DEPARTMENT OF FUTURES STUDIES

SCHOOL OF TECHNOLOGY, UNIVERSITY OF KERALA

Outcome Based Curriculum

M. Sc. Data Science

Syllabus effective from 2020 Admission onwards

Programme Specific Outcomes (PSOs)

PSO1	Understand principles and concepts of Data Science
PSO2	Develop necessary skills for data processing, statistical and computational analysis , visualization and document preparation
PSO3	Improve skills in solving scientific problems of interdisciplinary nature in academia and industry
PSO4	Students will become able to demonstrate a degree of mastery over the area as per the specialization of the program.
PSO5	Develop necessary mathematical and computational skill related to data science and allied areas
PSO6	Develop problem solving skill of interdisciplinary nature of present and futuristic problems
PSO7	Develop skills to act as data scientist in industry or academia

Structure of the Programme

Sem No.	Course Code	Name of the Course	Number of Credits
	Core Courses		
I	FDS-CC-511	Introduction to Data Science	2
	FDS-CC-512	Python for Data Science	3
	FDS-CC-513	Database Management Systems	3
	FDS-CC-514	Mathematical Foundations of Data Science	3
	FDS-CC-515	Statistical Foundations of Data Science	3
	FDS-CC-516	Lab 1: Data Base Management System	3
Total Credits for Semester I			**17**
	Core Courses		
II	FDS-CC-521	Introduction to Machine Learning	3
	FDS-CC-522	Parallel and Distributed Computing	3
	FDS-CC-523	Information Retrieval Techniques	3
	FDS-CC-524	Data Visualization and Presentation	2
	FDS-CC-525	Minor Project – Industry Based	3
	Internal Electives	Two electives	6
	FUS-DE-526(i)	Business Data Analytics	3
	FUS-DE-526(ii)	Time Series Analysis	3
	FUS-DE-526(iii)	Introduction to Big Data	3
	FUS-DE-527(i)	Basic Image Processing	3
	FUS-DE-527(ii)	Text Analytics	3
	FUS-DE-527(iii)	Operations Research	3
Total Credits for Semester II			**20**

III	Core Courses		
	FDS-CC-531	Advanced Machine Learning	3
	FDS-CC-532	Advanced graph and Network Analysis	3
	FDS-CC-533	Lab 3- Complex Network Analytics	3
	FDS-CC-534	Dissertation (Stage I)	4
	Internal Elect-ives	Two electives	6
	FDS-DE-535(i)	Fraud Analytics	3
	FDS-DE-535(ii)	Web Scraping and Analytics	3
	FDS-DE-535(iii)	Internet of Things in the Cloud	3
	FDS-DE-536(i)	Artificial Intelligence	3
	FDS-DE-536(ii)	Large Scale Optimization for Data Analytics	3
	FDS-DE-536(iii)	Models of Computations	3
Total Credits for Semester III			**19**
IV	FDS-CC-541	Dissertation (Stage II)	16
Extra Departmental Elective Courses			
I	FDS-GC-501	Foresight and Futures Research	2
III	FDS-GC-502	Parallel Programming with MPI	2
Generic Course			
Any semester	FDS-GC-503	Scientific Research Paper Writing	2

Codes: CC = Core; DE = Elective; P = Project; D = Dissertation; GC=Extra Departmental Elective.

Semester: I	Course Code: FDS-CC-511	Credits: 2

INTRODUCTION TO DATA SCIENCE

Course Outcomes : On completion of the course the student will be able to

CO	CO Statement	Mapping of COs		
		PO/ PSO	CL	KC
CO1	Understand evolution application of of data science	PSO 1, 4, 7	U	F,C
CO2	Understand different types of data		U	F,C
CO3	Develop creative thinking for information extraction from a given data set and gaining insights		CR	P
CO4	Understand big data and its management		U	F,C
CO5	Apply standard data analytics procedures		AP	P

CL- Cognitive Level: R-remember, U-understand, AP- Apply, AN- analyses, E- evaluate, CR- create,
KC- Knowledge Category: F-Factual, C- Conceptual, P-Procedural, M- Metacognitive)

Assessment Pattern (Internal & External)

Bloom's Category	Continuous Assessment Tests (percentage)			Terminal Examination (percentage)
	1	2	3	
Remember	10	20	20	20
Understand	30	20	20	20
Apply	40	40	40	40
Analyse				
Evaluate	10	10	10	10
Create				

COURSE CONTENT

MODULE I

Origins of data Science-Development-Popularization-Definition of DataScience Academic programs-Professional Organizations-Case Study- Mesh up of Disciplines- data Engineering-Acquiring - Ingesting -Transforming - Metadata -Storing - Retrieving - Scientific method-Reasoning Principles - Empirical Evidence -Hypothesis Testing -Repeatable Experiments –

MODULE II

Data Scientist-Thinking Like a Mathematician- Quantity -Structure - Space -Change – Thinking Like a Statistician - Collection - Organization - Analysis - Interpretation – Thinking Like a Programmer - Software Design - Programming Language - Source Code - Thinking Like a Visual Artist - Creative Process - Data Abstraction – Informationally Interesting -

MODULE III

What is Data-Data Point-Data Set-Data types- Data types in Mathematics, Statistics, Computer Science- Data types in R- Objects, Variables, Values and Vectors in R- Data sets and Data frames- Creating a data set in R- Talking to subject matter experts- looking for exception-exploring risks and uncertainty

MODULE IV

Big data definition, enterprise / structured data, social / unstructured data, unstructured data needs for analytics, Big data programming

MODULE V

Doing Data Science-Steps -Acquire-Parse-Filter-Mine-Represent-Refine-Interact-- Define the question- Define the ideal data set- Determine what data you can access- Obtain the data- Clean the data- Exploratory data analysis- Statistical prediction/modeling- Interpret results- Challenge results- Synthesize/write up results- Create reproducible codes- Distribute results to other people

MODULE VI

Case study on how to conduct or research of a data science problem

REFERENCES

1. O'Neil, C., & Schutt, R. (2013). Doing data science: Straight talk from the frontline. " O'Reilly Media, Inc.".
2. https://www.coursera.org/specializations/jhu-data-science
3. https://rpubs.com/rstober/steps-in-data-analysis

Semester: I	Course Code: FDS-CC-512	Credits: 3

PYTHON FOR DATA ANALYTICS

Course Outcomes : On completion of the course the student will be able to

CO	CO Statement	Mapping of COs		
		PO/ PSO	CL	KC
CO1	Understand programming syntax of Python for Data analysis	PSO 2, 4, 7	U	C
CO2	Carry out handling different type of data in Python		AP	P
CO3	Carry out various regressions in Python		AP	P
CO4	Carry out data analysis in python		AP	P
CO5	Write Python programmes		AP	P
CO6	Carry out data visualization python		AP	P
CO7	Evaluate different regression methods and its implement-ation in Python			

CL- Cognitive Level: R-remember, U-understand, AP- Apply, AN- analyses, E- evaluate, CR- create,
KC- Knowledge Category: F-Factual, C- Conceptual, P-Procedural, M- Metacognitive)

Assessment Pattern (Internal & External)

Bloom's Category	Continuous Assessment Tests (percentage)			Terminal Examination (percentage)
	1	2	3	
Remember	10	10	10	10
Understand	10	10	10	10
Apply	40	40	40	40
Analyse	20	20	20	20
Evaluate	20	20	20	20
Create				

COURSE CONTENT

MODULE I

Fundamentals of Python: Introduction to Python-Running Python Programs-Writing Python Code -Python Classes-Thinking about Objects-Class Variables and Methods-Managing Class Files

MODULE II

Working with Data: Data Types and Variables-Using Numeric Variables-Using String Variables-Dates and Times-Advanced Data and Time Management-Random Numbers-The Math Library-Class Instances-Creating Objects with Instance Data-Instance Methods-Managing Objects

MODULE III

Input and Output: Printing with Parameters-Getting Input from a User-String Formatting-Character Data-String Functions-Input Validation with "try / except",
Making Decisions: Logical Expressions-The "if" Statement-Logical Operators-More Complex Expressions

MODULE IV

Finding and Fixing Problems: Types of Errors-Troubleshooting Tools-Using the Python Debugger, Lists and Loops: Lists and Tuples-List Functions-"For" Loops-"While" Loops

MODULE V

Data Analysis with Pandas: Introduction, Pandas Series, DataFrames, Multi-index and index hierarchy, Working with Missing Data, Groupby Function, Merging, Joining and Concatenating DataFrames, Pandas Operations, Reading and Writing Files

MODULE VI

Regression Analysis using Python: Linear regression, Logistic regression, Ridge regression, Lasso regression, Polynomial regression, Stepwise regression

Data Visualization with Matplotlib

REFERENCES

1. McKinney, Wes. Python for Data Analysis. " O'Reilly Media, Inc.", 2013.
2. Sweigart, Al. Automate the boring stuff with Python: practical programming for total beginners. No Starch Press, 2015.
3. Albon, Chris. Machine learning with python cookbook: Practical solutions from preprocessing to deep learning. " O'Reilly Media, Inc.", 2018.
4. Beazley, David, and Brian K. Jones. Python Cookbook: Recipes for Mastering Python 3. " O'Reilly Media, Inc.", 2013.
5. Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems

ADDITIONAL REFERENCES

1. Matthes, Eric, Python Crash Course, and No Starch Press. "Introduction to Python." Introduction to Python: An open resource for students and teachers (2017).
2. Swaroop C. H A Byte of Python by. - https://python.swaroopch.com/
3. Perkovic, L. (2011). Introduction to computing using python: An application development focus. Wiley Publishing.
4. McKinney, W. (2012). Python for data analysis: Data wrangling with Pandas, NumPy, and IPython. " O'Reilly Media, Inc.".
5. Dierbach, C. (2012). Introduction to Computer Science Using Python: A Computational Problem-Solving Focus. Wiley Publishing.

Semester: I	Course Code: FDS–CC–513	Credits: 3

DATABASE MANAGEMENT SYSTEMS

Course Outcomes : On completion of the course the student will be able to

CO	CO Statement	Mapping of COs		
		PO/ PSO	CL	KC
CO1	Understand the basics of SQL and construct queries using SQL.		U	C
CO2	Understand the relational database design principles.		U	F, C
CO3	Understand the basic issues of transaction processing and concurrency control.		U	F, C
CO4	Understand database storage structures and access techniques.		U	F, C
CO5	Understand object oriented databases		U	C
CO6	Understand data warehousing		U	C
CO7	Understand MongoDB		U	C
CO8	Evaluate the nosql databases		E	C,P

CL- Cognitive Level: R-remember, U-understand, AP- Apply, AN- analyses, E- evaluate, CR- create,
KC- Knowledge Category: F-Factual, C- Conceptual, P-Procedural, M- Metacognitive)

Assessment Pattern (Internal & External)

Bloom's Category	Continuous Assessment Tests (percentage)			Terminal Examination (percentage)
	1	2	3	
Remember	10	10	10	10
Understand	10	10	10	10
Apply	40	40	40	40
Analyse	20	20	20	20
Evaluate	20	20	20	20
Create				

COURSE CONTENT

MODULE I

Introduction to File and Database systems- History- Advantages, disadvantages- Data views – Database Languages – DBA – Database Architecture – Data Models- Keys – Mapping Cardinalities

MODULE II

Relational Algebra and calculus – Query languages – SQL – Data definition – Queries in SQL – Updates – Views – Integrity and Security – triggers, cursor, functions, procedure – Embedded SQL – overview of QUEL, QBE.

MODULE III

Design Phases – Pitfalls in Design – Attribute types –ER diagram – Database Design for Banking Enterprise – Functional Dependence – Normalization (1NF, 2NF, 3NF, BCNF, 4NF, 5NF).File Organization – Organization of Records in files – Indexing and Hashing.

