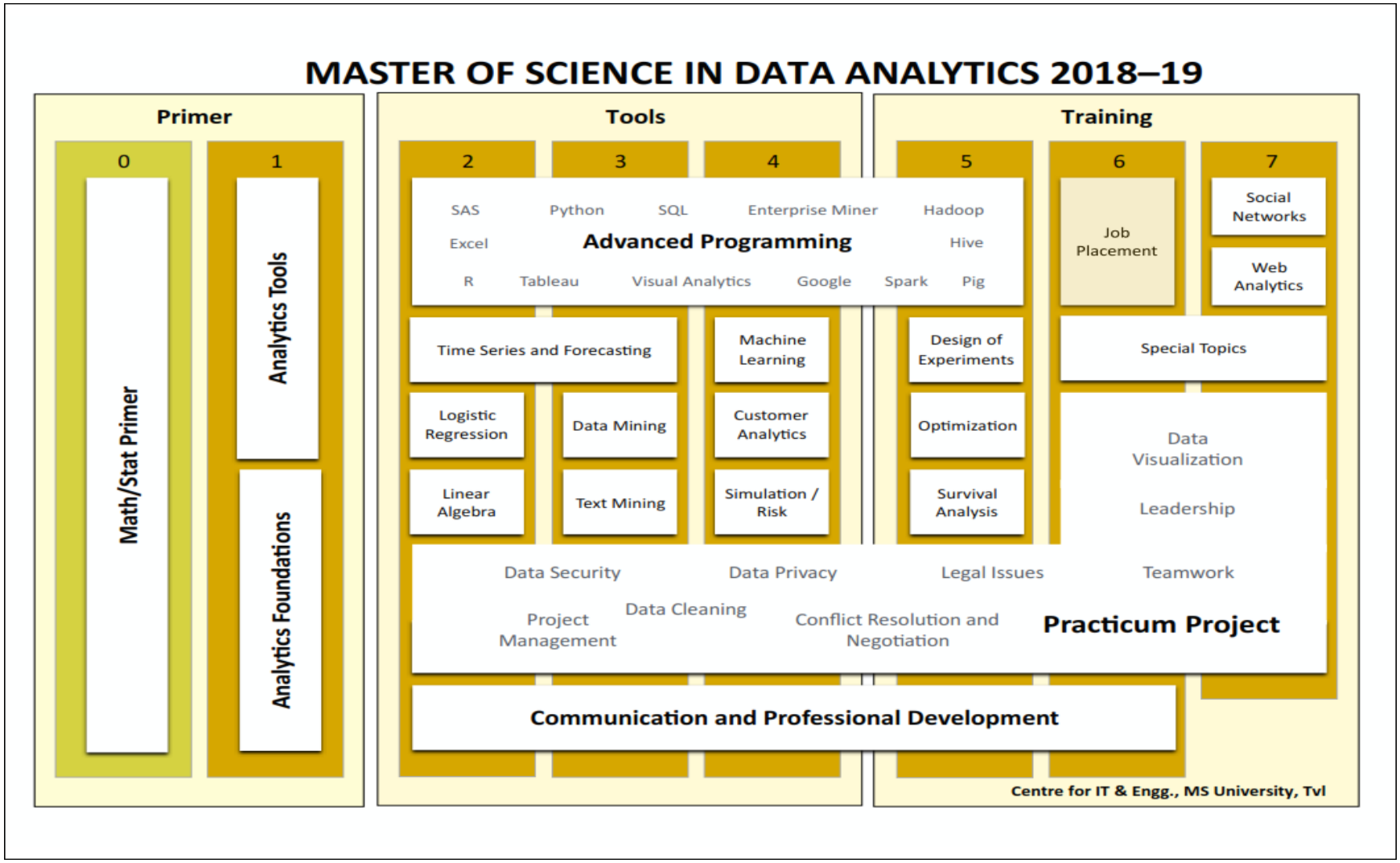
GA4. Career and Leadership Readiness: Technology savvy professional and, where relevant, fully accredited forward thinking and well informed tested and validated by work based experiences. Self-Transformed to gain specialist expertise in at least one of the domain areas of Data Analytics like Social Media Data Analytics, Marketing Data Analytics, Financial data Analytics etc. using Descriptive, Diagnostic, Predictive and Prescriptive approaches in Data Analytics.

GA5. **Intercultural and Ethical Competency:** Adept at operating in other cultures comfortable with different nationalities and social contexts; Able to determine and contribute to desirable social outcomes, demonstrated by study abroad or with an understanding of indigenous knowledges. Acquire knowledge in the necessary ethical skills as a Data Analytics professional or Decision Scientist

GA6. Self-awareness and Emotional Intelligence: A capacity for self-reflection and a willingness to engage in self-appraisal; Open to objective and constructive feedback from supervisors and peers; Able to negotiate difficult social situations, diffuse conflict and engage positively in purposeful debate. Understand the need for life-long learning as an IT professional.

PEO / GA	PEO1	PEO2	PEO3	PEO4	PEO5	PEO6
GA1						
GA2						
GA3						
GA4						
GA5						
GA6						

Course Framework for M.Sc. Data Analytics



Scheduling of Courses for M.Sc. DATA ANALYTICS – For Students admitted from the Academic year 2018-19.

Semester	Theory						Practical		Special Courses	Credit
1.	An Introduction to Data Analytics (4)	Mathematics for Data Analytics (4)	Statistical Learning for Data Analytics (4)	No SQL Database Management (4)	Data Visualization for Analytics using R and Python(4)		NoSQL Database Management Laboratory (2)	Data Visualization for Analytics Laboratory using R and Python (2)		24
2.	Machine Learning Algorithms for Data Analytics (4)	Descriptive and Discovery Analytics (4)	Elective 1 (3)	Elective 2 (3)	Elective 3 – (3)	Supportive Course 1 (3)	Machine Learning Laboratory (2)	Descriptive and Discovery Analytics Laboratory (2)		24
3.	Predictive and Prescriptive Analytics (4)	Professional ethics (4)	Elective4 (3)	Elective 5 - (3)		Supportive Course 2 (3)	Predictive and Prescriptive Analytics Laboratory (2)		Internship / Industrial Training (2)	21
4.	Elective 6 (3)		Elective7 (3)	Elective 8 (3)			Dissertation / Major Project work (12)			21

B.SCHEME FOR M.Sc. (BIG DATA ANALYTICS)

(For those who joined from the academic year 2018-2019 onwards)

Sem-ester	Title of the Subject	Status*	Hrs/ week	Credits	Maximum Marks			Passing Minimum Percentage	
					Inte-rnal	Exte-rnal	Total	Exter nal	Total
FIRST SEMESTER									
I	**MDAC11-**An Introduction to Data Analytics	C	4	4	25	75	100	50	50
I	**MDAC12-**Mathematics for Data Analytics	C	4	4	25	75	100	50	50
I	**MDAC13-**Statistical Learning for Data Analytics	C	4	4	25	75	100	50	50
I	**MDAC14-**No SQL Database Management	C	4	4	25	75	100	50	50
I	**MDAC15-**Data Visualization for Analytics using R and Python	C	4	4	25	75	100	50	50
I	**MDAL11-**NoSQL Database Management Laboratory	L	2	2	25	75	100	50	50
I	**MDAL12-**Data Visualization for Analytics Laboratory using R and Python	L	2	2	25	75	100	50	50
I Semester Total Credits				**24**					
SECOND SEMESTER									
II	Supportive Course I	S	3	3	25	75	100	50	50
II	**MDAC21-**Machine Learning Algorithms for Data Analytics	C	4	4	25	75	100	50	50
II	**MDAC22-**Descriptive and Discovery Analytics	C	4	4	25	75	100	50	50
II	Elective 1	E	3	3	25	75	100	50	50
II	Elective 2	E	3	3	25	75	100	50	50
II	Elective 3	E	3	3	25	75	100	50	50
II	**MDAL21-**Machine Learning Laboratory	L	2	2	25	75	100	50	50
II	**MDAL22-**Descriptive and Discovery Analytics Laboratory	L	2	2	25	75	100	50	50
II Semester Total Credits				**24**					
THIRD SEMESTER									
III	Supportive Course II	S	3	3	25	75	100	50	50

III	**MDAC31-**Predictive and Prescriptive Analytics	C	4	4	25	75	100	50	50
III	**MDAC32-**Ethics for Data Scientists	C	4	4	25	75	100	50	50
III	Elective 4	E	3	3	25	75	100	50	50
III	Elective 5	E	3	3	25	75	100	50	50
III	**MDAL31-**Predictive and Prescriptive Analytics Laboratory	L	2	2	25	75	100	50	50
III	**MDAI31-** Internship	P	5	2	25	75	100	50	50
III Semester Total Credits				**21**					
FOURTH SEMESTER									
IV	Elective 6	E	3	3	25	75	100	50	50
IV	Elective 7	E	3	3	25	75	100	50	50
IV	Elective8	E	3	3	25	75	100	50	50
IV	**MDAP41-**Major Project	P	10	12	25	75	100	50	50
IV Semester Total Credits				**21**					
OVERALL TOTAL CREDITS				**90**					

***C-Core, L-Lab, S-Supportive, E-Elective, P-Project**

Subjects for Electives 1,4									
Sl. No.	**Title of the Subject**	**Status**	**Hrs/week**	**Credits**	**Maximum Marks**			**Passing Minimum Percentage**	
					Inte-rnal	**Exte-rnal**	**Total**	**External**	**Total**
A1	**MDAEA-**Natural Language Processing for Data Analytics	E	3	3	25	75	100	50	50
A2	**MDAEB-** Big Data Analytics on Genomic Data	E	3	3	25	75	100	50	50
A3	**MDAEC-**Cloud Computing for Data Analytics	E	3	3	25	75	100	50	50
A4	**MDAED-**Programming with Mangodb for Data Analytics	E	3	3	25	75	100	50	50
Subjects for Electives 2,5,									
B1	**MDAEE-**Data Mining Essentials	E	3	3	25	75	100	50	50

	for Data Analytics								
B2	**MDAEF**-Latest Machine Algorithms for Data Analytics	E	3	3	25	75	100	50	50
B3	**MDAEG**-Big Data Security	E	3	3	25	75	100	50	50
B4	**MDAEH**-Programming with Cassandra	E	3	3	25	75	100	50	50
Subjects for Electives 3,6									
C1	**MDAEI**-HADOOP and Data Analytics	E	3	3	25	75	100	50	50
C2	**MDAEJ**-Storm for Data Analytics	E	3	3	25	75	100	50	50
C3	**MDAEK**-Spark for Data Analytics	E	3	3	25	75	100	50	50
C4	**MDAEL**-SPSS for Data Analytics	E	3	3	25	75	100	50	50
Subjects for Electives 7,8									
D1	**MDAEM**-Web Frameworks for Big Data Projects	E	3	3	25	75	100	50	50
D1	**MDAEN**-Real-Time Big Data Analytics	E	3	3	25	75	100	50	50
D2	**MDAEO**-Challenges and Research Issues in Big Data Projects	E	3	3	25	75	100	50	50
D4	**MDAEP**-Big Data in Computer Vision	E	3	3	25	75	100	50	50

M.Sc DATA ANALYTICS DEGREE PROGRAMME
C.SYLLABUS FOR M.Sc. in DATA ANALYTICS

SEMESTER I

LIST OF COURSES
(For The Candidates Admitted From 2017-18 Onwards)

Sl. No.	Subject code	Course name
1.	MDAC11	An Introduction to Data Analytics
2.	MDAC12	Mathematics for Data Analytics
3.	MDAC13	Statistical Learning for Data Analytics
4.	MDAC14	No SQL Database Management
5.	MDAC15	Data Visualization for Analytics using R and Python
6.	MDAL21	Machine Learning Laboratory using Matlab
7.	MDAL22	Descriptive and Discovery Analytics Laboratory

MDAC11	AN INTRODUCTION TO DATA ANALYTICS	Category	L	P	T	Credit
		PE	4	0	0	4

Preamble

To get introduced to the basic concepts in Data analytics. To understand the basic processing, storage and programming models for Data Analytics. To learn the different phases of Data Analytics Project.

Prerequisite

- Advanced Microsoft Excel Operations

Course Outcomes

On the successful completion of the course, students will be able to

Course Outcomes		Bloom's Level
CO1	Able to identify the scope of Data Analytics Project	Understanding
CO2	Able to define the basic hardware and programming requirements needed for a Data Analytics Project.	Apply
CO3	Able to plan the different phases of a Data Analytics Project	Understanding

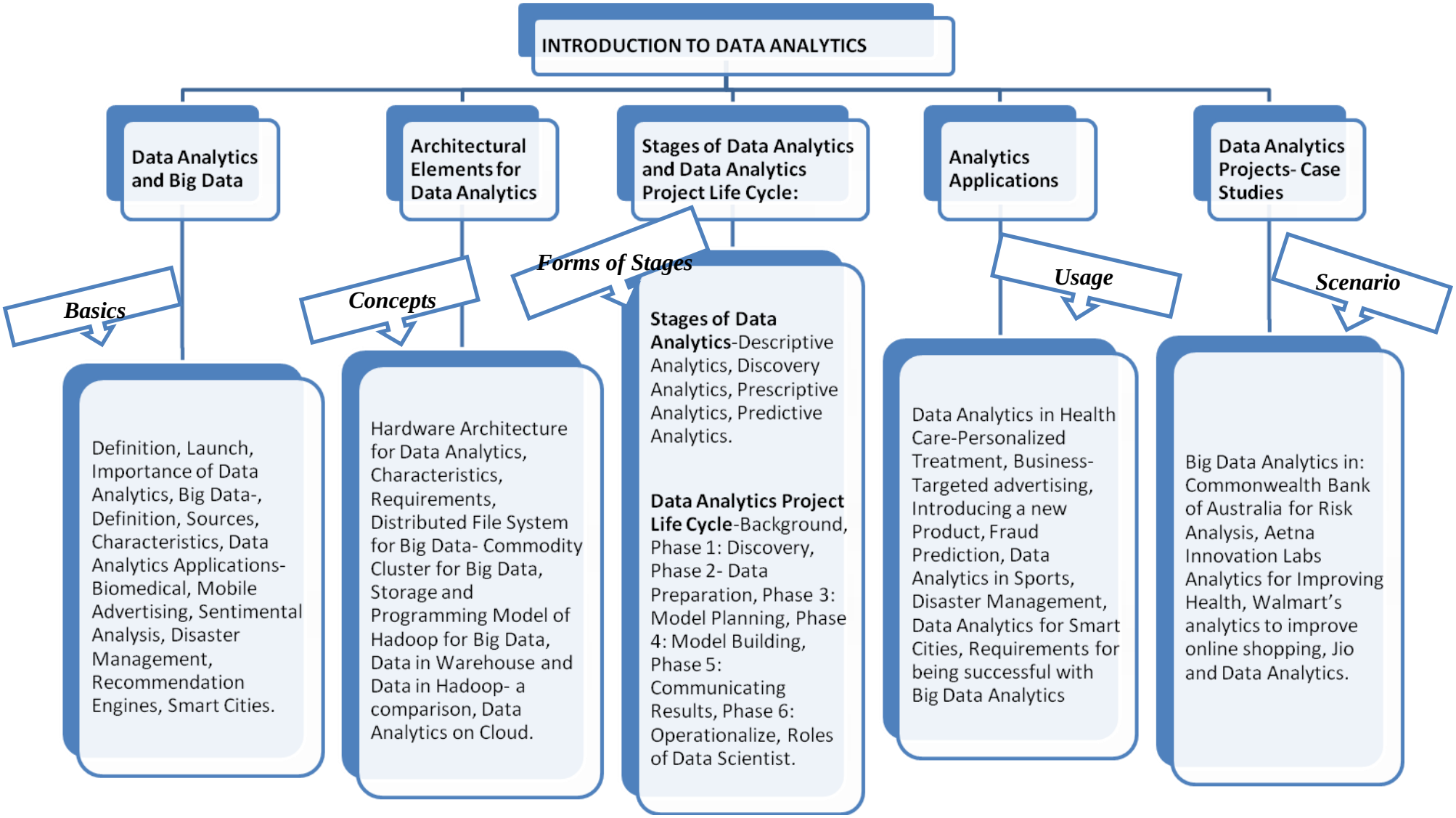
Mapping with Programme Outcomes

Cos	PSO1	PSO2	PSO3	PSO4	PSO5	PSO6
CO1	S			L		
CO2	M	S			L	
CO3		S	M			M

Assessment Pattern

Bloom's Category	Continuous Assessment Tests (25)			Terminal Examination (75)
Remember	5	5	5	20
Understand	10	10	10	20
Apply	5	5	5	10
Analyze	5	5	5	10
Evaluate	0	0	0	5
Create	0	0	0	0

Concept Map


INTRODUCTION TO DATA ANALYTICS
Data Analytics and Big Data
Architectural Elements for Data Analytics
Stages of Data Analytics and Data Analytics Project Life Cycle:
Analytics Applications
Data Analytics Projects- Case Studies
Basics
Concepts
Forms of Stages
Usage
Scenario
Definition, Launch, Importance of Data Analytics, Big Data-, Definition, Sources, Characteristics, Data Analytics Applications- Biomedical, Mobile Advertising, Sentimental Analysis, Disaster Management, Recommendation Engines, Smart Cities.
Hardware Architecture for Data Analytics, Characteristics, Requirements, Distributed File System for Big Data- Commodity Cluster for Big Data, Storage and Programming Model of Hadoop for Big Data, Data in Warehouse and Data in Hadoop- a comparison, Data Analytics on Cloud.
Stages of Data Analytics-Descriptive Analytics, Discovery Analytics, Prescriptive Analytics, Predictive Analytics.
Data Analytics Project Life Cycle-Background, Phase 1: Discovery, Phase 2- Data Preparation, Phase 3: Model Planning, Phase 4: Model Building, Phase 5: Communicating Results, Phase 6: Operationalize, Roles of Data Scientist.
Data Analytics in Health Care-Personalized Treatment, Business-Targeted advertising, Introducing a new Product, Fraud Prediction, Data Analytics in Sports, Disaster Management, Data Analytics for Smart Cities, Requirements for being successful with Big Data Analytics
Big Data Analytics in: Commonwealth Bank of Australia for Risk Analysis, Aetna Innovation Labs Analytics for Improving Health, Walmart's analytics to improve online shopping, Jio and Data Analytics.

Syllabus

UNIT I Data Analytics and Big Data:

Data Analytics- Definition, Launch, Importance of Data Analytics, Big Data-, Definition, Sources, Characteristics, Data Analytics Applications-Biomedical, Mobile Advertising, Sentimental Analysis, Disaster Management, Recommendation Engines, Smart Cities. (11hrs)

UNIT II Architectural Elements for Data Analytics:

Hardware Architecture for Data Analytics, Characteristics, Requirements, Distributed File System for Big Data- Commodity Cluster for Big Data, Storage and Programming Model of Hadoop for Big Data, Data in Warehouse and Data in Hadoop- a comparison, Data Analytics on Cloud. (10hrs)

UNIT III Stages of Data Analytics and Data Analytics Project Life Cycle:

Stages of Data Analytics-Descriptive Analytics, Discovery Analytics, Prescriptive Analytics, Predictive Analytics.

Data Analytics Project Life Cycle-Background, Phase 1: Discovery, Phase 2- Data Preparation, Phase 3: Model Planning, Phase 4: Model Building, Phase 5: Communicating Results, Phase 6: Operationalize, Roles of Data Scientist. (14hrs)

UNIT IV Analytics Applications : Data Analytics in Health Care-Personalized Treatment, Business-Targeted advertising, Introducing a new Product, Fraud Prediction, Data Analytics in Sports, Disaster Management, Data Analytics for Smart Cities, Requirements for being successful with Big Data Analytics. (12hrs)

UNIT V Data Analytics Projects- Case Studies: Big Data Analytics in: Commonwealth Bank of Australia for Risk Analysis, Aetna Innovation Labs Analytics for Improving Health, Walmart's analytics to improve online shopping, Jio and Data Analytics. (13hrs)

Books:

1. "A Guide to Big Data Analytics", Datameer.com.
2. Vignesh Prajapati, "Big Data Analytics with R and Hadoop", Packt Publications.
3. D. Dietrich, B.Heller, B.Yang, "Data Science and Big Data Analytics", EMC Education Services.
4. "Big Data Now", O'Reily Inc.
5. DeWitt, S. Madden, and M. Stonebraker, "A Comparison of Approaches to Large-Scale Data Analysis", SIGMOD Conference 2009.
6. Steven Cooper, "Data Science from Scratch", Data Science.
7. www.mercerindustries.com/wp-content/uploads/2015/02/Watson-Tutorial-Big-Data Business-Analytics
8. www.researchgate.net/publication/273961581_Big_Data_analytics_with_applications
9. http://cs.ulb.ac.be/conferences
10. www.businessesgrow.com/2016/12/06/big-data-case-studies
11. Eiman Al Nuaimi, Hind Al Neyadi, Nader Mohamed , 'Applications of big data to smart cities', Journal of Internet Services and Applications.
12. Samiya Khan1 , Kashish Ara Shakil and Mansaf Alam, 'Cloud-based Big Data Analytics – A Survey of Current Research and Future Directions
13. Bogdan Ionescu, Dan Ionescu, Cristian Gadea, Bogdan Solomon, and Mircea Trifan, "An Architecture and Methods for Big Data Analysis"

Course Contents Coverage and Lecture Schedule

S.No.	Topic	No. of Lectures
1.	**Data Analytics and Big Data:**	
1.1	Definition, Launch,	3
1.2	Importance of Data Analytics, Big Data-, Definition, Sources,	3
1.3	Characteristics, Data Analytics Applications	3
1.4	Biomedical, Mobile Advertising,	3
1.5	Sentimental Analysis, Disaster Management	3
1.6	Recommendation Engines, Smart Cities	3
2.	**Architectural Elements for Data Analytics:**	
2.1	Hardware Architecture for Data Analytics, Characteristics, Requirements,	2
2.2	Distributed File System for Big Data- Commodity Cluster for Big Data,	3
2.3	Storage and Programming Model of Hadoop for Big Data,	2
2.4	Data in Warehouse and Data in Hadoop- a comparison, Data Analytics on Cloud.	2
3.	**Stages of Data Analytics and Data Analytics Project Life Cycle:**	
3.1	**Stages of Data Analytics**-Descriptive Analytics, Discovery Analytics,	2
3.2	Prescriptive Analytics, Predictive Analytics.	2
3.3	**Data Analytics Project Life Cycle**-Background, Phase 1: Discovery, Phase 2- Data Preparation,	3
3.4	Phase 3: Model Planning, Phase 4: Model Building, Phase 5: Communicating Results,	2
3.5	Phase 6: Operationalize, Roles of Data Scientist.	2
4.	**Analytics Applications**	
4.1	Data Analytics in Health Care-Personalized Treatment, Business-Targeted advertising,.	3
4.2	Introducing a new Product, Fraud Prediction, Data Analytics in Sports,	3
4.3	Disaster Management, Data Analytics for Smart Cities, Requirements for being successful with Big Data Analytics	2
5.	**Data Analytics Projects- Case Studies:**	
5.1	Big Data Analytics in: Commonwealth Bank of Australia for Risk Analysis,	3
5.2	Aetna Innovation Labs Analytics for Improving Health,	3
5.3	Walmart's analytics to improve online shopping, Jio and Data Analytics.	3

MDAC12	MATHEMATICS FOR DATA ANALYTICS	Category	L	P	T	Credit
		PE	4	0	0	4

Preamble

To understand the need of Mathematical concepts in the different stages of Data Analytics. To acquire knowledge in different Optimization and Linear Algebra concepts towards inference and prediction stages of Data Analytics. To acquire skills in a math processing tool.

Prerequisite

- Foundations of Mathematics

Course Outcomes

On the successful completion of the course, students will be able to

Course Outcomes		Bloom's Level
CO1	Able to define the basic optimization and linear algebra concepts for prediction and inference	Understanding
CO2	Able to apply the right optimization technique at the Predictive stage of a Data Analytics project.	Apply
CO3	Able to acquire skills to program optimization and linear real world analytics problems in Octave.	Understanding

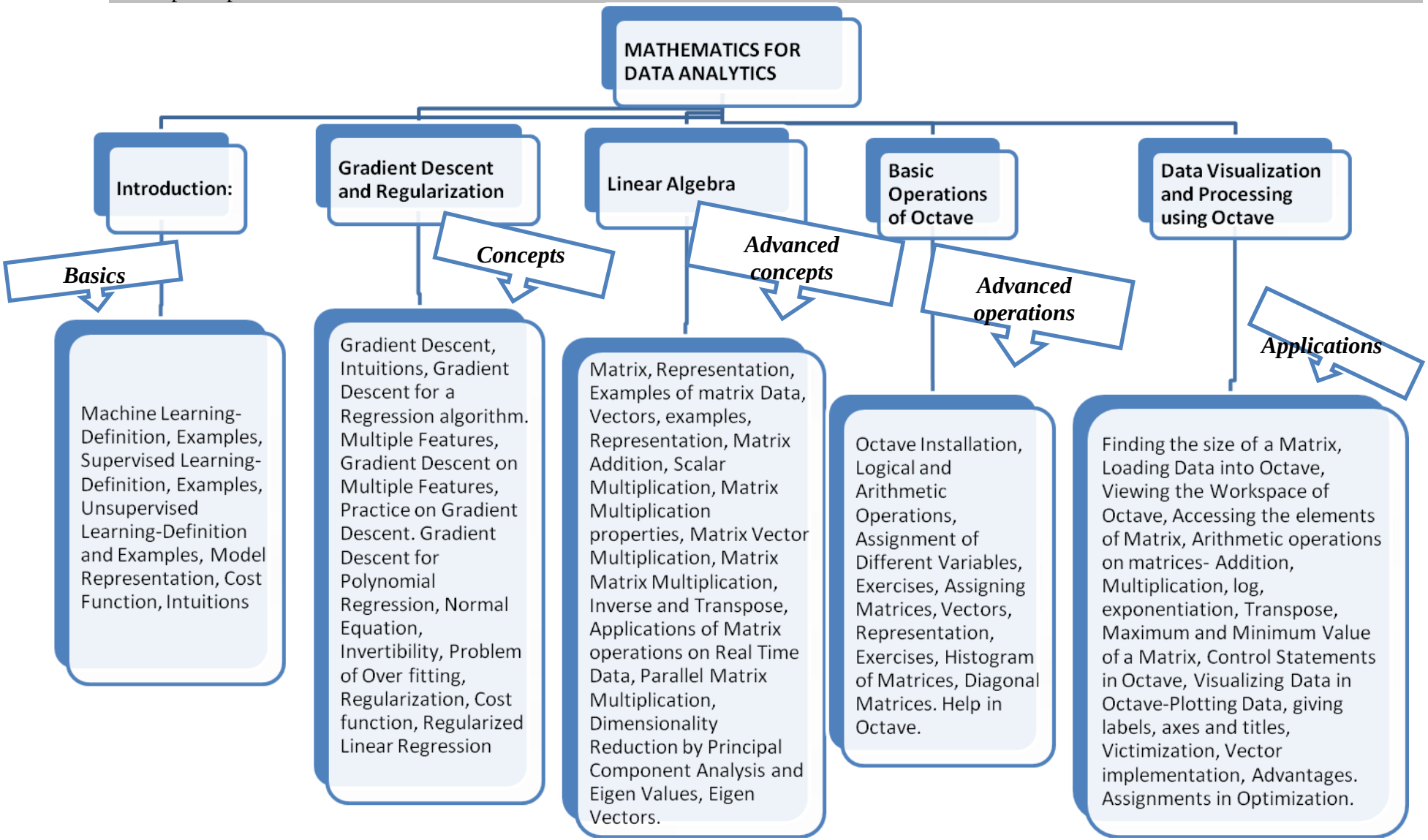
Mapping with Programme Outcomes

Cos	PSO1	PSO2	PSO3	PSO4	PSO5	PSO6
CO1	S					M
CO2	M				S	
CO3		S	M	L		

Assessment Pattern

Bloom's Category	Continuous Assessment Tests (25)			Terminal Examination (75)
Remember	5	5	5	20
Understand	10	10	10	20
Apply	5	5	5	10
Analyze	5	5	5	10
Evaluate	0	0	0	5
Create	0	0	0	0

Concept Map


MATHEMATICS FOR DATA ANALYTICS
Introduction:
Basics
Machine Learning-Definition, Examples, Supervised Learning-Definition, Examples, Unsupervised Learning-Definition and Examples, Model Representation, Cost Function, Intuitions
Gradient Descent and Regularization
Concepts
Gradient Descent, Intuitions, Gradient Descent for a Regression algorithm. Multiple Features, Gradient Descent on Multiple Features, Practice on Gradient Descent. Gradient Descent for Polynomial Regression, Normal Equation, Invertibility, Problem of Over fitting, Regularization, Cost function, Regularized Linear Regression
Linear Algebra
Advanced concepts
Matrix, Representation, Examples of matrix Data, Vectors, examples, Representation, Matrix Addition, Scalar Multiplication, Matrix Multiplication properties, Matrix Vector Multiplication, Matrix Matrix Multiplication, Inverse and Transpose, Applications of Matrix operations on Real Time Data, Parallel Matrix Multiplication, Dimensionality Reduction by Principal Component Analysis and Eigen Values, Eigen Vectors.
Basic Operations of Octave
Advanced operations
Octave Installation, Logical and Arithmetic Operations, Assignment of Different Variables, Exercises, Assigning Matrices, Vectors, Representation, Exercises, Histogram of Matrices, Diagonal Matrices. Help in Octave.
Data Visualization and Processing using Octave
Applications
Finding the size of a Matrix, Loading Data into Octave, Viewing the Workspace of Octave, Accessing the elements of Matrix, Arithmetic operations on matrices- Addition, Multiplication, log, exponentiation, Transpose, Maximum and Minimum Value of a Matrix, Control Statements in Octave, Visualizing Data in Octave-Plotting Data, giving labels, axes and titles, Victimization, Vector implementation, Advantages. Assignments in Optimization.

Syllabus

UNIT I Introduction: Machine Learning-Definition, Examples, Supervised Learning-Definition, Examples, Unsupervised Learning-Definition and Examples, Model Representation, Cost Function, Intuitions. (12hrs)

UNIT II Gradient Descent and Regularization: Gradient Descent, Intuitions, Gradient Descent for a Regression algorithm. Multiple Features, Gradient Descent on Multiple Features, Practice on Gradient Descent. Gradient Descent for Polynomial Regression, Normal Equation, Invertibility, Problem of Over fitting, Regularization, Cost function, Regularized Linear Regression (10hrs)

UNIT III Linear Algebra: Matrix, Representation, Examples of matrix Data, Vectors, examples, Representation, Matrix Addition, Scalar Multiplication, Matrix Multiplication properties, Matrix Vector Multiplication, Matrix Matrix Multiplication, Inverse and Transpose, Applications of Matrix operations on Real Time Data, Parallel Matrix Multiplication, Dimensionality Reduction by Principal Component Analysis and Eigen Values, Eigen Vectors. (11hrs)

UNIT IV Basic Operations of Octave: Octave Installation, Logical and Arithmetic Operations, Assignment of Different Variables, Exercises, Assigning Matrices, Vectors, Representation, Exercises, Histogram of Matrices, Diagonal Matrices. Help in Octave. (13hrs)

UNIT V Data Visualization and Processing using Octave: Finding the size of a Matrix, Loading Data into Octave, Viewing the Workspace of Octave, Accessing the elements of Matrix, Arithmetic operations on matrices- Addition, Multiplication, log, exponentiation, Transpose, Maximum and Minimum Value of a Matrix, Control Statements in Octave, Visualizing Data in Octave-Plotting Data, giving labels, axes and titles, Victimization, Vector implementation, Advantages. Assignments in Optimization. (14hrs)

Books:

1. Lectures of Professor Dr. Andrew Ng, Stanford University, Coursera.
2. Gene H.Golub, Charles F.Van Loan, “Matrix Computations”, John Hopkins University Press.
3. https://skymind.ai/wiki/eigenvector
4. Randolf H. Reiss, B.S, “Eigen Values and Eigen Vectors in Data dimension Reduction for Regression”, San Marcos, Texas.
5. Gilbert Strang, “Linear Algebra and its Applications”, Thomson Learning Inc., 4th Edition.

