தமிழ்நாடு மத்தியப் பல்கலைக்கழகம்

CENTRAL UNIVERSITY OF TAMIL NADU

तमिलनाडु केन्द्रीय विश्वविद्यालय

ESTABLISHED BY AN ACT OF PARLIAMENT IN 2009

SCHOOL OF MATHEMATICS AND COMPUTER SCIENCES

DEPARTMENT OF COMPUTER SCIENCE

PG DIPLOMA IN DATA SCIENCE

CURRICULUM AND SYLLABUS

(Academic Year 2023-24 onwards)

1. Overview

Central University of Tamil Nadu (CUTN) is an institution of higher education established by an Act of Parliament in 2009. The University aims at the forming of an enlightened society founded on a relentless pursuit of excellence through innovation in teaching-learning process, interdisciplinary studies and research. With a host of diversified integrated postgraduate, postgraduate and doctoral programmes, the University has been keen to stay abreast of the ever expanding academic frontiers. Within a brief period CUTN has established itself as a frontrunner in the pursuit of excellence.

1.1 School of Mathematics and Computer Sciences

The School of Mathematics and Computer Sciences aims to be in the forefront of teaching, research, and extension activities. This School presently comprises the department of Mathematics, Computer Science and Statistics & Applied Mathematics.

1.2 Department of Computer Science

The department of computer science is one among the science departments that offer M.Sc. and Ph.D. in computer science. The department was started in the academic year 2016-17. The department is equipped with five faculty members with doctorates and they are actively involved in research besides teaching.

The curriculum is designed to meet the standards expected from the IT industry and also facilitates to aspire research. The department offers a choice based credit system. The students are evaluated through a continuous assessment system. The department is equipped with a smart class room and two computer labs with 50 systems that caters the needs of M.Sc. students

VISION

To develop the department as a global leader in knowledge dissemination and to perform cutting edge research in computer science in compliance with international standards

MISSION

M1	To produce students by inculcating them with state-of-the-art education to make them technically competent in solving real time problems
M2	To kindle students as efficient IT professionals, researchers and entrepreneurs
M3	To enrich the students as social, ethical and value-based graduates to make significant contribution to the society

PROGRAM SPECIFIC OBJECTIVES

After five years of successful completion of the program, the students will be able to

PSO1	Ability to apply good analytical, design and implementation skills to formulate and solve scientific and business applications by promoting themselves with equipped ideas
PSO2	Equip students with profound knowledge and understanding in the theoretical foundations of computer science
PSO3	To mould themselves into successful personnel in IT industry / higher education and grow like entrepreneurs

PSO to MISSION Mapping

	M1	M2	M3
PSO1	3	2	1
PSO2	2	3	1
PSO3	3	3	3

Graduate Attributes of Post Graduate Diploma in Data Science

1. **Disciplinary Knowledge:** Content and pedagogical knowledge synchronised with the curriculum frameworks and policies
2. **Communication Skills:** Possess clarity in conveying the ideas
3. **Critical Thinking:** Capacity to apply analytical thought in the teaching and learning process
4. **Problem Solving:** Participate in the educational problem solving and applying the knowledge in the day-to-day professional endeavours.
5. **Cooperation:** Appreciate collaboration and cooperation among stakeholders of education.
6. **ICT Skills:** Selecting and integrating appropriate ICT skills for professional development.
7. **Ethics:** Doing what is right to society
8. **Self-Directed Learning:** Developing autonomy and self-regulation in teaching-learning and professional development.
9. **Reasoning:** Ability to interpret and draw the conclusion from qualitative/quantitative data with open-mindedness
10. **Creativity:** Ability to produce new ideas
11. **Societal and Environmental Concern:** Performing an act or solving a problem with respect to societal and environmental concern
12. **Lifelong Learning:** Understands the need for learning and practices it throughout life.

PROGRAM OBJECTIVES

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

PO1	**Core Knowledge:** Have ability to apply the knowledge of mathematics, probability and statistics and computer science in real-time applications
PO2	**Problem Analysis:** Possess clear understanding of the subject related concepts and contemporary issues and apply them to formulate, design, model, and analyze complex problems
PO3	**Design and Development of Product:** Acquire ability to design a component or a product that meet the specific needs with appropriate consideration for public health and safety, and the cultural, societal, and environmental considerations. Demonstrate knowledge and understanding of management skills related to project management
PO4	**Design and Conduct Experiments on Complex Problems:** Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions
PO5	**Modern Tool Usage:** Create, select, and apply appropriate techniques, resources, and modern IT tools including prediction and modelling to complex activities with an understanding of the limitations
PO6	**Individual and Teamwork:** Function effectively as an individual, and as a member or leader in diverse teams
PO7	**Life-long Learning:** Recognize the need for, and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change