MODULE IV

Transaction concept – state- Serializability – Recoverability- Concurrency Control – Locks- Two Phase locking – Deadlock handling – Transaction Management in Multi Databases.

MODULE V

Object-Oriented Databases- OODBMS- rules – ORDBMS- Complex Data types – Distributed databases – characteristics, advantages, disadvantages, rules- Homogenous and Heterogenous- Distributed data Storage – XML – Structure of XML Data – XML Document. Introduction to MongoDB , Overview of NoSQL.

MODULE VI

Introduction to data warehousing, evolution of decision support systems -Modeling a data warehouse, granularity in the data warehouse - Data warehouse life cycle, building a data warehouse, Data Warehousing Components, Data Warehousing Architecture - On Line Analytical Processing, Categorization of OLAP Tools

REFERENCES

1. Silberschatz, A., Korth, H. F., & Sudarshan, S. (1997). Database system concepts (Vol. 4). New York: McGraw-Hill. ADDITIONAL REFERENCES
2. Pratt, P. J., Adamski, J. J., & Adamski, J. J. (1994). Database systems: management and design. Cambridge, MA: Cti.
3. Shamkant, R. E., &Navathe, B. (2009). Fundamentals of Database Systems. (Elmasri, R., &Navathe, S. (2010). Fundamentals of database systems. Addison-Wesley Publishing Company.)
4. Majumdar, A. K., & Bhattacharyya, P. (1996). Database management systems. McGraw-Hill.
5. ISRD group, ‘Introduction to Database Management Systems’, TMH, 2008
6. Ramakrishnan, R., &Gehrke, J. (2000). Database management systems. McGraw Hill
7. Chodorow, K. (2013). MongoDB: The Definitive Guide: Powerful and Scalable Data Storage. " O'Reilly Media, Inc.".
8. Harrison, G. (2015). Next Generation Databases: NoSQLand Big Data. Apress.

Semester: I **Course Code: FDS–CC–514** **Credits: 3**

MATHEMATICAL FOUNDATIONS OF DATA SCIENCE

Course Outcomes :On completion of the course the student will be able to

CO	CO Statement	Mapping of COs		
		PO/ PSO	CL	KC
CO1	Understand the basic set theory and logic		U	C
CO2	Understand the fundamentals of linear algebra		U	F, C
CO3	Understand the mathematics of optimization		U	F, C
CO4	Understand linear optimization techniques		U	F, C
CO5	Understand basics of nonlinear optimization techniques		U	C
CO6	Apply linear optimization techniques		AP	P
CO7	Apply nonlinear optimization techniques		AP	P
CO8	Use optimization software packages		AP	P

CL- Cognitive Level: R-remember, U-understand, AP- Apply, AN- analyses, E- evaluate, CR- create,
KC- Knowledge Category: F-Factual, C- Conceptual, P-Procedural, M- Metacognitive)

Assessment Pattern (Internal & External)

Bloom's Category	Continuous Assessment Tests (percentage)			Terminal Examination (percentage)
	1	2	3	
Remember	10	10	10	10
Understand	10	10	10	10
Apply	40	40	40	40
Analyse	20	20	20	20
Evaluate	20	20	20	20
Create				

COURSE CONTENT

MODULE I:

Basics of Data Science: Introduction; Typology of problems; Logic and set theory- Importance of linear algebra, statistics and optimization from a data science perspective; Structured thinking for solving data science problems.

MODULE II

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

MODULE III:

Linear Optimization-: Linear programming: Mathematical Model, assumptions of linear programming, Solutions of linear programming problems – Graphical Method,Simplex method, Artificial Variable Method, Two phase Method, Big M Method, Applications, Duality ,Dual simplex method, Introduction to sensitivity analysis

MODULE IV:

Unconstrained optimization; Necessary and sufficiency conditions for optima; Gradient descent methods; Constrained optimization, KKT conditions; Introduction to non-gradient techniques; Introduction to least squares optimization; Optimization view of machine learning.

MODULE V:

Introduction to Data Science Methods:; Linear classification problems.

MODULE VI:

Basic Packages for Linear and Non-linear Optimization-Lab

REFERENCES

1. G. Strang (2016). Introduction to Linear Algebra, Wellesley-Cambridge Press, Fifth edition, USA.
2. Luenberger, D. G. (1997). Optimization by vector space methods. John Wiley & Sons.
3. O'Neil, C., & Schutt, R. (2013). Doing data science: Straight talk from the frontline. " O'Reilly Media, Inc.".

Semester: I **Course Code: FDS–CC–515** **Credits: 3**

STATISTICAL FOUNDATIONS OF DATA SCIENCE

Course Outcomes :On completion of the course the student will be able to

CO	CO Statement	Mapping of COs		
		PO/ PSO	CL	KC
CO1	Articulate and exemplify the basic knowledge in statistics	PSO 1, 4, 7	U	C
CO2	Articulate and exemplify the basic knowledge in probability theory		U	F, C
CO3	Articulate and exemplify the basic knowledge in distribution theory		U	F, C
CO4	Carry out hypothesis testing		AP	P
CO5	Explain design of experiments		U	C
CO6	Carry out various models of linear regressions		AP	P

CL- Cognitive Level: R-remember, U-understand, AP- Apply, AN- analyses, E- evaluate, CR- create,
KC- Knowledge Category: F-Factual, C- Conceptual, P-Procedural, M- Metacognitive)

Assessment Pattern (Internal & External)

Bloom's Category	Continuous Assessment Tests (percentage)			Terminal Examination (percentage)
	1	2	3	
Remember	10	10	10	10
Understand	10	10	10	10
Apply	40	40	40	40
Analyse	20	20	20	20
Evaluate	20	20	20	20
Create				

COURSE CONTENT

MODULE I:

Statistics and probability: Statistical measures, probability- definitions-conditional probability- Baye's theorem; Random variables; Probability distributions and density functions, standard distributions. Mathematical expectations and moments; Covariance and correlation;

MODULE II:

Sampling Distributions: Sampling, Estimation of parameters- properties of estimates- methods of estimation; Multi variate distribution – multi variate techniques

MODULE III:

Testing of hypothesis- Neyman-Pearson approach, basic concepts, parametric tests and non-parametric tests. Confidence (statistical) intervals; Correlation functions; White-noise process.

MODULE IV:

Design of Experiments: Principles of experimental design- standard designs- CRD,RBD,LSD – Fixed effect, random effect and mixed effect models -Factorial designs – Nested designs.

MODULE V:

Linear regression as an exemplar function approximation problem- Maximum Likelihood methods and models

MODULE VI:

R Package for basic Statistics –Lab

REFERENCES

1. Bendat, J. S., &Piersol, A. G. (2011). Random data: analysis and measurement procedures (Vol. 729). John Wiley & Sons.
2. Montgomery, D. C., &Runger, G. C. (2010). Applied statistics and probability for engineers. John Wiley & Sons.
3. Cleves, M., Gould, W., Gutierrez, R., & Marchenko, Y. (2008). An introduction to survival analysis using Stata. Stata press.

Semester: I **Course Code: FDS–CC–516** **Credits: 3**

LAB I –DATA BASE MANAGEMENT SYSTEM

Course Outcomes : On completion of the course the student will be able to

CO	CO Statement	Mapping of COs		
		PO/ PSO	CL	KC
CO1	Employ construct queries using SQL.	PSO 5	U	C
CO2	Articulate relational database principles and employing relational database operations		U	F, C
CO3	Organize the database storage structures		AP	p
CO4	Carry out basic operations of mongodb		AP	P
CO5	Carry out basics operations of casandra		AP	p
CO6	Implement a mini project in data warehousing casandra		AP	P

CL- Cognitive Level: R-remember, U-understand, AP- Apply, AN- analyses, E- evaluate, CR- create,
KC- Knowledge Category: F-Factual, C- Conceptual, P-Procedural, M- Metacognitive)

Assessment Pattern (Internal & External)

Bloom's Category	Continuous Assessment Tests (percentage)			Terminal Examination (percentage)
	1	2	3	
Remember				
Understand				
Apply	100	100	100	100
Analyse				
Evaluate				
Create				

COURSE CONTENT

MODULE 1

Designing a Database - Creating tables for Banks, Hospitals, Applying the constraints like Primary Key, Foreign key, NOT NULL to the tables.

MODULE II

Writing SQL statements for implementing ALTER, UPDATE and DELETE, Writing the queries to implement the joins, Writing the queries for implementing the following functions: MAX (), MIN (),AVG (),COUNT ()

MODULE III

Writing the queries to implement the concept of Integrity constraints Writing the queries to create the views, Performing the queries for triggers, Performing operations for insertion, updation and deletion using the referential integrity constraints

MODULE IV

A mini project for applying the above constructs MODULE V : Writing programs to CRUD – Create, Read, Update, Delete operations in Mongodb, Writing programs on database operations in Mongodb

MODULE VI

Hands on in data ware housing using CASANDRA, A mini project in CASANDRA

REFERENCES

1. Ivan, B., (2003). SQL/ PL/SQL, The Programming Language of Oracle'. BPB Publication, New Delhi.
2. Chodorow, K. (2013). MongoDB: The Definitive Guide: Powerful and Scalable Data Storage. " O'Reilly Media, Inc.".
3. Harrison, G. (2015). Next Generation Databases: NoSQLand Big Data. Apress.

Semester: II **Course Code: FDS-CC-521** **Credits: 3**

INTRODUCTION TO MACHINE LEARNING

Course Outcomes : On completion of the course the student will be able to

CO	CO Statement	Mapping of COs		
		PSO	CL	KC
CO1	Articulate the basic knowledge on machine learning		U	C
CO2	Carry out regression and classification analysis		AP	P
CO3	Demonstrate the ability analyse multivariate data and models		AP	p
CO4	Demonstrate the ability carry out dimensionality reduction analysis		AP	P
CO5	Carry out clustering analyses		AP	p

CL- Cognitive Level: R-remember, U-understand, AP- Apply, AN- analyses, E- evaluate, CR- create,
KC- Knowledge Category: F-Factual, C- Conceptual, P-Procedural, M- Metacognitive)

Assessment Pattern (Internal & External)

Bloom's Category	Continuous Assessment Tests (percentage)			Terminal Examination (percentage)
	1	2	3	
Remember	10	10	10	10
Understand	10	10	10	10
Apply	40	40	40	40
Analyse	20	20	20	20
Evaluate	20	20	20	20
Create				

COURSE CONTENT

MODULE I

Introduction: Machine Learning, Applications, Supervised Learning: Learning a Class from Examples, Vapnik - Chervonenkis (VC) Dimension, Probably Approximately Correct (PAC) Learning, Noise, Learning Multiple Classes,

MODULE II

Regression, Model Selection and Generalization, Dimensions of a Supervised Machine Learning Algorithm, Bayesian Decision Theory: Introduction, Classification, Losses and Risks, Discriminant Functions, Utility Theory, Association Rules. Parametric Methods: Introduction, Maximum Likelihood Estimation, Evaluating an Estimator- Bias and Variance, The Bayes' Estimator, Parametric Classification, Regression, Tuning Model Complexity: Bias/Variance Dilemma,

MODULE III

Model Selection Procedures, Multivariate Methods: Multivariate Data, Parameter Estimation, Estimation of Missing Values, Multivariate Normal Distribution, Multivariate Classification, Tuning Complexity, Discrete Features, Multivariate Regression.

MODULE IV

Dimensionality Reduction: Introduction, Subset Selection, Principal Components Analysis, Factor Analysis, Multidimensional Scaling, Linear Discriminant Analysis, Isomap, Locally Linear Embedding,

MODULE V

Clustering: Introduction, Mixture Densities, k-Means Clustering, Expectation- Maximization Algorithm, Mixtures of Latent Variable Models, Supervised Learning after Clustering, Hierarchical Clustering, Choosing the Number of Clusters

MODULE VI

Practicals and case study for basic machine learning I

REFERENCES

1. Alpaydin, E. (2009). Introduction to machine learning. MIT press.
2. Trevor, H., Robert, T., & JH, F. (2009). The elements of statistical learning: data mining, inference, and prediction.