Course Contents Coverage and Lecture Schedule

S.No.	Topic	No. of Lectures
1.	**Introduction**	
1.1	Machine Learning-Definition,	3
1.2	Supervised Learning-Definition, Examples,	3
1.3	Unsupervised Learning-Definition and Examples	3
1.4	Model Representation, Cost Function, Intuitions.	3
2.	**Gradient Descent and Regularization**	
2.1	Gradient Descent, Intuitions, Gradient Descent for a Regression algorithm.	3
2.2	Multiple Features, Gradient Descent on Multiple Features, Practice on Gradient Descent	3
2.3	Gradient Descent for Polynomial Regression, Normal Equation	3
2.4	Invertibility, Problem of Over fitting, Regularization, Cost function, Regularized Linear Regression	3
3.	**Linear Algebra:**	
3.1	Matrix, Representation, Examples of matrix Data, Vectors, examples, Representation, Matrix Addition,	3
3.2	Scalar Multiplication, Matrix Multiplication properties, Matrix Vector Multiplication,	3
3.3	Matrix Matrix Multiplication, Inverse and Transpose, Applications of Matrix operations on Real Time Data,	3
3.4	Parallel Matrix Multiplication, Dimensionality Reduction by Principal Component Analysis and Eigen Values, Eigen Vectors.	3
4.	**Basic Operations of Octave**	
4.1	Octave Installation, Logical and Arithmetic Operations, Assignment of Different Variables,	3
4.2	Exercises, Assigning Matrices, Vectors, Representation, Exercises, Histogram of Matrices, Diagonal Matrices. Help in Octave.	3
5.	**Data Visualization and Processing using Octave:**	
5.1	Finding the size of a Matrix, Loading Data into Octave, Viewing the Workspace of Octave,	3
5.2	Accessing the elements of Matrix, Arithmetic operations on matrices-Addition, Multiplication, log, exponentiation	3
5.3	Transpose, Maximum and Minimum Value of a Matrix, Control Statements in Octave,	3
5.4	Visualizing Data in Octave-Plotting Data, giving labels, axes and titles,	3
5.5	Victimization, Vector implementation, Advantages. Assignments in Optimization.	2

MDAC13	STATISTICAL LEARNING FOR DATA ANALYTICS	Category	L	P	T	Credit
		PE	4	0	0	4

Preamble

To understand the need of Statistical concepts in the different stages of Data Analytics. To acquire knowledge in different Descriptive, Inferential and Predictive concepts for relevant stages of Data Analytics. To acquire skills in using the right technique at the right stage of Data Analytics.

Prerequisite

- Basic Statistics

Course Outcomes

On the successful completion of the course, students will be able to

Course Outcomes		Level
CO1	Able to define the Descriptive, Inferential and Predictive concepts for relevant stages of Data Analytics	Understanding
CO2	Able to apply the right statistical technique right stage of a Data Analytics project	Apply

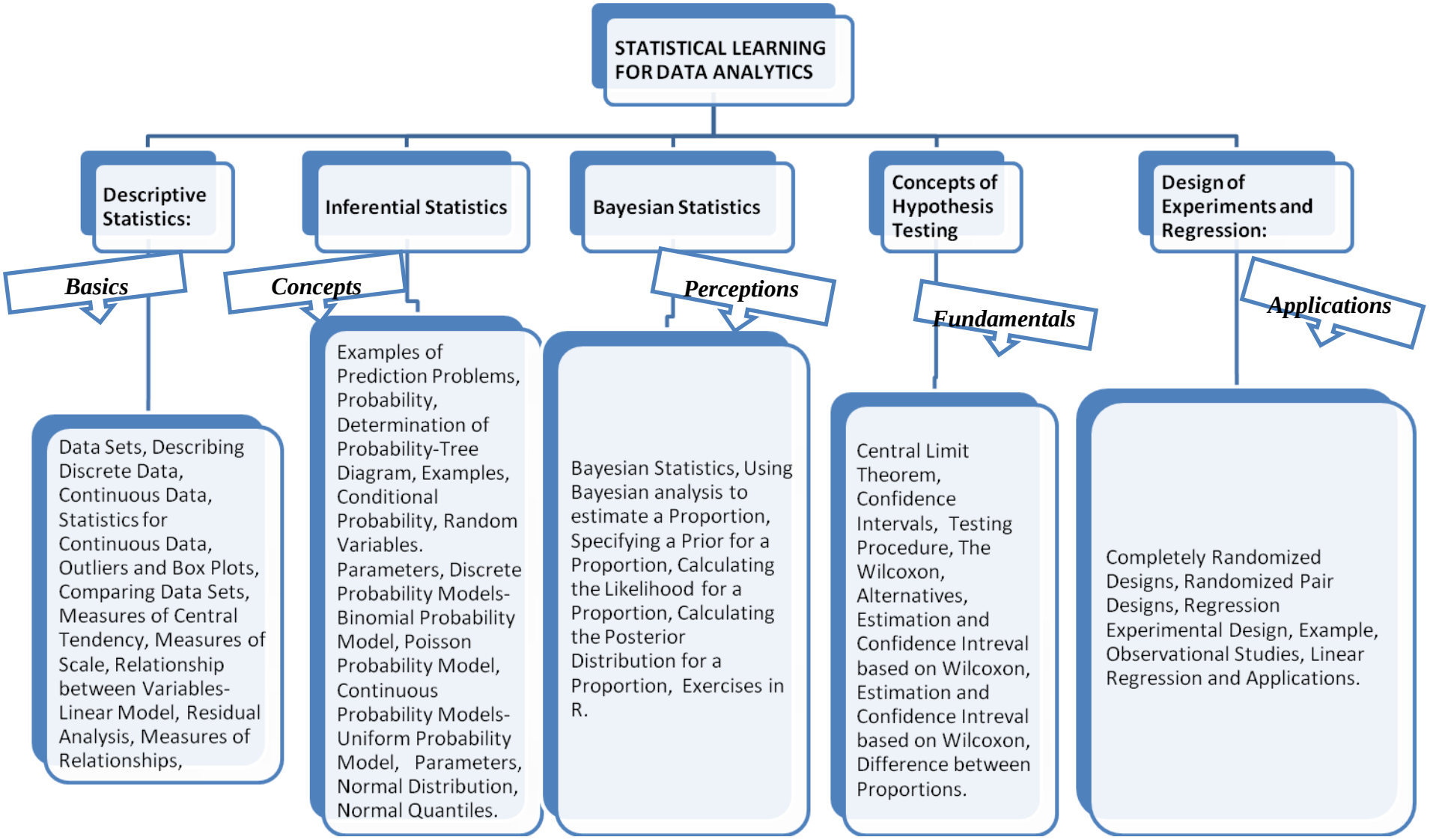
Mapping with Programme Outcomes

Cos	PSO1	PSO2	PSO3	PSO4	PSO5	PSO6
CO1	S	M		M	M	
CO2		M	S			L

Assessment Pattern

Category	Continuous Assessment Tests (25)			Terminal Examination (75)
Remember	5	5	5	20
Understand	10	10	10	20
Apply	5	5	5	10
Analyze	5	5	5	10
Evaluate	0	0	0	5
Create	0	0	0	0

Concept Map


STATISTICAL LEARNING FOR DATA ANALYTICS
Descriptive Statistics:
Inferential Statistics
Bayesian Statistics
Concepts of Hypothesis Testing
Design of Experiments and Regression:
Basics
Concepts
Perceptions
Fundamentals
Applications
Data Sets, Describing Discrete Data, Continuous Data, Statistics for Continuous Data, Outliers and Box Plots, Comparing Data Sets, Measures of Central Tendency, Measures of Scale, Relationship between Variables-Linear Model, Residual Analysis, Measures of Relationships,
Examples of Prediction Problems, Probability, Determination of Probability-Tree Diagram, Examples, Conditional Probability, Random Variables. Parameters, Discrete Probability Models-Binomial Probability Model, Poisson Probability Model, Continuous Probability Models-Uniform Probability Model, Parameters, Normal Distribution, Normal Quantiles.
Bayesian Statistics, Using Bayesian analysis to estimate a Proportion, Specifying a Prior for a Proportion, Calculating the Likelihood for a Proportion, Calculating the Posterior Distribution for a Proportion, Exercises in R.
Central Limit Theorem, Confidence Intervals, Testing Procedure, The Wilcoxon, Alternatives, Estimation and Confidence Intreval based on Wilcoxon, Estimation and Confidence Intreval based on Wilcoxon, Difference between Proportions.
Completely Randomized Designs, Randomized Pair Designs, Regression Experimental Design, Example, Observational Studies, Linear Regression and Applications.

Syllabus

UNIT I Descriptive Statistics: Data Sets, Describing Discrete Data, Continuous Data, Statistics for Continuous Data, Outliers and Box Plots, Comparing Data Sets, Measures of Central Tendency, Measures of Scale, Relationship between Variables- Linear Model, Residual Analysis, Measures of Relationships, (12hrs)

UNIT II Inferential Statistics: Examples of Prediction Problems, Probability, Determination of Probability-Tree Diagram, Examples, Conditional Probability, Random Variables. Parameters, Discrete Probability Models-Binomial Probability Model, Poisson Probability Model, Continuous Probability Models-Uniform Probability Model, Parameters, Normal Distribution, Normal Quantiles. (10hrs)

UNIT III Bayesian Statistics: Bayesian Statistics, Using Bayesian analysis to estimate a Proportion, Specifying a Prior for a Proportion, Calculating the Likelihood for a Proportion, Calculating the Posterior Distribution for a Proportion, Exercises in R. (11hrs)

UNIT IV Concepts of Hypothesis Testing: Central Limit Theorem, Confidence Intervals, Testing Procedure, The Wilcoxon, Alternatives, Estimation and Confidence Interval based on Wilcoxon, Estimation and Confidence Interval based on Wilcoxon, Difference between Proportions. (13hrs)

UNIT V Design of Experiments and Regression: Completely Randomized Designs, Randomized Pair Designs, Regression Experimental Design, Example, Observational Studies, Linear Regression and Applications. (14hrs)

Books:

1. A.Abebe, J.Daniels, J.W.Macean, Statistics and Data Analytics", Statistical Computation Lab, Western Michigan University, Kalamazoo.
2. Avril Coghlan,"A little Book for R using Bayesian Statistics", Wellcome Trust Sanger Institute, U.K.
3. "Kickstarting R"- cran.r-project.org/doc/contrib/Lemon-kickstart.
4. "Introduction to R" -cran.r-project.org/doc/manuals/R-intro.html.
5. "Bayesian Statistics" (product code M249/04) by the Open University, available from the Open University Shop.
6. Jim Albert, " Bayesian Computation with R".
7. Trevor Hasti, Robert Tibshirani, Jerome Friedman,"Data Mining, Inference and Statistics", 2nd Edition, Springer Series in Statistics.

Course Contents Coverage and Lecture Schedule

S.No.	Topic	No. of Lectures
1.	**Descriptive Statistics:**	
1.1	Data Sets, Describing Discrete Data	3
1.2	Continuous Data, Statistics for Continuous Data, Outliers and Box Plots, Comparing Data Sets	3
1.3	Measures of Central Tendency, Measures of Scale, Relationship between Variables	3
1.4	Linear Model, Residual Analysis, Measures of Relationships	2
2.	**Inferential Statistics**	
2.1	Examples of Prediction Problems, Probability, Determination of Probability-Tree Diagram	3
2.2	Examples, Conditional Probability, Random Variables. Parameters,	3
2.3	Discrete Probability Models-Binomial Probability Model, Poisson Probability Model	3
2.4	Continuous Probability Models-Uniform Probability Model, Parameters, Normal Distribution, Normal Quantiles.	1
3.	**Bayesian Statistics**	
3.1	Bayesian Statistics, Using Bayesian analysis to estimate a Proportion,	3
3.2	Specifying a Prior for a Proportion, Calculating the Likelihood for a Proportion,	3
3.3	Calculating the Posterior Distribution for a Proportion, Exercises in R	2
4.	**Concepts of Hypothesis Testing**	
4.1	Central Limit Theorem, Confidence Intervals, Testing Procedure	3
4.2	The Wilcoxon, Alternatives, Estimation and Confidence Intreval based on Wilcoxon	3
4.3	Estimation and Confidence Intreval based on Wilcoxon, Difference between Proportions.	3
5.	**Design of Experiments and Regression**	
5.1	Completely Randomized Designs	3
5.2	Randomized Pair Designs, Regression Experimental Design, Example,	3
5.3	Observational Studies, Linear Regression and Applications	3

MDAC15	NOSQL DATABASE MANAGEMENT	Category	L	P	T	Credit
		PE	4	0	0	4

Preamble

To understand the basics of Database Management. To establish the need for NoSQL Databases in Big Data Analytics.To understand the different types of NoSql databases and their uses.

Prerequisite

- Basics of database , MS Access

Course Outcomes

On the successful completion of the course, students will be able to

Course Outcomes		Bloom's Level
CO1	Able to define the basic concepts of Databases	Understanding
CO2	Able to identify the scope of NoSql Databases, advantages over conventional Databases.	Apply
CO3	Able to identify the right kind of NoSql Database for the right purpose while doing a Big Data Analytivs project.	Apply
CO4	Able to work with a particular NoSql tool.	Apply

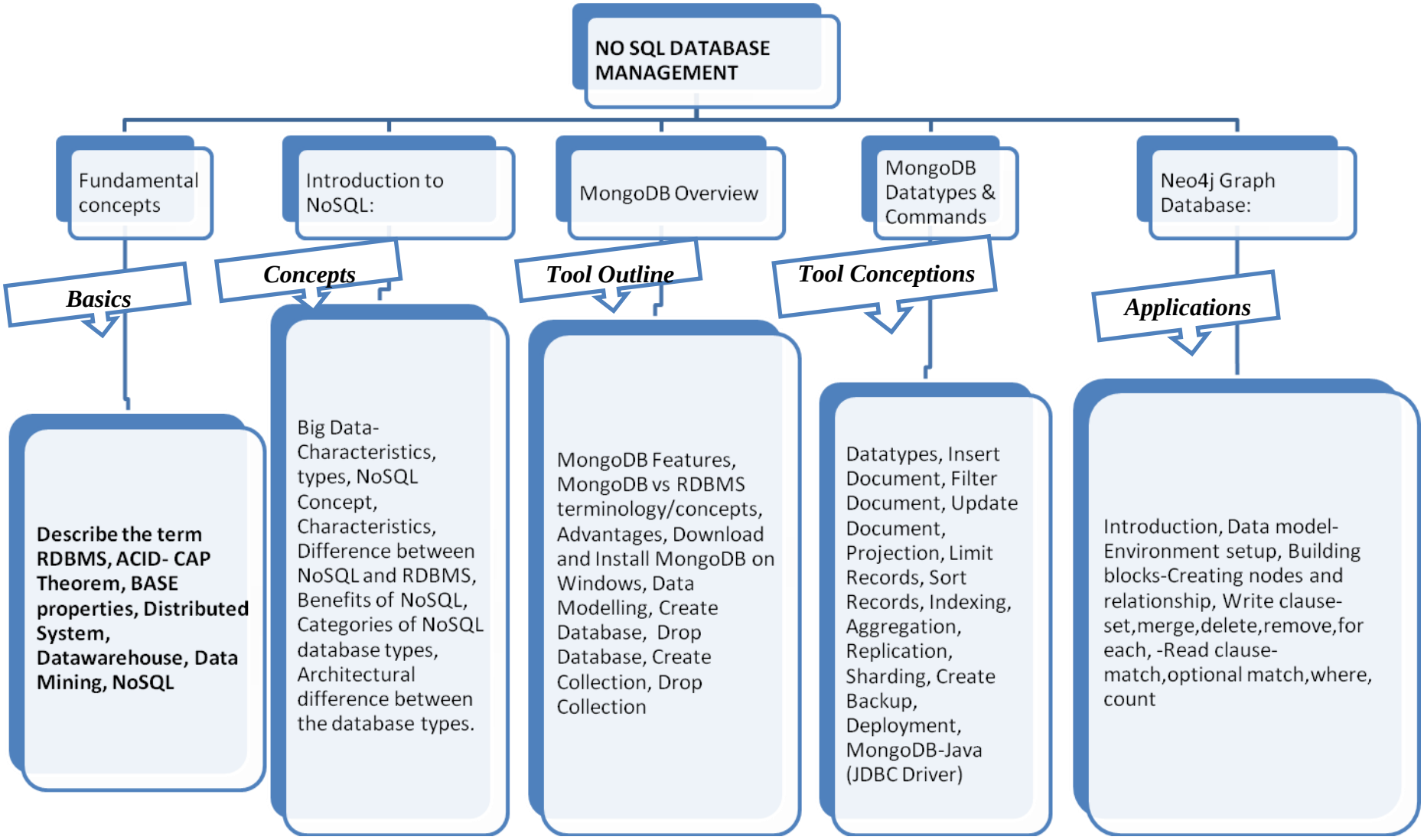
Mapping with Programme Outcomes

Cos	PSO1	PSO2	PSO3	PSO4	PSO5	PSO6
CO1		S				
CO2	S	M	S			
CO3						L
CO4			S			

Assessment Pattern

Bloom's Category	Continuous Assessment Tests (25)			Terminal Examination (75)
Remember	5	5	5	20
Understand	10	10	10	20
Apply	5	5	5	10
Analyze	5	5	5	10
Evaluate	0	0	0	5
Create	0	0	0	0

Concept Map


NO SQL DATABASE MANAGEMENT
Fundamental concepts
Introduction to NoSQL:
MongoDB Overview
MongoDB Datatypes & Commands
Neo4j Graph Database:
Basics
Concepts
Tool Outline
Tool Conceptions
Applications
Describe the term RDBMS, ACID- CAP Theorem, BASE properties, Distributed System, Datawarehouse, Data Mining, NoSQL
Big Data- Characteristics, types, NoSQL Concept, Characteristics, Difference between NoSQL and RDBMS, Benefits of NoSQL, Categories of NoSQL database types, Architectural difference between the database types.
MongoDB Features, MongoDB vs RDBMS terminology/concepts, Advantages, Download and Install MongoDB on Windows, Data Modelling, Create Database, Drop Database, Create Collection, Drop Collection
Datatypes, Insert Document, Filter Document, Update Document, Projection, Limit Records, Sort Records, Indexing, Aggregation, Replication, Sharding, Create Backup, Deployment, MongoDB-Java (JDBC Driver)
Introduction, Data model- Environment setup, Building blocks-Creating nodes and relationship, Write clause- set,merge,delete,remove,for each, -Read clause- match,optional match,where, count

Syllabus

UNIT I Fundamental concepts: Describe the term RDBMS, ACID- CAP Theorem, BASE properties, Distributed System, Datawarehouse, Data Mining, NoSQL (10hrs)

UNIT II Introduction to NoSQL: Big Data- Characteristics, types, NoSQL Concept, Characteristics, Difference between NoSQL and RDBMS, Benefits of NoSQL, Categories of NoSQL database types, Architectural difference between the database types. (11hrs)

.

UNIT III MongoDB Overview: MongoDB Features, MongoDB vs RDBMS terminology/concepts, Advantages, Download and Install MongoDB on Windows, Data Modelling, Create Database, Drop Database, Create Collection, Drop Collection (14hrs)

UNIT IV MongoDB Datatypes & Commands: Datatypes, Insert Document, Filter Document, Update Document, Projection, Limit Records, Sort Records, Indexing, Aggregation, Replication, Sharding, Create Backup, Deployment, MongoDB-Java (JDBC Driver) (13hrs)

UNIT V Neo4j Graph Database: Introduction, Data model-Environment setup, Building blocks-Creating nodes and relationship, Write clause-set,merge,delete,remove,for each, -Read clause-match,optional match,where, count. (12hrs)

Books:

1. eXist: A NoSQL Document Database and Application Platform by Erik Siegel, Adam Retter,

O'Reilly Media, 2014, ISBN: 978-1-44933-710-0

2. Professional NoSQL, Shashank Tiwari, Wrox-2011, ISBN: 047094224X,9780470942246

3. The Definitive Guide to MongoDB: The NoSQL Database for Cloud and Desktop Computing,

Eelco Plugge, DUPTim Hawkins, Peter Membrey, Apress, 2010, ISBN:

1430230517,9781430230519

4. https://www.tutorialspoint.com/neo4j/neo4j_data_model.htm

Course Contents Coverage and Lecture Schedule

S.No.	Topic	No. of Lectures
1	Fundamental concepts	
1.1	Describe the term RDBMS, ACID- CAP Theorem, BASE properties,	3
1.2	Distributed System, Datawarehouse, Data Mining, NoSQL	3
2	**Introduction to NoSQL**	
2.1	Big Data- Characteristics, types, NoSQL Concept, Characteristics	3
2.2	Difference between NoSQL and RDBMS, Benefits of NoSQL,	3
2.3	Categories of NoSQL database types, Architectural difference between the database types.	3
3	MongoDB Overview	
3.1	MongoDB Features, MongoDB vs RDBMS terminology/concepts	3
3.2	Advantages, Download and Install MongoDB on Windows,	2
3.3	Data Modelling, Create Database, Drop Database, Create Collection, Drop Collection	3
4	**MongoDB Datatypes & Commands**	
4.1	Datatypes, Insert Document, Filter Document, Update Document,)	3
4.2	Projection, Limit Records, Sort Records, Indexing, Aggregation,	3
4.3	Replication, Sharding, Create Backup, Deployment, MongoDB-Java (JDBC Driver	3
5.	**Neo4j Graph Database**	
5.1	Introduction, Data model-Environment setup, Building blocks-Creating nodes and relationship,	3
5.2	Write clause-set,merge,delete,remove,for each, -	3
5.3	Read clause-match,optional match,where, count.	3

MDAC15	DATA VISUALIZATION USING R AND PYTHON	Category	L	P	T	Credit
		PE	4	0	0	4

Preamble

To understand the basic concepts of Data Visualization. To understand the need for Data Visualization. To acquire skills in working with R and Python tools for Visualization..

Prerequisite

- Basic scientific graphs, charts

Course Outcomes

On the successful completion of the course, students will be able to

Course Outcomes		Bloom's Level
CO1	Able to identify the need for data Visualization in Data Analytics.	Understanding
CO2	Able to identify a right tool for Data Visualization based on the Data Analytics Problem.	Apply
CO3	Able to identify the right visualization technique to obtain information, knowledge and insight from particular Data-set.	Apply
CO4	Able to execute Visualization tasks in R and Python	Apply
CO5	Gain confidence in using any Visualization tool	Understanding

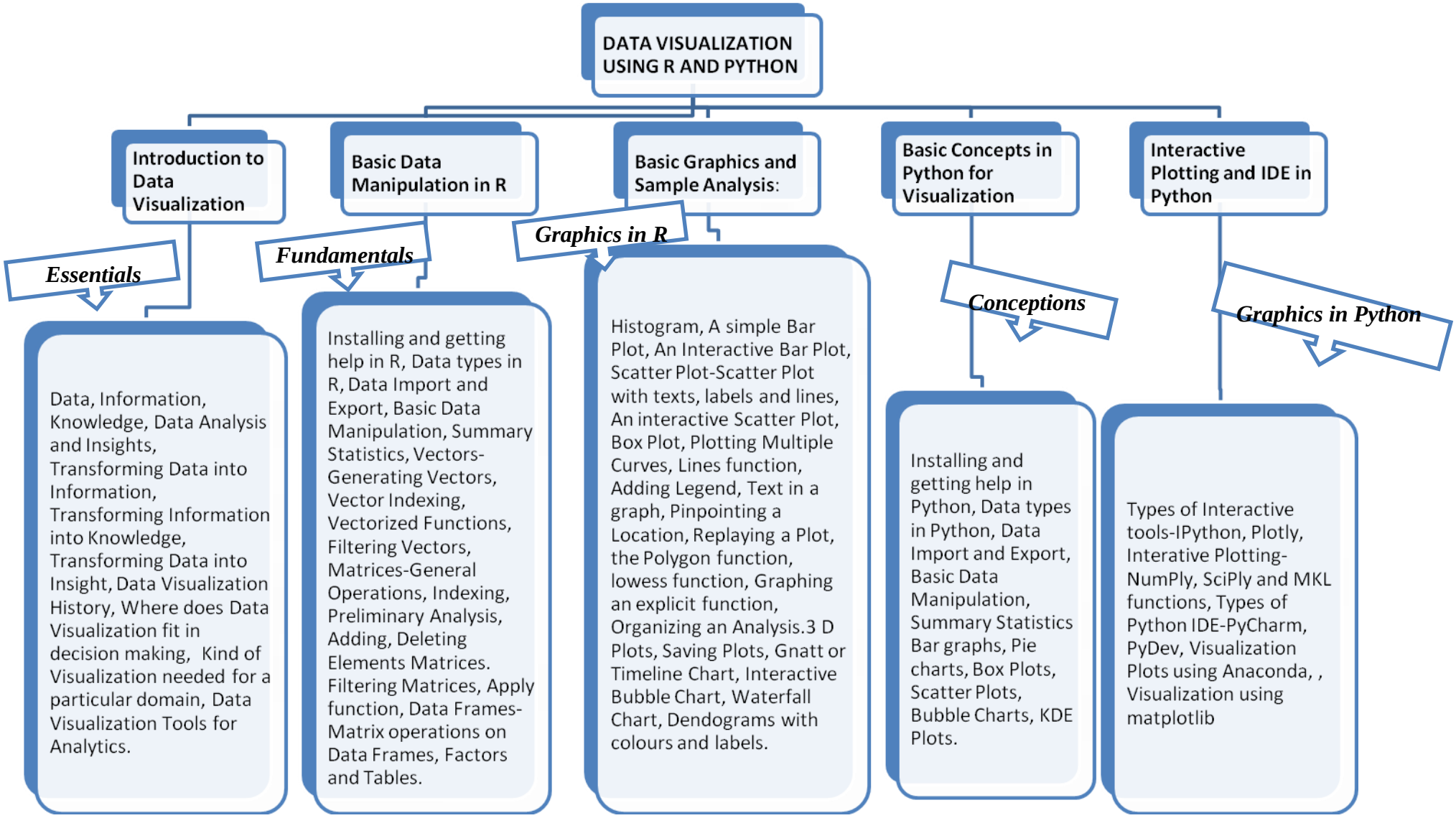
Mapping with Programme Outcomes

Cos	PSO1	PSO2	PSO3	PSO4	PSO5	PSO6
CO1		S				
CO2	S			L		
CO3		S	S			
CO4					M	
CO5		S	S			M

Assessment Pattern

Bloom's Category	Continuous Assessment Tests (25)			Terminal Examination (75)
Remember	5	5	5	20
Understand	10	10	10	20
Apply	5	5	5	10
Analyze	5	5	5	10
Evaluate	0	0	0	5
Create	0	0	0	0

Concept Map


DATA VISUALIZATION USING R AND PYTHON
Introduction to Data Visualization
Basic Data Manipulation in R
Basic Graphics and Sample Analysis:
Basic Concepts in Python for Visualization
Interactive Plotting and IDE in Python
Essentials
Fundamentals
Graphics in R
Conceptions
Graphics in Python
Data, Information, Knowledge, Data Analysis and Insights, Transforming Data into Information, Transforming Information into Knowledge, Transforming Data into Insight, Data Visualization History, Where does Data Visualization fit in decision making, Kind of Visualization needed for a particular domain, Data Visualization Tools for Analytics.
Installing and getting help in R, Data types in R, Data Import and Export, Basic Data Manipulation, Summary Statistics, Vectors-Generating Vectors, Vector Indexing, Vectorized Functions, Filtering Vectors, Matrices-General Operations, Indexing, Preliminary Analysis, Adding, Deleting Elements Matrices. Filtering Matrices, Apply function, Data Frames-Matrix operations on Data Frames, Factors and Tables.
Histogram, A simple Bar Plot, An Interactive Bar Plot, Scatter Plot-Scatter Plot with texts, labels and lines, An interactive Scatter Plot, Box Plot, Plotting Multiple Curves, Lines function, Adding Legend, Text in a graph, Pinpointing a Location, Replaying a Plot, the Polygon function, lowess function, Graphing an explicit function, Organizing an Analysis.3 D Plots, Saving Plots, Gnatt or Timeline Chart, Interactive Bubble Chart, Waterfall Chart, Dendograms with colours and labels.
Installing and getting help in Python, Data types in Python, Data Import and Export, Basic Data Manipulation, Summary Statistics Bar graphs, Pie charts, Box Plots, Scatter Plots, Bubble Charts, KDE Plots.
Types of Interactive tools-IPython, Plotly, Interative Plotting-NumPly, SciPly and MKL functions, Types of Python IDE-PyCharm, PyDev, Visualization Plots using Anaconda, , Visualization using matplotlib

Syllabus

UNIT 1: Introduction to Data Visualization: Data, Information, Knowledge, Data Analysis and Insights, Transforming Data into Information, Transforming Information into Knowledge, Transforming Data into Insight, Data Visualization History, Where does Data Visualization fit in decision making, Kind of Visualization needed for a particular domain, Data Visualization Tools for Analytics. (10hrs)

UNIT II: Basic Data Manipulation in R: Installing and getting help in R, Data types in R, Data Import and Export, Basic Data Manipulation, Summary Statistics, Vectors-Generating Vectors, Vector Indexing, Vectorized Functions, Filtering Vectors, Matrices-General Operations, Indexing, Preliminary Analysis, Adding, Deleting Elements Matrices. Filtering Matrices, Apply function, Data Frames-Matrix operations on Data Frames, Factors and Tables. (11hrs)

UNIT III Basic Graphics and Sample Analysis: Histogram, A simple Bar Plot, An Interactive Bar Plot, Scatter Plot-Scatter Plot with texts, labels and lines, An interactive Scatter Plot, Box Plot, Plotting Multiple Curves, Lines function, Adding Legend, Text in a graph, Pinpointing a Location, Replaying a Plot, the Polygon function, lowess function, Graphing an explicit function, Organizing an Analysis.3 D Plots, Saving Plots, Gnatt or Timeline Chart, Interactive Bubble Chart, Waterfall Chart, Dendograms with colours and labels. (14hrs)

UNIT IV Basic Concepts in Python for Visualization: Installing and getting help in Python, Data types in Python, Data Import and Export, Basic Data Manipulation, Summary Statistics Bar graphs, Pie charts, Box Plots, Scatter Plots, Bubble Charts, KDE Plots.
(12hrs)

UNIT V Interactive Plotting and IDE in Python: Types of Interactive tools-IPython, Plotly, Interative Plotting-NumPly, SciPly and MKL functions, Types of Python IDE-PyCharm, PyDev, Visualization Plots using Anaconda, , Visualization using matplotlib (13hrs)

Books:

1. Kirthi Raman, "Mastering Python Data Visualization", Packt Publishing Ltd., UK.
2. Atmajithsingh Gohil, "R Data Visualizing Codebook", Packt PublishingLtd., UK.
3. Brian S. Everitt and Torsten Hothorn, "A Handbook of Statistical Analyses Using R", 2nd Edition, Tayor and Francis Group.
4. Normann Mattloff, "The Art of R Programming", No Starch Press, San Fransisco.
5. Chambers, J. M. (1998), "Programming with Data", New York, USA: Springer.
6. Meyer, D., Zeileis, A., Karatzoglou, A., and Hornik, K. (2006), "Visualizing Categorical Data", URL http://CRAN.R-project.org, R package version 0.9-91.
7. Avril Coglan, "A Little Book of R for Time Series" Wellcome Trust Sanger Institute, U.K.

Course Contents Coverage and Lecture Schedule

S.No.	Topic	No. of Lectures
1.	**Introduction to Data Visualization**	
1.1	Data, Information, Knowledge, Data Analysis and Insights,	3
1.2	Transforming Data into Information, Transforming Information into Knowledge,	3
1.3	Transforming Data into Insight, Data Visualization History, Where does Data Visualization fit in decision making	2
1.4	Kind of Visualization needed for a particular domain,	3
1.5	Data Visualization Tools for Analytics.	3
2.	**Basic Data Manipulation in R**	
2.1	Installing and getting help in R, Data types in R, Data Import and Export,	3
2.2	Basic Data Manipulation, Summary Statistics, Vectors-Generating Vectors,	3
2.3	Vector Indexing, Vectorized Functions, Filtering Vectors, Matrices-General Operations, Indexing, Preliminary Analysis, Adding, Deleting Elements	3
2.4	Matrices. Filtering Matrices, Apply function, Data Frames-Matrix operations on Data Frames, Factors and Tables.	3
3.	**Basic Graphics and Sample Analysis**	
3.1	Histogram, A simple Bar Plot, An Interactive Bar Plot, Scatter Plot-Scatter Plot with texts, labels and lines,	3
3.2	An interactive Scatter Plot, Box Plot, Plotting Multiple Curves, Lines function, Adding Legend, Text in a graph, Pinpointing a Location,.	2
3.3	Replaying a Plot, the Polygon function, lowess function, Graphing an explicit function, Organizing an Analysis.	2
3.4	3 D Plots, Saving Plots, Gnatt or Timeline Chart, Interactive Bubble Chart, Waterfall Chart, Dendograms with colours and labels	3
4.	**Basic Concepts in Python for Visualization**	
4.1	Installing and getting help in Python, Data types in Python, Data Import and Export,	3
4.2	Basic Data Manipulation, Summary Statistics Bar graphs, Pie charts, Box Plots, Scatter Plots, Bubble Charts, KDE Plots.	3
5.	**Interactive Plotting and IDE in Python**	
5.1	Types of Interactive tools-IPython, Plotly, Interative Plotting-NumPly, SciPly and MKL functions,	3
5.2	Types of Python IDE-PyCharm, PyDev,	3
5.3	Visualization Plots using Anaconda, Visualization using matplotlib	3

MDAL11	NO SQL DATABASE MANAGEMENT LABORATORY

Exercises to improve your NoSQL Database Management skills

1. Design at least 10 SQL queries for suitable database application using SQL DML Statements: Insert, Select, Update, Delete Clauses using distinct, count, aggregation
2. Implement database with suitable example using MongoDB. 3. Implement all basic operations [create, drop database and collection].
4. Implement the basic operations in collection [insert, delete, update].
5. Use MongoDB to process semi structured/unstructured data collections.
6. Aggregation and indexing with suitable example using MongoDB.
7. Indexing and querying with MongoDB using suitable example.
8. Connectivity with MongoDB using any Java application.
9. Create and delete node, relationship, index using Neo4j.
10. Implement write [set, merge, delete, remove, for each] clause using Neo4j
11. Implement read [match, optional match, where, count] clause using Neo4j

MDAL12	DATA VISUALIZATION AND ANALYTICS LABORATORY USING R AND PYTHON

R and Python Programming Exercises

1. Software and Package Installation Steps.
2. Loading different kinds of data formats.
3. Using functions to find statistics of Data.
4. Exercises to work with Vectors and Matrices.
5. Selecting the right plot for a particular Data Set.
6. Scatter plot exercises with Text, Labels, Lines and Connecting Points.
7. Interactive Scatter Plots.
8. Relationship between two Continuous variables of a Dataset using Scatter Plot
9. Plotting categorical variables using different Bar Plots.
10. Interactive Bar Plots.
11. Histograms, merging Histograms.
12. Inferring from different Visualizations for possible predictions.
13. Comparing different datasets using suitable visualization technique.