Semester: II **Course Code: FDS-CC-522** **Credits: 3**

PARALLEL AND DISTRIBUTED COMPUTING

Course Outcomes : On completion of the course the student will be able to

CO	CO Statement	Mapping of COs		
		PSO	CL	KC
CO1	Explain the range of requirements, functionality, design tradeoffs, memory hierarchy and cost-performance tradeoffs related to parallel/distributed systems, cluster computing, grid computing, supercomputing, cloud computing that modern parallel/distributed systems have to address.	PSO 5	U	C
CO2	Develop Parallel programmes		CR	P
CO3	Evaluate the requirement of a parallel program and suggest a suitable methodology		E	C

CL- Cognitive Level: R-remember, U-understand, AP- Apply, AN- analyses, E- evaluate, CR- create,
KC- Knowledge Category: F-Factual, C- Conceptual, P-Procedural, M- Metacognitive)

Assessment Pattern (Internal & External)

Bloom's Category	Continuous Assessment Tests (percentage)			Terminal Examination (percentage)
	1	2	3	
Remember	10	10	10	10
Understand	10	10	10	10
Apply	20	20	20	20
Analyse	20	20	20	20
Evaluate	20	20	20	20
Create	20	20	20	20

COURSE CONTENT

MODULE I

Distributed System Models and Enabling Technologies: Scalable Computing over the Internet, Technologies for Network-Based Systems, System Models for Distributed and Cloud Computing, Software Environments for Distributed Systems and Clouds, Performance, Security, and Energy Efficiency Computer Clusters for Scalable Parallel Computing: Clustering for Massive Parallelism, Computer Clusters and MPP Architectures, Design Principles of Computer Clusters, Cluster Job and Resource Management

MODULE II

Virtual Machines and Virtualization of Clusters and Data Centers: Implementation Levels of Virtualization, Virtualization Structures/Tools and Mechanisms, Virtualization of CPU, Memory, and I/O Devices, Virtual Clusters and Resource Management, Virtualization for Data- Center Automation . Cloud Platform Architecture over Virtualized Data Centers: Cloud Computing and Service Models, Data-Center Design and Interconnection Networks, Architectural Design of Compute and Storage Clouds, Cloud Security and Trust Management

MODULE III

Service-Oriented Architectures for Distributed Computing: Services and Service- Oriented Architecture, Message-Oriented Middleware, Discovery, Registries, Metadata, and Databases, Workflow in Service-Oriented Architectures, Programming on Amazon AWS and Microsoft Azure, Emerging Cloud Software Environments

MODULE IV

Cloud Programming and Software Environments: Features of Cloud and Grid Platforms, Parallel and Distributed Programming Paradigms, Programming Support of Google App Engine,

MODULE V

Grid Computing Systems and Resource Management: Grid Architecture and Service Modeling, Grid Projectand Grid Systems Built, Grid Resource Management and Brokering, Peer-to-Peer Computing and Overlay Networks: Peer-to-Peer Computing Systems, P2P Overlay Networks and Properties Ubiquitous Clouds and the Internet of Things: Cloud Trends in Supporting Ubiquitous Computing, Performance of Distributed Systems and the Cloud, Enabling Technologies for the Internet of Things

MODULE VI

Programming scalable systems: Programming using MPI paradigm Programming shared-address space systems: OpenMP, Cilk Plus Programming heterogeneous systems: CUDA and OpenCL, OpenACC and OpenMP (4.0)

REFERENCES

1. Hwang, K., Dongarra, J., & Fox, G. C. (2013). Distributed and cloud computing: from parallel processing to the internet of things. Morgan Kaufmann.
2. Grama, A., Kumar, V., Gupta, A., &Karypis, G. (2003). Introduction to parallel computing. Pearson Education.
3. Quinn, M. J. (2003). Parallel Programming in C with MPI and OpenMP, Mc-Graw Hill.
4. Wen-mei, W. H. (2010). Programming massively parallel processors. Morgan Kaufmann.
5. Gropp, W. D., Gropp, W., Lusk, E., Skjellum, A., & Lusk, A. D. F. E. E. (1999). Using MPI: portable parallel programming with the message-passing interface (Vol. 1). MIT press.

Semester: II **Course Code: FDS–CC–523** **Credits: 3**

INFORMATION RETRIEVAL TECHNIQUES

Course Outcomes : On completion of the course the student will be able to

CO	CO Statement	Mapping of COs		
		PSO	CL	KC
CO1	Articulate the basic knowledge in information retrieval		U	C
CO2	Understand the techniques of information retrieval		U	P
CO3	Carry out web search		AP	P
CO4	Carry out link analysis		AP	P
CO5	Carry out information filtering and text minining		AP	P

CL- Cognitive Level: R-remember, U-understand, AP- Apply, AN- analyses, E- evaluate, CR- create, KC- Knowledge Category: F-Factual, C- Conceptual, P-Procedural, M- Metacognitive)

Assessment Pattern (Internal & External)

Bloom's Category	Continuous Assessment Tests (percentage)			Terminal Examination (percentage)
	1	2	3	
Remember	10	10	10	10
Understand	10	10	10	10
Apply	20	20	20	20
Analyse	20	20	20	20
Evaluate	20	20	20	20
Create	20	20	20	20

COURSE CONTENT

MODULE I

Introduction- HistoryofIR –Components of IR-Issues– Open source Search engine Frameworks- Theimpactofthewebon IR-Theroleofartificialintelligence(AI)inIR–IRVersus Web Search- Components of a Search engine – Characterizing the web

MODULE II

Boolean and vector-space retrieval models - Term weighting- TF-IDF weighting – cosine similarity – Preprocessing-Invertedindices-efficientprocessingwithsparsevectors– LanguageModelbased IR – Probabilistic IR– Latent Semantic Indexing- Relevance feedback and query expansion.

MODULE III

Web search overview, web structure, the user, paid placement, search engine optimization/spam. Web size measurement-search engine optimization/spam –Web Search Architectures-crawling- meta-crawlers-Focused Crawling-web indexes –-Near-duplicate detection- Index Compression– XML retrieval

MODULE IV

Link Analysis – hubs and authorities–Page Rank and HITS algorithms- Searching and Ranking– Relevance Scoring and ranking for Web–Similarity-Hadoop & MapReduce-Evaluation- Personalizedsearch-Collaborative filtering and content-based recommendation of documents and products–handling "invisible" Web- Snippet generation, Summarization, Question Answering, Cross- Lingual Retrieval

MODULE V

Information filtering; organization and relevance feedback – Text Mining- Text classification and clustering-Categorization algorithms: naive Bayes; decision trees; and nearest neighbor-Clustering

algorithms:agglomerativeclustering;k-means;expectationmaximization(EM). MODULE VI: Case Study presentation of IR

REFERENCES

1. Schütze, H., Manning, C. D., & Raghavan, P. (2008). Introduction to information retrieval (Vol. 39). Cambridge University Press.
2. Ricardo Baeza-Yates &Berthier Ribeiro-Neto (2011). Modern Information Retrieval: The Concepts and Technology behind Search 2nd Edition, ACM Press Books.
3. Bruce Croft, Donald Metzler and Trevor Strohman (2009). Search Engines: Information Retrieval in Practice, 1st Edition Addison Wesley.
4. Mark Levene(2010) An Introduction to Search Engines and Web Navigation, 2nd Edition Wiley.

ADDITIONAL REFERENCES

1. Stefan Buettcher, Charles L. A. Clarke, Gordon V. Cormack(2010). Information Retrieval: Implementing and Evaluating Search Engines, The MIT Press.
2. Grossman, D. A., & Frieder, O. (2012). Information retrieval: Algorithms and heuristics (Vol. 15). Springer Science & Business Media.
3. Manu Konchady(2008) “Building Search Applications: Lucene, Ling Pipe”,First Edition, Gate Mustru Publishing.
4. www.nptel.ac.in

Semester: II **Course Code: FDS-CC-524** **Credits: 2**

DATA VISUALIZATION AND PRESENTATION

Course Outcomes : On completion of the course the student will be able to

CO	CO Statement	Mapping of COs		
		PSO	CL	KC
CO1	Articulate and exemplify the basic knowledge on data visualization.	PSO 4, 5	U	C
CO2	Demonstrate the use data visualization tools		AP	P
CO3	Demonstrate the ability to develop visualization of output based on requirement		AP	P
CO4	Differentiate various visualization strategies		E	P
CO5	Demonstrate the ability to employ ggplot in R		AP	P
CO6	Development of computer code for visualization		CR	P

CL- Cognitive Level: R-remember, U-understand, AP- Apply, AN- analyses, E- evaluate, CR- create, KC- Knowledge Category: F-Factual, C- Conceptual, P-Procedural, M- Metacognitive)

Assessment Pattern (Internal & External)

Bloom's Category	Continuous Assessment Tests (percentage)			Terminal Examination (percentage)
	1	2	3	
Remember	10	10	10	10
Understand	10	10	10	10
Apply	20	20	20	20
Analyse	20	20	20	20
Evaluate	20	20	20	20
Create	20	20	20	20

COURSE CONTENT

MODULE I

Purpose of visualization, visual perception, cognitive issues- evaluation as well as other theory and design principles behind information visualization

MODULE II

Multidimensional visualization, tree visualization, graph visualization.

MODULE III

Time series data visualization techniques

MODULE IV

Understanding analytics output and their usage, basic interaction techniques such as selection and distortion, evaluation,

MODULE V

Examples of information visualization applications and systems, user tasks and analysis-visualisation packages

MODULE VI

Grammar of graphics using R-Construct/Deconstruct a graphic into a data- order of accuracy of perceptual tasks and its impact and Case study presentations and lab based on R package of Data Visualizations.

REFERENCES

1. Wickham, H. (2016). ggplot2: elegant graphics for data analysis. Springer.
2. Keen, K. J. (2010). Graphics for statistics and data analysis with R. CRC Press.
3. Cook, D., Swayne, D. F., &Buja, A. (2007). Interactive and dynamic graphics for data analysis: with R and GGobi. Springer Science & Business Media.
4. Dalgaard, P. (2008). Introductory statistics with R. Springer Science & Business Media.
5. Verzani, J. (2014). Using R for introductory statistics. CRC Press.
6. Murrell, P. (2016). R graphics. CRC Press.
7. Cleveland, W. S. (1993). Visualizing data. Hobart Press.
8. Tufte, E. R., Goeler, N. H., & Benson, R. (1990). Envisioning information (Vol. 126).
9. Cheshire, CT: Graphics press.
10. Tufte, E., & Graves-Morris, P. (2014). The visual display of quantitative information.; 1983.

Semester: II	Course Code: FDS–CC–525	Credits: 3

MINOR PROJECT - INDUSTRY BASED

Objective: The aim of this course is to address a data science problem of practical relevance in an industry or an organization of similar type.

COURSE DESCRIPTION

- To identify an existing/new data science problem in an industry/organization and approach the problem using the tools and technologies of data science
- To study and analyse a data science problem in an industry/organization and suggest possible solution and submit the report under the supervision and guidance from the concerned officer/executive of the industrial concern.

REFERENCES

All literature relevant to the chosen Industrial Problem as suggested by the advisor of the industry and required for the student.