SEMESTER II

LIST OF COURSES
(For The Candidates Admitted From 2017-18 Onwards)

Sl. No.	Subject code	Course name
1.		Supportive Course
2.	MDAC21	Descriptive and Discovery Analytics
3.	MDAC22	Machine Learning for Data Analytics
4.		Elective 1
5.		Elective 2
6.		Elective 3
7.	MDAL21	Machine Learning Laboratory
8.	MDAL22	Descriptive and Discovery Analytics Laboratory

MDAC 21- DESCRIPTIVE AND DISCOVERY ANALYTICS	Category	L	P	Credit
	PE	4	0	4

Preamble

To get the knowledge of descriptive analytics and its role in future data, acquire knowledge in different Descriptive and Discovery analytics concepts, get the idea of which model to apply for descriptive and discovery analytics.

Prerequisite

Statistical Learning for Data Analytics (Basics of Descriptive Statistics)

Course Outcomes

On the successful completion of the course, students will be able to

Course Outcomes		Level
CO1	Able to identify the scope, role and application of Descriptive Analytics.	Understanding
CO2	Able to handle raw data using Python.	Apply
CO3	Able to understand various descriptive and discovery analytics techniques.	Understand
CO4	Able to identify the right kind of analytics model for the right purpose while doing a Data Analytics project.	Apply

Mapping with Programme Outcomes

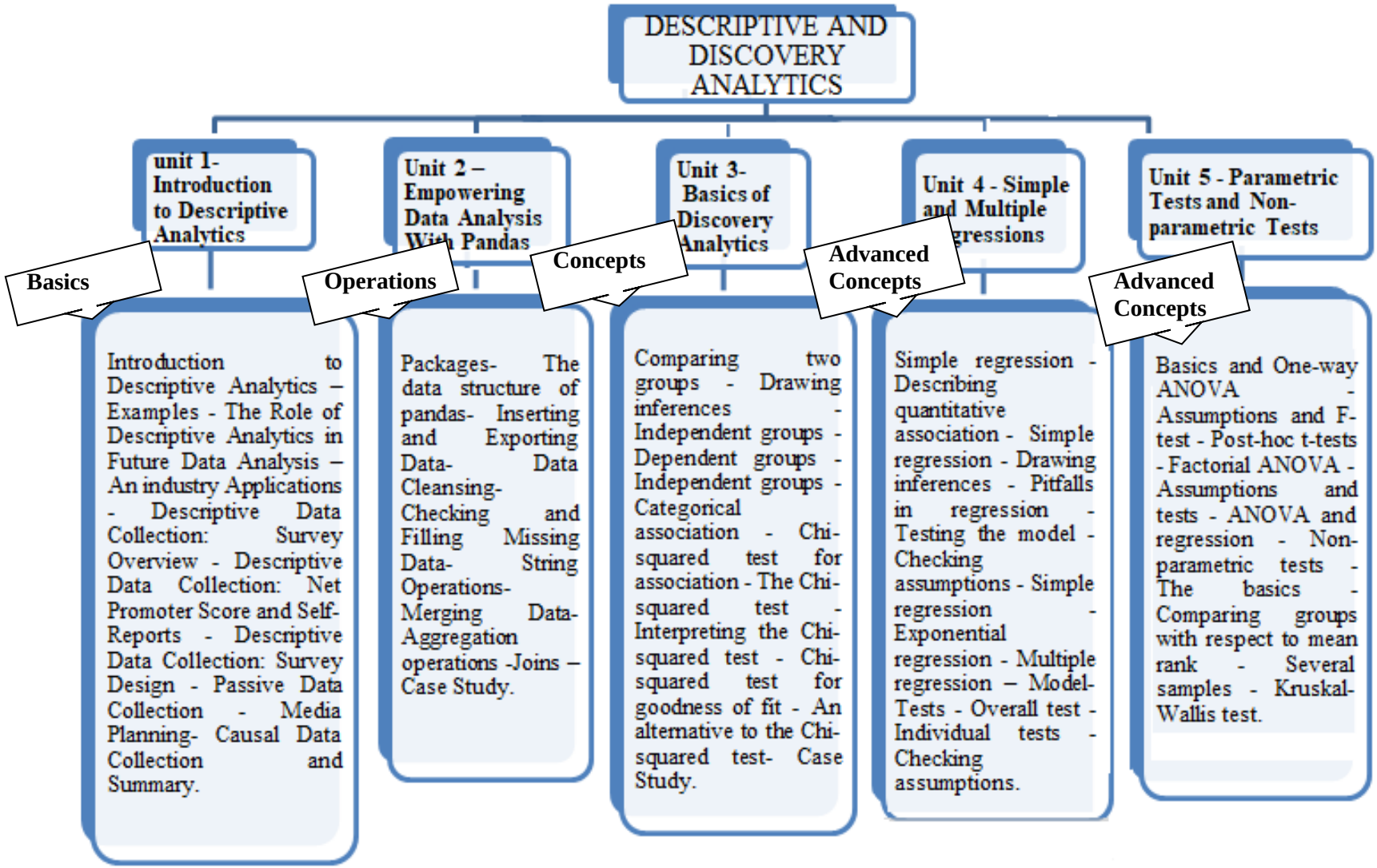
Cos	PSO1	PSO2	PSO3	PSO4	PSO5	PSO6
CO1	S	M	S			
CO2	M					

CO3		M	L			
CO4					M	L

Assessment Pattern

Category	Continuous Assessment Tests (25)			Terminal Examination (75)
Remember	5	5	5	20
Understand	10	10	10	20
Apply	5	5	5	10
Analyze	5	5	5	10
Evaluate				5
Create				

Concept Map


DESCRIPTIVE AND DISCOVERY ANALYTICS
unit 1- Introduction to Descriptive Analytics
Unit 2 – Empowering Data Analysis With Pandas
Unit 3- Basics of Discovery Analytics
Unit 4 - Simple and Multiple gressions
Unit 5 - Parametric Tests and Non-parametric Tests
Basics
Operations
Concepts
Advanced Concepts
Advanced Concepts
Introduction to Descriptive Analytics – Examples - The Role of Descriptive Analytics in Future Data Analysis – An industry Applications - Descriptive Data Collection: Survey Overview - Descriptive Data Collection: Net Promoter Score and Self-Reports - Descriptive Data Collection: Survey Design - Passive Data Collection - Media Planning- Causal Data Collection and Summary.
Packages- The data structure of pandas- Inserting and Exporting Data- Data Cleansing- Checking and Filling Missing Data- String Operations- Merging Data- Aggregation operations -Joins – Case Study.
Comparing two groups - Drawing inferences - Independent groups - Dependent groups - Independent groups - Categorical association - Chi-squared test for association - The Chi-squared test - Interpreting the Chi-squared test - Chi-squared test for goodness of fit - An alternative to the Chi-squared test- Case Study.
Simple regression - Describing quantitative association - Simple regression - Drawing inferences - Pitfalls in regression - Testing the model - Checking assumptions - Simple regression - Exponential regression - Multiple regression – Model-Tests - Overall test - Individual tests - Checking assumptions.
Basics and One-way ANOVA - Assumptions and F-test - Post-hoc t-tests - Factorial ANOVA - Assumptions and tests - ANOVA and regression - Non-parametric tests - The basics - Comparing groups with respect to mean rank - Several samples - Kruskal-Wallis test.

Syllabus

Unit 1 - Introduction to Descriptive Analytics

Introduction to Descriptive Analytics – Examples - The Role of Descriptive Analytics in Future Data Analysis – An industry Applications - Descriptive Data Collection: Survey Overview - Descriptive Data Collection: Net Promoter Score and Self-Reports - Descriptive Data Collection: Survey Design - Passive Data Collection - Media Planning- Causal Data Collection and Summary.

Unit 2 – Empowering data analysis with pandas

Packages- The data structure of pandas- Inserting and Exporting Data- Data Cleansing- Checking and Filling Missing Data- String Operations- Merging Data- Aggregation operations - Joins – Case Study.

Unit 3 – Basics of Discovery Analytics

Comparing two groups - Drawing inferences - Independent groups - Dependent groups - Independent groups - Categorical association - Chi-squared test for association - The Chi-squared test - Interpreting the Chi-squared test - Chi-squared test for goodness of fit - An alternative to the Chi-squared test- Case Study.

Unit 4 - Simple and Multiple Regressions

Simple regression - Describing quantitative association - Simple regression - Drawing inferences - Pitfalls in regression - Testing the model - Checking assumptions - Simple regression - Exponential regression - Multiple regression – Model- Tests - Overall test - Individual tests - Checking assumptions.

Unit 5 - Parametric Tests and Non-parametric Tests

Basics and One-way ANOVA - Assumptions and F-test - Post-hoc t-tests - Factorial ANOVA - Assumptions and tests - ANOVA and regression - Non-parametric tests - The basics - Comparing groups with respect to mean rank - Several samples - Kruskal-Wallis test.

References:

1. https://www.coursera.org/learn/wharton-customer-analytics
2. http://www.dataversity.net/fundamentals-descriptive-analytics
3. Samir Madhavan, “Mastering Python for Data Science”, Packt, 2015.
4. https://www.coursera.org/learn/inferential-statistics

Course Contents and Lecture Schedule

S.No.	Topic	No. of Lectures
1	Introduction to Descriptive Analytics	
1.1	Introduction to Descriptive Analytics – Examples	1
1.2	The Role of Descriptive Analytics in Future Data Analysis – An industry Applications	1
1.3	Descriptive Data Collection: Survey Overview - Descriptive Data Collection: Net Promoter Score and Self-Reports	2
1.4	Descriptive Data Collection: Survey Design	2
1.5	Passive Data Collection , Media Planning	2
1.6	Causal Data Collection and Summary	1
2	Empowering data analysis with pandas	
2.1	Packages	1
2.2	The data structure of pandas	1
2.3	Inserting and Exporting Data	1
2.4	Data Cleansing	1
2.5	Checking and Filling Missing Data	1
2.6	String Operations	1
2.7	Merging Data- Aggregation operations -	2
2.8	Joins – Case Study.	1
3	Basics of Discovery Analytics	
3.1	Comparing two groups - Drawing inferences	2
3.2	Independent groups - Dependent groups	1
3.3	Categorical association - Chi-squared test for association	1
3.4	The Chi-squared test - Interpreting the Chi-squared test	2
3.5	Chi-squared test for goodness of fit	2
3.6	An alternative to the Chi-squared test- Case Study	2
4	Simple and Multiple Regressions	
4.1	Simple regression - Describing quantitative association	1
4.2	Simple regression - Drawing inferences	2
4.3	Pitfalls in regression - Testing the model	2
4.4	Checking assumptions	1
4.5	Simple regression - Exponential regression	2
4.6	Multiple regression – Model	2
4.7	Tests - Overall test	2
4.8	Individual tests - Checking assumptions	2
5	Parametric Tests and Non-parametric Tests	
5.1	Basics and One-way ANOVA	1
5.2	Assumptions and F-test	1
5.3	Post-hoc t-tests - Factorial ANOVA	1
5.4	Assumptions and tests	2

5.5	ANOVA and regression	1
5.6	Non-parametric tests	1
5.7	The basics- Comparing groups with respect to mean rank	2
5.8	Several samples - Kruskal-Wallis test	2

MDAC22 MACHINE LEARNING ALGORITHMS FOR DATA ANALYTICS

Category	L	P	Credit
PE	4	0	4

Preamble

To know the importance of Machine Learning, to learn basic Machine learning algorithms, and to understand how to apply the learning algorithms for various prediction problems.

Prerequisite

Introduction to data analytics

Course Outcomes

On the successful completion of the course, students will be able to

Course Outcomes		Level
CO1	Know the importance of Machine Learning	Understanding
CO2	Learn basic Machine learning algorithms	Understanding
CO3	Understand how to apply the learning algorithms for various prediction problems	Apply

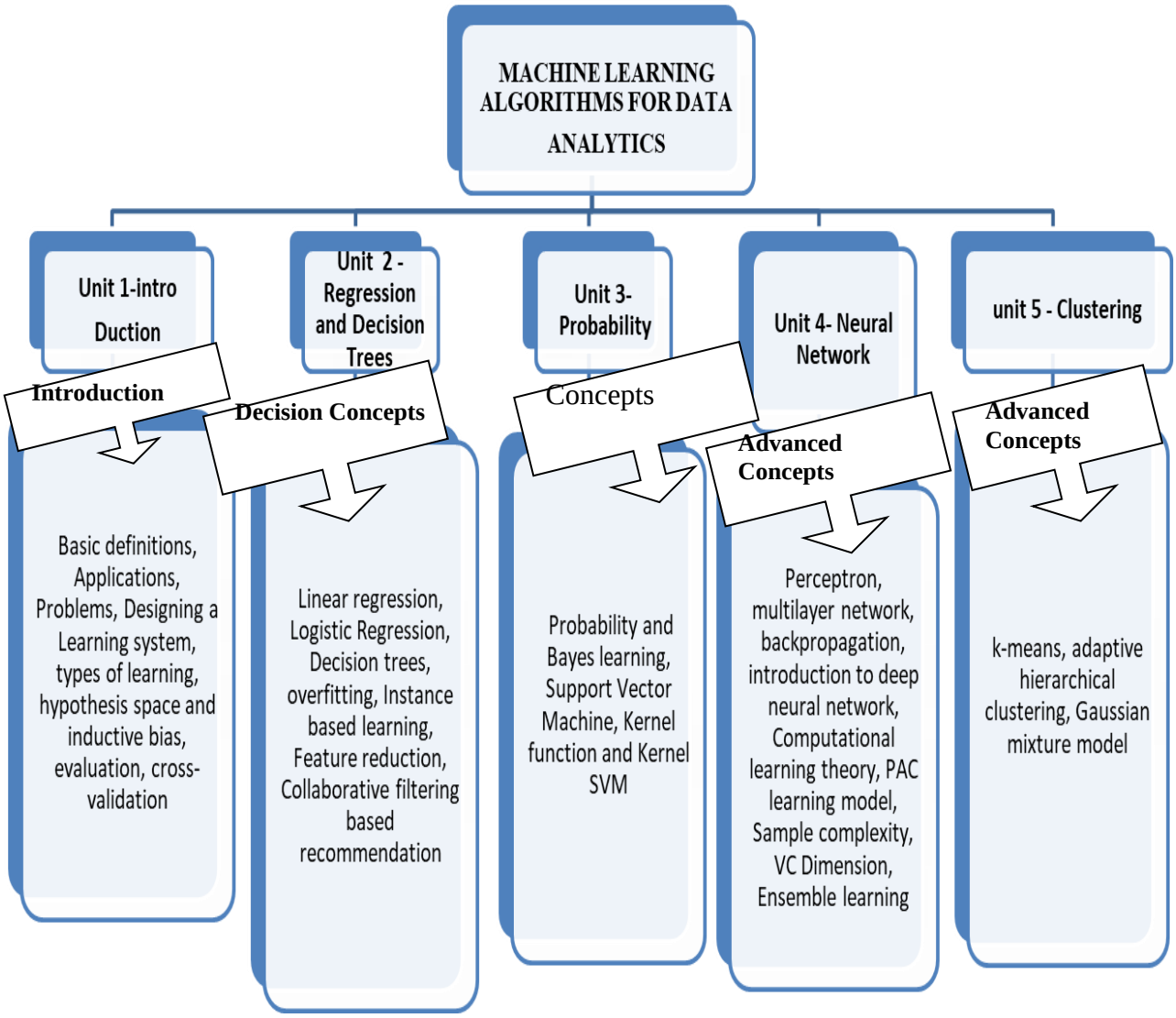
Mapping with Programme Outcomes

Cos	PSO1	PSO2	PSO3	PSO4	PSO5	PSO6
CO1	S	S	M	L	S	L
CO2	S	S	M	L	S	L
CO3	M	S	M	L	S	M

Assessment Pattern

Category	Continuous Assessment Tests (25)			Terminal Examination (75)
Remember	5	5	5	20
Understand	10	10	10	20
Apply	5	5	5	10
Analyze	5	5	5	10
Evaluate				5
Create				

Concept Map


MACHINE LEARNING ALGORITHMS FOR DATA ANALYTICS
Unit 1-intro Duction
Introduction
Basic definitions, Applications, Problems, Designing a Learning system, types of learning, hypothesis space and inductive bias, evaluation, cross-validation
Unit 2 - Regression and Decision Trees
Decision Concepts
Linear regression, Logistic Regression, Decision trees, overfitting, Instance based learning, Feature reduction, Collaborative filtering based recommendation
Unit 3- Probability
Concepts
Probability and Bayes learning, Support Vector Machine, Kernel function and Kernel SVM
Unit 4- Neural Network
Advanced Concepts
Perceptron, multilayer network, backpropagation, introduction to deep neural network, Computational learning theory, PAC learning model, Sample complexity, VC Dimension, Ensemble learning
unit 5 - Clustering
Advanced Concepts
k-means, adaptive hierarchical clustering, Gaussian mixture model

Syllabus

Unit:I Introduction

Basic definitions, Applications, Problems, Designing a Learning system, types of learning, hypothesis space and inductive bias, evaluation, cross-validation,

Unit II Regression and Decision Trees

Linear regression, Logistic Regression, Decision trees, overfitting, Instance based learning, Feature reduction, Collaborative filtering based recommendation,

Unit: III Probability

Probability and Bayes learning, Support Vector Machine, Kernel function and Kernel SVM

Unit: IV Neural network

Perceptron, multilayer network, backpropagation, introduction to deep neural network, Computational learning theory, PAC learning model, Sample complexity, VC Dimension, Ensemble learning

Unit: V Clustering

k-means, adaptive hierarchical clustering, Gaussian mixture model.

References:

1.Machine Learning. Tom Mitchell. First Edition, McGraw- Hill, 1997.
2. Introduction to Machine Learning Edition 2, by Ethem Alpaydin

URLS:

1. A Course in Machine Learning - http://ciml.info/dl/v0_8/ciml-v0_8-all.pdf
2. http://alex.smola.org/drafts/thebook.pdf
3.https://www.cs.ubbcluj.ro/~gabis/ml/ml-books/McGrawHill%20-%20Machine%20Learning%20-Tom%20Mitchell.pdf
4https://www.analyticsvidhya.com/blog/2017/09/common-machine-learning-algorithms/
5.https://www.cs.huji.ac.il/~shais/UnderstandingMachineLearning/understanding-machine-learning-theory-algorithms.pdf
6.https://www.analyticsvidhya.com/blog/2016/04/complete-tutorial-tree-based-modeling-scratch-in-python/#one
7.https://doc.lagout.org/Others/Data%20Mining/Data%20Mining%20and%20Predictive%20Analytics%20%5BLarose%20%26%20Larose%202015-03-16%5D.pdf
8. http://auapps.american.edu/alberto/www/analytics/ISLRLectures.html

Course Contents and Lecture Schedule

S.No.	Topic	No. of Lectures
1.	**INTRODUCTION**	
1. 1	Basic definitions	1
1.2	Applications in machine learning	1
1.3	Problems	1
1.4	Designing a Learning system	
1.5	Types of learning	4
1.6	Hypothesis space and inductive bias	
1.7	Evaluation	1
1.8	Cross-validation	1
2.	**REGRESSION AND DECISION TREES**	
2.1	Linear regression	2
2.2	Logistic Regression	2
2.3	Decision trees	1
2.4	Overfitting	2
2.5	Instance based learning	2
2.6	Feature reduction	2
2.7	Collaborative filtering based recommendation	1
3.	**PROBABILITY**	
3.1	Probability and Bayes learning	1
3.2	Support Vector Machine	2
3.3	Kernel function	2
3.4	Kernel SVM	2
4.	**NEURAL NETWORK**	
4.1	Perceptron	1
4.2	Multilayer network & Backpropagation	2
4.3	Introduction to deep neural network	1
4.4	Computational learning theory	2
4.5	PAC learning model	1
4.6	Sample complexity	1
4.7	VC Dimension	1
4.8	Ensemble learning	2
5	**CLUSTERING**	
5.1	K-means	2
5.2	Adaptive hierarchical clustering	2
5.3	Gaussian mixture model	3

MDAL21

MACHINE LEARNING LABORATORY

Lab Exercises using Python and R

1. Working with Python-Tutorials
2. Working with R- tutorials
3. Understanding Machine Learning Problems, Training Dataset, Test Data Set.
4. Execute Linear Regression in Python and R using suitable Training and Testing data set for predicting the cost of a flat.
5. Execute Logistic Regression in Python and R using suitable Training and Testing data to predict discrete outputs.
6. Execute Decision trees in Python and R using suitable Training and Testing data set for making suitable predictions.
7. Execute Support Vector Machine in Python and R using suitable Training and Testing data set for making suitable predictions.
8. Execute K means clustering in Python and R using suitable Training and Testing data set for making suitable predictions.

References:

1.https://www.youtube.com/watch?v=2uCXIbkbDSE
2. https://www.youtube.com/watch?v=eDrhZb2onWY
3.https://www.analyticsvidhya.com/blog/2017/09/common-machine-learning-algorithms/
4.https://www.analyticsvidhya.com/blog/2016/04/complete-tutorial-tree-based-modeling-scratch-in-python/#one
5. https://www.analyticsvidhya.com/blog/2014/10/support-vector-machine-simplified/

MDAL22

DESCRIPTIVE AND DISCOVERY ANALYTICS LABORATORY USING PYTHON

Python Programming Exercises

1. Write a program to import and export data files of different data format.
2. Write a program to apply your own or another library's functions to Pandas objects.
3. Write a program to identify and handle missing data values
4. Write a program for data normalization
5. Write a program to convert common standard data format.
6. Write a program to summarize the data
7. Write a program to implement ANOVA.
8. Write a program to implement correlation
9. Write a program to perform various SQL operations using pandas.
10. Write a program to Model Evaluation Using Visualization
11. Write a program to Simple Linear Regression.
12. Write a program to Multiple Linear Regression model
13. Write a program to R-squared and MSE for In-Sample Evaluation
14. Write a program to perform Model Evaluation.

SEMESTER III

LIST OF COURSES
(For The Candidates Admitted From 2017-18 Onwards)

Sl. No.	Subject code	Course name
1.		Supportive Course II
2.	**MDAC31**	Predictive and Prescriptive Analytics
3.	**MDAC32**	Ethics for Data Scientists
4.		Elective 4
5.		Elective 5
6.	**MDAL31**	Predictive and Prescriptive Analytics Laboratory
7.	**MDAI31**	Internship

MDAC31 PREDICTIVE AND PRESCRIPTIVE ANALYTICS	Category	L	P	Credit
	PE	4	0	4

Preamble

To get introduced to the need of Predictive and Prescriptive Analytics, understand the concepts of textmining and time forecasting concepts for predictive analytics and to practice predictive techniques using Rapid Miner.

Prerequisite

Introduction to data Analytics

Course Outcomes

On the successful completion of the course, students will be able to

Course Outcomes		Level
CO1	Able to understand the need of Predictive Analytics	Understanding
CO2	Understand data mining and time forecasting concepts for making Predictions	Understand and apply
CO3	Apply data mining and time forecasting techniques for predictions	Apply

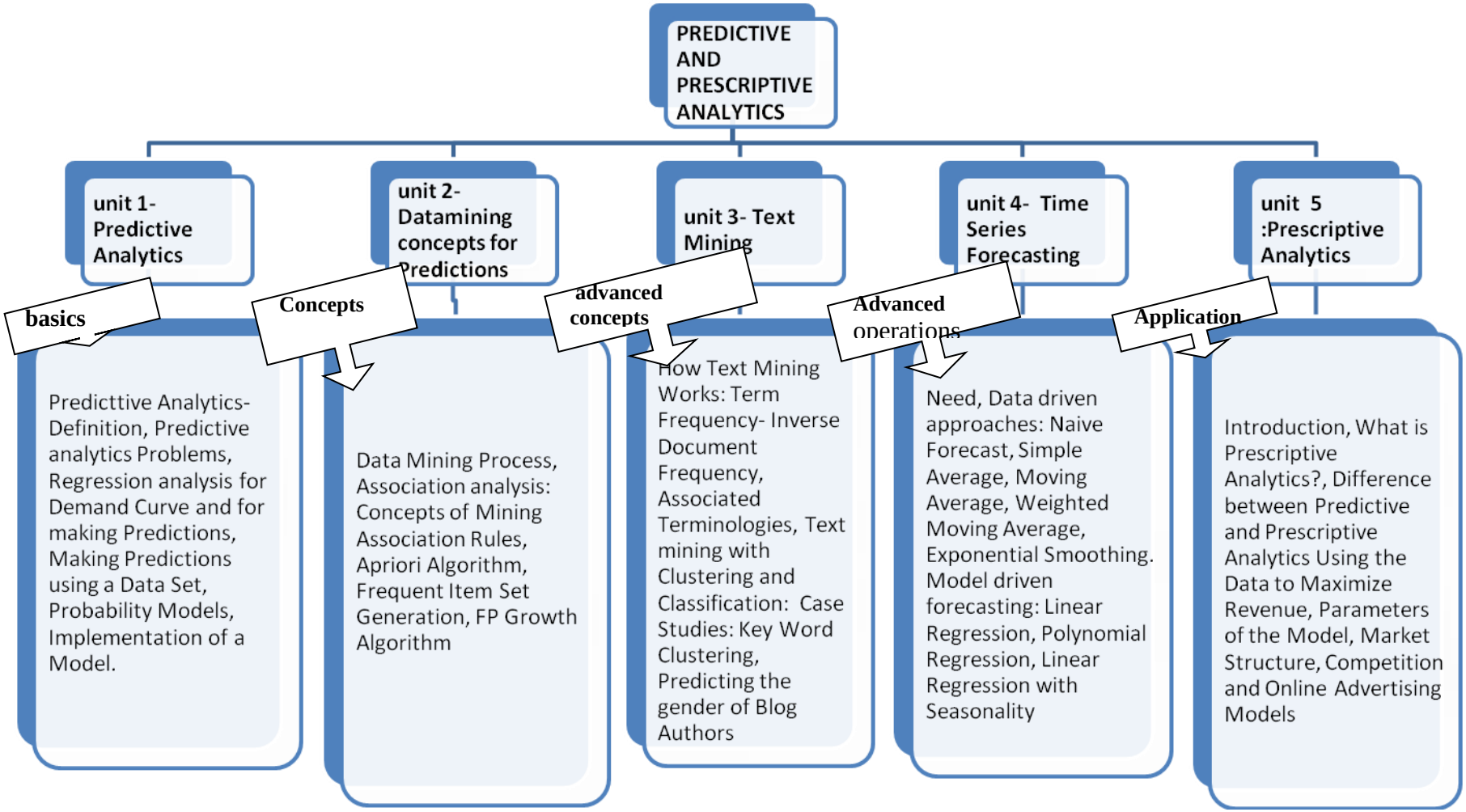
Mapping with Programme Outcomes

Cos	PSO1	PSO2	PSO3	PSO4	PSO5	PSO6
CO1	M	M				
CO2	M	M		L		
CO3		S	M		M	M

Assessment Pattern

Category	Continuous Assessment Tests (25)			Terminal Examination (75)
Remember	5	5	5	20
Understand	10	10	10	20
Apply	5	5	5	10
Analyze	5	5	5	10
Evaluate				5
Create				5

Concept Map


PREDICTIVE AND PRESCRIPTIVE ANALYTICS
unit 1- Predictive Analytics
unit 2- Datamining concepts for Predictions
unit 3- Text Mining
unit 4- Time Series Forecasting
unit 5 :Prescriptive Analytics
basics
Concepts
advanced concepts
Advanced operations
Application
Predicttive Analytics- Definition, Predictive analytics Problems, Regression analysis for Demand Curve and for making Predictions, Making Predictions using a Data Set, Probability Models, Implementation of a Model.
Data Mining Process, Association analysis: Concepts of Mining Association Rules, Apriori Algorithm, Frequent Item Set Generation, FP Growth Algorithm
How Text Mining Works: Term Frequency- Inverse Document Frequency, Associated Terminologies, Text mining with Clustering and Classification: Case Studies: Key Word Clustering, Predicting the gender of Blog Authors
Need, Data driven approaches: Naive Forecast, Simple Average, Moving Average, Weighted Moving Average, Exponential Smoothing. Model driven forecasting: Linear Regression, Polynomial Regression, Linear Regression with Seasonality
Introduction, What is Prescriptive Analytics?, Difference between Predictive and Prescriptive Analytics Using the Data to Maximize Revenue, Parameters of the Model, Market Structure, Competition and Online Advertising Models

Syllabus

Unit I: Predictive Analytics: Predicttive Analytics-Definition, Predictive analytics Problems, Regression analysis for Demand Curve and for making Predictions, Making Predictions using a Data Set, Probability Models, Implementation of a Model.

Unit: II Datamining concepts for Predictions: Data Mining Process, Association analysis: Concepts of Mining Association Rules, Apriori Algorithm, Frequent Item Set Generation, FP Growth Algorithm,

Unit: III Text Mining: How Text Mining Works: Term Frequency- Inverse Document Frequency, Associated Terminologies, Text mining with Clustering and Classification: Case Studies: Key Word Clustering, Predicting the gender of Blog Authors.

Unit:IV Time Series Forecasting: Need, Data driven approaches: Naive Forecast, Simple Average, Moving Average, Weighted Moving Average, Exponential Smoothing. Model driven forecasting: Linear Regression, Polynomial Regression, Linear Regression with Seasonality,

Unit:V Prescriptive Analytics: Introduction, What is Prescriptive Analytics?, Difference between Predictive and Prescriptive Analytics Using the Data to Maximize Revenue, Parameters of the Model, Market Structure, Competition and Online Advertising Models

Books:

1. Predictive Analytics and Data Mining Concepts and Practice with RapidMiner, 'Vijay Kotu, Bala Deshpande, PhD', Elsevier Publication.

URLs

1. https://www.coursera.org/learn/wharton-customer-analytics
2. https://www.analyticsvidhya.com/blog/2014/10/support-vector-machine-simplified/
3.https://www.neuraldesigner.com/blog/6_Applications_of_predictive_analytics_in_business_intelligence
4.https://doc.lagout.org/Others/Data%20Mining/Data%20Mining%20and%20Predictive%20Analytics%20%5BLarose%20%26%20Larose%202015-03-16%5D.pdf

.

Course Contents and Lecture Schedule

S.No.	Topic	No. of Lectures
1.	**Predictive Analytics**	
1.1	Predicttive Analytics-Definition, Predictive analytics Problems,	2
1.2	Regression analysis for Demand Curve and for making Predictions,	2
1.3	Making Predictions using a Data Set, Probability Models, Implementation of a Model.	2
2.	**Datamining concepts for Predictions**	
2.1	Data Mining Process,	3
2.2	Association analysis: Concepts of Mining Association Rules,	3
2.3	Apriori Algorithm, Frequent Item Set Generation,	3
2.4	FP Growth Algorithm,	3
3.	**Text Mining**	
3.1	Associated Terminologies,	3
3.2	How Text Mining Works: Term Frequency- Inverse Document Frequency,	3
3.3	Text mining with Clustering and Classification: Case Studies: Key Word Clustering,	5
3.4	Predicting the gender of Blog Authors.	
4.	**Time Series Forecasting**	
4.1	Need, Data driven approaches: Naive Forecast	3
4.2	Simple Average, Moving Average, Weighted Moving Average, Exponential Smoothing.	3
4.3	Model driven forecasting: Linear Regression,	3
4.4	Polynomial Regression,	4
4.5	Linear Regression with Seasonality,	
5	**Prescriptive Analytics**	
5.1	Introduction, What is Prescriptive Analytics?, Difference between Predictive and Prescriptive Analytics	3
5.2	Using the Data to Maximize Revenue, Parameters of the Model,	4
5.3	Market Structure, Competition and Online Advertising Models	

MDAC32 ETHICS FOR DATA SCIENTISTS	Category	L	P	Credit
	PE	4	0	4

Preamble

To understand the benefits and drawbacks of using data while using them for making predictions by understanding the structure of Ethics, law and societal values.