Semester: II **Course Code: FDS-DE-526(i)** **Credits: 3**

BUSINESS DATA ANALYTICS

Course Outcomes : On completion of the course the student will be able to

CO	CO Statement	Mapping of COs		
		PSO	CL	KC
CO1	Articulate and exemplify the basic knowledge on business data analytics		U	C
CO2	Demonstrate the use analytic techniques		AP	P
CO3	Demonstrate the ability to solve issues related to over-fitting		AP	P
CO4	Evaluation of models developed		E	P
CO5	Perform text mining		AP	P
CO6	Identify strength and weakness of analytics techniques		AN	P

CL- Cognitive Level: R-remember, U-understand, AP- Apply, AN- analyses, E- evaluate, CR- create,
KC- Knowledge Category: F-Factual, C- Conceptual, P-Procedural, M- Metacognitive)

Assessment Pattern (Internal & External)

Bloom's Category	Continuous Assessment Tests (percentage)			Terminal Examination (percentage)
	1	2	3	
Remember	10	10	10	10
Understand	10	10	10	10
Apply	20	20	20	20
Analyse	20	20	20	20
Evaluate	20	20	20	20
Create	20	20	20	20

COURSE CONTENT

MODULE I

Introduction – Ubiquity of Data Opportunities, Data Science and Data Driven decisionMaking, Data Processing and Big Data, Data Analytic Thinking, Data Science and Data Mining.Business Problems and Data Science Solutions -Fundamental concepts, From Business Problems to Data Mining Tasks, Supervised Versus Unsupervised Methods, The Data Mining Process

MODULE II

Other Analytics Techniques and Technologies. Introduction to Predictive Modeling: Fundamental concepts, Models, Induction, and Prediction, Supervised Segmentation, Visualizing Segmentations, Trees as Sets of Rules, Probability Estimation. Fitting a Model to Data – Fundamental concepts, Classification via Mathematical Functions, Regression via Mathematical Functions, Class Probability Estimation and Logistic "Regression", Nonlinear Functions, Support Vector Machines, and Neural Networks.

MODULE III

Overfitting - Fundamental concepts: Generalization, Overfitting, Overfitting Examined, From Holdout Evaluation to Cross-Validation, Learning Curves, Overfitting Avoidance and Complexity Control. Similarity, Neighbors, and Clusters - Fundamental concepts, Similarity and Distance, Nearest-Neighbor Reasoning, Important Technical Details Relating to Similarities and Neighbors, Clustering - Hierarchical Clustering, Clustering Around Centroids, Understanding the Results of Clustering. Solving a Business Problem versus Data Exploration

MODULE IV

Decision Analytic Thinking I: Fundamental concepts, Evaluating Classifiers, Generalizing Beyond Classification, A Key Analytical Framework: Expected Value, Evaluation, Baseline Performance, and Implications for Investments in Data, Visualizing Model Performance - Fundamental concepts, Ranking Instead of Classifying, Profit Curves, ROC Graphs and Curves, The Area Under the ROC Curve (AUC), Cumulative Response and Lift Curves. Evidence and Probabilities - Fundamental concepts, Combining Evidence, Applying Bayes' Rule to Data Science, A Model of Evidence ``Lift".

MODULE V

Representing and Mining Text - Fundamental concepts, Representation – Bag of Words, Term Frequency, Measuring Sparseness: Inverse Document Frequency, Combining Them: TFIDF, The Relationship of IDF to Entropy; Beyond Bag of Words – N-gram Sequences, Named Entity Extraction, Topic Models. Decision Analytic Thinking II: Fundamental concept, Targeting the Best Prospects for a Charity Mailing – The Expected Value Framework: Decomposing the Business Problem and Recomposing the Solution Pieces, A Brief Digression on Selection Bias.

MODULE VI

Other Data Science Tasks and Techniques - Fundamental concepts, Co-occurrences and Associations, Profiling, Link Prediction and Social Recommendation, Data Reduction, Latent Information, and Movie Recommendation, Bias, Variance, and Ensemble Methods, Data-Driven Causal Explanation.
Explanation.

REFERENCES

Provost, F., & Fawcett, T. (2013). Data Science for Business: What you need to know about data mining and data-analytic thinking. " O'Reilly Media, Inc.".

ADDITIONAL REFERENCE

Lander, J. P. (2014). R for everyone: advanced analytics and graphics. Pearson Education.

Semester: II **Course Code: FDS-DE-526(ii)** **Credits: 3**

TIME SERIES ANALYSIS

Course Outcomes : On completion of the course the student will be able to

CO	CO Statement	Mapping of COs		
		PSO	CL	KC
CO1	Articulate and exemplify the basic knowledge on time series analyse		U	C
CO2	Demonstrate the use linear methods		AP	P
CO3	Perform time series forecasting		AP	P
CO4	Discriminate stationary and non-stationary characteristics		E	P
CO5	Evaluate time series models		E	P
CO6	Write computer codes for time series analyses		CR	P

CL- Cognitive Level: R-remember, U-understand, AP- Apply, AN- analyses, E- evaluate, CR- create,
KC- Knowledge Category: F-Factual, C- Conceptual, P-Procedural, M- Metacognitive)

Assessment Pattern (Internal & External)

Bloom's Category	Continuous Assessment Tests (percentage)			Terminal Examination (percentage)
	1	2	3	
Remember	10	10	10	10
Understand	10	10	10	10
Apply	20	20	20	20
Analyse	20	20	20	20
Evaluate	20	20	20	20
Create	20	20	20	20

COURSE CONTENT

MODULE I

Stochastic process and its main characteristics: Stochastic process. Time series as a discrete stochastic process. Stationarity. Main characteristics of stochastic processes (means, autocovariation and autocorrelation functions). Stationary stochastic processes. Stationarity as the main characteristic of stochastic component of time series. Wold decomposition. Lag operator.

MODULE II

Autoregressive-moving average models ARMA (p,q) Moving average models MA(q). Condition of invertability. Autoregressive models AR(p). Yull-Worker equations. Stationarity conditions. Autoregressive-moving average models ARMA (p,q). Coefficients estimation in autoregressive models.

MODULE III

Coefficient estimation in ARMA (p) processes: Quality of adjustment of time series models. AIC information criterion. BIC information criterion. "Portmonto"-statistics. Box-Jenkins methodology to identification of stationary time series models.

MODULE IV

Forecasting in the framework of Box-Jenkins model: Forecasting, trend and seasonality in Box-Jenkins model. Implementation using R software packages.

MODULE V

Non-stationary time series: Non-stationary time series. Time series with non- stationary variance. Non-stationary mean. ARIMA (p,d,q) models. The use of Box-Jenkins methodology to determination of order of integration.

MODULE VI

Fitting Models to Data: Model identification, Parameter estimation, Model diagnostics and model selection, Forecasting time series

REFERENCES

1. Wei, W. W. S. (2006). Time Series Analysis Univariate and Multivariate Methods, 2nd Edition. Addison Wesley
2. Andrew C. Harvey(1993). Time Series Models. Harvester Wheatsheaf
3. P. J. Brockwell, R. A. Davis.(1996) Introduction to Time Series and Forecasting. Springer. Enders, W. (2008). Applied econometric time series. John Wiley & Sons.
4. Hamilton, J. D. (1994). Time series analysis (Vol. 2, pp. 690-696). Princeton, NJ: Princeton University Press.
5. Brockwell, P. J., Davis, R. A., & Fienberg, S. E. (1991). Time Series: Theory and Methods: Theory and Methods. Springer Science & Business Media.
6. Shumway, R. &Stoer, D.(2006). Time Series Analysis and Its Applications with R Examples, Springer.

Semester: II | Course Code: FDS-DE-526(iii) | Credits: 3

INTRODUCTION TO BIG DATA

Course Outcomes : On completion of the course the student will be able to

CO	CO Statement	Mapping of COs		
		PSO	CL	KC
CO1	Articulate and exemplify the basic knowledge on big data analyses	PSO 2, 5	U	C
CO2	Demonstrate the use of MapReduce framework		AP	P
CO3	Write programmes for big data analytics		CR	P
CO4	Discriminate big data architecture		E	P
CO5	Articulate basic knowledge on NoSQL		E	P
CO6	Propose big data analytic methodologies		CR	P

CL- Cognitive Level: R-remember, U-understand, AP- Apply, AN- analyses, E- evaluate, CR- create,
KC- Knowledge Category: F-Factual, C- Conceptual, P-Procedural, M- Metacognitive)

Assessment Pattern (Internal & External)

Bloom's Category	Continuous Assessment Tests (percentage)			Terminal Examination (percentage)
	1	2	3	
Remember	10	10	10	10
Understand	10	10	10	10
Apply	20	20	20	20
Analyse	20	20	20	20
Evaluate	20	20	20	20
Create	20	20	20	20

COURSE CONTENT

MODULE I

Big Data – Introduction, Structuring Big Data, Elements of Big data, Big data analytics, Big data applications. Big Data in business context, Technologies for handling big data –Distributed and Parallel computing for Big Data, Data Models, Computing Models, Introducing Hadoop – HDFS and MapReduce.

MODULE II

Understanding Analytics and Big data – Comparison of Reporting and Analysis, Types of Analytics, Analytical approaches. Hadoop EcoSystem, Hadoop Distributed file system, HDFS architecture, MapReduce, Hadoop YARN, Introducing HBase, Hive and Pig

MODULE III

MapReduce framework, Techniques to Optimize MapReduce, Uses of MapReduce, Role of HBase in Big data processing, Processing Data with MapReduce – Framework, Developing simple MapReduce Application.

MODULE IV

MapReduce execution and Implementing MapReduce Programs, YARN Architecture – Limitations of MapReduce, Advantages of YARN, Working of YARN, YARN Schedulers,
Configurations, Commands, Containers

MODULE V

Introduction to Mahout – Machine Learning, Clustering, Classification, Mahout
Algorithms, Environment for Mahout. Introduction to NoSQL.

MODULE VI

Overview of High Value BD Use Cases and Examples

REFERENCES

1. DT Editorial Services , “Big Data, Black Book: Covers Hadoop 2, MapReduce, Hive, YARN, Pig, R and Data Visualization Wiley India

ADDITIONAL REFERENCES

1. Lublinsky, B., Smith, K. T., &Yakubovich, A. (2013). Professional hadoop solutions. John Wiley & Sons.
2. Chris Eaton,DirkDeroos(2012) , “Understanding Big data ”, McGraw Hill.
3. Sima Acharya, Subhashini Chellappan, Big Data and Analytics, Willey.
4. Tom White(2012), “Hadoop: The definitive Guide”, O Reilly.
5. Vignesh Prajapati (2013), “Big Data Analytics with R and Hadoop”, Packet Publishing.
6. Kulkarni, P., Joshi, S., & Brown, M. S. (2016). Big data analytics. PHI Learning Pvt. Ltd.

Semester: II | Course Code: FDS-DE-527(i) | Credits: 3

BASIC IMAGE PROCESSING

Course Outcomes : On completion of the course the student will be able to

CO	CO Statement	Mapping of COs		
		PSO	CL	KC
CO1	Articulate and exemplify the basic knowledge on image processing	PSO 5	U	C
CO2	Demonstrate the use of filtering techniques		AP	P
CO3	Propose image restoration and reconstruction techniques		CR	P
CO4	Formulate image compression techniques		CR	P
CO5	Formulate image segmentation approach		CR	P

CL- Cognitive Level: R-remember, U-understand, AP- Apply, AN- analyses, E- evaluate, CR- create,
KC- Knowledge Category: F-Factual, C- Conceptual, P-Procedural, M- Metacognitive)

Assessment Pattern (Internal & External)

Bloom's Category	Continuous Assessment Tests (percentage)			Terminal Examination (percentage)
	1	2	3	
Remember	10	10	10	10
Understand	10	10	10	10
Apply	20	20	20	20
Analyse	20	20	20	20
Evaluate	20	20	20	20
Create	20	20	20	20

COURSE CONTENT

MODULE I

Introduction, Digital Image Fundamentals: elements of visual perception, light and electromagnetic spectrum, image sensing and acquisition, image sampling and quantization, some basic relationship between pixels. Intensity Transformations: Basics of intensity transformations, some basic intensity transformation functions, histogram processing.

MODULE II

Spatial Filtering: fundamentals of spatial filtering, smoothing and sharpening filters. Frequency domain Filtering: Background, preliminary concepts, sampling, Fourier transforms and DFT, 2-D DFT and properties, frequency domain filtering, low pass filters, high pass filters, implementation.

MODULE III

Image restoration and Reconstruction: Noise models, restoration in the presence of noise, linear-positive invariant degradations, inverse filtering, Wiener filtering, constrained least square filtering, geometric mean filter.

MODULE IV

Image Compression: fundamentals, basic compression methods.
Morphological Image Processing: preliminaries, erosion and dilation, opening and closing, basic morphological algorithms.

MODULE V

Image Segmentation: fundamentals, point, line and edge detection, thresholding, region based segmentation, use of motion in segmentation.

MODULE VI

Natural language processing - tokenziation, part-of-speech tagging, chunking, syntax parsing and named entity recognition. Document representation : representing unstructured text documents with appropriate format- structure to support automated text mining algorithms. Assigning a text document to classes/ categories. supervised text categorization algorithms, Naive Bayes, k Nearest Neighbor (kNN) and Logistic Regression

REFERENCES

1. Jain, A. K. (1989). Fundamentals of digital image processing. Englewood Cliffs, NJ: Prentice Hall,.
2. Pratt, W. K. (2007). Digital image processing: PIKS Scientific inside (Vol. 4). Hoboken, New Jersey: Wiley-interscience.
3. Aggarwal, C. C., &Zhai, C. (Eds.). (2012). Mining text data. Springer Science & Business Media.
4. Jurafsky, D. (2000). Speech & language processing. Pearson Education India.
5. Gonzalez, R. C., & Woods, R. E. (2002). Digital image processing.

Semester: II | Course Code: FDS-DE-527(ii) | Credits: 3

TEXT ANALYTICS

Course Outcomes : On completion of the course the student will be able to

CO	CO Statement	Mapping of COs		
		PSO	CL	KC
CO1	Articulate and exemplify the basic knowledge on text analytics	PSO 4, 5	U	C
CO2	Understand the sources and limitations of text and social media data.		U	C
CO3	understand the structural, syntactical, semantic elements of textual data.		U	C
CO4	understand the structural and social aspects of social media networks.		U	C
CO5	Become familiar with core practice communities, publications, and organizations focusing on text and social media analytics and the research questions they are engaged in.		U	C
CO6	use common text mining and social media analytics tools to gather managerial insights.		AP	P
CO7	develop and present network visualization for social media data.		CR	P

CL- Cognitive Level: R-remember, U-understand, AP- Apply, AN- analyses, E- evaluate, CR- create, KC- Knowledge Category: F-Factual, C- Conceptual, P-Procedural, M- Metacognitive)

Assessment Pattern (Internal & External)

Bloom's Category	Continuous Assessment Tests (percentage)			Terminal Examination (percentage)
	1	2	3	
Remember	10	10	10	10
Understand	10	10	10	10
Apply	20	20	20	20
Analyse	20	20	20	20
Evaluate	20	20	20	20
Create	20	20	20	20

COURSE CONTENT

MODULE I

Natural language. Language, syntax and structure. Language semantics. Natural language processing. Foundations of Natural Language Processing and preparing text data

MODULE II

Understanding Text and Processing: Text tokenization. Text normalization. Cleaning text. Understanding structure and syntax.

MODULE III

Text Similarity and Clustering: Information retrieval. Text similarity and similarity measures. Common distance measures: Hamming distance, Manhattan distance, Euclidian distance, Levenshtein Edit Distance. Document clustering.

MODULE IV

Introduction to Sentiment Analysis: Defining the sentiment analysis problem – objective and tasks. Understanding affect, emotion, mood, and opinion. Preparing data for analysis. Supervised and unsupervised learning. Classification using lexicon-based approach.

MODULE V

Introduction to Social Media Analytics: Introduction. Social media and social media networks. Social media data – structured and unstructured data. Applications.

MODULE VI

Social Media Data Analysis and Visualization: Collecting and extracting social media data. Statistical analysis of data. Extracting useful patterns. Network analysis. Creating network graphs. Node importance – key influencers. Modeling network dynamics and growth.

REFERENCES

1. Steven Struhl: Practical Text Analytics: Interpreting Text and Unstructured Data for Business Intelligence. 1st edition. Kogun Page (2015).
2. Bing Liu: Sentiment Analysis: Mining Opinions, Sentiments, and Emotions. 1st edition. Cambridge University Press (2015).
3. Marco Bonzanini: Mastering Social Media Mining with Python. 1st edition. Packt Publishing (2016).

ADDITIONAL REFERENCES

1. Abdul-Mageed, M. (2016). Sentiment Analysis.
2. Bird, S., Klein, E., &Loper, E. (2009). Natural language processing with Python: analyzing text with the natural language toolkit. O'Reilly Media, Inc.
3. Bayley, R., Cameron, R., & Lucas, C. (2013). The Oxford handbook of sociolinguistics. Oxford University Press
4. Taylor, J. R. (Ed.). (2015). The Oxford handbook of the word. Oxford University Press.

Semester: II	Course Code: FDS-DE-527(iii)	Credits: 3

OPERATION RESEARCH

Course Outcomes : On completion of the course the student will be able to

CO	CO Statement	Mapping of COs		
		PSO	CL	KC
CO1	Articulate the basic knowledge on Operations Research	PSO 4, 5	U	C
CO2	Solve leaner programming problems		AP	P
CO3	Solve Integer programming problems		AP	P
CO4	Solve TSP and mixed integer programming problem		AP	P
CO5	Simulate queueing models		AP	P
CO6	Articulate and solve inventory models		AP	P
CO7	Analyse real world situations/problems for building and solving		E	P

CL- Cognitive Level: R-remember, U-understand, AP- Apply, AN- analyses, E- evaluate, CR- create, KC- Knowledge Category: F-Factual, C- Conceptual, P-Procedural, M- Metacognitive)

Assessment Pattern (Internal & External)

Bloom's Category	Continuous Assessment Tests (percentage)			Terminal Examination (percentage)
	1	2	3	
Remember	10	10	10	10
Understand	30	30	30	30
Apply	40	40	40	40
Analyse	10	10	10	10
Evaluate	10	10	10	10
Create				

COURSE CONTENT

MODULE I

Linear programming: Mathematical Model, assumptions of linear programming, Solutions of linear programming problems – Graphical Method, Simplex method, Artificial Variable Method, Two phase Method, Big M Method, Applications, Duality ,Dual simplex method, Introduction to sensitivity analysis

MODULE II

Special types of Linear programming problems- Transportation Problem – Mathematical formulation of Transportation Problem, Basic feasible solution in TP, Degeneracy in TP, Initial basic feasible solutions to TP, Matrix Minima Method, Row Minima Method, Column Minima Method, Vogel's Approximation Method, Optimal Solution to TP, MODI Method, Stepping Stone Method, Assignment problems – Definition, Hungarian Method

MODULE III

Integer Programming: Pure Integer Programming, Mixed Integer Programming, Solution Methods – Cutting plane method, branch and bound method. Binary Integer Linear programming- Travelling salesman problems – Iterative method, Branch and bound method

MODULE IV

Dynamic programming: Deterministic and Probabilistic Dynamic programming. Linear programming by dynamic programming approach.

MODULE V

Queuing Model: Elements and Characteristics of queuing systems., Classification of queuing systems –Structures of Basic Queuing System, Definition and classification of stochastic processes- discrete-time Markov Chains – Continuous Markov Chains-The classical system-Poisson Queuing Sysetm – M/M/1: ∞/FIFO, M/M/1: ∞/SIRO, M/M/1: N/FIFO , Birth Death Queuing Systems, Pure Birth system, Pure Death system, M/M/C: N/FIFO, M/M/C:
C/FIFO

MODULE VI

Inventory Models: Deterministic inventory models, economic order quantity and its extensions - Stochastic inventory models, setting safety stocks - Reorder point order quantity models

REFERENCES

1. JK Sharma(2009), Operations Research – Theory and Applications, 4th Ed, Mc Millan Publishing.
2. Taha(2007), Operations Research, 8th Ed., Mc Millan Publishing Company
3. Kantiswaroop, PK Guptha, &Manmohan(2007),Operation Research", 13th Ed, Sulthan Chand &Sons.
4. Beightler C. S, & Philips D T (2009), 'Foundations of optimisation', 2nd Ed., Prentice Hall.
5. Mc Millan Claude Jr(1979), 'Mathematical Programming', 2nd Ed. Wiley Series.
6. Srinath L.S, 'Linear Programming', East-West, New Delhi.

Semester: III	Course Code: FDS-CC-531	Credits: 3

ADVANCED MACHINE LEARNING

Course Outcomes : On completion of the course the student will be able to

CO	CO Statement	Mapping of COs		
		PSO	CL	KC
CO1	Explain advanced machine learning models	PSO4, 5	U	C
CO2	Explain deep learning		U	C
CO3	Understand CNN, reinforcement learning		U	P
CO4	Understand learning algorithms		U	P
CO5	Apply advanced machine learning techniques		AP	P
CO6	Evaluate real-world problems for recommend suitable models		E	P
CO7	Create computer codes for ML models		CR	p

CL- Cognitive Level: R-remember, U-understand, AP- Apply, AN- analyses, E- evaluate, CR- create, KC- Knowledge Category: F-Factual, C- Conceptual, P-Procedural, M- Metacognitive)

Assessment Pattern (Internal & External)

Bloom's Category	Continuous Assessment Tests (percentage)			Terminal Examination (percentage)
	1	2	3	
Remember	10	10	10	10
Understand	20	20	20	20
Apply	40	40	40	40
Analyse	10	10	10	10
Evaluate	10	10	10	10
Create	10	10	10	10

COURSE CONTENT

MODULE I

Multilayer Perceptrons: Introduction, The Perceptron, Training a Perceptron, Learning Boolean Functions, Multilayer Perceptrons, Backpropagation Algorithm, Training Procedures, Competitive Learning, Radial Basis Function, Introduction to Kernel Machines, Optimal Separating Hyperplane, The Nonseparable Case-Soft Margin Hyperplane, v-SVM, Kernel Trick, Vectorial Kernels, Defining Kernels

MODULE II

What is deep learning? DL successes -Intro to neural networks (Kate) cost functions, hypotheses and tasks; training data; maximum likelihood based cost, cross entropy, MSE cost; feed-forward net-works; MLP, sigmoid units; neuroscience inspiration; Learning in neural networks (Kate) output vs hidden layers; linear vs nonlinear networks;

MODULE III

Back propagation (Kate) learning via gradient descent; recursive chain rule (backpropagation); if time: bias-variance trade off, regularization; output units: linear, softmax; hidden units: tanh, RELU

MODULE IV

Deep learning strategies I (Brian) (e.g., GPU training, regularization,etc); Deep learning strategies II (Brian) (e.g., RLUs, dropout, etc)-SCC/TensorFlow overview (Katia Oleinik) How to use the SCC cluster; introduction to Tensor flow.

MODULE V

CNNs (Kate) Convolutional neural networks -CNNs II (Kate) -Deep Belief Nets I (Brian) probabilistic methods-Deep Belief Nets II (Brian) -RNNs I (Sarah) Recurrent neural networks -RNNs II (Kate)-Other DNN variants (Kate) (e.g. attention, memory networks, etc.)