Prerequisite

Introduction to Data Analytics

Course Outcomes

On the successful completion of the course, students will be able to

Course Outcomes		Level
CO1	Able to understand the benefits and drawback of data.	Understanding
CO2	Understand ethical and Societal values while using data.	Understand and apply
CO3	Apply data ethics while practicing Data Science.	Apply

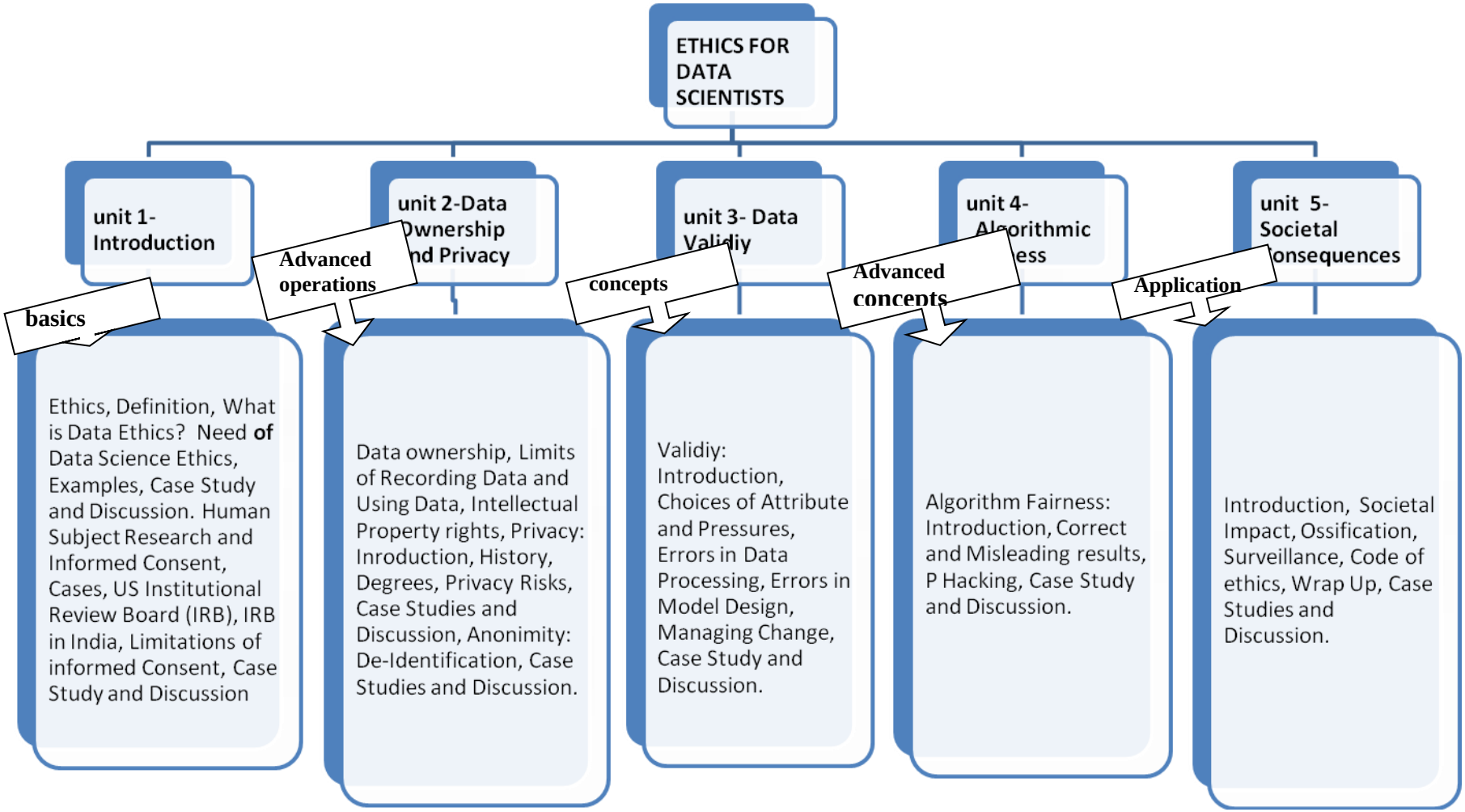
Mapping with Programme Outcomes

Cos	PSO1	PSO2	PSO3	PSO4	PSO5	PSO6
CO1	M	M	L	S	S	M
CO2	M	M	L	S	S	M
CO3	S	S	L	S	S	M

Assessment Pattern

Category	Continuous Assessment Tests (25)			Terminal Examination (75)
Remember	5	5	5	20
Understand	10	10	10	20
Apply	5	5	5	10
Analyze	5	5	5	10
Evaluate				5
Create				5

Concept Map


ETHICS FOR DATA SCIENTISTS
unit 1- Introduction
unit 2-Data Ownership nd Privacy
unit 3- Data Validiy
unit 4- Algorithmic ess
unit 5- Societal onsequences
basics
Advanced operations
concepts
Advanced concepts
Application
Ethics, Definition, What is Data Ethics? Need of Data Science Ethics, Examples, Case Study and Discussion. Human Subject Research and Informed Consent, Cases, US Institutional Review Board (IRB), IRB in India, Limitations of informed Consent, Case Study and Discussion
Data ownership, Limits of Recording Data and Using Data, Intellectual Property rights, Privacy: Inroduction, History, Degrees, Privacy Risks, Case Studies and Discussion, Anonimity: De-Identification, Case Studies and Discussion.
Validiy: Introduction, Choices of Attribute and Pressures, Errors in Data Processing, Errors in Model Design, Managing Change, Case Study and Discussion.
Algorithm Fairness: Introduction, Correct and Misleading results, P Hacking, Case Study and Discussion.
Introduction, Societal Impact, Ossification, Surveillance, Code of ethics, Wrap Up, Case Studies and Discussion.

Syllabus

Unit I: Introduction: Ethics, Definition, What is Data Ethics? Need **of** Data Science Ethics, Examples, Case Study and Discussion. Human Subject Research and Informed Consent, Cases, US Institutional Review Board (IRB), IRB in India, Limitations of informed Consent, Case Study and Discussion.

Unit: II Data Ownership and Privacy: Data ownership, Limits of Recording Data and Using Data, Intellectual Property rights, Privacy: Inroduction, History, Degrees, Privacy Risks, Case Studies and Discussion, Anonimity: De-Identification, Case Studies and Discussion.

Unit: III Data Validiy: Validiy: Introduction, Choices of Attribute and Pressures, Errors in Data Processing, Errors in Model Design, Managing Change, Case Study and Discussion.

Unit:IV Algorithmic Fairness: Algorithm Fairness: Introduction, Correct and Misleading results, P Hacking, Case Study and Discussion.

Unit:V Societal Consequences: Introduction, Societal Impact, Ossification, Surveillance, Code of ethics, Wrap Up, Case Studies and Discussion.

Books:

1. Ethics and Data Science, ' DJ Patil, Hilary Mason, Mike Loukides', O'Reilly Media, Inc.

URLs

MOOCS: https://courses.edx.org/courses/course-v1:MichiganX+DS101x+1T2018/courseware/94ac457869964552a69a3f37ba579954/671f4645836145eea658edbf9298be64/

.

Course Contents and Lecture Schedule

S.No.	Topic	No. of Lectures
1.	**Introduction**	
1.1	Ethics, Definition, What is Data Ethics?	2
1.2	Need **of** Data Science Ethics, Examples, Case Study and Discussion.	2
1.3	Human Subject Research and Informed Consent, Cases, US Institutional Review Board (IRB), IRB in India,	2
1.4	Limitations of informed Consent,	2
1.5	Case Study and Discussion	2
2.	**Data Ownership and Privacy**	
2.1	Data ownership, Limits of Recording Data and Using Data	3
2.2	Intellectual Property rights,	3
2.3	Privacy: Inroduction, History, Degrees.	3
2.4	Privacy Risks, Case Studies and Discussion,	3
2.5	Anonimity: Introduction, De-Identification,	2
2.6	Case Studies and Discussion	2
3.	**Data Validiy**	
3.1	Validiy: Introduction, Choices of Attribute and Pressures,	3
3.2	Errors in Data Processing, Errors in Model Design,	3
3.3	Managing Change	5
3.4	Case Study and Discussion.	
4.	**Algorithmic Fairness**	
4.1	Algorithm Fairness: Introduction,	3
4.2	Correct and Misleading results.	3
4.3	P Hacking,	3
4.4	Case Study and Discussion	2
5	**Societal Consequences**	
5.1	Introduction, Societal Impact,	3
5.2	Ossification, Surveillance,	4
5.3	Code of ethics, Wrap Up,	
5.4	Case Studies and Discussion	2

MDAL31 PREDICTIVE AND PRESCRIPTIVE ANALYTICS LABORATORY

Exercises

1. Getting familiar with Rapid Miner.
2. The retrieval of association rules from a data set is implemented through the FP-Growth algorithm in RapidMiner.
3. Execute the **Crawl Web** operator of Rapid Miner to allow setting up of simple crawling rules and based on these rules to store the crawled pages in a directory for further processing.
4. Execute the **GetPage** operator of Rapid Miner to retrieve a single page and stores the content as an example set and the **Get Pages** operator to access multiple pages identified by their URLs contained in an input file.
5. Predict the content of Blog Authors using the contents of Blog using Rapid Miner.
6. Model a Time series to forecast profits using Rapid Miner for the next twelve months for a dataset which has the monthly profit of a particular product from January 2009 to June 2010. The process consists of the following three steps: (1) set up windowing; (2) train the model with several different algorithms; and (3) generate the forecasts.

References:

Predictive Analytics and Data Mining Concepts and Practice with RapidMiner, 'Vijay Kotu, Bala Deshpande, PhD', Elsevier Publication. Exercise: 2, Chapter: 6 Pages 211-214, Exercises: 3,4 Chapter 9 Pages: 285, Exercises: 5 Chapter 9 Pages 287-302, Exercises; 6.Chapter 10-Pages: 318-326.

MDAI31 INTERNSHIP

The student can pursue the Internship in an IT company for not more than 2 weeks to acquire Data Analytics Skill Set or can opt for a MOOC course for the same. To assessment for the Internship either done in the Industry or through a MOOC is done by a Viva Voce conducted for the candidate by a Department Committee chaired by the Head of the department with two other faculty members from the department. On successful completion of the Viva Voce, the CIA is carried out for 25% marks from the Presentation on the Internship and the External Assessment is carried for 75% marks from the Final Internship Report and Viva-Voce.

SEMESTER IV

LIST OF COURSES
(For The Candidates Admitted From 2017-18 Onwards)

Sl. No.	Subject code	Course name
1.		Elective 6
2.		Elective 7
3.		Elective8
4.	**MDAP41**	Major Project

MDAP41 MAJOR PROJECT

The Student can carry out the final Major project work in 4th semester in an IT company, the student can get permission from the concerned Project Supervisor and Head of the Department after submitting the Acceptance Letter from the IT Company. For Major Project, the CIA is carried out for 25% marks and the External Assessment is for 75% marks from the Final Project Presentation, Project Report, and Viva-Voce.

LIST OF ELECTIVES FOR ELECTIVES 1,4
(For The Candidates Admitted From 2017-18 Onwards)

Sl. No.	Subject code	Course name
1.	MDAEA	Natural Language Processing for Data Analytics
2.	MDAEB	Big Data Analytics on Genomic Data
3.	MDAEC	Cloud Computing for Data Analytics
4.	MDAED	Programming with Mangodb for Data Analytics

MDAEA NATURAL LANGUAGE PROCESSING FOR DATA ANALYTICS

Category	L	P	Credit
PE	3	0	3

Preamble

To .Extract information from Unstructured Text, Analyze linguistic structure in text and apply the techniques for text based analytics

Prerequisite

Introduction to machine learning

Course Outcomes

On the successful completion of the course, students will be able to

Course Outcomes		Level
CO1	Extract information from Unstructured Text	Understanding
CO2	Analyze linguistic structure in text.	Understanding
CO3	Apply the techniques for text based analytics	Apply

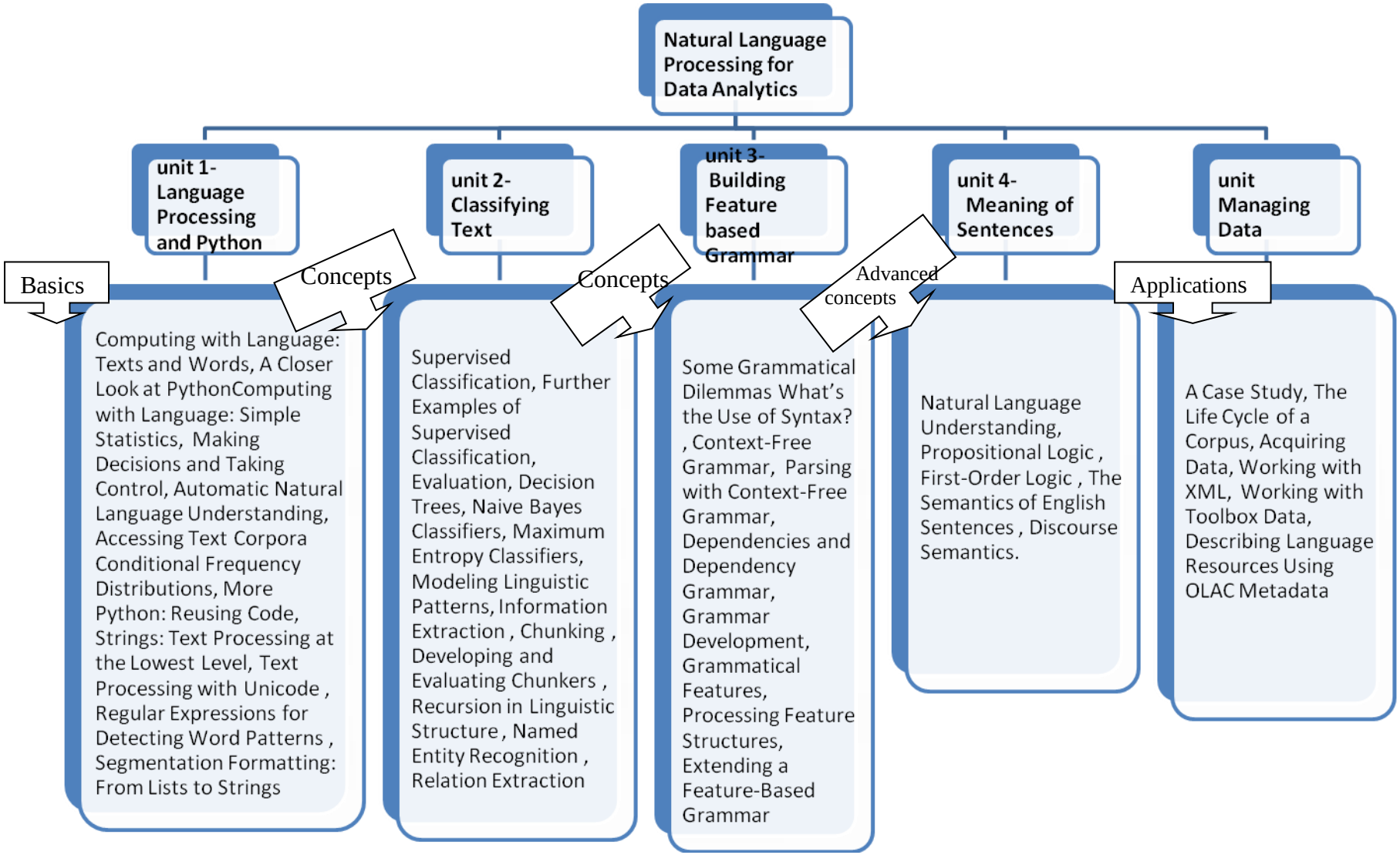
Mapping with Programme Outcomes

Cos	PSO1	PSO2	PSO3	PSO4	PSO5	PSO6
CO1	S	S	M	L	S	L
CO2	S	S	M	L	S	L
CO3	S	S	M	L	S	L

Assessment Pattern

Category	Continuous Assessment Tests (25)			Terminal Examination (75)
Remember	5	5	5	20
Understand	10	10	10	20
Apply	5	5	5	10
Analyze	5	5	5	10
Evaluate				5
Create				

Concept Map


Natural Language Processing for Data Analytics
unit 1- Language Processing and Python
unit 2- Classifying Text
unit 3- Building Feature based Grammar
unit 4- Meaning of Sentences
unit Managing Data
Basics
Concepts
Concepts
Advanced concepts
Applications
Computing with Language: Texts and Words, A Closer Look at PythonComputing with Language: Simple Statistics, Making Decisions and Taking Control, Automatic Natural Language Understanding, Accessing Text Corpora Conditional Frequency Distributions, More Python: Reusing Code, Strings: Text Processing at the Lowest Level, Text Processing with Unicode , Regular Expressions for Detecting Word Patterns , Segmentation Formatting: From Lists to Strings
Supervised Classification, Further Examples of Supervised Classification, Evaluation, Decision Trees, Naive Bayes Classifiers, Maximum Entropy Classifiers, Modeling Linguistic Patterns, Information Extraction , Chunking , Developing and Evaluating Chunkers , Recursion in Linguistic Structure , Named Entity Recognition , Relation Extraction
Some Grammatical Dilemmas What's the Use of Syntax? , Context-Free Grammar, Parsing with Context-Free Grammar, Dependencies and Dependency Grammar, Grammar Development, Grammatical Features, Processing Feature Structures, Extending a Feature-Based Grammar
Natural Language Understanding, Propositional Logic , First-Order Logic , The Semantics of English Sentences , Discourse Semantics.
A Case Study, The Life Cycle of a Corpus, Acquiring Data, Working with XML, Working with Toolbox Data, Describing Language Resources Using OLAC Metadata

Syllabus

Unit: I Language Processing and Python: Computing with Language: Texts and Words, A Closer Look at Python: Texts as Lists of Words, Computing with Language: Simple Statistics, Making Decisions and Taking Control, Automatic Natural Language Understanding, Accessing Text Corpora Conditional Frequency Distributions, More Python: Reusing Code, Lexical Resources, WordNet, Accessing Text from the Web and from Disk , Strings: Text Processing at the Lowest Level, Text Processing with Unicode , Regular Expressions for Detecting Word Patterns , Useful Applications of Regular Expressions , Normalizing Text , Regular Expressions for Tokenizing Text , Segmentation Formatting: From Lists to Strings

Unit: II Classifying Text: Supervised Classification, Further Examples of Supervised Classification, Evaluation, Decision Trees, Naive Bayes Classifiers, Maximum Entropy Classifiers, Modeling Linguistic Patterns, Information Extraction , Chunking , Developing and Evaluating Chunkers , Recursion in Linguistic Structure , Named Entity Recognition , Relation Extraction,

Unit: III Building Feature based Grammar: Some Grammatical Dilemmas What's the Use of Syntax? , Context-Free Grammar, Parsing with Context-Free Grammar, Dependencies and Dependency Grammar, Grammar Development, Grammatical Features, Processing Feature Structures, Extending a Feature-Based Grammar

Unit: IV Meaning of Sentences: Natural Language Understanding, Propositional Logic , First-Order Logic , The Semantics of English Sentences , Discourse Semantics.

Unit: V Managing Data: A Case Study, The Life Cycle of a Corpus, Acquiring Data, Working with XML, Working with Toolbox Data, Describing Language Resources Using OLAC Metadata

References:

Natural Language Processing with Python, 'Steven Bird, Ewan Klein, Edward Loper ', O'Reilly Media Inc.

Course Contents and Lecture Schedule

S.No.	Topic	No. of Lectures
1.	**Language Processing and Python**	
1.1	Computing with Language: Texts and Words, A Closer Look at Python: Texts as Lists of Words, ,	3
1.2	Computing with Language: Simple Statistics, Making Decisions and Taking Control, Automatic Natural Language Understanding,	3
1.3	Accessing Text Corpora Conditional Frequency Distributions, More Python: Reusing Code, Lexical Resources, WordNet, Accessing Text from the Web and from Disk ,	3
1.4	Strings: Text Processing at the Lowest Level, Text Processing with Unicode , Regular Expressions for Detecting Word Patterns	
1.5	Useful Applications of Regular Expressions , Normalizing Text , Regular Expressions for Tokenizing Text , Segmentation Formatting: From Lists to Strings	2
2.	**Classifying Text**	
2.1	Supervised Classification, Further Examples of Supervised Classification, Evaluation	4
2.2	Decision Trees, Naive Bayes Classifiers, Maximum Entropy Classifiers,	3
2.3	**:**, Modeling Linguistic Patterns, Information Extraction , Chunking , Developing and Evaluating Chunkers	2
2.4	Recursion in Linguistic Structure , Named Entity Recognition , Relation Extraction,	3
3.	**Building Feature based Grammar**	
3.1	Some Grammatical Dilemmas What's the Use of Syntax?	3
3.2	Context-Free Grammar, Parsing with Context-Free Grammar, Dependencies and Dependency Grammar,	2
3.3	Grammar Development, Grammatical Features,	3
3.4	Processing Feature Structures, Extending a Feature-Based Grammar	3
4.	**Meaning of Sentences**	
4.1	Natural Language Understanding,	2
4.2	Propositional Logic ,	2
4.3	First-Order Logic ,	2
4.4	The Semantics of English Sentences	2
4.5	Discourse Semantics.	2
5	**Managing Data**	
5.1	A Case Study, The Life Cycle of a Corpus,	3
5.2	Acquiring Data, Working with XML,	3
5.3	Working with Toolbox Data,	3
5.4	Describing Language Resources Using OLAC Metadata	3

MDAEB	**BIG DATA ANALYTICS ON GENOMIC DATA**	Category	L	P	Credit
		PE	3	0	3

Preamble

To articulate the main concepts of gene molecular structure, learn algorithms to do gene sequencing and other relevant operations, implement suitable algorithms for gene based analytics.

Prerequisite

Basic knowledge of Genetic

Course Outcomes

On the successful completion of the course, students will be able to

Course Outcomes		Level
CO1	Understand the gene molecular structure	Understanding
CO2	Learn algorithms to do gene sequencing and other relevant operations	Understanding
CO3	Implement suitable algorithms for gene based analytics.	Apply

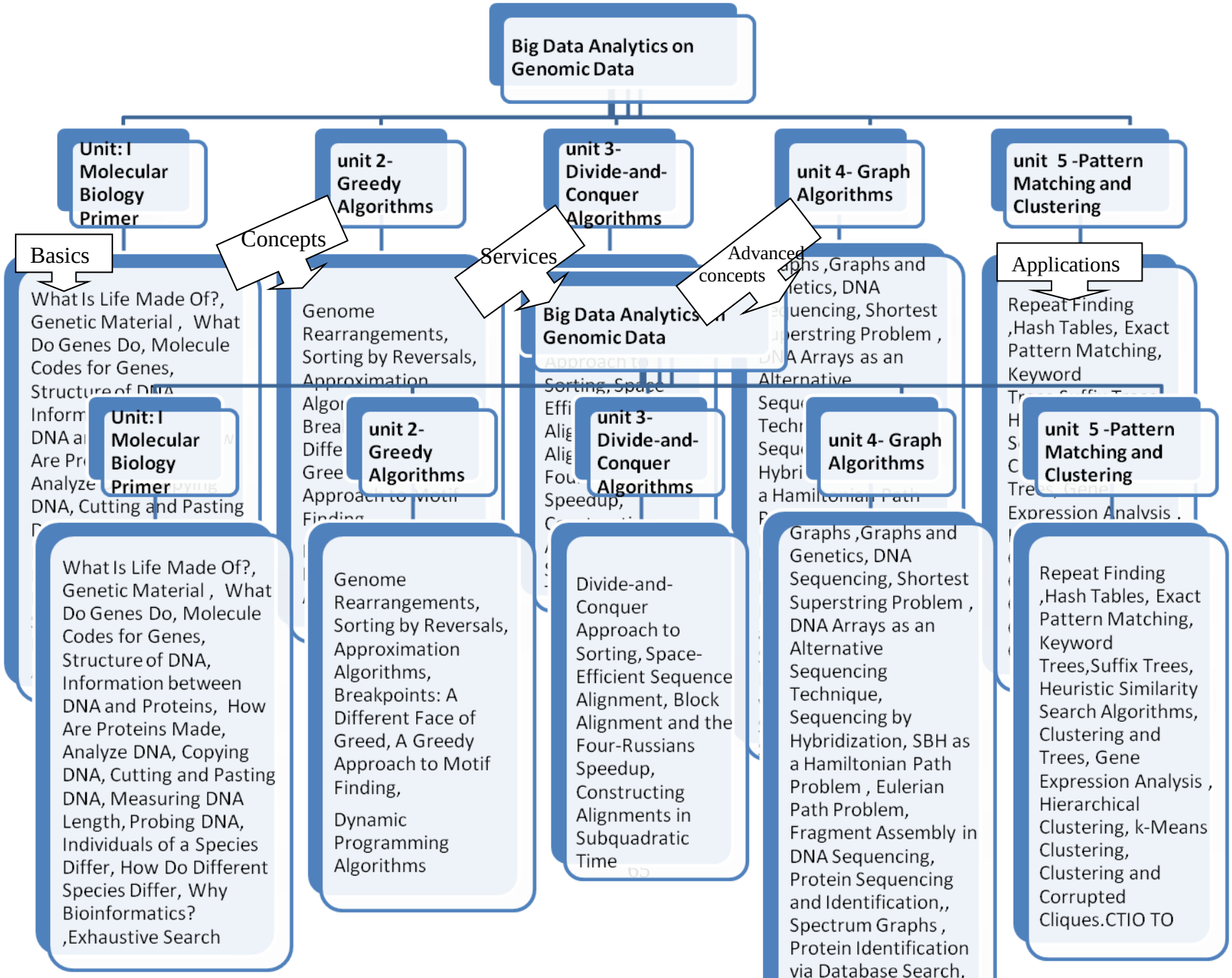
Mapping with Programme Outcomes

Cos	PSO1	PSO2	PSO3	PSO4	PSO5	PSO6
CO1	S	S	S	L	L	L
CO2	S	S	S	L	L	L
CO3	S	S	S	L	L	L

Assessment Pattern

Category	Continuous Assessment Tests (25)			Terminal Examination (75)
Remember	5	5	5	20
Understand	10	10	10	20
Apply	5	5	5	10
Analyze	5	5	5	10
Evaluate				5
Create				

Concept Map


Big Data Analytics on Genomic Data
Unit: I Molecular Biology Primer
unit 2- Greedy Algorithms
unit 3- Divide-and-Conquer Algorithms
unit 4- Graph Algorithms
unit 5 -Pattern Matching and Clustering
Basics
Concepts
Services
Advanced concepts
Applications
What Is Life Made Of?, Genetic Material , What Do Genes Do, Molecule Codes for Genes, Structure of DNA, Information between DNA and Proteins, How Are Proteins Made, Analyze DNA, Copying DNA, Cutting and Pasting DNA, Measuring DNA Length, Probing DNA, Individuals of a Species Differ, How Do Different Species Differ, Why Bioinformatics? ,Exhaustive Search
Genome Rearrangements, Sorting by Reversals, Approximation Algorithms, Breakpoints: A Different Face of Greed, A Greedy Approach to Motif Finding,
Dynamic Programming Algorithms
Big Data Analytics on Genomic Data
Divide-and-Conquer Approach to Sorting, Space-Efficient Sequence Alignment, Block Alignment and the Four-Russians Speedup, Constructing Alignments in Subquadratic Time
Graphs ,Graphs and Genetics, DNA Sequencing, Shortest Superstring Problem , DNA Arrays as an Alternative Sequencing Technique, Sequencing by Hybridization, SBH as a Hamiltonian Path Problem , Eulerian Path Problem, Fragment Assembly in DNA Sequencing, Protein Sequencing and Identification,, Spectrum Graphs , Protein Identification via Database Search,
Repeat Finding ,Hash Tables, Exact Pattern Matching, Keyword Trees,Suffix Trees, Heuristic Similarity Search Algorithms, Clustering and Trees, Gene Expression Analysis , Hierarchical Clustering, k-Means Clustering, Clustering and Corrupted Cliques.CTIO TO

Syllabus

Unit: I Molecular Biology Primer: What Is Life Made Of?, Genetic Material , What Do Genes Do, Molecule Codes for Genes, Structure of DNA, Information between DNA and Proteins, How Are Proteins Made, Analyze DNA, Copying DNA, Cutting and Pasting DNA, Measuring DNA Length, Probing DNA, Individuals of a Species Differ, How Do Different Species Differ, Why Bioinformatics?

Exhaustive Search: Restriction Mapping, Impractical Restriction Mapping Algorithms, A Practical Restriction Mapping Algorithm, Regulatory Motifs in DNA Sequences, Profiles, The Motif Finding Problem, Search Trees, Finding Motifs, Finding a Median String

Unit: II Greedy Algorithms: Genome Rearrangements, Sorting by Reversals, Approximation Algorithms, Breakpoints: A Different Face of Greed, A Greedy Approach to Motif Finding,

Dynamic Programming Algorithms : The Power of DNA Sequence Comparison, The Change Problem Revisited, The Manhattan Tourist Problem, Edit Distance and Alignments, Longest Common Subsequences, Global Sequence Alignment, Scoring Alignments, Local Sequence Alignment, Alignment with Gap Penalties, Multiple Alignment, Gene Prediction, Statistical Approaches to Gene Prediction, Similarity-Based Approaches to Gene Prediction, Spliced Alignment.

Unit: III Divide-and-Conquer Algorithms: Divide-and-Conquer Approach to Sorting, Space-Efficient Sequence Alignment, Block Alignment and the Four-Russians Speedup, Constructing Alignments in Subquadratic Time,

Unit: IV Graph Algorithms: Graphs ,Graphs and Genetics, DNA Sequencing, Shortest Superstring Problem , DNA Arrays as an Alternative Sequencing Technique, Sequencing by Hybridization, SBH as a Hamiltonian Path Problem , SBH as an Eulerian Path Problem, Fragment Assembly in DNA Sequencing, Protein Sequencing and Identification, The Peptide Sequencing Problem, Spectrum Graphs , Protein Identification via Database Search, Spectral Convolution, Spectral Alignment

Unit: V Pattern Matching and Clustering : Repeat Finding ,Hash Tables, Exact Pattern Matching,
Keyword Trees,Suffix Trees, Heuristic Similarity Search Algorithms, Clustering and Trees, Gene Expression Analysis , Hierarchical Clustering, k-Means Clustering, Clustering and Corrupted Cliques.

References:

1. Bioinformatics algorithms, ‘Neil c. Jones and pavel a. Pevzner’, A Bradford Book
The MIT Press.
2. http://www.bioalgorithms.info

Course Contents and Lecture Schedule

S.No.	Topic	No. of Lectures
1.	**Molecular Biology Primer**	
1.1	What Is Life Made Of?, Genetic Material ,	2
1.2	What Do Genes Do, Molecule Codes for Genes, Structure of DNA, Information between DNA and Proteins,	2
1.3	How Are Proteins Made, Analyze DNA, Copying DNA, Cutting and Pasting DNA,	2
1.4	Measuring DNA Length, Probing DNA,	
1.5	Individuals of a Species Differ, How Do Different Species Differ?	3
1.6	Why Bioinformatics?	
1.7	**Exhaustive Search:** Restriction Mapping, Impractical Restriction Mapping Algorithms	4
1.8	A Practical Restriction Mapping Algorithm, Regulatory Motifs in DNA Sequences, Profiles, The Motif Finding Problem, Search Trees, Finding Motifs, Finding a Median String	3
2.	**Greedy Algorithms**	
2.1	Genome Rearrangements, Sorting by Reversals, Approximation Algorithms, Breakpoints: A Different Face of Greed, A Greedy Approach to Motif Finding,	3
2.2	The Power of DNA Sequence Comparison, The Change Problem Revisited, The Manhattan Tourist Problem	3
2.3	, Edit Distance and Alignments, Longest Common Subsequences, Global Sequence Alignment,	3
2.4	Scoring Alignments, Local Sequence Alignment, Alignment with Gap Penalties, Multiple Alignment,	3
2.5	Gene Prediction, Statistical Approaches to Gene Prediction,	3
2.6	Similarity-Based Approaches to Gene Prediction, Spliced Alignment	2
3.	**Divide-and-Conquer Algorithms**	
3.1	Divide-and-Conquer Approach to Sorting,	3
3.2	Space-Efficient Sequence Alignment,	2
3.3	Block Alignment and the Four-Russians Speedup	3
3.4	, Constructing Alignments in Subquadratic Time,	1
4.	**Graph Algorithms**	
4.1	Graphs ,Graphs and Genetics, DNA Sequencing, Shortest Superstring Problem ,	2
4.2	DNA Arrays as an Alternative Sequencing Technique, Sequencing by Hybridization, SBH as a Hamiltonian Path Problem ,	2
4.3	SBH as an Eulerian Path Problem, Fragment Assembly in DNA Sequencing, Protein Sequencing and Identification,	2
4.4	The Peptide Sequencing Problem, Spectrum Graphs , Protein Identification via Database Search,	2
4.5	Spectral Convolution, Spectral Alignment	2
5	**Pattern Matching and Clustering**	

5.1	Repeat Finding ,Hash Tables, Exact Pattern Matching, Keyword Trees,Suffix Trees,	3
5.2	Heuristic Similarity Search Algorithms, Clustering and Trees, Gene Expression Analysis , Hierarchical Clustering, k-Means Clustering, Clustering and Corrupted Cliques.	5
5.3	Heuristic Similarity Search Algorithms, Clustering and Trees, Gene Expression Analysis.	3
5.4	Hierarchical Clustering, k-Means Clustering, Clustering and Corrupted Cliques.	3

MDAEC CLOUD COMPUTING FOR DATA ANALYTICS	Category	L	P	Credit
	PE	3	0	3

Preamble

To articulate the main concepts, key technologies, strengths, and limitations of cloud computing and the possible applications for state-of-the-art cloud computing and to learn the core issues of cloud computing such as security, privacy, and interoperability.

Prerequisite

Introduction to networking

Course Outcomes

On the successful completion of the course, students will be able to

Course Outcomes		Level
CO1	Able to understand the main concepts, key technologies, strengths, and limitations of cloud computing	Understanding
CO2	Able to know the core issues of cloud computing such as security, privacy, and interoperability	Understanding
CO3	Able to generate new ideas and innovations in cloud computing	Apply

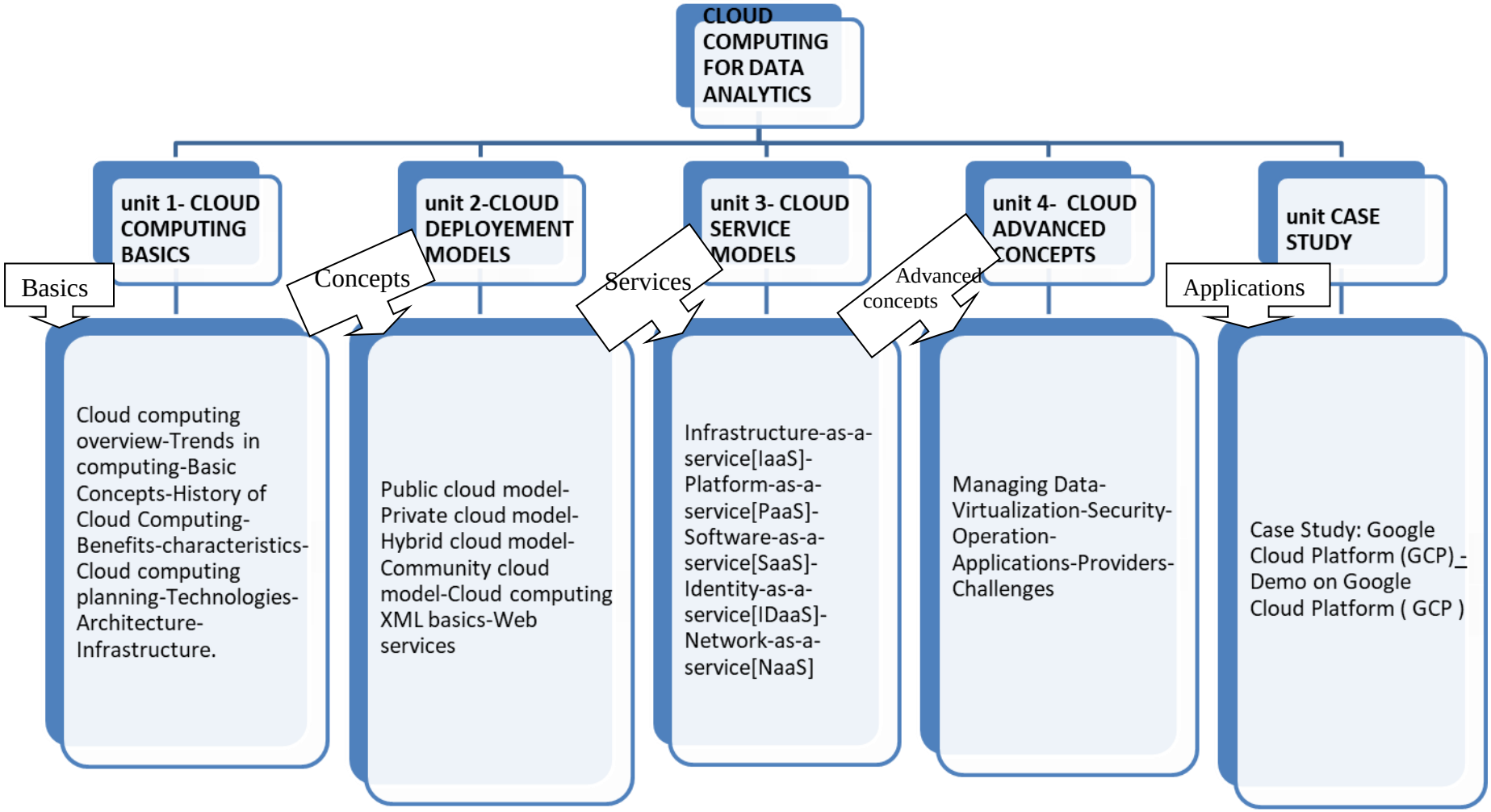
Mapping with Programme Outcomes

Cos	PSO1	PSO2	PSO3	PSO4	PSO5	PSO6
CO1	M	M				
CO2	M	M		L		
CO3		S	M		M	M

Assessment Pattern

Category	Continuous Assessment Tests (25)			Terminal Examination (75)
Remember	5	5	5	20
Understand	10	10	10	20
Apply	5	5	5	10
Analyze	5	5	5	10
Evaluate				5
Create				

Concept Map


CLOUD COMPUTING FOR DATA ANALYTICS
unit 1- CLOUD COMPUTING BASICS
unit 2-CLOUD DEPLOYEMENT MODELS
unit 3- CLOUD SERVICE MODELS
unit 4- CLOUD ADVANCED CONCEPTS
unit CASE STUDY
Basics
Concepts
Services
Advanced concepts
Applications
Cloud computing overview-Trends in computing-Basic Concepts-History of Cloud Computing-Benefits-characteristics-Cloud computing planning-Technologies-Architecture-Infrastructure.
Public cloud model-Private cloud model-Hybrid cloud model-Community cloud model-Cloud computing XML basics-Web services
Infrastructure-as-a-service[IaaS]-Platform-as-a-service[PaaS]-Software-as-a-service[SaaS]-Identity-as-a-service[IDaaS]-Network-as-a-service[NaaS]
Managing Data-Virtualization-Security-Operation-Applications-Providers-Challenges
Case Study: Google Cloud Platform (GCP) - Demo on Google Cloud Platform (GCP)

Syllabus

UNIT-I-CLOUD COMPUTING BASICS

Cloud computing overview-Trends in computing-Basic Concepts-History of Cloud Computing-Benefits-characteristics-Cloud computing planning-Technologies-Architecture-Infrastructure.

UNIT-II CLOUD DEPLOYEMENT MODELS

Public cloud model-Private cloud model-Hybrid cloud model-Community cloud model-Cloud computing XML basics-Web services.

UNIT-III CLOUD SERVICE MODELS

Infrastructure-as-a-service[IaaS]-Platform-as-a-service[PaaS]-Software-as-a-service [SaaS]-Identity-as-a-service [IDaaS]- Network-as-a-service [NaaS]

UNIT-IV CLOUD ADVANCED CONCEPTS

Managing Data-Virtualization-Security-Operation-Applications-Providers-Challenges

UNIT-V CASE STUDY

Case Study: Google Cloud Platform (GCP) - Demo on Google Cloud Platform (GCP)

References:

1. Cloud Computing: Principles and Paradigms, Editors: Rajkumar Buyya, James Broberg, Andrzej M. Goscinski, Wiley,2011
2. Enterprise Cloud Computing - Technology, Architecture, Applications, Gautam Shroff, Cambridge University Press, 2010
3. Cloud Computing Bible, Barrie Sosinsky, Wiley-India, 2010
4. Cloud Security: A Comprehensive Guide to Secure Cloud Computing, Ronald L. Krutz, Russell Dean Vines, Wiley- India,2010
5. https://nptel.ac.in/

Course Contents and Lecture Schedule

S.No.	Topic	No. of Lectures
5.	**CLOUD COMPUTING BASICS**	
1.1	*Cloud computing overview*	1
1.2	Trends in computing	2
1.3	Basic Concepts	1
1.4	History of Cloud Computing	
1.5	Benefits	2
1.6	characteristics	
1.7	Cloud computing planning	1
1.8	Technologies	3
1.9	Architecture	2
1.10	Infrastructure	
6.	**CLOUD DEPLOYEMENT MODELS**	
2.1	Public cloud model	2
2.2	Private cloud model	1
2.3	Hybrid cloud model	2
2.4	Community cloud model	3
2.5	Cloud computing XML basics	2
2.6	Web services	2
7.	**CLOUD SERVICE MODELS**	
3.1	Infrastructure-as-a-service[IaaS]	1
3.2	Platform-as-a-service[PaaS]	2
3.3	Software-as-a-service[SaaS]	2
3.4	Identity-as-a-service[IDaaS]	1
3.5	Network-as-a-service[NaaS]	2
8.	**CLOUD ADVANCED CONCEPTS**	
4.1	Managing Data	2
4.2	Virtualization	1
4.3	Security	1
4.4	Operation	2
4.5	Applications	1
4.6	Providers	3
4.7	Challenges	2
5	**CASE STUDY**	
5.1	Case Study: Google Cloud Platform (GCP)	3
5.2	Demo on Google Cloud Platform (GCP)	5

MDAED - PROGRAMMING WITH MONGODB FOR DATA ANALYTICS	Category	L	P	Credit
	PE	3	0	3

Preamble

To understand the NoSQL databases, design goals, operations performing with java driver and to learn GSON and java EE environment and creating application with wildfly

Prerequisite

Introduction to mongodb

Course Outcomes

On the successful completion of the course, students will be able to

Course Outcomes		Level
CO1	Able to understand the NoSQL databases, design goals, operations performing with java driver	Understanding
CO2	Able to learn GSON and java EE environment	Apply
CO3	Able to generate application with wildfly	Apply

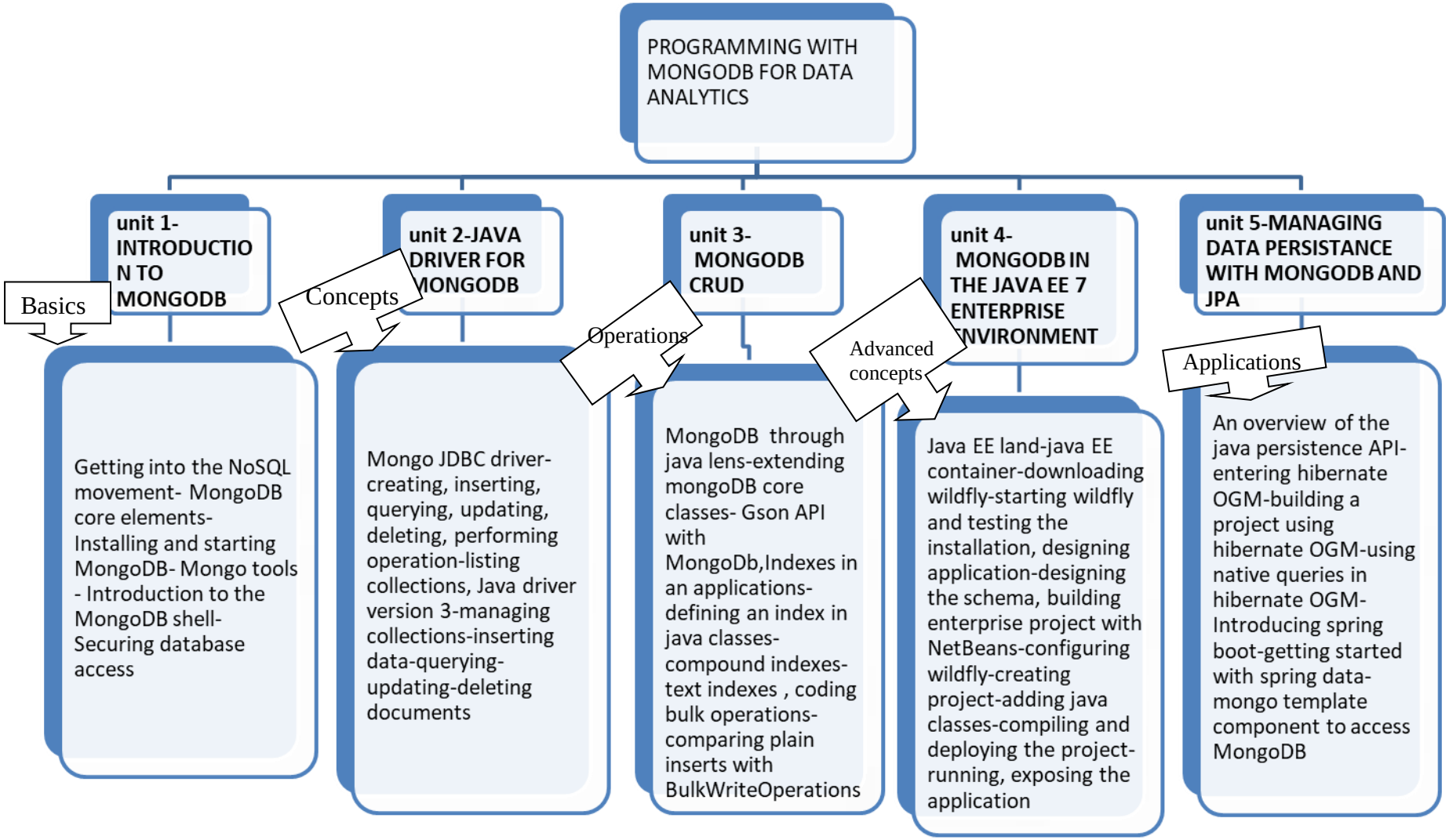
Mapping with Programme Outcomes

Cos	PSO1	PSO2	PSO3	PSO4	PSO5	PSO6
CO1	M	M				
CO2	M	M		L		
CO3		M			M	M

Assessment Pattern

Category	Continuous Assessment Tests (25)			Terminal Examination (75)
Remember	5	5	5	20
Understand	10	10	10	20
Apply	5	5	5	10
Analyze	5	5	5	10
Evaluate				5
Create				

Concept Map


PROGRAMMING WITH MONGODB FOR DATA ANALYTICS
unit 1- INTRODUCTION TO MONGODB
unit 2-JAVA DRIVER FOR MONGODB
unit 3- MONGODB CRUD
unit 4- MONGODB IN THE JAVA EE 7 ENTERPRISE ENVIRONMENT
unit 5-MANAGING DATA PERSISTANCE WITH MONGODB AND JPA
Basics
Concepts
Operations
Advanced concepts
Applications
Getting into the NoSQL movement- MongoDB core elements- Installing and starting MongoDB- Mongo tools - Introduction to the MongoDB shell- Securing database access
Mongo JDBC driver- creating, inserting, querying, updating, deleting, performing operation-listing collections, Java driver version 3-managing collections-inserting data-querying-updating-deleting documents
MongoDB through java lens-extending mongoDB core classes- Gson API with MongoDb,Indexes in an applications- defining an index in java classes- compound indexes- text indexes , coding bulk operations- comparing plain inserts with BulkWriteOperations
Java EE land-java EE container-downloading wildfly-starting wildfly and testing the installation, designing application-designing the schema, building enterprise project with NetBeans-configuring wildfly-creating project-adding java classes-compiling and deploying the project-running, exposing the application
An overview of the java persistence API-entering hibernate OGM-building a project using hibernate OGM-using native queries in hibernate OGM-Introducing spring boot-getting started with spring data-mongo template component to access MongoDB

Syllabus

UNIT - I: INTRODUCTION TO MONGODB

Getting into the NoSQL movement- MongoDB core elements- Installing and starting MongoDB- Mongo tools - Introduction to the MongoDB shell- Securing database access

UNIT - II: JAVA DRIVER FOR MONGODB

Mongo JDBC driver-creating, inserting, querying, updating, deleting, performing operation-listing collections, Java driver version 3-managing collections-inserting data-querying-updating-deleting documents

UNIT - III: MONGODB CRUD

MongoDB through java lens-extending mongoDB core classes- Gson API with MongoDb,Indexes in an applications-defining an index in java classes-compound indexes-text indexes , coding bulk operations-comparing plain inserts with BulkWriteOperations

UNIT - IV: MONGODB IN THE JAVA EE 7 ENTERPRISE ENVIRONMENT

Java EE land-java EE container-downloading wildfly-starting wildfly and testing the installation, designing application-designing the schema, building enterprise project with NetBeans-configuring wildfly-creating project-adding java classes-compiling and deploying the project-running, exposing the application

UNIT - V: MANAGING DATA PERSISTANCE WITH MONGODB AND JPA

An overview of the java persistence API-entering hibernate OGM-building a project using hibernate OGM-using native queries in hibernate OGM-Introducing spring boot-getting started with spring data-mongo template component to access MongoDB

References:

1. MongoDb for java developers- Francesco Marchioni, PACKT publishing 2015.
2. Java persistance with hibernate-Christian Bauer,Gavin King, 2007
3. http://www.packtpub. com/jboss-as-5-development/book
4. https://nptel.ac.in/

Course Contents and Lecture Schedule

S.No.	Topic	No. of Lectures
1.	**INTRODUCTION TO MONGODB**	
1.1	Getting into the NoSQL movement	1
1.2	MongoDb core elements	1
1.3	Installing and starting MongoDB	2
1.4	Mongo tools	
1.5	Introduction to the MongoDB shell	2
1.6	Securing database access	
2.	**JAVA DRIVER FOR MONGODB**	
2.1	MongoDb JDBC driver	2
2.2	Performing operation	
2.3	Listing collection	2
2.4	Java driver version 3	1
2.5	Managing collections	2
2.6	Inserting data,quering,updating,deleting documents	2
3.	**MONGODB CRUD**	
3.1	MongoDb through java lens	1
3.2	Extending mongoDB core classes	2
3.3	Indexing in an applications	2
3.4	Defining an index in java classes	1
3.5	Coding bulk operations	2
4.	**MONGODB IN THE JAVA EE 7 ENTERPRISE ENVIRONMENT**	
4.1	Java EE land and container	2
4.2	Downloading wildfly	1
4.3	Designing application	2
4.4	Configuring wildfly	2
4.5	Adding java classes	1
4.6	Complying and deploying the project	3
4.7	Running, exposing the application	1
5	**MANAGING DATA PERSISTANCE WITH MONGODB AND JPA**	
5.1	An overview of the java persistence API	1
5.2	Entering hibernate OGM	2
5.3	Building a project using hibernate OGM	2
5.4	Using native queries in hibernate OGM	2
5.5	Introducing spring boot	
5.6	Getting started with spring data	2
5.7	Mongo template component to access mongodb	1

LIST OF ELECTIVES FOR ELECTIVES 2,5
(For The Candidates Admitted From 2017-18 Onwards)

Sl. No.	Subject code	Course name
1.	MDAEE	Data Mining Essentials for Data Analytics
2.	MDAEF	Latest Machine Algorithms for Data Analytics
3.	MDAEG	Big Data Security
4.	MDAEH	Programming with Cassandra

MDAEE-DATA MINING ESSENTIALS FOR DATA ANALYTICS	Category	L	P	Credit
	PE	3	0	3

Preamble

To Learn the concepts of data base technology evolutionary path which has lead to the need for data mining and its application , Examine the types of the data to be mined and present a general classification of tasks and primitives to integrates a data mining system .

Prerequisite

Introduction to DBMS

Course Outcomes

On the successful completion of the course, students will be able to

Course Outcomes		Level
CO1	Learn the concepts of data base technology evolutionary path which has lead to the need for data mining and its application.	Understanding
CO2	Examine the types of the data to be mined and present a general classification of tasks and primitives to integrates a data mining system .	Understanding
CO3	Apply preprocessing statistical methods for any given raw data.	Apply

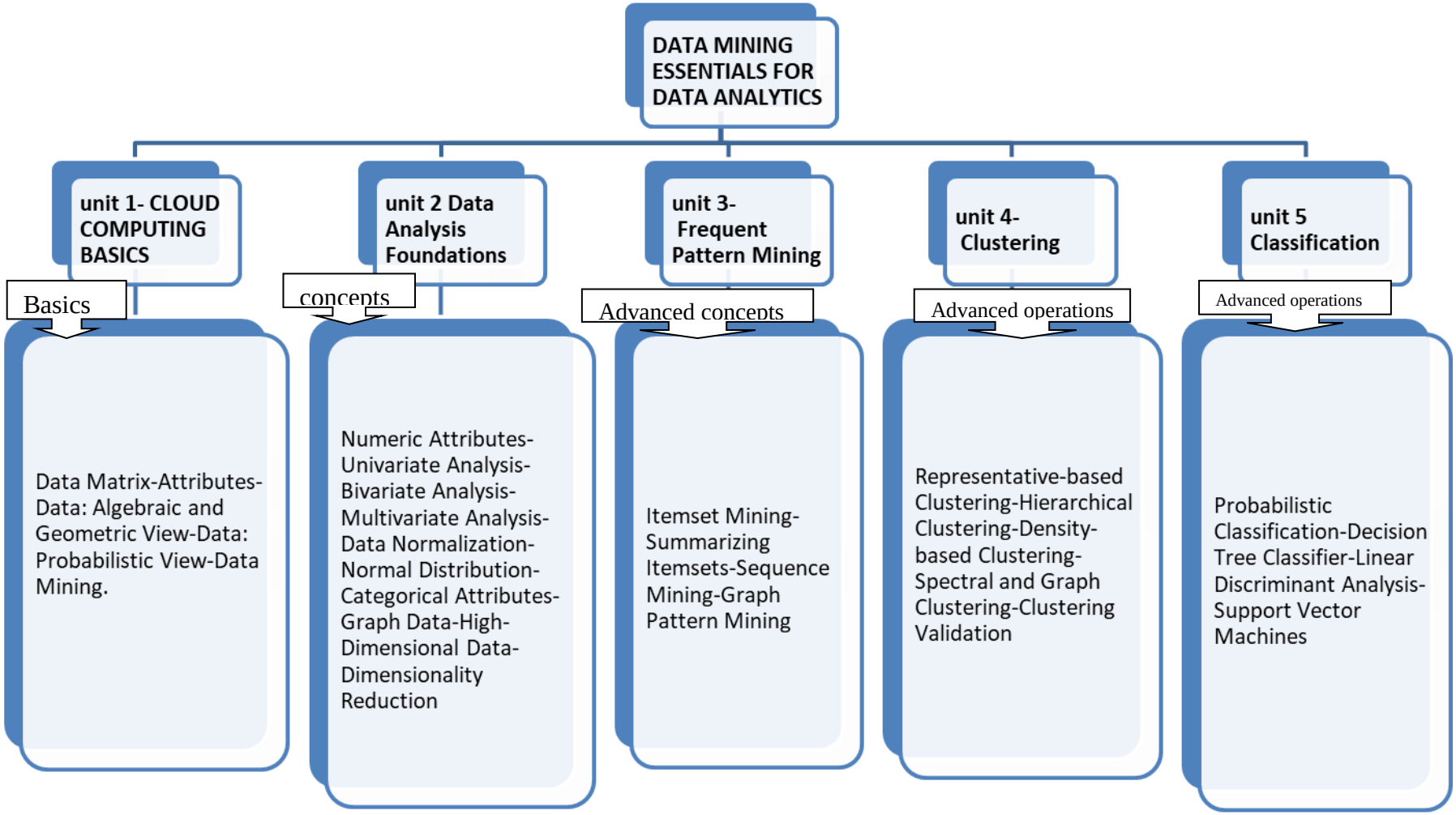
Mapping with Programme Outcomes

Cos	PSO1	PSO2	PSO3	PSO4	PSO5	PSO6
CO1	M	M				
CO2	M	M	M			
CO3		S	M		M	M

Assessment Pattern

Category	Continuous Assessment Tests (25)			Terminal Examination (75)
Remember	5	5	5	20
Understand	10	10	10	20
Apply	5	5	5	10
Analyze	5	5	5	10
Evaluate				5
Create				

Concept Map


DATA MINING ESSENTIALS FOR DATA ANALYTICS
unit 1- CLOUD COMPUTING BASICS
unit 2 Data Analysis Foundations
unit 3- Frequent Pattern Mining
unit 4- Clustering
unit 5 Classification
Basics
concepts
Advanced concepts
Advanced operations
Advanced operations
Data Matrix-Attributes-Data: Algebraic and Geometric View-Data: Probabilistic View-Data Mining.
Numeric Attributes-Univariate Analysis-Bivariate Analysis-Multivariate Analysis-Data Normalization-Normal Distribution-Categorical Attributes-Graph Data-High-Dimensional Data-Dimensionality Reduction
Itemset Mining-Summarizing Itemsets-Sequence Mining-Graph Pattern Mining
Representative-based Clustering-Hierarchical Clustering-Density-based Clustering-Spectral and Graph Clustering-Clustering Validation
Probabilistic Classification-Decision Tree Classifier-Linear Discriminant Analysis-Support Vector Machines

Syllabus

UNIT-I-Data Mining and Analysis

Data Matrix-Attributes-Data: Algebraic and Geometric View-Data: Probabilistic View-Data Mining

UNIT-II Data Analysis Foundations

Numeric Attributes-Univariate Analysis-Bivariate Analysis-Multivariate Analysis-Data Normalization-Normal Distribution-Categorical Attributes-Graph Data-High-Dimensional Data-Dimensionality Reduction

UNIT-III Frequent Pattern Mining

Itemset Mining-Summarizing Itemsets-Sequence Mining-Graph Pattern Mining

UNIT-IV Clustering

Representative-based Clustering-Hierarchical Clustering-Density-based Clustering-Spectral and Graph Clustering-Clustering Validation

UNIT-V Classification

Probabilistic Classification-Decision Tree Classifier-Linear Discriminant Analysis-Support Vector Machines

References:

1. Mohammed J. Zaki, Wagner Meira, Jr., Data Mining and Analysis: Fundamental Concepts and Algorithms, Cambridge University Press, May 2014. ISBN: 9780521766333.
2. https://repo.palkeo.com/algo/information-retrieval/Data%20mining%20and%20analysis.pdf/

Course Contents and Lecture Schedule

S.No.	Topic	No. of Lectures
1.	**DATA MINING AND ANALYSIS**	
1.1	Data Matrix	1
1.2	Attributes	1
1.3	Data: Algebraic and Geometric View	2
1.4	Data: Probabilistic View	
1.5	Data Mining	1
2.	**DATA ANALYSIS FOUNDATIONS**	
2.1	Numeric Attributes	2
2.2	Univariate Analysis	2
2.3	Bivariate Analysis	2
2.4	Multivariate Analysis	2
2.5	Data Normalization	2
2.6	Normal Distribution	
2.7	Categorical Attributes	1
2.8	Graph Data	
2.9	High-Dimensional Data	2
2.10	Dimensionality Reduction	
3.	**FREQUENT PATTERN MINING**	
3.1	Itemset Mining	1
3.2	Summarizing Itemsets	2
3.3	Sequence Mining	2
3.4	Graph Pattern Mining	2
4.	**CLUSTERING**	
4.1	Representative- based Clustering	2
4.2	Hierarchical Clustering	2
4.3	Density- based Clustering	2
4.4	Spectral and Graph Clustering	2
4.5	Clustering Validation	1
5	**CLASSIFICATION**	
5.1	Probabilistic Classification	2
5.2	Decision Tree Classifier	2
5.3	Linear Discriminant Analysis	2
5.4	Support Vector Machines	2

MDAEF LATEST MACHINE LANGUAGE FOR DATA ANALYTICS	Category	L	P	Credit
	PE	3	0	3

Preamble

To understand the importance of Machine Learning and to develop key machine learning skill set.To know the scope of Machine Learning for future.

Prerequisite

Introduction to machine learning

Course Outcomes

On the successful completion of the course, students will be able to

Course Outcomes		Level
CO1	Able to understand the importance of Machine Learning	Understanding
CO2	Able to develop key machine learning skill set	Understanding
CO3	Able to know the scope of Machine Learning for future	Apply

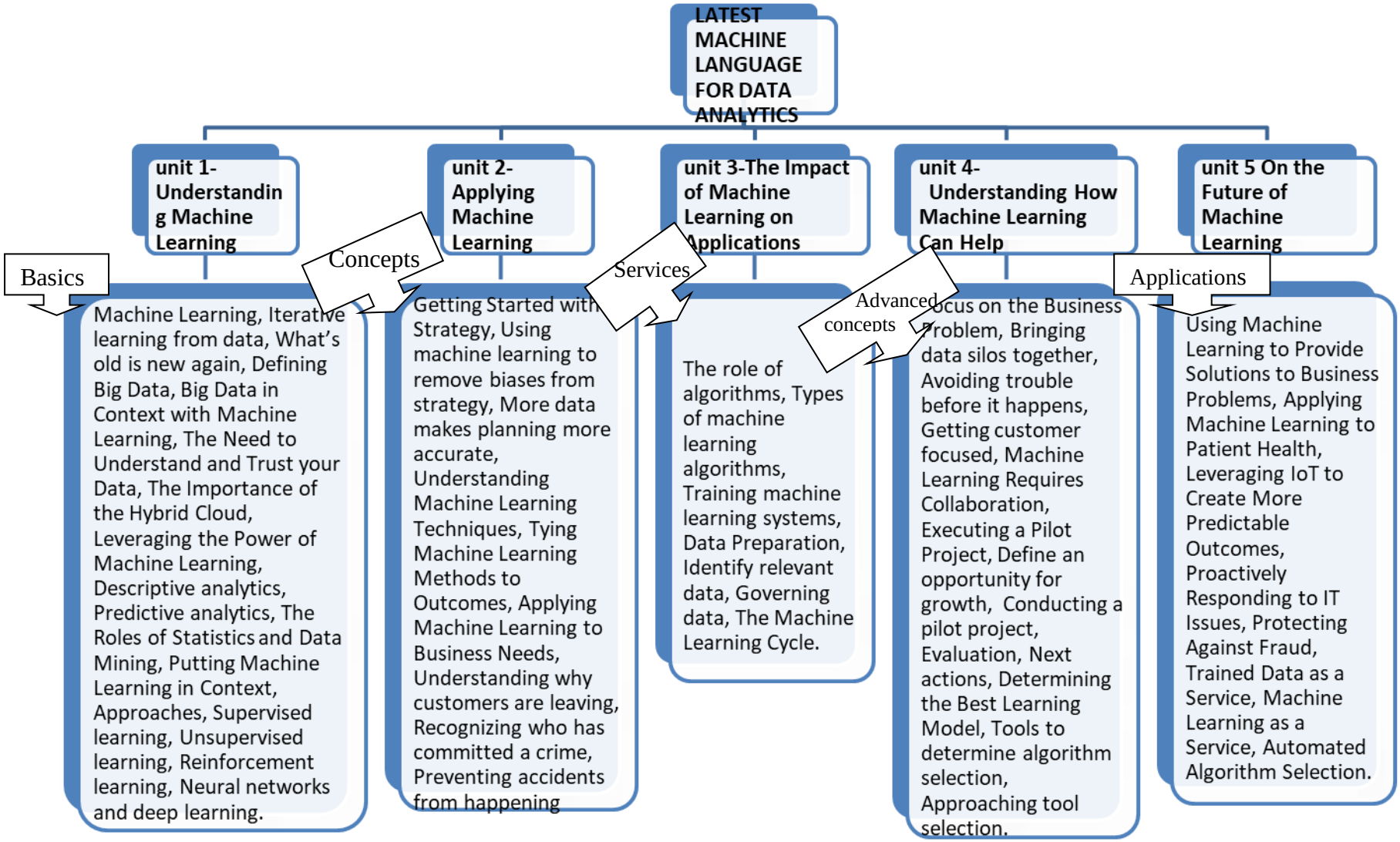
Mapping with Programme Outcomes

Cos	PSO1	PSO2	PSO3	PSO4	PSO5	PSO6
CO1	M	M	M	L	M	L
CO2	S	S	S	L	S	L
CO3	M	M	M	L	S	L

Assessment Pattern

Category	Continuous Assessment Tests (25)			Terminal Examination (75)
Remember	5	5	5	20
Understand	10	10	10	20
Apply	5	5	5	10
Analyze	5	5	5	10
Evaluate				5
Create				

Concept Map


LATEST MACHINE LANGUAGE FOR DATA ANALYTICS
unit 1- Understanding Machine Learning
unit 2- Applying Machine Learning
unit 3-The Impact of Machine Learning on Applications
unit 4- Understanding How Machine Learning Can Help
unit 5 On the Future of Machine Learning
Basics
Concepts
Services
Advanced concepts
Applications
Machine Learning, Iterative learning from data, What's old is new again, Defining Big Data, Big Data in Context with Machine Learning, The Need to Understand and Trust your Data, The Importance of the Hybrid Cloud, Leveraging the Power of Machine Learning, Descriptive analytics, Predictive analytics, The Roles of Statistics and Data Mining, Putting Machine Learning in Context, Approaches, Supervised learning, Unsupervised learning, Reinforcement learning, Neural networks and deep learning.
Getting Started with Strategy, Using machine learning to remove biases from strategy, More data makes planning more accurate, Understanding Machine Learning Techniques, Tying Machine Learning Methods to Outcomes, Applying Machine Learning to Business Needs, Understanding why customers are leaving, Recognizing who has committed a crime, Preventing accidents from happening
The role of algorithms, Types of machine learning algorithms, Training machine learning systems, Data Preparation, Identify relevant data, Governing data, The Machine Learning Cycle.
ocus on the Business roblem, Bringing data silos together, Avoiding trouble before it happens, Getting customer focused, Machine Learning Requires Collaboration, Executing a Pilot Project, Define an opportunity for growth, Conducting a pilot project, Evaluation, Next actions, Determining the Best Learning Model, Tools to determine algorithm selection, Approaching tool selection.
Using Machine Learning to Provide Solutions to Business Problems, Applying Machine Learning to Patient Health, Leveraging IoT to Create More Predictable Outcomes, Proactively Responding to IT Issues, Protecting Against Fraud, Trained Data as a Service, Machine Learning as a Service, Automated Algorithm Selection.