MODULE VI

Overview of reinforcement learning: the agent environment framework, successes of reinforcement learning-Bandit problems and online learning-Markov decision processes-Returns, and value functions-Solution methods: dynamic programming-Solution methods: Monte Carlo learning- Solution methods: Temporal difference learning-Eligibility traces-Value function approximation (function approximation)-Models and planning (table lookup case)

REFERENCES

1. Learning, D. (2016). Ian Goodfellow, YoshuaBengio, Aaron Courville.
2. Sutton, R. S., &Barto, A. G. (1998). Introduction to reinforcement learning (Vol. 135). Cambridge: MIT press.
3. Szepesvári, C. (2010). Algorithms for reinforcement learning. Synthesis lectures on artificial intelligence and machine learning, 4(1), 1-103.
4.

ADDITIONAL REFERENCES

Alpaydin, E. (2010). Introduction to machine learning/EthemAlpaydin.

Semester: III	Course Code: FDS-CC-532	Credits: 3

ADVANCED GRAPH AND NETWORK ANALYSIS

Course Outcomes : On completion of the course the student will be able to

CO	CO Statement	Mapping of COs		
		PSO	CL	KC
CO1	Articulate the basic knowledge in graph theory	PSO4, 5	U	C
CO2	Interpret network metrics and properties		U	C
CO3	Analyse graph algorithms		AN	C,P
CO4	Analyse graph visualization algorithms		AN	C
CO5	Analyse graph models		AN	C
CO6	Carry out network analysis		AP	P

CL- Cognitive Level: R-remember, U-understand, AP- Apply, AN- analyses, E- evaluate, CR- create, KC- Knowledge Category: F-Factual, C- Conceptual, P-Procedural, M- Metacognitive)

Assessment Pattern (Internal & External)

Bloom's Category	Continuous Assessment Tests (percentage)			Terminal Examination (percentage)
	1	2	3	
Remember	20	20	20	20
Understand	30	30	30	30
Apply	30	30	30	30
Analyse	20	20	20	20
Evaluate				
Create				

COURSE CONTENT

MODULE I

Quick review of graph theory concepts, Random networks, Scale-free networks, Small world, Preferential attachment, fitness, resilience against random attacks

MODULE II

Network metrics and properties: degree, clustering coefficient, diameter, density, shortest paths, centralities, communities, influence detection techniques and measures of user influence

MODULE III

Graph Algorithms: Shortest Paths, Minimal Spanning Trees, Graphs Searching, BFS, DFS algorithms, Adjacency Matrix, Eigenvalues and Graph Spectra, Computing Eigenvalues and Eigenvectors

MODULE IV

Network visualization: graph formats; graph drawing; graph layout methods, Community detection, The Nature of Communities, Overlapping Communities, Clique, Clique percolation, Clique percolation algorithms

MODULE V

Graph models: Erdos-Renyi random models (for comparison); small-world model; models of scale free networks, preferential attachment, and Barabasi-Alberto model (other models as science advances).

MODULE VI

Citation Networks-Detailed analysis of citation networks

REFERENCES

1. Van Steen, M. (2010). Graph theory and complex networks. An introduction, 144.
2. Newman, M. (2018). Networks. Oxford university press.
3. Pascual, M., & Dunne, J. A. (Eds.). (2006). Ecological networks: linking structure to dynamics in food webs. Oxford University Press.
4. Newman, M., Barabasi, A. L., & Watts, D. J. (2011). The structure and dynamics of networks (Vol. 19). Princeton University Press.
5. Lovász, L. (2012). Large networks and graph limits (Vol. 60). American Mathematical Soc..
6. http://www.cs.cornell.edu/~lwang/Wang10TAMC.pdf?attredirects=0
7. http://www.cs.usyd.edu.au/~visual/valacon/

Semester: III	Course Code: FDS-CC-533	Credits: 3

LAB 3 – COMPLEX NETWORK ANALYTICS

Course Outcomes : On completion of the course the student will be able to

CO	CO Statement	Mapping of COs		
		PSO	CL	KC
CO1	Carryout basic transformation and visualization	PSO4, 5	U	C
CO2	Understand network data and representations		U	C
CO3	Execute graph algorithms		AN	C,P
CO4	Analyse graph visualization algorithms		AN	C
CO5	Analyse bipartite networks		AN	C
CO6	Analyse real-world networks		AN	P

Assessment Pattern (Internal & External)

Bloom's Category	Continuous Assessment Tests (percentage)			Terminal Examination (percentage)
	1	2	3	
Remember	20	20	20	20
Understand	30	30	30	30
Apply	30	30	30	30
Analyse	20	20	20	20
Evaluate				
Create				

CL- Cognitive Level: R-remember, U-understand, AP- Apply, AN- analyses, E- evaluate, CR- create,
KC- Knowledge Category: F-Factual, C- Conceptual, P-Procedural, M- Metacognitive)

COURSE CONTENT

MODULE I

Introduction to basic transformation and visualization tools- loading, manipulating, visualizing and saving network data.

MODULE II

Network data and representations, Adjacency matrix and properties, Weighted, directed, undirected networks.

MODULE III

Graph Algorithms: Shortest Paths, Minimal Spanning Trees, Graphs Searching, BFS, DFS algorithms

MODULE IV

Measures of centrality, PageRank, Hubs and Authorities, clustering coefficient, diameter, density, cores, cliques

MODULE V

Affiliations-Bipartite networks analysis, Clustering, Community detection, propagation models

MODULE VI

Analysis of real-world networks

REFERENCES

1. Mej, N. (2010). Networks: an introduction.
2. De Nooy, W., Mrvar, A., &Batagelj, V. (2018). Exploratory social network analysis with Pajek. Cambridge University Press.
3. Prell, C. (2012). Social network analysis: History, theory and methodology. Sage.
4. http://snap.stanford.edu/
5. http://mrvar.fdv.uni-lj.si/pajek/
6. https://gephi.org/
7. https://sci2.cns.iu.edu/user/index.php

Semester: III	Course Code: FDS-CC-534	Credits: 4

DISSERTATION (STAGE I)

Course Outcomes : On completion of the course the student will be able to

CO	CO Statement	Mapping of COs		
		PSO	CL	KC
CO1	Analyze of the existing literature and identifying knowledge gap	PS03 PSO6 PSO6	AN	F, C
CO2	Formulate appropriate methodology to address the issues		AP	F,C,P
CO3	Develop suitable solutions / suggestions to possible improvements		CR	C, P

CL- Cognitive Level: R-remember, U-understand, AP- Apply, AN- analyses, E- evaluate, CR- create,
KC- Knowledge Category: F-Factual, C- Conceptual, P-Procedural, M- Metacognitive)

COURSE CONTENT

Assign the student to develop a research plan and schedule for the semester/session and use this plan as the basis for assignments and assessment of the student's performance.

An exhaustive review of literature is to be done and place the problem suitably in the overall realm of research arena so that the exact gap identified. The student should have a clear idea of the objectives, tools and methodology for the problem at hand.

REFERENCES

Necessary literature relevant to the chosen Research Problem as suggested by the advisor

Semester: III | Course Code: FDS-DE-535(i) | Credits: 3

FRAUD ANALYTICS

Course Outcomes : On completion of the course the student will be able to

CO	CO Statement	Mapping of COs		
		PSO	CL	KC
CO1	Articulate the basic knowledge on analysis of fraud	PSO 4, 5	U	C
CO2	Articulate fraud detection models		U	C
CO3	Develop automation process of fraud detention		AP	P
CO4	Compare different models		AN	C
CO5	Formulate and evaluate fraud detection		CR	P

CL- Cognitive Level: R-remember, U-understand, AP- Apply, AN- analyses, E- evaluate, CR- create,
KC- Knowledge Category: F-Factual, C- Conceptual, P-Procedural, M- Metacognitive)

Assessment Pattern (Internal & External)

Bloom's Category	Continuous Assessment Tests (percentage)			Terminal Examination (percentage)
	1	2	3	
Remember	10	10	10	10
Understand	10	10	10	10
Apply	20	20	20	20
Analyse	20	20	20	20
Evaluate	20	20	20	20
Create	20	20	20	20

COURSE CONTENT

MODULE I

Formulation and evaluation of fraud detection - Fraud detection using data analysis

MODULE II

Obtain and cleanse the data for fraud detection-preprocess data for fraud detection - sampling, missing values, outliers, categorisation etc.- Explain characteristics and components of the data and assess its completeness

MODULE III

Identify known fraud symptoms -and use digital analysis to identify unknown fraud symptoms—Fraud detection models using supervised analytics (logistic regression, decision trees, neural networks, ensemble models, etc.

MODULE IV

Automating fraud detection process -Fraud detection models using unsupervised analytics (hierarchical clustering, non-hierarchical clustering, k-means, self organizing maps, etc.

MODULE V

Fraud detection models using social network analytics (homophily, featurization, egonets, PageRank, bigraphs etc. -Verification of results and understand how to prosecute fraud.

MODULE VI

Fraud detection and prevention-case studies

REFERENCES

Nigrini, M. J. (2011). Forensic analytics: methods and techniques for forensic accounting investigations (Vol. 558). John Wiley & Sons.

ADDITIONAL REFERENCES

Baesens, B., Van Vlasselaer, V., & Verbeke, W. (2015). Fraud analytics using descriptive, predictive, and social network techniques: a guide to data science for fraud detection. John Wiley & Sons.

Semester: III | Course Code: FDS-DE-535(ii) | Credits: 3

WEB SCRAPING AND ANALYTICS

Course Outcomes : On completion of the course the student will be able to

CO	CO Statement	Mapping of COs		
		PSO	CL	KC
CO1	Articulate and exemplify the basic knowledge on web scraping	PSO 4, 5	U	C
CO2	Develop Python codes for we scraping		U	C
CO3	Apply available APIs for collecting data		AP	P
CO4	Differentiate type of web data		AN	C
CO5	Implement web scraping using Python and Java		AP	P

CL- Cognitive Level: R-remember, U-understand, AP- Apply, AN- analyses, E- evaluate, CR- create,
KC- Knowledge Category: F-Factual, C- Conceptual, P-Procedural, M- Metacognitive)

Assessment Pattern (Internal & External)

Bloom's Category	Continuous Assessment Tests (percentage)			Terminal Examination (percentage)
	1	2	3	
Remember	10	10	10	10
Understand	10	10	10	10
Apply	20	20	20	20
Analyse	20	20	20	20
Evaluate	20	20	20	20
Create	20	20	20	20

COURSE CONTENT

MODULE I

Introduction to Python II (parsing&using libraries) Using Python to scrape the web I (regex & other-libraries)

MODULE II

Using Python to scrape the web II (data cleaning) Text Mining with Python (nltk)

MODULE III

Regular Expressions- Extracting Data With Regular Expressions Networks and Sockets- protocol-sthat web browsers use to retrieve documents and web applications use to interactwith Application Program Interfaces (APIs)

MODULE IV

Use Python to retrieve data from web sites and APIs over the Internet.- Reading Web
Data From Python - Scraping HTML Data How to retrieve and parse XML (eXtensible Markup Lan-guage) data- eXtensible Markup Language - Extracting Data from XML

MODULE V

API's / Web Services using the JavaScript Object Notation (JSON) data format

MODULE VI

Practical case studies of web scrapping

REFERENCES

1. Severance, C. R., Blumenberg, S., & Hauser, E. (2016). Python for Everybody:Exploring Data in Python 3. CreateSpace Independent Publishing Platform.
2. Munzert, S., Rubba, C., Meißner, P., &Nyhuis, D. (2014). Automated data collection withR: A practical guide to web scraping and textmining. John Wiley& Sons.

Semester: III Course Code: FDS-DE-535(iii) Credits: 3

INTERNET OF THINGS IN THE CLOUD

Course Outcomes : On completion of the course the student will be able to

CO	CO Statement	Mapping of COs		
		PSO	CL	KC
CO1	Articulate and exemplify the basic knowledge on text analytics	PSO 4	U	C
CO2	Understand the sources and limitations of text and social media data.		U	C
CO3	understand the structural, syntactical, semantic elements of textual data.		U	C
CO4	understand the structural and social aspects of social media networks.		U	C
CO5	Become familiar with core practice communities, publications, and organizations focusing on text and social media analytics and the research questions they are engaged in.		U	C
CO6	use common text mining and social media analytics tools to gather managerial insights.		AP	P
CO7	develop and present network visualization for social media data.		CR	P

CL- Cognitive Level: R-remember, U-understand, AP- Apply, AN- analyses, E- evaluate, CR- create,
KC- Knowledge Category: F-Factual, C- Conceptual, P-Procedural, M- Metacognitive)

Assessment Pattern (Internal & External)

Bloom's Category	Continuous Assessment Tests (percentage)			Terminal Examination (percentage)
	1	2	3	
Remember	10	10	10	10
Understand	10	10	10	10
Apply	20	20	20	20
Analyse	20	20	20	20
Evaluate	20	20	20	20
Create	20	20	20	20

COURSE CONTENT

MODULE I

Introduction to Python II (parsing&using libraries) Using Python to scrape the web I (regex & other libraries)

MODULE II

Using Python to scrape the web II (data cleaning) Text Mining with Python (nltk)

MODULE III

Regular Expressions- Extracting Data With Regular Expressions Networks and Sockets- protocol-sthat web browsers use to retrieve documents and web applications use to interactwith Application Program Interfaces (APIs)

MODULE IV

Use Python to retrieve data from web sites and APIs over the Internet.- Reading Web Data From Python - Scraping HTML Data How to retrieve and parse XML (eXtensible Markup Language) data- eXtensible Markup Language - Extracting Data from XML

MODULE V

API's / Web Services using the JavaScript Object Notation (JSON) data format

MODULE VI

Practical case studies of web scrapping

REFERENCES

1. Severance, C. R., Blumenberg, S., & Hauser, E. (2016). Python for Everybody:Exploring
2. Data in Python 3. CreateSpace Independent Publishing Platform.
3. Munzert, S., Rubba, C., Meißner, P., &Nyhuis, D. (2014). Automated data collection withR: A practical guide to web scraping and textmining. John Wiley& Sons.

ADDITIONAL REFERENCES

1. Hersent, O., Boswarthick, D., &Elloumi, O. (2012). The internet of things: Applications to the Smart Grid and Building. John Wiley & Sons.
2. Hersent, O., Boswarthick, D., &Elloumi, O. (2011). The internet of things: Key applications and protocols. John Wiley & Sons

Semester: III | Course Code: FDS-DE-536(i) | Credits: 3

ARTIFICIAL INTELLIGENCE

Course Outcomes : On completion of the course the student will be able to

CO	CO Statement	Mapping of COs		
		PSO	CL	KC
CO1	Articulate and exemplify the basic knowledge artificial intelligence		U	C
CO2	Understand the basics of knowledge representation		U	C
CO3	Use AI programming languages		AP	P
CO4	Use the methods of AI implementation		AP	P
CO5	Recommend AI strategies based on applications		E	p

CL- Cognitive Level: R-remember, U-understand, AP- Apply, AN- analyses, E- evaluate, CR- create,
KC- Knowledge Category: F-Factual, C- Conceptual, P-Procedural, M- Metacognitive)

Assessment Pattern (Internal & External)

Bloom's Category	Continuous Assessment Tests (percentage)			Terminal Examination (percentage)
	1	2	3	
Remember	20	20	20	20
Understand	30	30	30	30
Apply	40	40	40	40
Analyse	10	10	10	10
Evaluate				
Create				

COURSE CONTENT

MODULE I

Introduction:Artificial Intelligence, AI Problems, AI Techniques, The Level of the Model, Criteria For Success. Defining the Problem as a State Space Search, Problem Characteristics,Production Systems, Search: Issues in The Design of Search Programs, Un-Informed Search, BFS, DFS; Heuristic Search Techniques: Generate-And-Test, Hill Climbing, Best-First Search, A*Algorithm, Problem Reduction, AO*Algorithm, Constraint Satisfaction, Means-Ends Analysis.

MODULE II

Knowledge Representation:Procedural Vs Declarative Knowledge, Representations & Approaches to Knowledge Representation, Forward Vs Backward Reasoning, Matching Techniques, Partial Matching, Fuzzy Matching Algorithms and RETE Matching Algorithms; Logic Based Programming-AI Programming languages: Overview of LISP, Search Strategies in LISP, Pattern matching in LISP , An Expert system Shell in LISP, Over view of Prolog, Production System using Prolog

MODULE III

Symbolic Logic:Propositional Logic, First Order Predicate Logic: RepresentingInstance and is-a Relationships, Computable Functions and Predicates, Syntax & Semantics of FOPL, Normal Forms, Unification &Resolution, Representation Using Rules, Natural Deduction; Structured Representa-

tions of Knowledge: Semantic Nets, Partitioned Semantic Nets, Frames, Conceptual Dependency, Conceptual Graphs, Scripts, CYC;.

MODULE IV

Reasoning under Uncertainty: Introduction to Non-Monotonic Reasoning, Truth Maintenance Systems, Logics for Non-Monotonic Reasoning, Model and Temporal Logics; Statistical Reasoning: Bayes Theorem, Certainty Factors and Rule-Based Systems, Bayesian Probabilistic Inference, Bayesian Networks, Dempster-Shafer Theory, Fuzzy Logic: Crisp Sets ,Fuzzy Sets, Fuzzy Logic Control, Fuzzy Inferences & Fuzzy Systems.

MODULE V

Experts Systems:Overview of an Expert System, Structure of an Expert Systems, Different Types of Expert Systems-Rule Based, Model Based, Case Based and Hybrid Expert Systems, Knowledge Acquisition and Validation Techniques, Black Board Architecture, Knowledge Building System Tools, Expert System Shells, Fuzzy Expert systems.

MODULE VI

Machine Learning:Knowledge and Learning, Learning by Advise, Examples, Learning in problem Solving, Symbol Based Learning, Explanation Based Learning, Version Space, ID3 Decision Based Induction Algorithm, Unsupervised Learning, Reinforcement Learning, Supervised Learning: Perceptron Learning, Back propagation Learning, Competitive Learning, Hebbian Learning.

REFERENCES

1. Artificial Intelligence, George F Luger, Pearson Education Publications
2. Artificial Intelligence, Elaine Rich and Knight, Mcgraw-Hill Publications
3. Introduction to Artificial Intelligence & Expert Systems, Patterson, PHI
4. Multi Agent systems-a modern approach to Distributed Artificial intelligence, Weiss.G, MIT Press.
5. Artificial Intelligence : A modern Approach, Russell and Norvig, Printice Hall

Semester: III | Course Code: FDS-DE-536(ii) | Credits: 3

LARGE SCALE OPTIMIZATION FOR DATA ANALYTICS

Course Outcomes : On completion of the course the student will be able to

CO	CO Statement	Mapping of COs		
		PSO	CL	KC
CO1	Articulate and exemplify the basic knowledge optimization theory	PSO 5	U	C
CO2	Discriminate various optimization methods		E	C
CO3	Compare different optimization algorithms		E	P
CO4	Access the issue related to scaling up		E	C

CL- Cognitive Level: R-remember, U-understand, AP- Apply, AN- analyses, E- evaluate, CR- create,
KC- Knowledge Category: F-Factual, C- Conceptual, P-Procedural, M- Metacognitive)

Assessment Pattern (Internal & External)

Bloom's Category	Continuous Assessment Tests (percentage)			Terminal Examination (percentage)
	1	2	3	
Remember	20	20	20	20
Understand	30	30	30	30
Apply	40	40	40	40
Analyse	10	10	10	10
Evaluate				
Create				

COURSE CONTENT

MODULE I

Unconstrained nonlinear optimization theory and algorithms-Linear and nonlinear regression, logistic regression-Numerical solution of linear systems-Stochastic gradient descent - Deep neural networks

MODULE II

Introduction to constrained nonlinear optimization theory-Quadratic programs (example: support vector machines)

MODULE III

Gradient methods-Frank-Wolfe and Projected gradient methods Subgradient methods-Proximal gradient methods-Nesterov's accelerated methods-Mirror descent method

MODULE IV

Large-scale numerical linear algebra (Conjugate gradient, Power methods, Lanczos methods)-Douglas-Rachford splitting and Alternating direction methods of multiplier (ADMM)- Primal-dual proximal methods- Quasi-Newton methods / BFGS-Stochastic gradient descent (SGD)

MODULE V

Global geometry: saddle point characterization-Escaping saddle points-Solving quadratic systems of equations

MODULE VI

Low-rank matrix recovery and completion-Implicit regularization-Neural networks

REFERENCES

1. Bertsekas, D. P., & Scientific, A. (2015). Convex optimization algorithms. Belmont: Athena Scientific.
2. Bubeck, S. (2015). Convex optimization: Algorithms and complexity. Foundations and Trends® in Machine Learning, 8(3-4), 231-357.
3. Beck, A. (2017). First-Order Methods in Optimization (Vol. 25). SIAM.

ADDITIONAL REFERENCES

1. Trevor Hastie and Robert Tibshirani(2014). An Introduction to Statistical Learning (with Applications in R), Gareth James, Daniela Witten, , Springer, 2013, ISBN 978-1-4614-7137-0.
2. Rardin, R. L. (2001). Optimization in Operations Research, ISBN-13: 978-0-13-438455-9

Semester: III | Course Code: FDS-DE-536(iii) | Credits: 3

MODELS OF COMPUTATIONS

Course Outcomes :On completion of the course the student will be able to

CO	CO Statement	Mapping of COs		
		PSO	CL	KC
CO1	Articulate and exemplify the basic knowledge on computation models	PSO 4, 5	U	C
CO2	Articulate and exemplify the basic knowledge of quantum computing		U	C
CO3	Articulate and exemplify the basic knowledge nature inspired algorithms		U	C
CO4	Articulate and exemplify the basic knowledge social computing		U	C
CO5	Articulate and exemplify the basic knowledge evolutionary algorithms		E	P
CO6	Access various approach of computations			

CL- Cognitive Level: R-remember, U-understand, AP- Apply, AN- analyses, E- evaluate, CR- create,
KC- Knowledge Category: F-Factual, C- Conceptual, P-Procedural, M- Metacognitive)

Assessment Pattern (Internal & External)

Bloom's Category	Continuous Assessment Tests (percentage)			Terminal Examination (percentage)
	1	2	3	
Remember	20	20	20	20
Understand	30	30	30	30
Apply	30	30	30	30
Analyse	10	10	10	10
Evaluate	10	10	10	10
Create				

COURSE CONTENT

MODULE I

TURING MACHINE MODEL : Turing Machine Logic, Proof, Computability

MODULE II

QUANTUM COMPUTATION : Quantum Computing History, Postulates of Quantum Theory, Dirac Notation, the Quantum Circuit Model, Simple Quantum Protocols: Teleportation, Superdense Coding, Foundation Algorithms

MODULE III

Nature-Inspired Computing Optimization and Decision Support Techniques, Evolutionary Algorithms, Swarm Intelligence, Benchmarks and Testing

MODULE IV

Social Computing Online communities, Online discussions, Twitter, Social Networking Systems,

Web 2.0, social media, Crowdsourcing, Facebook, blogs, wikis, social recommendations, Collective intelligence

MODULE V

Evolutionary Computing Introduction to Genetic Algorithms, Genetic Operators and Parameters, Genetic Algorithms in Problem Solving, Theoretical Foundations of Genetic Algorithms, Implementation Issues

MODULE VI

Case study of different models of computations

REFERENCES

1. Boyd, D. (2014). It's complicated: The social lives of networked teens. Yale University Press.
2. Fleck, M. M. (2013). Building Blocks for Theoretical Computer Science (Version 1.3).