Syllabus

Unit: I **Understanding Machine Learning:** What Is Machine Learning?, Iterative learning from data, What's old is new again, Defining Big Data, Big Data in Context with Machine Learning, The Need to Understand and Trust your Data, The Importance of the Hybrid Cloud, Leveraging the Power of Machine Learning, Descriptive analytics, Predictive analytics, The Roles of Statistics and Data Mining with Machine Learning, Putting Machine Learning in Context, Approaches to Machine Learning, Supervised learning, Unsupervised learning, Reinforcement learning, Neural networks and deep learning.

Unit: II Applying Machine Learning: Getting Started with a Strategy, Using machine learning to remove biases from strategy, More data makes planning more accurate, Understanding Machine Learning Techniques, Tying Machine Learning Methods to Outcomes, Applying Machine Learning to Business Needs, Understanding why customers are leaving, Recognizing who has committed a crime, Preventing accidents from happening.

Unit: III The Impact of Machine Learning on Applications: The role of algorithms, Types of machine learning algorithms, Training machine learning systems, Data Preparation, Identify relevant data, Governing data, The Machine Learning Cycle.

Unit: IV Understanding How Machine Learning Can Help: Focus on the Business Problem, Bringing data silos together, Avoiding trouble before it happens, Getting customer focused, Machine Learning Requires Collaboration, Executing a Pilot Project, Define an opportunity for growth, Conducting a pilot project, Evaluation, Next actions, Determining the Best Learning Model, Tools to determine algorithm selection, Approaching tool selection.

Unit: V On the Future of Machine Learning: Using Machine Learning to Provide Solutions to Business Problems, Applying Machine Learning to Patient Health, Leveraging IoT to Create More Predictable Outcomes, Proactively Responding to IT Issues, Protecting Against Fraud, Trained Data as a Service, Machine Learning as a Service, Automated Algorithm Selection.

References:

1.Machine Learning, 'Judith Hurwitz and Daniel Kirsch', IBM Limited Edition, A Wiley Brand.

For Case Studies

1. medium.com/inside-machinelearning
2. www. udacity.com/course/intro-to-machine-learning--ud120.
3. www.galvanize.com/san-francisco/data-science.
4. www.edx. org/course/machine-learning-data-science-analyticscolumbiax-
5. https://cognitiveclass.ai to build data science and cognitive computing skills

Course Contents and Lecture Schedule

S.No.	Topic	No. of Lectures
1.	**Understanding Machine Learning**	
1.1	What Is Machine Learning?	1
1.2	Iterative learning from data	2
1.3	What's old is new again	2
1.4	Defining Big Data	
1.5	Big Data in Context with Machine Learning	2
1.6	The Need to Understand and Trust your Data	
1.7	The Importance of the Hybrid Cloud	2
1.8	Leveraging the Power of Machine Learning	2
1.9	Descriptive analytics	2
1.10	Predictive analytics	
1.11	The Roles of Statistics and Data Mining with Machine Learning	2
1.12	Putting Machine Learning in Context	
1.13	Approaches to Machine Learning	2
1.14	Supervised learning, Unsupervised learning	
1.15	Reinforcement learning	1
1.16	Neural networks and deep learning	
2.	**Applying Machine Learning**	
2.1	Getting Started with a Strategy	2
2.2	Using machine learning to remove biases from strategy	2
2.3	More data makes planning more accurate	2
2.4	Understanding Machine Learning Techniques	2
2.5	Tying Machine Learning Methods to Outcomes	2
2.6	Applying Machine Learning to Business Needs	
2.7	Understanding why customers are leaving	1
2.8	Recognizing who has committed a crime	1
2.9	Preventing accidents from happening	
3.	**The Impact of Machine Learning on Applications**	
3.1	The role of algorithms	1
3.2	Types of machine learning algorithms	2
3.3	Training machine learning systems	
3.4	Data Preparation	1
3.5	Identify relevant data	2
3.6	Governing data	2
3.7	The Machine Learning Cycle	1
4.	**Understanding How Machine Learning Can Help**	
4.1	Focus on the Business Problem	2
4.2	Bringing data silos together	
4.3	Avoiding trouble before it happens	2
4.4	Getting customer focused	
4.5	Machine Learning Requires Collaboration	2

4.6	Executing a Pilot Project	
4.7	Define an opportunity for growth	4
4.8	Conducting a pilot project	
4.9	Evaluation	2
4.10	Next actions	
4.11	Determining the Best Learning Model	2
4.12	Tools to determine algorithm selection	2
4.13	Approaching tool selection	2
5	**On the Future of Machine Learning**	
5.1	Using Machine Learning to Provide Solutions to Business Problems	2
5.2	Applying Machine Learning to Patient Health	2
5.3	Leveraging IoT to Create More Predictable Outcomes	2
5.4	Proactively Responding to IT Issues	
5.5	Protecting Against Fraud	2
5.6	Trained Data as a Service	2
5.7	Machine Learning as a Service	2
5.8	Automated Algorithm Selection	

MDAEG- BIG DATA SECURITY

Category	L	P	Credit
PE	3	0	3

Preamble

To articulate the main concepts, key technologies, strengths, and limitations of cloud computing and the possible applications for state-of-the-art cloud computing and to learn the core issues of cloud computing such as security, privacy, and interoperability.

Prerequisite

Introduction to networking

Course Outcomes

On the successful completion of the course, students will be able to

Course Outcomes		Level
CO1	Able to understand the main concepts, key technologies, strengths, and limitations of cloud computing	Understanding
CO2	Able to know the core issues of cloud computing such as security, privacy, and interoperability	Understanding
CO3	Able to to generate new ideas and innovations in cloud computing	Apply

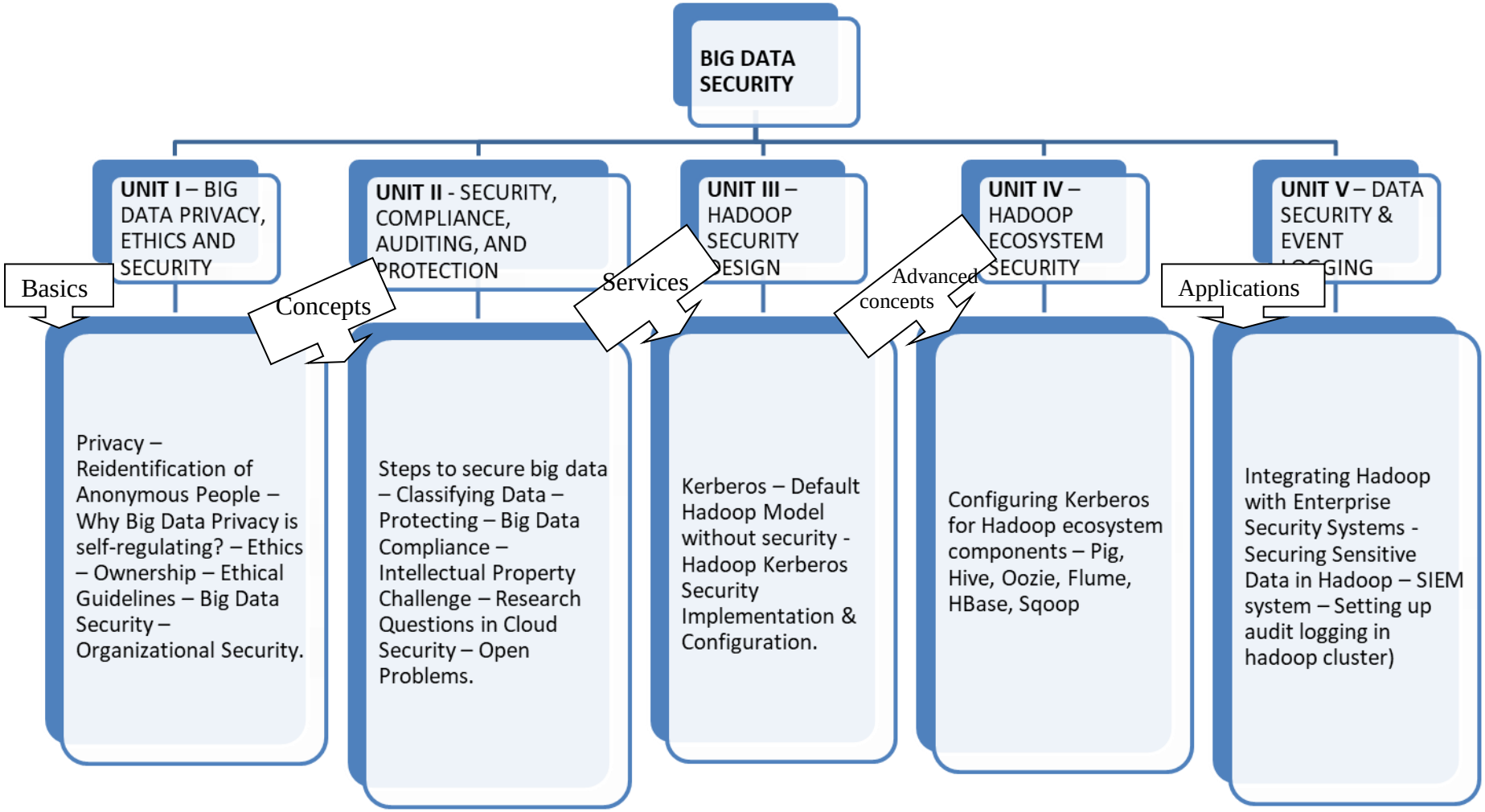
Mapping with Programme Outcomes

Cos	PSO1	PSO2	PSO3	PSO4	PSO5	PSO6
CO1	M	M				
CO2	M	M		L		
CO3		S	M		M	M

Assessment Pattern

Category	Continuous Assessment Tests (25)			Terminal Examination (75)
Remember	5	5	5	20
Understand	10	10	10	20
Apply	5	5	5	10
Analyze	5	5	5	10
Evaluate				5
Create				

Concept Map


BIG DATA SECURITY
UNIT I – BIG DATA PRIVACY, ETHICS AND SECURITY
UNIT II - SECURITY, COMPLIANCE, AUDITING, AND PROTECTION
UNIT III – HADOOP SECURITY DESIGN
UNIT IV – HADOOP ECOSYSTEM SECURITY
UNIT V – DATA SECURITY & EVENT LOGGING
Basics
Concepts
Services
Advanced concepts
Applications
Privacy – Reidentification of Anonymous People – Why Big Data Privacy is self-regulating? – Ethics – Ownership – Ethical Guidelines – Big Data Security – Organizational Security.
Steps to secure big data – Classifying Data – Protecting – Big Data Compliance – Intellectual Property Challenge – Research Questions in Cloud Security – Open Problems.
Kerberos – Default Hadoop Model without security - Hadoop Kerberos Security Implementation & Configuration.
Configuring Kerberos for Hadoop ecosystem components – Pig, Hive, Oozie, Flume, HBase, Sqoop
Integrating Hadoop with Enterprise Security Systems - Securing Sensitive Data in Hadoop – SIEM system – Setting up audit logging in hadoop cluster)

Syllabus

UNIT I – BIG DATA PRIVACY, ETHICS AND SECURITY Privacy – Reidentification of Anonymous People – Why Big Data Privacy is self-regulating? – Ethics – Ownership – Ethical Guidelines – Big Data Security – Organizational Security.

UNIT II - SECURITY, COMPLIANCE, AUDITING, AND PROTECTION Steps to secure big data – Classifying Data – Protecting – Big Data Compliance – Intellectual Property Challenge – Research Questions in Cloud Security – Open Problems.

UNIT III – HADOOP SECURITY DESIGN Kerberos – Default Hadoop Model without security - Hadoop Kerberos Security Implementation & Configuration.

UNIT IV – HADOOP ECOSYSTEM SECURITY Configuring Kerberos for Hadoop ecosystem components – Pig, Hive, Oozie, Flume, HBase, Sqoop.

UNIT V – DATA SECURITY & EVENT LOGGING Integrating Hadoop with Enterprise Security Systems - Securing Sensitive Data in Hadoop – SIEM system – Setting up audit logging in hadoop cluster

References:

1. Mark Van Rijmenam, "Think Bigger: Developing a Successful Big Data Strategy for Your Business", Amazon, 1 edition, 2014.
2. Frank Ohlhorst John Wiley & Sons, "Big Data Analytics: Turning Big Data into Big Money", John Wiley & Sons, 2013.
3. Sherif Sakr, "Large Scale and Big Data: Processing and Management", CRC Press, 2014.
4. Sudeesh Narayanan, "Securing Hadoop", Packt Publishing, 2013.
5. Ben Spivey, Joey Echeverria, "Hadoop Security Protecting Your Big Data Problem", O'Reilly Media, 2015.

Course Contents and Lecture Schedule

S.No.	Topic	No. of Lectures
1.	**BIG DATA PRIVACY, ETHICS AND SECURITY**	
1.1	*Privacy*	1
1.2	Reidentification of Anonymous People	2
1.3	Why Big Data Privacy is self-regulating?	2
1.4	Ethics	
1.5	Ownership	2
1.6	Ethical Guidelines	
1.7	Big Data Security	2
1.8	Organizational Security	1
2.	**SECURITY, COMPLIANCE, AUDITING, AND PROTECTION**	
2.1	Steps to secure big data	2
2.2	Classifying Data	1
2.3	Protecting	1
2.4	Big Data Compliance	2
2.5	Intellectual Property Challenge	2
2.6	Research Questions in Cloud Security	2
2.7	Open Problems	2
3.	**HADOOP SECURITY DESIGN**	
3.1	Kerberos	2
3.2	Default Hadoop Model without security	2
3.3	Hadoop Kerberos Security Implementation & Configuration	3
4.	**HADOOP ECOSYSTEM SECURITY**	
4.1	Configuring Kerberos for Hadoop ecosystem components	2
4.2	Pig	1
4.3	Hive	2
4.4	Oozie	2
4.5	Flume	2
4.6	HBase, Sqoop	4
5	**DATA SECURITY & EVENT LOGGING**	
5.1	Integrating Hadoop with Enterprise Security Systems	1
5.2	Securing Sensitive Data in Hadoop	2
5.3	SIEM system	2
5.4	Setting up audit logging in hadoop cluster	2

MDAEH-PROGRAMMING WITH CASSANDRA	Category	L	P	Credit
	PE	3	0	3

Preamble

To Understand the difference between RDBMS and Cassandra and to do cassandra Installation, Store and retrieve data in efficient way,linking java and cassandra.

Prerequisite

Introduction to DBMS

Course Outcomes

On the successful completion of the course, students will be able to

Course Outcomes		Level
CO1	Able to Understand the difference between RDBMS and Cassandra	Understanding
CO2	Able to do cassandra Installation, Store and retrieve data in efficient way	Understanding
CO3	Able to Linking Java and cassandra	Apply

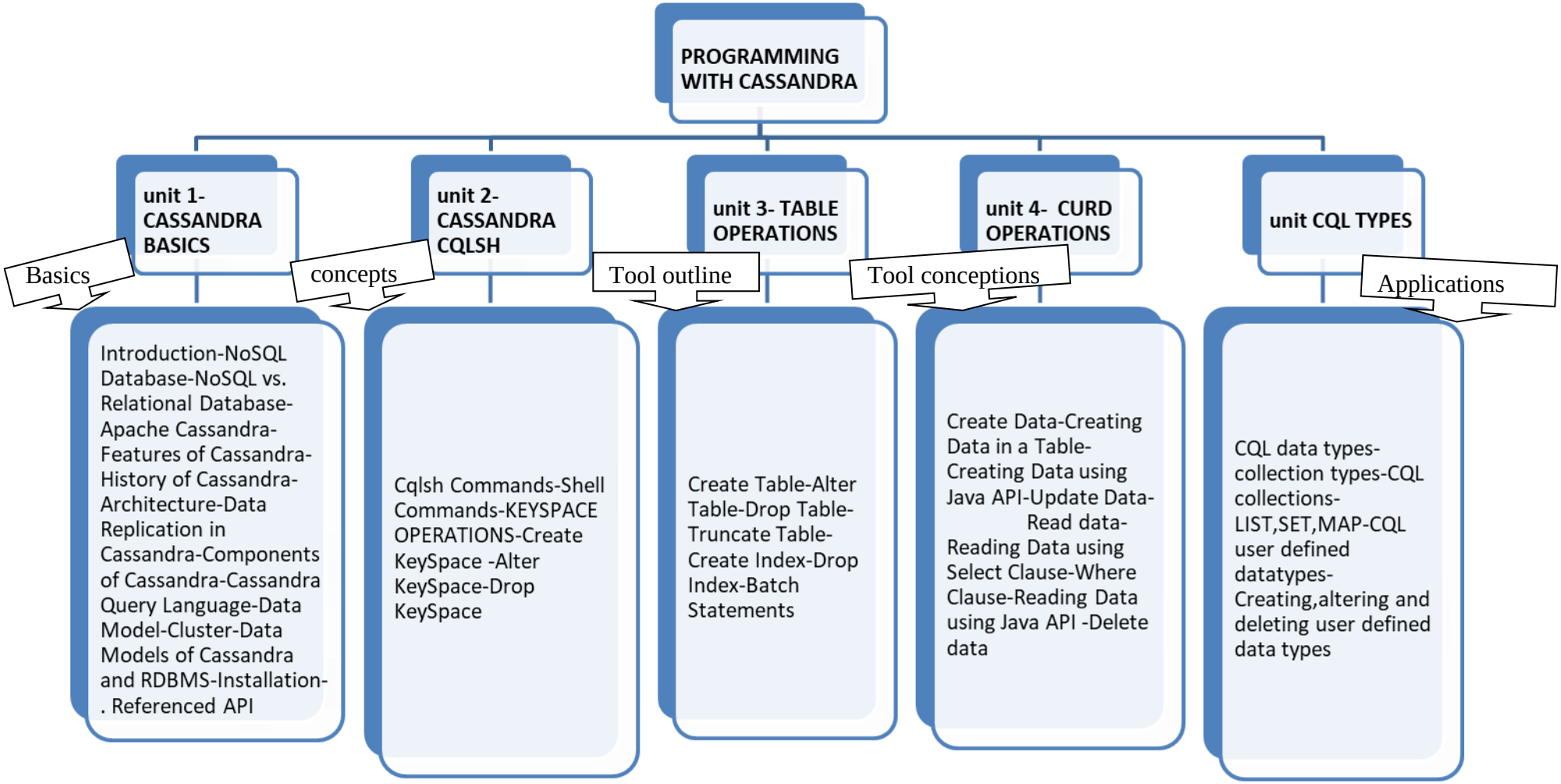
Mapping with Programme Outcomes

Cos	PSO1	PSO2	PSO3	PSO4	PSO5	PSO6
CO1	M		S			
CO2	M	M	L			
CO3		S	M		M	M

Assessment Pattern

Category	Continuous Assessment Tests (25)			Terminal Examination (75)
Remember	5	5	5	20
Understand	10	10	10	20
Apply	5	5	5	10
Analyze	5	5	5	10
Evaluate				5
Create				

Concept Map


PROGRAMMING WITH CASSANDRA
unit 1- CASSANDRA BASICS
unit 2- CASSANDRA CQLSH
unit 3- TABLE OPERATIONS
unit 4- CURD OPERATIONS
unit CQL TYPES
Basics
concepts
Tool outline
Tool conceptions
Applications
Introduction-NoSQL Database-NoSQL vs. Relational Database-Apache Cassandra-Features of Cassandra-History of Cassandra-Architecture-Data Replication in Cassandra-Components of Cassandra-Cassandra Query Language-Data Model-Cluster-Data Models of Cassandra and RDBMS-Installation-. Referenced API
Cqlsh Commands-Shell Commands-KEYSPACE OPERATIONS-Create KeySpace -Alter KeySpace-Drop KeySpace
Create Table-Alter Table-Drop Table-Truncate Table-Create Index-Drop Index-Batch Statements
Create Data-Creating Data in a Table-Creating Data using Java API-Update Data-Read data-Reading Data using Select Clause-Where Clause-Reading Data using Java API -Delete data
CQL data types-collection types-CQL collections-LIST,SET,MAP-CQL user defined datatypes-Creating,altering and deleting user defined data types

Syllabus

UNIT-I-CASSANDRA BASICS

Introduction-NoSQL Database-NoSQL vs. Relational Database-Apache Cassandra-Features of Cassandra-History of Cassandra-Architecture-Data Replication in Cassandra-Components of Cassandra-Cassandra Query Language-Data Model-Cluster-Data Models of Cassandra and RDBMS-Installation-. Referenced API

UNIT-II CASSANDRA CQLSH

Cqlsh Commands-Shell Commands-KEYSPACE OPERATIONS-Create KeySpace -Alter KeySpace-Drop KeySpace

UNIT-III TABLE OPERATIONS

Create Table-Alter Table-Drop Table-Truncate Table- Create Index-Drop Index-Batch Statements

UNIT-IV CURD OPERATIONS

Create Data-Creating Data in a Table-Creating Data using Java API-Update Data- Read data-Reading Data using Select Clause-Where Clause-Reading Data using Java API -Delete data

UNIT-V CQL TYPES

CQL data types-collection types-CQL collections-LIST,SET,MAP-CQL user defined datatypes-Creating,altering and deleting user defined data types

References:

1.Mastering Apache Cassandra 3.x by Aaron Ploetz (Author), Tejaswi Malepati (Author), Nishant Neeraj (Author)
2.https://www.gocit.vn/files/Cassandra.The.Definitive.Guide-www.gocit.vn.pdf
3.https://www.guru99.com/cassandra-tutorial.html
4.https://www.javatpoint.com/cassandra-setup-and-installation

Course Contents and Lecture Schedule

S.No.	Topic	No. of Lectures
1.	**CASSANDRA BASICS**	
1.1	*Introduction*	1
1.2	NoSQL Database	2
1.3	NoSQL vs. Relational Database	2
1.4	Apache Cassandra	
1.5	Features of Cassandra	2
1.6	History of Cassandra	
1.7	Architecture	2
1.8	Data Replication in Cassandra	2
1.9	Components of Cassandra-	2
1.10	Cassandra Query Language	
1.11	Data Model	
1.12	Cluster	2
1.13	Data Models of Cassandra and RDBMS	2
1.14	Installation	2
1.15	*Referenced API*	
2.	**CASSANDRA CQLSH**	
2.1	Cqlsh Commands	2
2.2	Shell Commands	2
2.3	KEYSPACE OPERATIONS	2
2.4	Create KeySpace	2
2.5	Alter KeySpace	2
2.6	Drop KeySpace	
3.	**TABLE OPERATIONS**	
3.1	Create Table	2
3.2	Alter Table	
3.3	Drop Table	
3.4	Truncate Table	2
3.5	Create Index	2
3.6	Drop Index	2
3.7	Batch Statements	
4.	**CURD OPERATIONS**	
4.1	Create Data	2
4.2	Creating Data in a Table	2
4.3	Creating Data using Java API	
4.4	Update Data	2
4.5	Read data	

4.6	Reading Data using Select Clause	2
4.7	Where Clause	
4.8	Reading Data using Java API	2
4.9	Delete data	
5.	**CQL TYPES**	
5.1	CQL data types	2
5.2	collection types	2
5.3	CQL collections	2
5.4	LIST,SET,MAP	2
5.5	CQL user defined datatypes	2
5.6	Creating,altering and deleting user defined data types	2

LIST OF ELECTIVES FOR ELECTIVES 3,6
(For The Candidates Admitted From 2017-18 Onwards)

Sl. No.	Subject code	Course name
1.	**MDAEI**	HADOOP and Data Analytics
2.	**MDAEJ**	Storm for Data Analytics
3.	**MDAEK**	Spark for Data Analytics
4.	**MDAEL**	SPSS for Data Analytics

MDAEI	**HADOOP AND DATA ANALYTICS**	Category	L	P	Credit
		PE	3	0	3

Preamble

To get introduced to the basic of **Apache Hadoop,** a free, open source, Java-based programming framework and also Learn about Hadoop's architecture and it's core components, such as Map Reduce and the Hadoop Distributed File System (HDFS).

Prerequisite

Introduction to data Analystics

Course Outcomes

On the successful completion of the course, students will be able to

Course Outcomes		Level
CO1	Able to understand Hadoop's architecture and core components	Understanding
CO2	Able to know Hadoop clusters	Apply
CO3	Able to apply Hadoop ecosystem such as Oozie, Sqoop, Flume	Apply

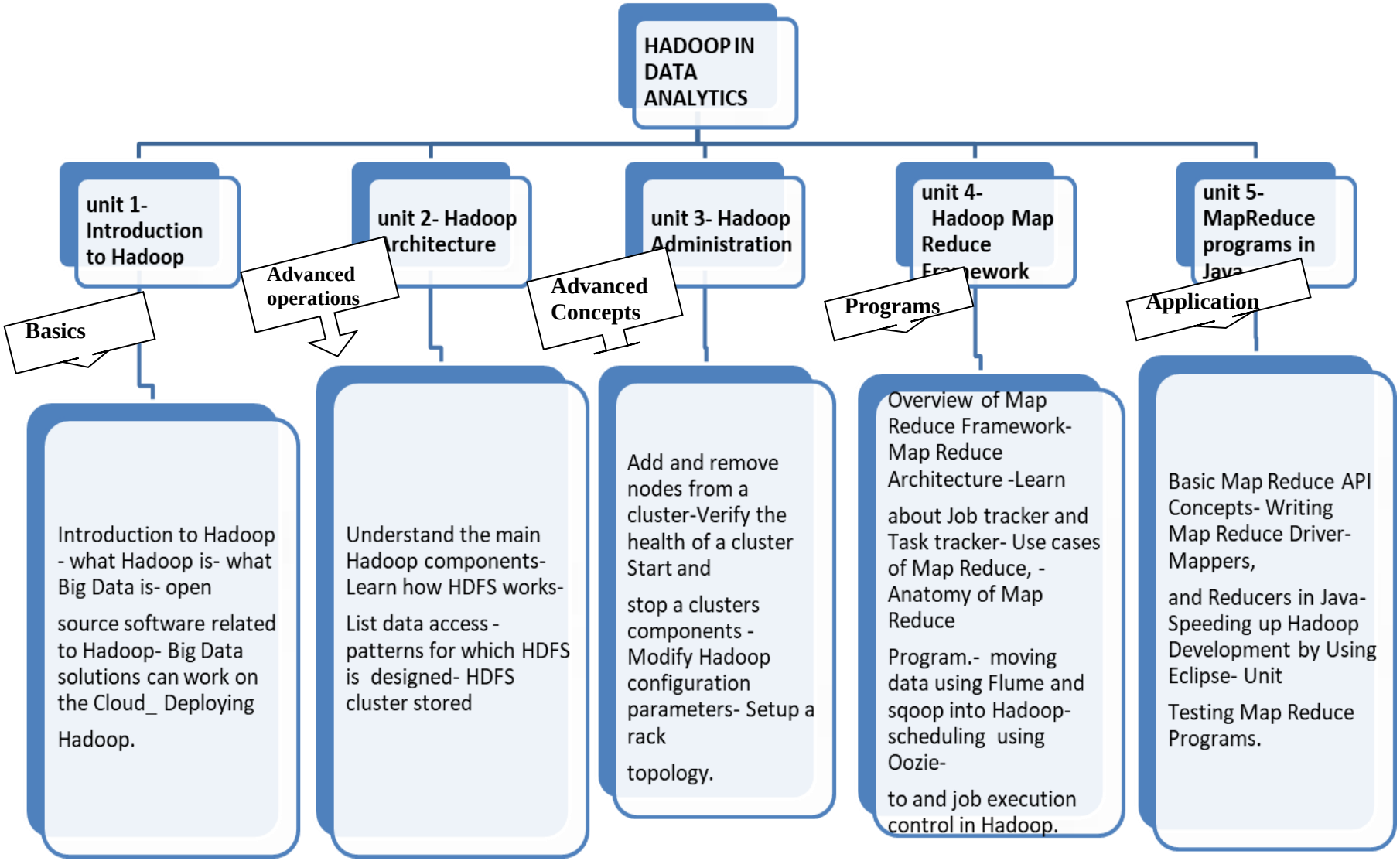
Mapping with Programme Outcomes

Cos	PSO1	PSO2	PSO3	PSO4	PSO5	PSO6
CO1	M	M				
CO2	M	M		L		
CO3		S	M		M	M

Assessment Pattern

Category	Continuous Assessment Tests (25)			Terminal Examination (75)
Remember	5	5	5	20
Understand	10	10	10	20
Apply	5	5	5	10
Analyze	5	5	5	10
Evaluate				5
Create				

Concept Map


HADOOP IN DATA ANALYTICS
unit 1- Introduction to Hadoop
unit 2- Hadoop Architecture
unit 3- Hadoop Administration
unit 4- Hadoop Map Reduce Framework
unit 5- MapReduce programs in Java
Basics
Advanced operations
Advanced Concepts
Programs
Application
Introduction to Hadoop - what Hadoop is- what Big Data is- open source software related to Hadoop- Big Data solutions can work on the Cloud_ Deploying Hadoop.
Understand the main Hadoop components- Learn how HDFS works- List data access - patterns for which HDFS is designed- HDFS cluster stored
Add and remove nodes from a cluster-Verify the health of a cluster Start and stop a clusters components - Modify Hadoop configuration parameters- Setup a rack topology.
Overview of Map Reduce Framework- Map Reduce Architecture -Learn about Job tracker and Task tracker- Use cases of Map Reduce, - Anatomy of Map Reduce Program.- moving data using Flume and sqoop into Hadoop- scheduling using Oozie- to and job execution control in Hadoop.
Basic Map Reduce API Concepts- Writing Map Reduce Driver- Mappers, and Reducers in Java- Speeding up Hadoop Development by Using Eclipse- Unit Testing Map Reduce Programs.

Syllabus

Unit 1 - Introduction To Hadoop

Introduction to Hadoop - what Hadoop is- what Big Data is- open source software related to Hadoop- Big Data solutions can work on the Cloud_ Deploying Hadoop.

Unit 2 - Hadoop Architecture

Understand the main Hadoop components-Learn how HDFS works- List data access - patterns for which HDFS is designed- HDFS cluster stored

Unit 3 - Hadoop Administration

Add and remove nodes from a cluster-Verify the health of a cluster Start and stop clusters components - Modify Hadoop configuration parameters- Setup a rack topology.

Unit 4 - Hadoop Map Reduce Framework

Overview of Map Reduce Framework- Map Reduce Architecture - Learn about Job tracker and Task tracker- Use cases of Map Reduce, - Anatomy of Map Reduce Program.- moving data using Flume and sqoop into Hadoop- scheduling using Oozie- to and job execution control in Hadoop.

Unit 5 - MapReduce programs

Basic Map Reduce API Concepts- Writing Map Reduce Driver- Mappers and Reducers in Java-Speeding up Hadoop Development by Using Eclipse- Unit Testing Map Reduce Programs

Books:

1. **Hadoop Beginner's Guide** by Garry Turkington Published by Packt Publishing Ltd. Livery Place35 Livery Street Birmingham B3 2PB, UK.ISBN 978-1-84951-7-300 www.packtpub.com
2. **Hadoop Illuminated** by Mark Kerzner and Sujee Maniyam [https://github.com/hadoop-illuminated/hadoop-book]

References:

1. https://www.youtube.com/watch?v=-65WgvIJ5xo&feature=youtu.be
2. https://www.youtube.com/watch?v=PS5QSGAoLNw&feature=youtu.be
3. https://www.youtube.com/watch?v=8AtrYcqO5ho&feature=youtu.be
4. 4.https://www.youtube.com/watch?v=iJmJhxIsmb8&feature=youtu.be
5. https://www.youtube.com/watch?v=Gd1sVPOYzuk&feature=youtu.be
6. https://www.youtube.com/watch?v=LtGliUam-_U&feature=youtu.be
7. https://www.youtube.com/watch?v=sVrSx4zt8ho&feature=youtu.be

Course Contents and Lecture Schedule

S.No.	Topic	No. of Lectures
1	**Introduction to Hadoop**	
1.1	what Hadoop is	1
1.2	what Big Data is	2
1.3	open source software related to Hadoop	2
1.4	Big Data solutions can work on the Cloud	2
1.5	Deploying Hadoop	
2	Understand the main Hadoop components	
2.1	Learn how HDFS works	2
2.2	List data access	2
2.3	patterns for which HDFS is designed	2
2.4	HDFS cluster stored	1
3	**Hadoop Administration**	
3.1	Add and remove nodes from a cluster	2
3.2	The health of a cluster Start and stop a clusters components	2
3.3	Modify Hadoop configuration parameters	2
3.4	Setup a rack topology.	1
4	Hadoop Map Reduce Framework	
4.1	Overview of Map Reduce Framework	2
4.2	Map Reduce Architecture	2
4.3	Learn about Job tracker and Task tracker, Anatomy of Map Reduce	2
4.4	Use cases of Map Reduce	
4.5	moving data using Flume and sqoop into Hadoop	2
4.6	scheduling using Oozie- to and job execution control in Hadoop.	2
5	Map Reduce programs	2
5.1	Basic Map Reduce API Concepts	2
5.2	Writing Map Reduce Driver	2
5.3	Mappers and Reducers in Java	2
5.4	Speeding up Hadoop Development by Using Eclipse	2
5.5	Unit Testing Map Reduce Programs	2

MDAEJ	**STORM FOR DATA ANALYSTICS**	Category	L	P	Credit
		PE	3	0	3

Preamble

To get introduced to make a career in Big Data Analytics using Apache Storm framework. It will give you enough understanding on creating and deploying a Storm cluster in a distributed environment.