ADDITIONAL REFERENCES

1. Nielsen, M. A., & Chuang, I. L. (2014). Quantum computation and quantum information.
2. Mitchell, M. (1998). An introduction to genetic algorithms. MIT press.

Semester: IV	Course Code: FDS-DE-541	Credits: 16

DISSERTATION (STAGE II)

AIM: The student will continue the development (with the advisor's guidance) of the research problem selected in the Dissertation (Stage I) and complete the work using appropriate tools and methods, algorithms, experiment etc.

COURSE DESCRIPTION

- To discover and pursue a unique topic of research in order to construct new knowledge
- To design and conduct an original research project.
- To develop skills in designing a discipline specific research methodology.
- To develop a working knowledge of relevant literature in the discipline
- To practice scientific writing and learn how to participate in the peer review process
- To be able to discuss research and other topics with academics in your field

COURSE CONTENT

Assign the student to develop the research plan and question as a continuation of Dissertation (stage I) and schedule for the semester/session and use this plan as the basis for assignments and assessment of the student's performance.

The Dissertation (Stage II) must contain the detailed procedures for data collection/ survey/methods, theory and tools to be developed. The student should present the results/output and analysis of the study before finalizing the report. The final report is to be prepared by incorporating the suggestions after the presentations.

REFERENCES

Necessary literature relevant to the chosen Research Problem as suggested by the advisor

Semester: I | Course Code: FDS-501-501 | Credits: 2

FORESIGHT AND FUTURES RESEARCH

Course Outcomes : On completion of the course the student will be able to

CO	CO Statement	Mapping of COs		
		PSO	CL	KC
CO1	Understand the critical concepts in futures studies	PSO2, 4, 5	U	F, C
CO2	Understand Basic Concepts and six Basic Pillars of Futures Studies	PSO2,5,6	U	F,C,P
CO3	Apply methods of foresight and futures research	PSO2, 6	AP	C, P
CO4	Evaluate various aspects of scenario building	PSO4,6	E	C, P
CO5	Evaluate the implications of scenario building and Opportunity Analysis	PSO2,4,6	E	C, P
CO6	Evaluate suitability of foresight methods for given situation	PSO2,4,6	E	C, P

CL- Cognitive Level: R-remember, U-understand, AP- Apply, AN- analyses, E- evaluate, CR- create,
KC- Knowledge Category: F-Factual, C- Conceptual, P-Procedural, M- Metacognitive)

Assessment Pattern (Internal & External)

Bloom's Category	Continuous Assessment Tests (percentage)			Terminal Examination (percentage)
	1	2	3	
Remember	10	10	10	10
Understand	30	20	20	20
Apply	40	40	40	40
Analyse	10	10	10	10
Evaluate	10	10	10	10
Create				

COURSE CONTENT

MODULE I

Review futures basics- Become acquainted with key concepts, terms, and perspectives of the futures field.- Introduce some of the publications and organizations in the field and discuss futures skills- Identify the major events and founders in the history of future studies - Discuss what it means to be a futurist- Role of a futurist

MODULE II

Six Basic Concepts and six Basic Pillars of Futures Studies- Prominent Futures Schools

MODULE III

Environmental Monitoring, Scanning-Identify and monitor key quantities and conditions that indicate the baseline occurring -Environmental scanning as the means to keep current on change within a specific domain- Practice environmental scanning and assess how well each scanning hit fits the ideal criteria -Use scanning hits as the empirical support for alternative plausible futures

MODULE IV

Creativity- Scenarios -The key concepts, terms, and approaches to creativity- A standard approach to creative problem solving and a variety of other Approaches - Scenario Theory- Key concepts, terms, and criteria of good scenarios-The use of scenarios in professional practice -A review of different ways that people can build scenarios Strengths and weaknesses of various scenario approaches

MODULE V

Implications and Opportunity Analysis - Drawing out the implications of scenarios using, brainstorming, futures wheels, expert panels and judgment- Identify challenges of specific scenarios for specific domains- Use those challenges to prioritize strategic issues that should be addressed- See change as resulting from trends, events, issues and images- Pick up the weak signals of coming change through environmental scanning - Setting up a good scanning system- Progress/Decline: That social change is generally an improvement of the human condition; or alternatively, that societies have descended from a better period to today. Development: That social change moves in a definite direction, but that the direction is neutral—not necessarily any better or any worse than the past, just different.

MODULE VI

Case Studies on Scenarios and other Futures Techniques

REFERENCES

1. Bill Ralston & Ian Wilson, The Scenario Planning Handbook
2. Jerry Glenn & Ted Gordon (eds.), Futures Research Methodology V3.0, CD, Washington DC, Millennium Project, 2012.
3. Pero Micic, The Five Futures Glasses, 2010
4. Thomas Chermack, Scenario Planning in Organizations: How to Create, Use, and Assess Scenarios, 2011
5. Trevor Noble (2000), Social Theory and Social Change, New York: St Martin’s Press.

ONLINE SOURCES

1. https://en.wikiversity.org/wiki/Introduction_to_Futures_Studies
2. http://www.csudh.edu/global_options/IntroFS.HTML
3. https://cals.arizona.edu/futures/

Semester: III	Course Code: FDS-GC-502	Credits: 2

PARALLEL PROGRAMMING WITH MPI

Course Outcomes : On completion of the course the student will be able to

CO	CO Statement	Mapping of COs		
		PSO	CL	KC
CO1	Understand the concepts of parallel programming	PSO3,4	U	F,C
CO2	Understand different types of Parallel Systems	PSO5	U	F,P
CO3	Create parallel programs using MPI	PSO4,5	CR	F,C,P
CO4	Rewrite Sequential Program to Parallel Program	PSO5,6,7	CR	F,C,P

CL- Cognitive Level: R-remember, U-understand, AP- Apply, AN- analyses, E- evaluate, CR- create,
KC- Knowledge Category: F-Factual, C- Conceptual, P-Procedural, M- Metacognitive)

Assessment Pattern (Internal & External)

Bloom's Category	Continuous Assessment Tests (percentage)			Terminal Examination (percentage)
	1	2	3	
Remember	10	10	10	10
Understand	10	10	10	10
Apply	40	40	40	40
Analyse	10	10	10	10
Evaluate	10	10	10	10
Create	20	20	20	20

COURSE CONTENT

MODULE I

Introduction to Parallel Processing -What is Parallel Processing ?-The Goals of Parallel Processing -Pros and Cons of Parallel Processing -Sequential Limits -Why Parallel Processing -Simplified Examples -Applications -History of Supercomputing

MODULE II

Types of Parallel Systems-SISD - Single Instruction stream over a Single Data Stream Single Instruction stream over a Single Data Stream -MISD - Multiple Instruction stream over a Single Data stream -SIMD - Single Instruction, Multiple Data Stream MIMD - Multiple Instruction, Multiple Data Stream

MODULE III

What is MPI -Basic Idea of MPI -When Use MPI -Getting started with LAM -MPI Commands -MPI Environment -MPI Functions Specifications -MPI Datatypes

MODULE IV

Parallel Programming using MPI -Communication Strategies-Point to Point Communication –Collective Communication -Performance Evaluation

MODULE V

Demonstration and Writing of Simple MPI Programs

MODULE VI

Rewrite Sequential Program to Parallel Program, Strategies to rewrite Sequential Program

REFERENCES

1. Ananth Grama, Anshul Gupta, George Karypis, Vipin Kumar, Introduction to Parallel Computing. Second Edition. Addison-Wesley, 2003 (ISBN 0 201 64865 2).
2. Gropp, William, Ewing Lusk, and Anthony Skjellum. Using MPI: Portable Parallel Programming with the Message Passing Interface. 2nd ed. Cambridge, MA: MIT Press, 1999. ISBN: 9780262571326.
3. Gropp, William, Steven Huss-Lederman, Andrew Lumsdaine, Ewing Lusk, Bill Nitzberg, William Saphir, and Marc Snir. MPI: The Complete Reference (Vol. 2) - The MPI-2 Extensions. 2nd ed. Cambridge, MA: MIT Press, 1998. ISBN: 9780262571234.
4. Selim G . Akl, The Design and Analysis of Parallel Algorithms, Prentice-Hall, Inc. 1989.
5. Snir, Marc, and William Gropp. MPI: The Complete Reference (2-volume set). 2nd ed. Cambridge, MA: MIT Press, 1998. ISBN: 9780262692168.
6. Snir, Marc, and William Gropp. Using MPI-2: Advanced Features of the Message Passing Interface. Cambridge, MA: MIT Press, 1999. ISBN: 9780262571333.

ADDITIONAL REFERENCES

1. http://openmp.org/wp/
2. http://www.mpi-forum.org/
3. https://computing.llnl.gov/tutorials/mpi/
4. https://computing.llnl.gov/tutorials/parallel_comp/

Semester: Any	Course Code: FDS-GC-503	Credits: 2

SCIENTIFIC RESEARCH PAPER WRITING

Course Outcomes : On completion of the course the student will be able to

CO	CO Statement	Mapping of COs		
		PSO	CL	KC
CO1	Distinguish different types of research, their audiences and how research material might be effectively presented	PSO3	U	F,C
CO2	Prepare scientific and technical papers, and presentations.	PSO3	CR	F,P
CO3	Format documents and presentations to optimize their visual appeal when viewed in-press, as a podcast or audio/video file format on the internet, or through personal presentations to an audience	PSO3	CR	F,C,P
CO4	Effectively use LaTeX for professional documents	PSO3	CR	F,C,P
CO5	Accept constructive criticism and use reviewers' comments to improve quality and clarity of written reports and presentations.	PSO3	U	F,C

CL- Cognitive Level: R-remember, U-understand, AP- Apply, AN- analyses, E- evaluate, CR- create,
KC- Knowledge Category: F-Factual, C- Conceptual, P-Procedural, M- Metacognitive)

Assessment Pattern (Internal & External)

Bloom's Category	Continuous Assessment Tests (percentage)			Terminal Examination (percentage)
	1	2	3	
Remember	10	10	10	10
Understand	10	10	10	10
Apply	40	40	40	40
Analyse	10	10	10	10
Evaluate	10	10	10	10
Create	20	20	20	20

COURSE CONTENT

MODULE I

Introduction into Research, Types of scientific communication with examples, Scientific Literature, Searching the scientific literature, Plagiarism and how to avoid it

MODULE II

Beginning to write - Establishing your constraints, Organizing your writing, Preparing outlines, Standard formats for scientific papers, research projects and theses, Style guides. Content - Creating a literature review, Preparing other sections of a research report (abstract, introduction, materials and methods, results and discussion, conclusions), Including and summarizing research data.

MODULE III

Style and grammar - Scientific writing style, Passive vs. active voice Avoiding excessive wording, Grammar, Avoiding misuse of words, When to use footnotes. Reference citations - How to use references - Within the text - How to make lists of references, Citation Network Analysis.

MODULE IV

Revising - Dealing with revisions, Accepting criticism, Making sense of reviewers' comments, Making the changes, What to do if you don't agree with reviewers' comments. Other communication - Other types of scientific writing like research proposals, creating a fact sheet/bulletin, articles for popular press, memos, letters and emails.

MODULE V

Computer skills - LaTex.

MODULE VI

Presentations - Organization and formats for posters, Oral Presentations, Designing and preparing slides for an oral presentation.

REFERENCES

1. How to Write and Publish a Scientific Paper. 6 Edition. Authors: Robert A. Day and Barbara Gastel. ISBN: 0-313-33040-9
2. Alley, M. 2003. The Craft of Scientific Presentations: Critical steps to succeed and critical errors to avoid. Springer, NY. 241 pages. ISBN:0-387-95555-0.
3. Lamport, Leslie. LATEX: a document preparation system: user's guide and reference manual. Addison-wesley, 1994.

ADDITIONAL REFERENCES

1. https://www.sharelatex.com/