Prerequisite

Basic knowledge of Java and Linux

Course Outcomes

On the successful completion of the course, students will be able to

Course Outcomes		Level
CO1	Able to apply Introduction to STORM	Understanding
CO2	Able to apply Storm application programming	Apply
CO3	Able to apply storm configuration, monitoring and tuning	Apply

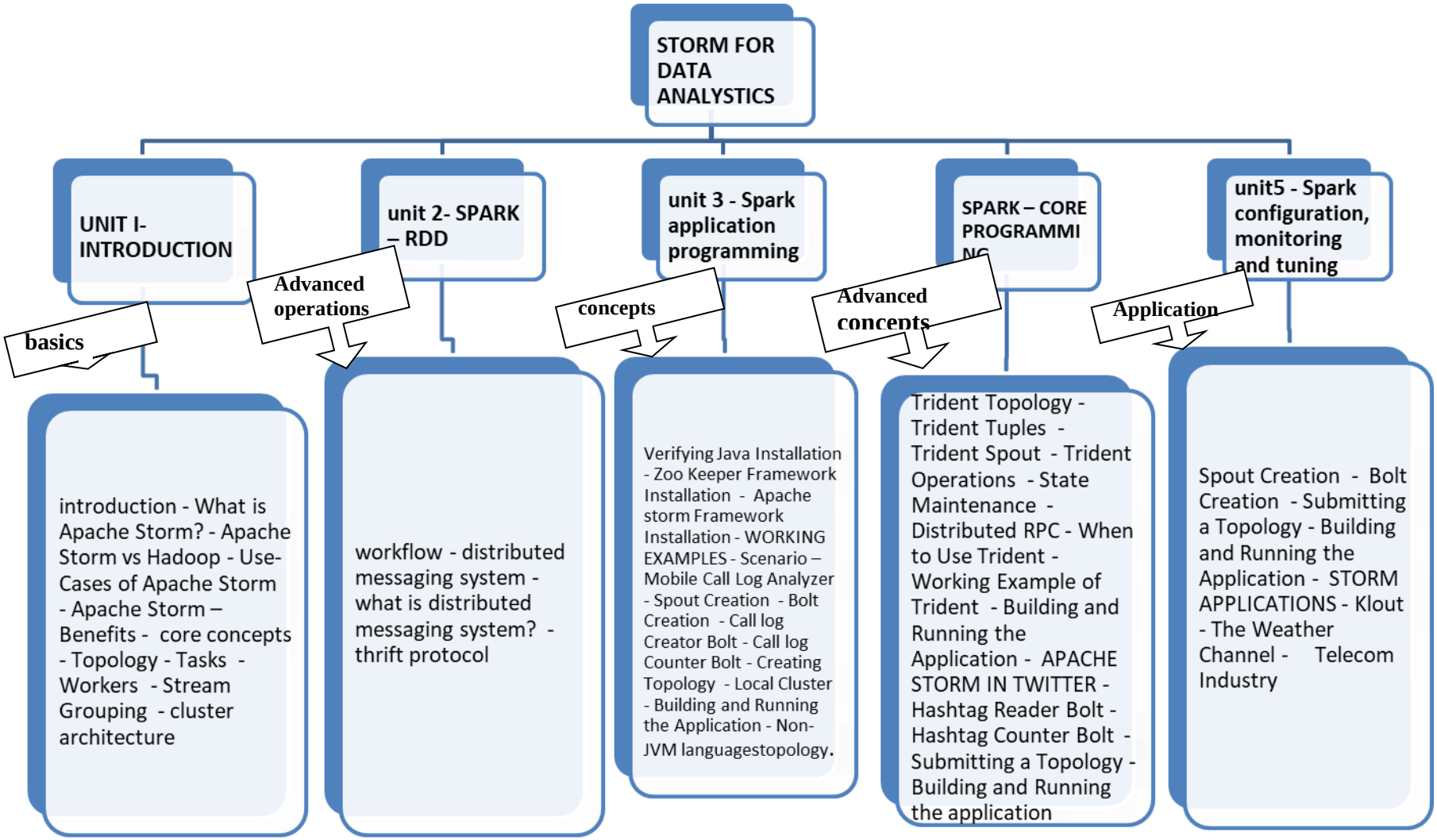
Mapping with Programme Outcomes

Cos	PSO1	PSO2	PSO3	PSO4	PSO5	PSO6
CO1	M	M				
CO2	M	M		L		
CO3		S	M		M	M

Assessment Pattern

Category	Continuous Assessment Tests (25)			Terminal Examination (75)
Remember	5	5	5	20
Understand	10	10	10	20
Apply	5	5	5	10
Analyze	5	5	5	10
Evaluate				5
Create				5

Concept Map


STORM FOR DATA ANALYSTICS
UNIT I- INTRODUCTION
unit 2- SPARK – RDD
unit 3 - Spark application programming
SPARK – CORE PROGRAMMI NG
unit5 - Spark configuration, monitoring and tuning
basics
Advanced operations
concepts
Advanced concepts
Application
introduction - What is Apache Storm? - Apache Storm vs Hadoop - Use-Cases of Apache Storm - Apache Storm – Benefits - core concepts - Topology - Tasks - Workers - Stream Grouping - cluster architecture
workflow - distributed messaging system - what is distributed messaging system? - thrift protocol
Verifying Java Installation - Zoo Keeper Framework Installation - Apache storm Framework Installation - WORKING EXAMPLES - Scenario – Mobile Call Log Analyzer - Spout Creation - Bolt Creation - Call log Creator Bolt - Call log Counter Bolt - Creating Topology - Local Cluster - Building and Running the Application - Non-JVM languagestopology.
Trident Topology - Trident Tuples - Trident Spout - Trident Operations - State Maintenance - Distributed RPC - When to Use Trident - Working Example of Trident - Building and Running the Application - APACHE STORM IN TWITTER - Hashtag Reader Bolt - Hashtag Counter Bolt - Submitting a Topology - Building and Running the application
Spout Creation - Bolt Creation - Submitting a Topology - Building and Running the Application - STORM APPLICATIONS - Klout - The Weather Channel - Telecom Industry

Syllabus

UNIT I-INTRODUCTION

introduction - What is Apache Storm? - Apache Storm vs Hadoop - Use-Cases of Apache Storm - Apache Storm – Benefits - core concepts - Topology - Tasks - Workers - Stream Grouping - cluster architecture

UNIT II- DISTRIBUTED MESSAGING SYSTEM

workflow - distributed messaging system - what is distributed messaging system? - thrift protocol

UNIT III- INSTALLATION

Verifying Java Installation - Zoo Keeper Framework Installation - Apache storm Framework Installation - WORKING EXAMPLES - Scenario – Mobile Call Log Analyzer - Spout Creation - Bolt Creation - Call log Creator Bolt - Call log Counter Bolt - Creating Topology - Local Cluster - Building and Running the Application - Non-JVM languages

UNIT IV: APACHE TRIDENT

Trident Topology - Trident Tuples - Trident Spout - Trident Operations - State Maintenance - Distributed RPC - When to Use Trident - Working Example of Trident - Building and Running the Application - APACHE STORM IN TWITTER - Hashtag Reader Bolt - Hashtag Counter Bolt - Submitting a Topology - Building and Running the application

UNIT V-APACHE STORM IN YAHOO FINANCE

Spout Creation - Bolt Creation - Submitting a Topology - Building and Running the Application - STORM APPLICATIONS - Klout - The Weather Channel - Telecom Industry

Books:

1. Mastering Apache Storm by Ankit Jain.
2. Storm Real-Time Processing Cookbook,Author: Quinton Anderson,Publisher: Packt Publishing.

Course Contents and Lecture Schedule

S.No.	Topic	No. of Lectures
1	**INTRODUCTION**	
1.1	Introduction	1
1.2	What Is Apache Storm?	2
1.3	Apache Storm Vs Hadoop	1
1.4	Use-Cases Of Apache Storm	2
1.5	Spark Built On Hadoop	2
1.6	Components Of Spark	2
2	**DISTRIBUTED MESSAGING SYSTEM**	
2.1	Workflow	1
2.2	Distributed Messaging System	2
2.3	What Is Distributed Messaging System?	2
2.4	Thrift Protocol	2
3	**UNIT III- INSTALLATION**	
	Verifying Java	2
3.1	Installation	2
3.2	Zoo Keeper Framework Installation	2
3.3	Apache storm Framework Installation	2
3.4	WORKING EXAMPLES	
3.5	Scenario	2
3.6	Mobile Call Log Analyzer	2
3.7	Spout Creation	2
3.8	Bolt Creation and Call log Creator Bolt	2
3.9	Actions	2
3.10	Local Cluster	2
3.11	Non-JVM languages ge	2
4	**APACHE TRIDENT** Trident Topology	
4.1	Trident Tuples	1
4.2	Trident Spout	2
4.3	Trident Operations	2
4.4	State Maintenance	2
4.5	Distributed RPC - When to Use Trident	2
5	**UNIT V-APACHE STORM IN YAHOO FINANCE**	
5.1	Spout Creation	2
5.2	Building and Running the Application	2
5.3	STORM APPLICATIONS	2

MDAEK	SPARK FOR DATA ANALYSTICS	Category	L	P	Credit
		PE	3	0	3

Preamble

To get introduced to Learn the fundamentals of Spark**,** the technology that is revolutionizing the analytics and big data world

Prerequisite

Introduction to data Analystics

Course Outcomes

On the successful completion of the course, students will be able to

Course Outcomes		Level
CO1	Able to apply Introduction to Spark	Understanding
CO2	Able to apply Spark application programming	Apply
CO3	Able to apply Spark configuration, monitoring and tuning	Apply

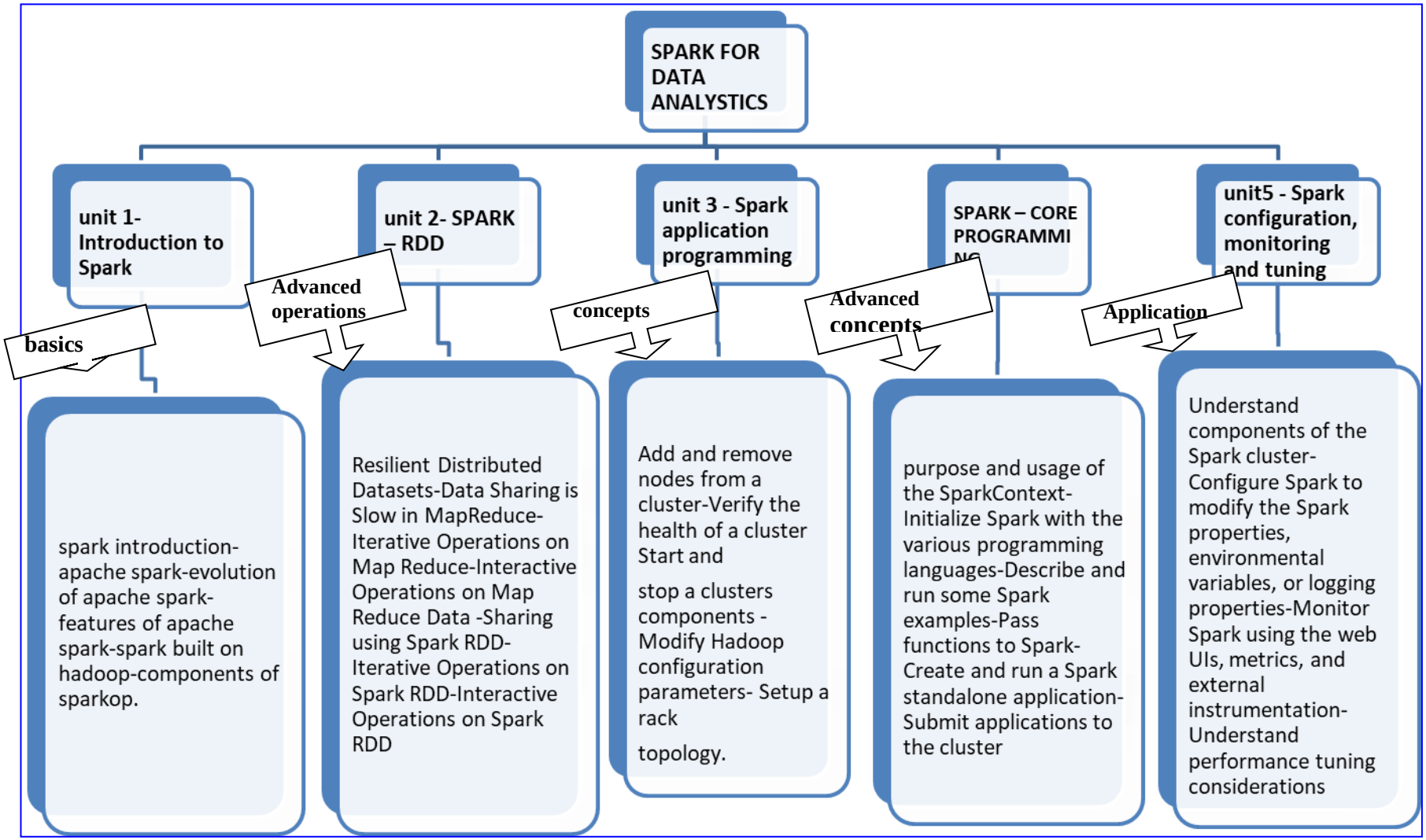
Mapping with Programme Outcomes

Cos	PSO1	PSO2	PSO3	PSO4	PSO5	PSO6
CO1	M	M				
CO2	M	M		L		
CO3		S	M		M	M

Assessment Pattern

Category	Continuous Assessment Tests (25)			Terminal Examination (75)
Remember	5	5	5	20
Understand	10	10	10	20
Apply	5	5	5	10
Analyze	5	5	5	10
Evaluate				5
Create				5

Concept Map


SPARK FOR DATA ANALYSTICS
unit 1- Introduction to Spark
unit 2- SPARK – RDD
unit 3 - Spark application programming
SPARK – CORE PROGRAMMI
unit5 - Spark configuration, monitoring and tuning
basics
Advanced operations
concepts
Advanced concepts
Application
spark introduction-apache spark-evolution of apache spark-features of apache spark-spark built on hadoop-components of sparkop.
Resilient Distributed Datasets-Data Sharing is Slow in MapReduce-Iterative Operations on Map Reduce-Interactive Operations on Map Reduce Data -Sharing using Spark RDD-Iterative Operations on Spark RDD-Interactive Operations on Spark RDD
Add and remove nodes from a cluster-Verify the health of a cluster Start and
stop a clusters components - Modify Hadoop configuration parameters- Setup a rack
topology.
purpose and usage of the SparkContext-Initialize Spark with the various programming languages-Describe and run some Spark examples-Pass functions to Spark-Create and run a Spark standalone application-Submit applications to the cluster
Understand components of the Spark cluster-Configure Spark to modify the Spark properties, environmental variables, or logging properties-Monitor Spark using the web UIs, metrics, and external instrumentation-Understand performance tuning considerations

Syllabus

Unit 1-Introduction to Spark Spark introduction-apache spark-evolution of apache spark-features of apache spark-spark built on hadoop-components of spark

Unit 2- SPARK – RDD Resilient Distributed Datasets-Data Sharing is Slow in MapReduce-Iterative Operations on Map Reduce-Interactive Operations on Map Reduce Data -Sharing using Spark RDD-Iterative Operations on Spark RDD-Interactive Operations on Spark RDD

Unit 3 - Spark application programming Purpose and usage of the SparkContext-Initialize Spark with the various programming languages-Describe and run some Spark examples-Pass functions to Spark-Create and run a Spark standalone application-Submit applications to the cluster

Unit 4 SPARK – CORE PROGRAMMING Spark Shell-RDD-Transformations-Actions-Programming with RDD-UN Persist the Storage.

Unit5 - Spark configuration, monitoring and tuning Understand components of the Spark cluster-Configure Spark to modify the Spark properties, environmental variables, or logging properties-Monitor Spark using the web UIs, metrics, and external instrumentation-Understand performance tuning considerations

Books:

1. LearningSpark by Holden Karau, Andy Konwinski, Patrick Wendell, and Matei Zaharia
2. Spark: The Definitive Guide by Matei Zaharia, Bill Chambers, Publisher: O'Reilly Media, Inc,Release Date: February 2018,ISBN: 9781491912201

Course Contents and Lecture Schedule

S.No.	Topic	No. of Lectures
1	**Introduction to Spark**	
1.2	spark introduction	2
1.3	apache spark	2
1.4	evolution of apache spark	2
1.5	features of apache spark	2
1.6	spark built on hadoop	2
1.7	components of spark	2
2	**SPARK – RDD**	
2.1	Resilient Distributed Datasets	12
2.2	Data Sharing is Slow in MapReduce	1
2.3	Iterative Operations on Map Reduce	1
2.4	Interactive Operations on Map Reduce	1
2.5	Sharing using Spark RDD	2
2.6	Iterative Operations on Spark RDD	2
3	**Spark application programming**	
3.1	purpose and usage of the Spark Context	1
3.2	Initialize Spark with the various programming languages	2
3.3	Describe and run some Spark examples-	2
3.4	Pass functions to Spark	2
3.5	Create and run a Spark standalone application2	
3.6	Submit applications to the cluster	2
4	**SPARK – CORE PROGRAMMING**	
4.1	Spark Shell-RDD.	2
4.2	Transformations	
4.3	Actions	2
4.4	Programming with RDD	2
4.5	UN Persist the Storage	2
5	**Spark configuration, monitoring and tuning**	2
5.1	Understand components of the Spark cluster	2
5.2	Configure Spark to modify the Spark properties	2
5.3	environmental variables, or logging properties-	2
5.4	Monitor Spark using the web UIs	2

MDAEL SPSS FOR DATA ANALYTICS	Category	L	P	Credit
	PE	3	0	3

Preamble

To get introduced to the concept of more efficient and accurate complex predictive analysis and Integrate R and Python Programming with SPSS statistics more efficient

Prerequisite

Introduction to data Analystics

Course Outcomes

On the successful completion of the course, students will be able to

Course Outcomes		Level
CO1	Able to understand SPSS architecture and core components	Understanding
CO2	Able to know R with SPSS	Apply
CO3	Able to know Python with SPSS	Apply

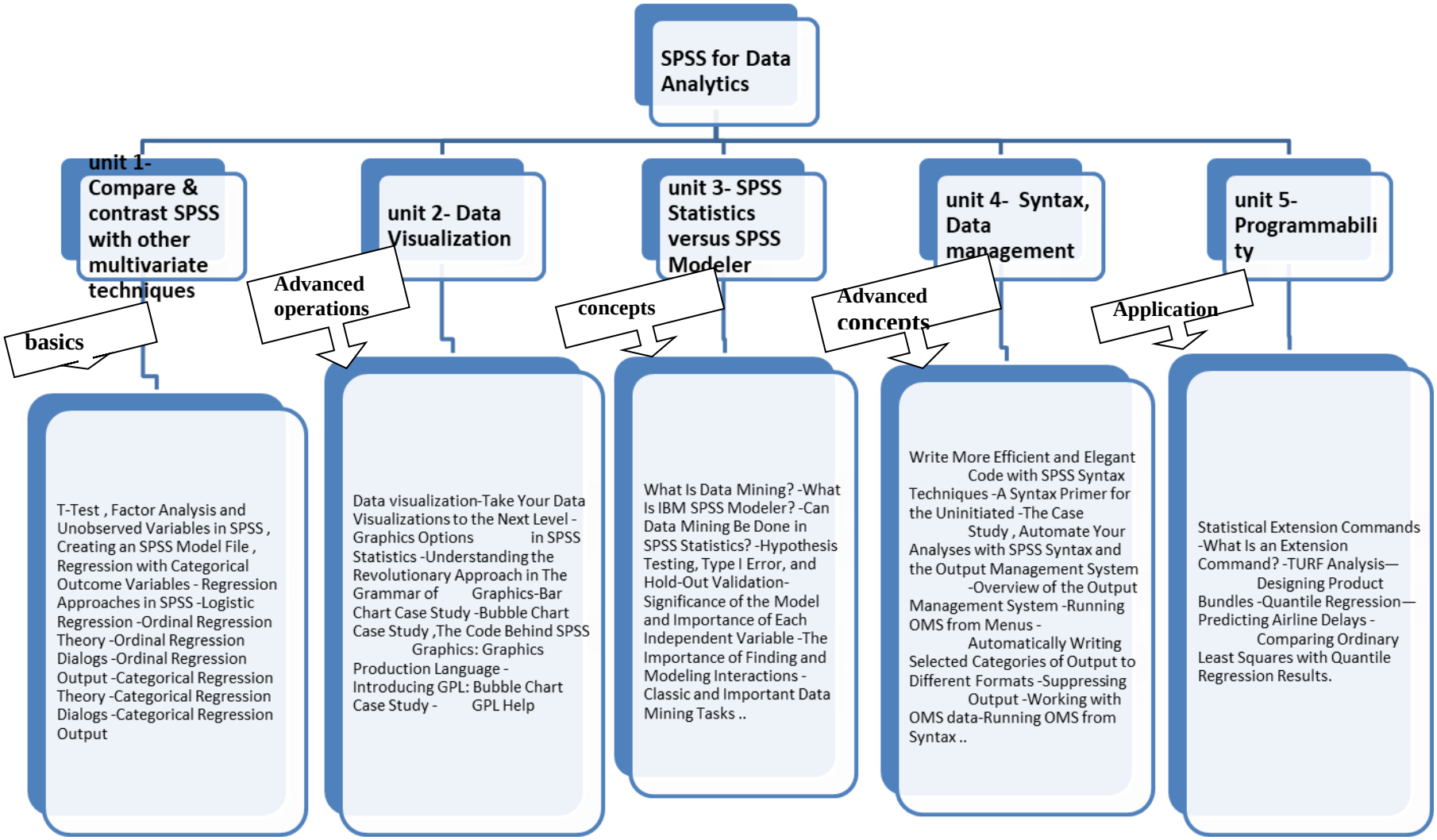
Mapping with Programme Outcomes

Cos	PSO1	PSO2	PSO3	PSO4	PSO5	PSO6
CO1	M	M				
CO2	M	M		L		
CO3		S	M		M	M

Assessment Pattern

Category	Continuous Assessment Tests (25)			Terminal Examination (75)
Remember	5	5	5	20
Understand	10	10	10	20
Apply	5	5	5	10
Analyze	5	5	5	10
Evaluate				5
Create				5

Concept Map


SPSS for Data Analytics
unit 1- Compare & contrast SPSS with other multivariate techniques
unit 2- Data Visualization
unit 3- SPSS Statistics versus SPSS Modeler
unit 4- Syntax, Data management
unit 5- Programmability
basics
Advanced operations
concepts
Advanced concepts
Application
T-Test , Factor Analysis and Unobserved Variables in SPSS , Creating an SPSS Model File , Regression with Categorical Outcome Variables - Regression Approaches in SPSS -Logistic Regression -Ordinal Regression Theory -Ordinal Regression Dialogs -Ordinal Regression Output -Categorical Regression Theory -Categorical Regression Dialogs -Categorical Regression Output
Data visualization-Take Your Data Visualizations to the Next Level - Graphics Options in SPSS Statistics -Understanding the Revolutionary Approach in The Grammar of Graphics-Bar Chart Case Study -Bubble Chart Case Study ,The Code Behind SPSS Graphics: Graphics Production Language - Introducing GPL: Bubble Chart Case Study - GPL Help
What Is Data Mining? -What Is IBM SPSS Modeler? -Can Data Mining Be Done in SPSS Statistics? -Hypothesis Testing, Type I Error, and Hold-Out Validation- Significance of the Model and Importance of Each Independent Variable -The Importance of Finding and Modeling Interactions - Classic and Important Data Mining Tasks ..
Write More Efficient and Elegant Code with SPSS Syntax Techniques -A Syntax Primer for the Uninitiated -The Case Study , Automate Your Analyses with SPSS Syntax and the Output Management System -Overview of the Output Management System -Running OMS from Menus - Automatically Writing Selected Categories of Output to Different Formats -Suppressing Output -Working with OMS data-Running OMS from Syntax ..
Statistical Extension Commands -What Is an Extension Command? -TURF Analysis— Designing Product Bundles -Quantile Regression— Predicting Airline Delays - Comparing Ordinary Least Squares with Quantile Regression Results.

Syllabus

Unit 1: Compare & Contrast SPSS with other multivariate techniques

T-Test , Factor Analysis and Unobserved Variables in SPSS , Creating an SPSS Model File , Regression with Categorical Outcome Variables - Regression Approaches in SPSS -Logistic Regression -Ordinal Regression Theory -Ordinal Regression Dialogs -Ordinal Regression Output -Categorical Regression Theory -Categorical Regression Dialogs -Categorical Regression Output.

Unit 2: Data Visualization

Data visualization-Take Your Data Visualizations to the Next Level - Graphics Options in SPSS Statistics -Understanding the Revolutionary Approach in The Grammar of Graphics-Bar Chart Case Study -Bubble Chart Case Study ,The Code Behind SPSS Graphics: Graphics Production Language - Introducing GPL: Bubble Chart Case Study -GPL Help

Unit 3: SPSS Statistics versus SPSS Modeler

What Is Data Mining? -What Is IBM SPSS Modeler? -Can Data Mining Be Done in SPSS Statistics? -Hypothesis Testing, Type I Error, and Hold-Out Validation-Significance of the Model and Importance of Each Independent Variable -The Importance of Finding and Modeling Interactions -Classic and Important Data Mining Tasks .

Unit 4: Syntax , Data Management

Write More Efficient and Elegant Code with SPSS Syntax Techniques -A Syntax Primer for the Uninitiated -The Case Study , Automate Your Analyses with SPSS Syntax and the Output Management System -Overview of the Output Management System -Running OMS from Menus - Automatically Writing Selected Categories of Output to Different Formats - Suppressing Output -Working with OMS data-Running OMS from Syntax .

Unit 5: Programmability

Statistical Extension Commands -What Is an Extension Command? -TURF Analysis—Designing Product Bundles -Quantile Regression—Predicting Airline Delays -Comparing Ordinary Least Squares with Quantile Regression Results .

Books:

1. SPSS, Keith Statistics for Data Analysis and Visualization by Jason Verlen, Andrew Wheeler, Jon Peck, Jesus Salcedo McCormick.
2. A Handbook of Statistical Analyses using SPSS by Sabine Landau and Brian S. Everitt.
3. Discovering Statistics using SPSS by Field, Andy.

Course Contents and Lecture Schedule

S.No.	Topic	No. of Lectures
1.	**Compare & Contrast SPSS with other multivariate techniques**	
1.1	T-Test	1
1.2	Factor Analysis and Unobserved Variables in SPSS	1
1.3	Creating an SPSS Model File	2
1.4	Regression Approaches in SPSS	
1.5	Logistic Regression	2
1.6	Ordinal Regression Theory	
1.7	Ordinal Regression Dialogs	
1.8	Ordinal Regression Output	2
1.9	Categorical Regression Theory	
1.1	Categorical Regression Dialogs	1
1.1	Categorical Regression Output.	
2.	Data Visualization	
2.1	Take Your Data Visualizations to the Next Level	1
2.2	Graphics Options in SPSS Statistics	1
2.3	Understanding the Revolutionary Approach in The Grammar of Graphics	1
2.4	Bar Chart Case Study	1
2.5	Bubble Chart Case Study	
2.6	Graphics Production Language	2
2.7	Introducing GPL	
2.8	Bubble Chart Case Study	2
2.9	GPL Help	
3.	**SPSS Statistics versus SPSS Modeler**	
3.1	What Is Data Mining?	2
3.2	What Is IBM SPSS Modeler?	2
3.3	Can Data Mining Be Done in SPSS Statistics?	
3.4	Hypothesis Testing	2
3.5	Type I Error and Hold Out Validation	
3.6	Significance of the Model and Importance of Each Independent Variable	1
3.7	The Importance of Finding and Modeling Interactions	
3.8	Classic and Important Data Mining Tasks	2
4.	**Syntax , Data Management**	
4.1	Write More Efficient and Elegant Code with SPSS Syntax Techniques	2
4.2	A Syntax Primer for the Uninitiated	2
4.3	The Case Study	1
4.4	Automate Your Analyses with SPSS Syntax and the Output Management System	2
4.5	Overview of the Output Management System -Running OMS from Menus	
4.6	Automatically Writing Selected Categories of Output to Different Formats	2
4.7	Suppressing Output	

4.8	Working with OMS data	1
4.9	Running OMS from Syntax	1
5	**Programmability**	
5.1	Statistical Extension Commands	1
5.2	What Is an Extension Command?	2
5.3	TURF Analysis	
5.4	Designing Product Bundles	2
5.5	Quantile Regression	
5.6	Predicting Airline Delays	1
5.7	Comparing Ordinary Least Squares with Quantile Regression Results	1

LIST OF ELECTIVES FOR ELECTIVES 7,8
(For The Candidates Admitted From 2017-18 Onwards)

Sl. No.	Subject code	Course name
1.	**MDAEM**	Web Frameworks for Big Data Projects
2.	**MDAEN**	Real-Time Big Data Analytics
3.	**MDAEO**	Challenges and Research Issues in Big Data Projects
4.	**MDAEP**	Big Data in Computer Vision

MDAEM WEB FRAMEWORKS FOR BIG DATA PROJECTS	Category	L	P	Credit
	PE	3	0	3

Preamble

To understand and become familiar with the front end and user interface of few of the tools used for Data analytics.

Prerequisite

Introduction to Data Analytics

Course Outcomes

On the successful completion of the course, students will be able to

Course Outcomes		Level
CO1	Able to understand the frond end of different data analytics tools	Understanding
CO2	Work with the User Interface of Data Analytics tools for various purposes	Understand and apply
CO3	Apply the tools to do Data Analytics Projects	Apply

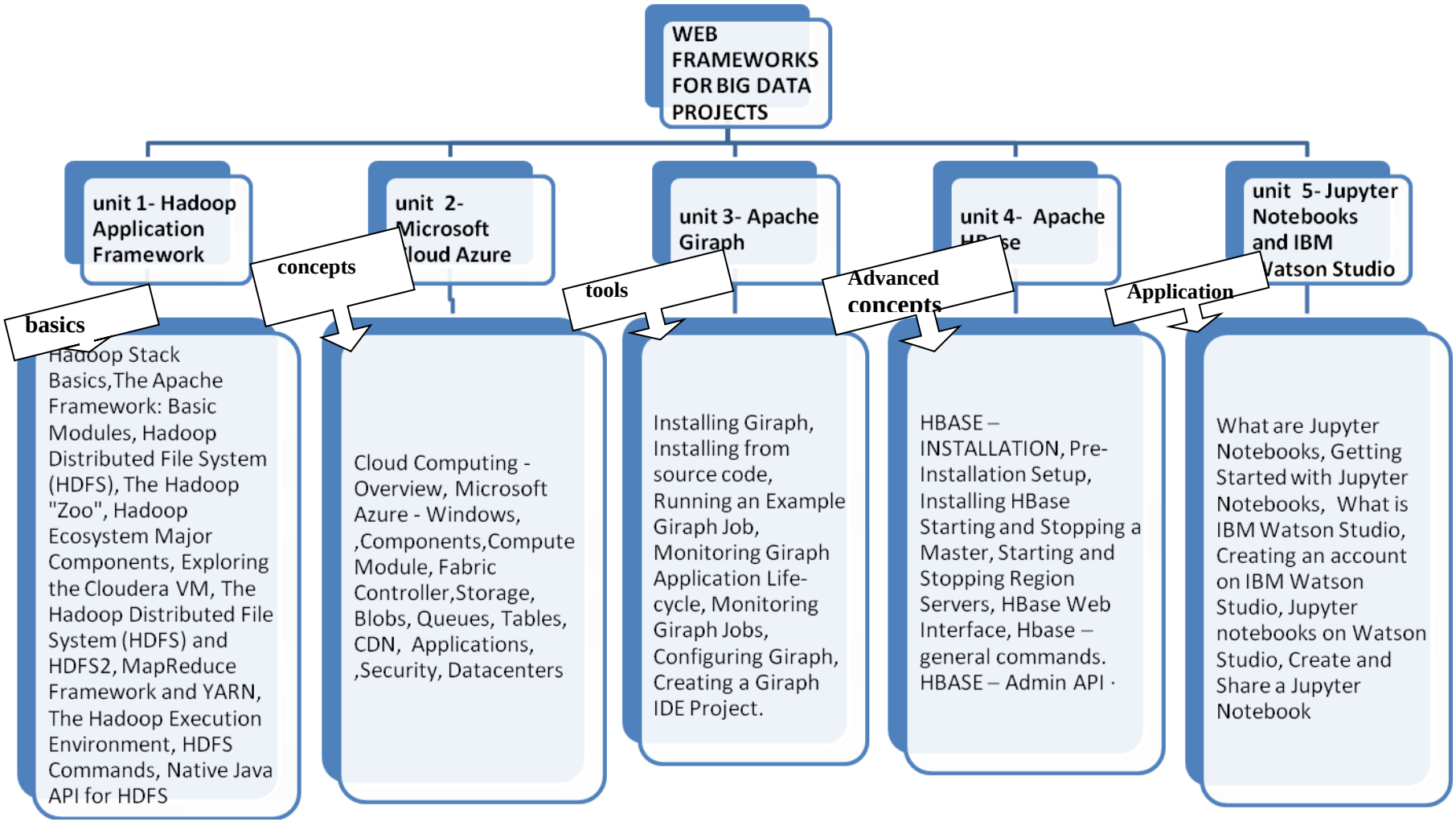
Mapping with Programme Outcomes

Cos	PSO1	PSO2	PSO3	PSO4	PSO5	PSO6
CO1	S	S	S	M	M	M
CO2	S	S	S	M	M	M
CO3	S	S	S	M	S	M

Assessment Pattern

Category	Continuous Assessment Tests (25)			Terminal Examination (75)
Remember	5	5	5	20
Understand	10	10	10	20
Apply	5	5	5	10
Analyze	5	5	5	10
Evaluate				5
Create				5

Concept Map


WEB FRAMEWORKS FOR BIG DATA PROJECTS
unit 1- Hadoop Application Framework
unit 2- Microsoft loud Azure
unit 3- Apache Giraph
unit 4- Apache se
unit 5- Jupyter Notebooks and IBM Watson Studio
basics
concepts
tools
Advanced concepts
Application
Hadoop Stack Basics,The Apache Framework: Basic Modules, Hadoop Distributed File System (HDFS), The Hadoop "Zoo", Hadoop Ecosystem Major Components, Exploring the Cloudera VM, The Hadoop Distributed File System (HDFS) and HDFS2, MapReduce Framework and YARN, The Hadoop Execution Environment, HDFS Commands, Native Java API for HDFS
Cloud Computing - Overview, Microsoft Azure - Windows, ,Components,Compute Module, Fabric Controller,Storage, Blobs, Queues, Tables, CDN, Applications, ,Security, Datacenters
Installing Giraph, Installing from source code, Running an Example Giraph Job, Monitoring Giraph Application Life-cycle, Monitoring Giraph Jobs, Configuring Giraph, Creating a Giraph IDE Project.
HBASE – INSTALLATION, Pre-Installation Setup, Installing HBase Starting and Stopping a Master, Starting and Stopping Region Servers, HBase Web Interface, Hbase – general commands. HBASE – Admin API ·
What are Jupyter Notebooks, Getting Started with Jupyter Notebooks, What is IBM Watson Studio, Creating an account on IBM Watson Studio, Jupyter notebooks on Watson Studio, Create and Share a Jupyter Notebook

Syllabus

Unit:I Hadoop Application Framework: Hadoop Stack Basics,The Apache Framework: Basic Modules, Hadoop Distributed File System (HDFS), The Hadoop "Zoo", Hadoop Ecosystem Major Components, Exploring the Cloudera VM, The Hadoop Distributed File System (HDFS) and HDFS2, MapReduce Framework and YARN, The Hadoop Execution Environment, HDFS Commands, Native Java API for HDFS

Unit: II Microsoft Cloud Azure: Cloud Computing - Overview, Microsoft Azure - Windows, ,Components,Compute Module, Fabric Controller,Storage, Blobs, Queues, Tables, CDN, Applications, ,Security, Datacenters.

Unit: III Apache Giraph.: Installing Giraph, Installing from source code, Running an Example Giraph Job, Monitoring Giraph Application Life-cycle, Monitoring Giraph Jobs, Configuring Giraph, Creating a Giraph IDE Project.

Unit:IV Apache HBase: HBASE ─ INSTALLATION, Pre-Installation Setup, Installing HBase Starting and Stopping a Master, Starting and Stopping Region Servers, HBase Web Interface, Hbase ─ general commands. HBASE ─ Admin API ·

Unit:V Jupyter Notebooks and IBM Watson Studio: What are Jupyter Notebooks, Getting Started with Jupyter Notebooks, What is IBM Watson Studio, Creating an account on IBM Watson Studio, Jupyter notebooks on Watson Studio, Create and Share a Jupyter Notebook

Books:

1. Large Scale Graph Processing Using Apache Giraph, 'Sherif Sakr, Faisal Moeen, Ibrahim Abdelaziz, Zuhair Khayyat', Springer International Publishing AG, http:// dx.doi.org/10.1007/978-3-319-47431-1

URLs

https://www.coursera.org/learn/hadoop
https://www.tutorialspoint.com/microsoft_azure/
https://www.tutorialspoint.com/hbase/hbase_tutorial.pdf
http:// dx.doi.org/10.1007/978-3-319-47431-1
https://www.coursera.org/learn/open-source-tools-for-data-science

Course Contents and Lecture Schedule

S.No.	Topic	No. of Lectures
1.	**Hadoop Application Framework**	
1.1	Hadoop Stack Basics,The Apache Framework: Basic Modules, Hadoop Distributed File System (HDFS), The Hadoop "Zoo",	3
1.2	Hadoop Ecosystem Major Components, Exploring the Cloudera VM,	3
1.3	The Hadoop Distributed File System (HDFS) and HDFS2, MapReduce Framework and YARN,	2
1.4	The Hadoop Execution Environment	2
1.5	HDFS Commands, Native Java API for HDFS	3
2.	**Microsoft Cloud Azure**	
2.1	Cloud Computing - Overview,	3
2.2	Microsoft Azure - Windows, ,Components,Compute Module, Fabric Controller,Storage,	3
2.3	Blobs	3
2.4	,Security	3
2.5	Datacenters	2
3.	**Apache Giraph**	
3.1	Installing Giraph,	3
3.2	Installing from source code, Running an Example Giraph Job,	3
3.3	Monitoring Giraph Application Life-cycle,	5
3.4	Monitoring Giraph Jobs,	
3.5	Configuring Giraph, Creating a Giraph IDE Project	3
4.	**Apache HBase**	
4.1	HBASE ─ INSTALLATION,	3
4.2	Pre-Installation Setup, Installing HBase	3
4.3	Starting and Stopping a Master, Starting and Stopping Region Servers,	3
4.4	HBase Web Interface, Hbase ─ general commands.	3
4.5	HBASE ─ Admin API	2
5	**Jupyter Notebooks and IBM Watson Studio**	
5.1	What are Jupyter Notebooks, Getting Started with Jupyter Notebooks,	3
5.2	What is IBM Watson Studio,	4
5.3	Creating an account on IBM Watson Studio,	
5.4	Jupyter notebooks on Watson Studio,	2
5.5	Create and Share a Jupyter Notebook	2

MDAEN REAL-TIME BIG DATA ANALYTICS	Category	L	P	Credit
	PE	3	0	3

Preamble

To understand streaming real time data from APIs and the issues, challenges associated while accessing data from Real Time Data Sources.

Prerequisite

Introduction to Data Analytics

Course Outcomes

On the successful completion of the course, students will be able to

Course Outcomes		Level
CO1	Know how real-time data can be streamed.	Understanding
CO2	Understand the challenges and issues in using online data.	Understand and apply
CO3	Keep in mind the privacy issues while using real-time data.	Apply

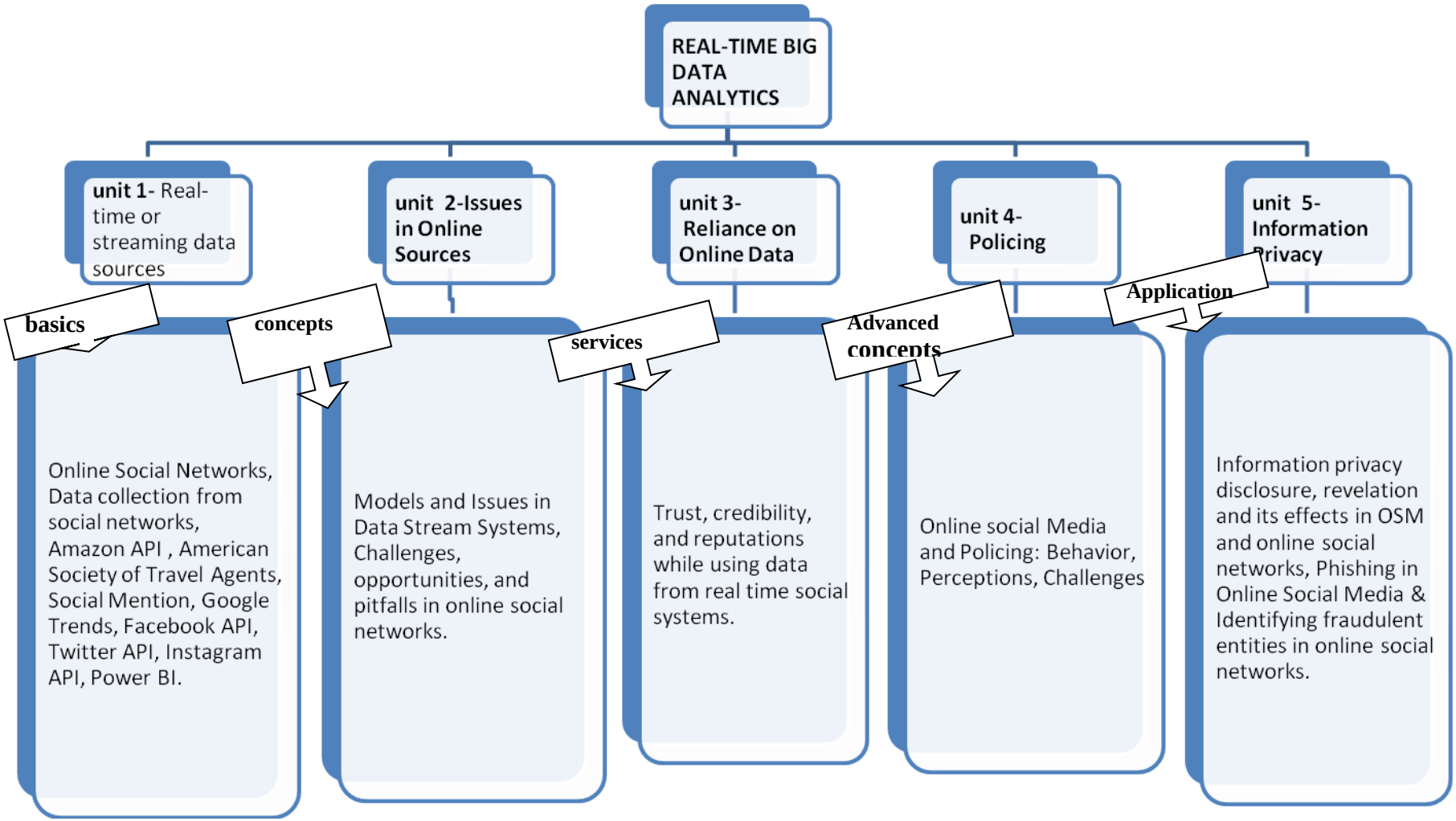
Mapping with Programme Outcomes

Cos	PSO1	PSO2	PSO3	PSO4	PSO5	PSO6
CO1	S	S	S	M	M	M
CO2	S	S	S	M	M	M
CO3	S	S	L	M	S	M

Assessment Pattern

Category	Continuous Assessment Tests (25)			Terminal Examination (75)
Remember	5	5	5	20
Understand	10	10	10	20
Apply	5	5	5	10
Analyze	5	5	5	10
Evaluate				5
Create				5

Concept Map


REAL-TIME BIG DATA ANALYTICS
unit 1- Real-time or streaming data sources
unit 2-Issues in Online Sources
unit 3- Reliance on Online Data
unit 4- Policing
unit 5- Information Privacy
basics
concepts
services
Advanced concepts
Application
Online Social Networks, Data collection from social networks, Amazon API , American Society of Travel Agents, Social Mention, Google Trends, Facebook API, Twitter API, Instagram API, Power BI.
Models and Issues in Data Stream Systems, Challenges, opportunities, and pitfalls in online social networks.
Trust, credibility, and reputations while using data from real time social systems.
Online social Media and Policing: Behavior, Perceptions, Challenges
Information privacy disclosure, revelation and its effects in OSM and online social networks, Phishing in Online Social Media & Identifying fraudulent entities in online social networks.

Syllabus

Unit:I **Real-time or streaming data sources:** Online Social Networks, Data collection from social networks, Amazon API , American Society of Travel Agents, Social Mention, Google Trends, Facebook API, Twitter API, Instagram API, Power BI.

Unit: II Issues in Online Sources: Models and Issues in Data Stream Systems, Challenges, opportunities, and pitfalls in online social networks.

Unit: III Reliance on Online Data : Trust, credibility, and reputations while using data from real time social systems.

Unit:IV Policing: Online social Media and Policing: Behavior, Perceptions, Challenges

Unit:V Information Privacy: Information privacy disclosure, revelation and its effects in OSM and online social networks, Phishing in Online Social Media & Identifying fraudulent entities in online social networks.

Books:

1. Online Social Media and Police in India:Behavior, Perceptions, Challenges, 'Niharika Sachdeva, Ponnurangam Kumaraguru', Cybersecurity Education and Research Centre.
Available:
https://www.researchgate.net/publication/260678980_Online_Social_Media_and_Police_in_India_Behavior_Perceptions_Challenges [accessed Dec 16 2018].
2. Nptel course on Privacy and security in Online Social Media by Pandurangan Kumaraguru.
3. Models and Issues in Data Stream Systems Brian Babcock Shivnath Babu Mayur Datar Rajeev Motwani Jennifer Wido, Department of Computer Science Stanford University.

URLS

1. https://docs.microsoft.com/en-us/power-bi/service-real-time-streaming
2. https://infolab.usc.edu/csci599/Fall2002/paper/DML2_streams-issues.pdf

Course Contents and Lecture Schedule

S.No.	Topic	No. of Lectures
1.	**Real-time or streaming data sources**	
1.1	Online Social Networks, Data collection from social networks,	3
1.2	Amazon API , American Society of Travel Agents,	3
1.3	Social Mention, Google Trends, Facebook API,	2
1.4	Twitter API, Instagram API,	2
1.5	Power BI.	3
2.	**Issues in Online Sources**	
2.1	Models and Issues in Data Stream Systems,	3
2.2	Challenges, opportunities.	3
2.3	pitfalls in online social networks.	3
3.	**Reliance on Online Data**	
3.1	Trust while using data from real time social systems.	3
3.2	Credibility while using data from real time social systems.	3
3.3	Reputations while using data from real time social systems.	5
4.	**Policing**	
4.1	Online social Media and Policing	3
4.2	Behavior	3
4.3	Perceptions	3
4.4	Challenges	3
5	**Information Privacy**	
5.1	Information privacy disclosure,	3
5.2	revelation and its effects in OSM and online social networks,	4
5.3	Phishing in Online Social Media	
5.4	Identifying fraudulent entities in online social networks.	2

MDAEO CHALLENGES AND RESEARCH ISSUES IN BIG DATA PROJECTS	Category	L	P	Credit
	PE	3	0	3

Preamble

To understand challenges and Research Issues in Big Data Projects and to prepare towards facing those issues.

Prerequisite

Introduction to Data Analytics

Course Outcomes

On the successful completion of the course, students will be able to

Course Outcomes		Level
CO1	Know the challenges of Data Analytics.	Understanding
CO2	Understand the Research Issues in Analytics	Understand
CO3	Keep in mind the issues for taking up the Data Projects	Apply

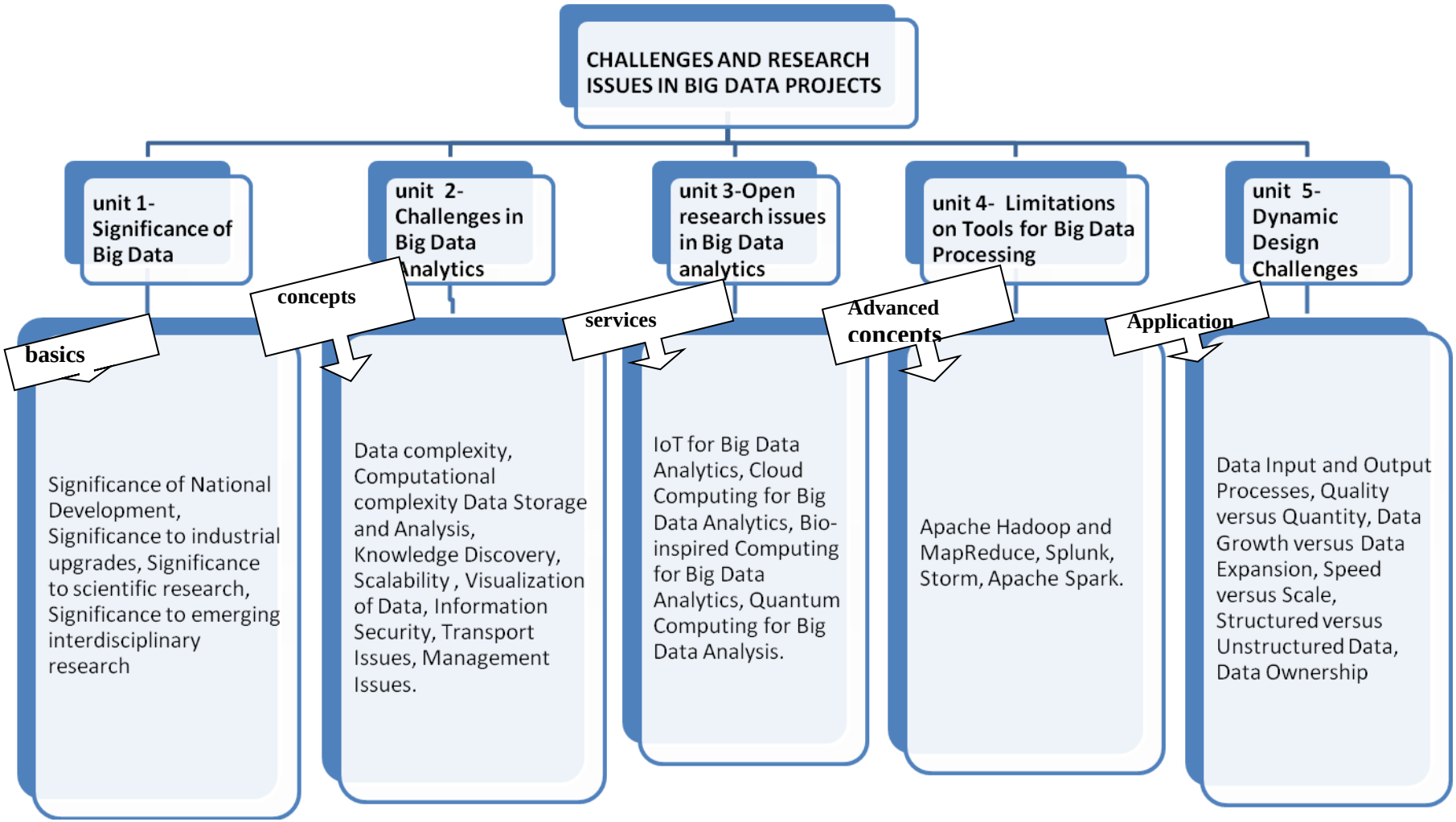
Mapping with Programme Outcomes

Cos	PSO1	PSO2	PSO3	PSO4	PSO5	PSO6
CO1	S	M	L	S	M	M
CO2	S	M	L	S	M	M
CO3	S	M	L	S	S	M

Assessment Pattern

Category	Continuous Assessment Tests (25)			Terminal Examination (75)
Remember	5	5	5	20
Understand	10	10	10	20
Apply	5	5	5	10
Analyze	5	5	5	10
Evaluate				5
Create				5

Concept Map


CHALLENGES AND RESEARCH ISSUES IN BIG DATA PROJECTS
unit 1- Significance of Big Data
unit 2- Challenges in Big Data Analytics
unit 3-Open research issues in Big Data analytics
unit 4- Limitations on Tools for Big Data Processing
unit 5- Dynamic Design Challenges
basics
concepts
services
Advanced concepts
Application
Significance of National Development, Significance to industrial upgrades, Significance to scientific research, Significance to emerging interdisciplinary research
Data complexity, Computational complexity Data Storage and Analysis, Knowledge Discovery, Scalability , Visualization of Data, Information Security, Transport Issues, Management Issues.
IoT for Big Data Analytics, Cloud Computing for Big Data Analytics, Bio-inspired Computing for Big Data Analytics, Quantum Computing for Big Data Analysis.
Apache Hadoop and MapReduce, Splunk, Storm, Apache Spark.
Data Input and Output Processes, Quality versus Quantity, Data Growth versus Data Expansion, Speed versus Scale, Structured versus Unstructured Data, Data Ownership

Syllabus

Unit:I Significance of Big Data: Significance of National Development, Significance to industrial upgrades, Significance to scientific research, Significance to emerging interdisciplinary research

Unit: II Challenges in Big Data Analytics: Data complexity, Computational complexity Data Storage and Analysis, Knowledge Discovery, Scalability , Visualization of Data, Information Security, Transport Issues, Management Issues.

Unit: III Open research issues in Big Data analytics: IoT for Big Data Analytics, Cloud Computing for Big Data Analytics, Bio-inspired Computing for Big Data Analytics, Quantum Computing for Big Data Analysis.

Unit:IV Limitations on Tools for Big Data Processing: Apache Hadoop and MapReduce, Splunk, Storm, Apache Spark.

Unit:V Dynamic Design Challenges: Data Input and Output Processes, Quality versus Quantity, Data Growth versus Data Expansion, Speed versus Scale, Structured versus Unstructured Data, Data Ownership

Books:

1. Significance and Challenges of Big Data Research ☆ Xiaolong Jin a,∗, Benjamin W. Wah a,b, Xueqi Cheng a, Yuanzhuo Wang, Elsevier Publications
2. A Survey on Big Data Analytics: Challenges, Open , Research Issues and Tools, 'D. P. Acharjya, Kauser Ahmed P', International Journal of Advanced Computer Science and Applications, Vol. 7, No. 2, 2016.
3. Big Data: Issues and Challenges Moving Forward, ' Stephen Kaisler, Frank Armour, J. Alberto Espinosa, William Money',
4. Addressing Big Data Issues in Scientific Data Infrastructure, 'Yuri Demchenko, Paola Grosso, Cees de Laat'

URLS

1. https://infolab.usc.edu/csci599/Fall2002/paper/DML2_streams-issues.pdf
https://infolab.usc.edu/csci599/Fall2002/paper/DML2_streams-issues.pdf

Course Contents and Lecture Schedule

S.No.	Topic	No. of Lectures
1.	**Significance of Big Data**	
1.1	Significance of National Development,	3
1.2	Significance to industrial upgrades	3
1.3	Significance to scientific research,	2
1.4	Significance to emerging interdisciplinary research	2
2.	**Challenges in Big Data Analytics**	
2.1	Data complexity, Computational complexity Data Storage and Analysis,	3
2.2	Knowledge Discovery, Scalability ,	3
2.3	Visualization of Data, Information Security,	3
2.4	Transport Issues, Management Issues	
3.	**Open research issues in Big Data analytics**	
3.1	IoT for Big Data Analytics,	3
3.2	Cloud Computing for Big Data Analytics,	3
3.3	Bio-inspired Computing for Big Data Analytics,	3
3.4	Quantum Computing for Big Data Analysis.	2
4.	**Limitations on Tools for Big Data Processing**:	
4.1	Apache Hadoop and MapReduce,.	3
4.2	Splunk,	3
4.3	Storm,	2
4.4	Apache Spark	2
5	**Dynamic Design Challenges**	
5.1	Data Input and Output Processes,	3
5.2	Quality versus Quantity,	4
5.3	Data Growth versus Data Expansion,	
5.4	Speed versus Scale, Structured versus Unstructured Data, Data Ownership	3

MDAEP VISION	BIG DATA IN COMPUTER VISION	Category PE	L 3	P 0	Credit 3

Preamble

To understand stages of a Computer Vision System, use of Computer Vision in Big data based applications and the application of Machine learning for various recognition applications.

Prerequisite

Introduction to Data Analytics

Course Outcomes

On the successful completion of the course, students will be able to

Course Outcomes		Level
CO1	Understand stages of a Computer Vision System	Understanding
CO2	Understand Vision in Big data based applications	Understand
CO3	Understand and applyMachine learning for various recognition applications	Apply

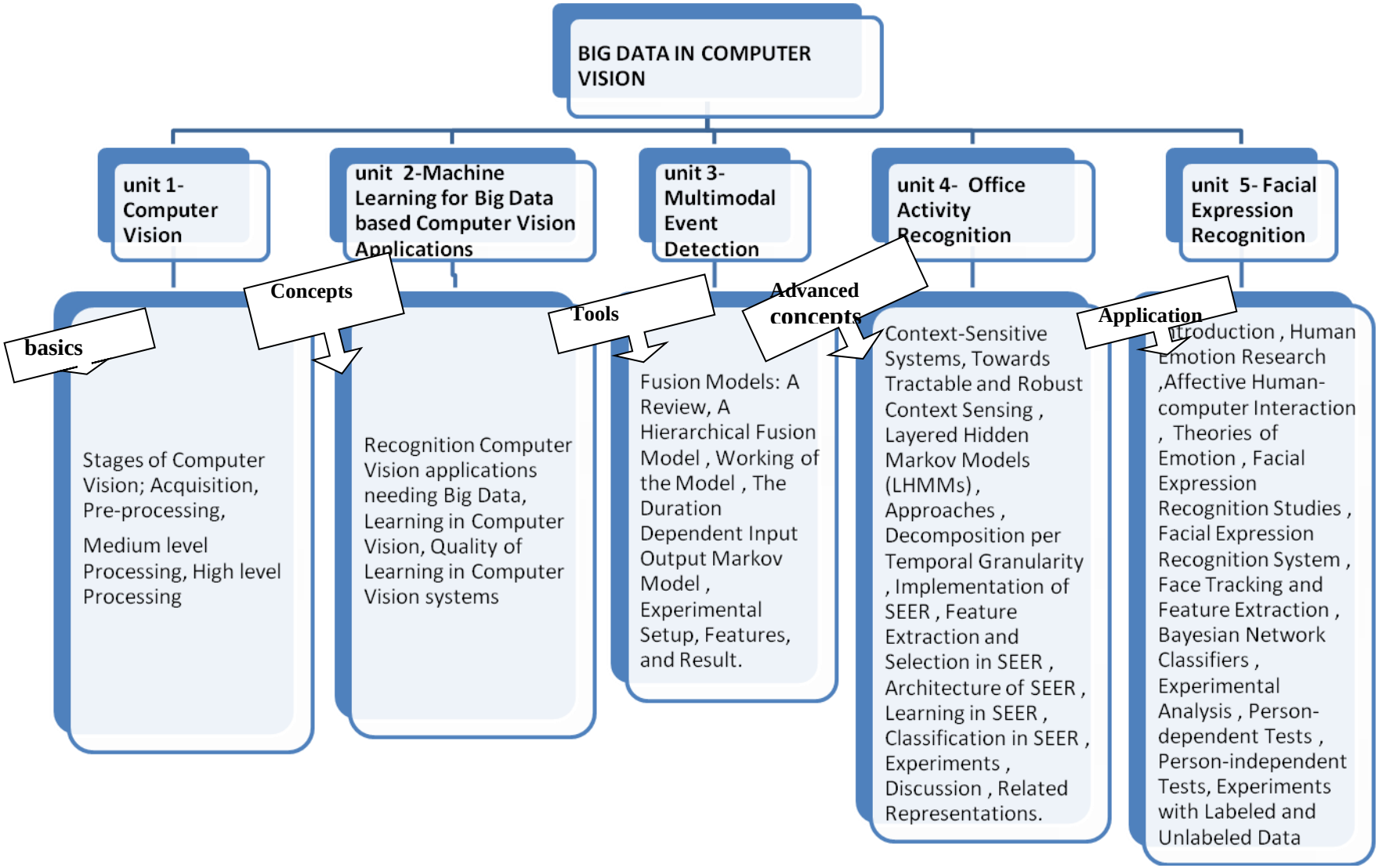
Mapping with Programme Outcomes

Cos	PSO1	PSO2	PSO3	PSO4	PSO5	PSO6
CO1	S	S	M	L	S	M
CO2	S	S	M	L	S	M
CO3	S	S	S	L	S	M

Assessment Pattern

Category	Continuous Assessment Tests (25)			Terminal Examination (75)
Remember	5	5	5	20
Understand	10	10	10	20
Apply	5	5	5	10
Analyze	5	5	5	10
Evaluate				5
Create				5

Concept Map


BIG DATA IN COMPUTER VISION
unit 1- Computer Vision
unit 2-Machine Learning for Big Data based Computer Vision Applications
unit 3- Multimodal Event Detection
unit 4- Office Activity Recognition
unit 5- Facial Expression Recognition
basics
Concepts
Tools
Advanced concepts
Application
Stages of Computer Vision; Acquisition, Pre-processing,
Medium level Processing, High level Processing
Recognition Computer Vision applications needing Big Data, Learning in Computer Vision, Quality of Learning in Computer Vision systems
Fusion Models: A Review, A Hierarchical Fusion Model , Working of the Model , The Duration Dependent Input Output Markov Model , Experimental Setup, Features, and Result.
Context-Sensitive Systems, Towards Tractable and Robust Context Sensing , Layered Hidden Markov Models (LHMMs) , Approaches , Decomposition per Temporal Granularity , Implementation of SEER , Feature Extraction and Selection in SEER , Architecture of SEER , Learning in SEER , Classification in SEER , Experiments , Discussion , Related Representations.
ltroduction , Human Emotion Research ,Affective Human-computer Interaction , Theories of Emotion , Facial Expression Recognition Studies , Facial Expression Recognition System , Face Tracking and Feature Extraction , Bayesian Network Classifiers , Experimental Analysis , Person-dependent Tests , Person-independent Tests, Experiments with Labeled and Unlabeled Data

Syllabus

Unit:I Computer Vision: Stages of Computer Vision; Acquisition, Pre-processing, Medium level Processing, High level Processing.

Unit: II Machine Learning for Big Data based Computer Vision Applications: Recognition Computer Vision applications needing Big Data, Learning in Computer Vision, Quality of Learning in Computer Vision systems.

Unit: III Multimodal Event Detection: Fusion Models: A Review, A Hierarchical Fusion Model , Working of the Model , The Duration Dependent Input Output Markov Model , Experimental Setup, Features, and Result.

Unit:IV Office Activity Recognition: Context-Sensitive Systems, Towards Tractable and Robust Context Sensing , Layered Hidden Markov Models (LHMMs) , Approaches , Decomposition per Temporal Granularity , Implementation of SEER , Feature Extraction and Selection in SEER , Architecture of SEER , Learning in SEER , Classification in SEER , Experiments , Discussion , Related Representations.

Unit:V Facial Expression Recognition: Introduction , Human Emotion Research ,Affective Human-computer Interaction , Theories of Emotion , Facial Expression Recognition Studies , Facial Expression Recognition System , Face Tracking and Feature Extraction , Bayesian Network Classifiers: Learning the "Structure" of the Facial Features , Experimental Analysis , Experimental Results with Labeled Data , Person-dependent Tests , Person-independent Tests, Experiments with Labeled and Unlabeled Data

Books:

1.Computer vision and Artificial Intelligence techniques Applied to Robot Soccer, 'Alexander Baratella, Mauricio Gomes', International Journal of Innovative Computing, Information and Control.
2. Machine Learning in Computer Vision by N. Sebe , Ashutosh Garg and Thomas S. Huang, Published by Springer.
3. Computer Vision: Algorithms and Applications, ' Richard Szeliski', Published by Springer.

URLS

1. https://pdfs.semanticscholar.org/05de/dee0bcf1f0e73876537c7b1ee27ad769f695.pdf
https://infolab.usc.edu/csci599/Fall2002/paper/DML2_streams-issues.pdf
2. https://www.iotforall.com/computer-vision-applications-in-daily-life/
3. http://szeliski.org/Book/.

Course Contents and Lecture Schedule

S.No.	Topic	No. of Lectures
1.	**Computer Vision:**	
1.1	Stages of Computer Vision; Acquisition	3
1.2	Pre-processing	3
1.3	Medium level Processing,	3
1.4	High level Processing	3
2.	**Machine Learning for Big Data based Computer Vision Applications**	
2.1	Recognition Computer Vision applications needing Big Data	3
2.2	Learning in Computer Vision,	3
2.3	Quality of Learning in Computer Vision systems.,	3
3.	**Multimodal Event Detection**	
3.1	Fusion Models: A Review, A Hierarchical Fusion Model ,	3
3.2	Working of the Model ,	3
3.3	The Duration Dependent Input Output Markov Model,	3
3.4	Experimental Setup, Features and Result.	3
4.	**Office Activity Recognition**	
4.1	Context-Sensitive Systems,	3
4.2	Towards Tractable and Robust Context Sensing	3
4.3	Layered Hidden Markov Models (LHMMs) , Approaches ,	2
4.4	Decomposition per Temporal Granularity	2
4.5	Implementation of SEER , Feature Extraction and Selection in SEER ,	2
4.6	Architecture of SEER , Learning in SEER ,.	3
4.7	Classification in SEER , Experiments , Discussion , Related Representations	3
5	**Facial Expression Recognition**	
5.1	Introduction , Human Emotion Research ,	3
5.2	Affective Human-computer Interaction , Theories of Emotion ,	4
5.3	Facial Expression Recognition Studies , Facial Expression Recognition System , Face Tracking and Feature Extraction ,	
5.4	Bayesian Network Classifiers: Learning the “Structure” of the Facial Features	3
	Experimental Analysis , Experimental Results with Labeled Data	3