PO to PSO Mapping

	PSO1	PSO2	PSO3
PO1	3	3	2
PO2	3	3	3
PO3	3	3	3
PO4	3	3	2
PO5	3	1	2
PO6	1	2	3
PO7	3	3	3

2. Regulations

2.1 Name of the programme	**Post Graduate Diploma in Data Science**
2.2 Duration of the programme	**One Year (Two Semesters)**
2.3 Specialisation	**Data Science**
2.4 Total intake	**9**
2.5 Eligibility Criteria: Bachelor's degree in Computer Applications / Computer Sciences/Mathematics/ or any Bachelor's degree in Science stream with Mathematics as one of the subjects for at least four semesters. Candidates should have secured a minimum of 60% marks or 6.5 CGPA (on a 10-point scale) in the qualifying degree examination for General Category, 55% marks or 6.0 CGPA (on a 10 point scale) for EWS/OBC (Non-creamy layer) and 50% aggregate marks or 5.5 CGPA (on a 10-point scale) for SC/ST candidates.	
2.6 Medium of Instruction	English
2.7 Attendance	As per the University rules
2.8 Scheme of Examination: Each theory course shall have 40 marks towards continuous internal assessment and 60 marks towards end-semester examination. Lab courses shall have only continuous internal assessment of 100 marks, Internship and Dissertation shall have only continuous internal assessment of 100 marks.	
2.9 Miscellaneous: All other matters not referred specifically in these regulations shall be governed as per the ordinances of the university as revised from time to time.	

PROGRAM STRUCTURE

No	Course Code	Course Title	Course Type	Credit	Hours / Week	Marks	
						Int.	Ext.
SEMESTER – I							
1	CSC4011	Discrete Mathematics	CC	3	3	40	60
2	CSC4012	Statistics for Data Science	CC	3	3	40	60
3	CSC4013	Big Data Analytics	CC	4	4	40	60
4	CSC4014	Machine Learning	CC	3	3	40	60
5	CSC4015	Introduction to Programming using Python	CC	3	3	40	60
6	CSC4016	Big Data Analytics Lab	CCP	2	4	100	
7	CSC4017	Machine Learning Lab	CCP	2	4	100	
8	CSC4018	Python Programming Lab	CCP	2	4	100	
TOTAL				**22**	**28**		
SEMESTER – II							
1	CSC4021	Principles of Deep Learning	CC	4	4	40	60
2	CSC4022	Data Visualization	CC	4	4	40	60
3	CSC4023	Deep Learning Lab	CCP	2	4	100	
4	CSC4024	Data Visualization Lab	CCP	2	4	100	
5	CSC4025	Project / Dissertation	Project	12	12	100	
7	CSCEC##	Elective – 1 / MOOCS	DSE	4	4	40	60
8	CSCVA##	Value Added Course	VAC	--	2		
TOTAL				**28**	**34**		

TOTAL CREDITS = 50

CC → Core Course	CCP → Core Course Practical	OE → Open Elective
DSE → Discipline Specific Elective	VAC → Value Added Course	

PG DIPLOMA IN DATA SCIENCE CREDIT DISTRIBUTION BASED ON COURSE TYPE			
S. NO.	**COURSE COMPONENTS / NAME OF THE COURSE**	**CREDITS**	**PERCENTAGE**
1	Core Courses (CC)	22	40.74
2	Core Courses Practical (CCP)	10	18.52
3	Discipline Specific Elective (DSE)	04	7.42
4	Open Elective (OE)	03	5.55
5	Internship / Field Visit / Training	03	5.55
6	Project	12	22.22
	Total	54	100.00
6	Value Added Course	02	

DISCIPLINE SPECIFIC ELECTIVES

COURSE CODE	COURSE TITLE	CREDITS
CSCEC04	Data Warehousing and Data Mining	4
CSCEC06	Natural Language Processing	4
CSCEC38	Web Analytics	4
CSCEC39	Multivariate Analysis	4
CSCEC40	Graphs Algorithms and Mining	4
CSCEC41	Optimization Techniques	4

SEMESTER – I

CSC4011	DISCRETE MATHEMATICS	3 Credits

a. Course Outcome (CO)

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

	Course Outcome	Level
CO1	Understand the concepts of Sets, relations, functions and Boolean algebra with applications	Apply
CO2	Introduce the concepts of mathematical logic	Understand
CO3	Describe mathematical induction and probability	Apply
CO4	Understand basic properties of graphs and related discrete structures, and be able to relate these to practical examples	Analyse
CO5	Describe the use of group theory and its applications	Skill

b. Syllabus

Unit – 1	SETS – RELATIONS – FUNCTIONS – BOOLEAN ALGEBRA	9 hrs

Set Operations – Representation and Properties of Relations – Equivalence Relations – Partially Ordering. Functions – one-one – onto bijective – composition of relations and functions – inverse functions. Boolean Algebra and Boolean Functions and its representations – Simplifications of Boolean Functions.

Unit – 2	MATHEMATICAL LOGIC	9 hrs

Propositional and Predicate Logic – Propositional Equivalences – Normal Forms – Predicates and Quantifiers – Nested Quantifiers – Rules of Inference.

Unit – 3	COUNTING – MATHEMATICAL INDUCTION AND DISCRETE PROBABILITY	9 hrs

Basics of Counting – Pigeonhole Principle – Permutations and Combinations – Inclusion-Exclusion Principle – Mathematical Induction – Probability – Bayes Theorem.

Unit – 4	GROUP THEORY	9 hrs

Groups – Subgroups – Semi Groups – Product and Quotients of Algebraic Structures – Isomorphism – Homomorphism – Automorphism – Rings - Integral Domains – Fields – Applications of Group Theory – Polya's theory of counting – Introduction to Error-Correcting Codes. Discrete Geometry: Some basic definitions – Ham-Sandwich theorem.

Unit – 5	GRAPH THEORY	9 hrs

Simple Graph – Multigraph – Weighted Graph – Paths and Circuits – Shortest Paths in Weighted Graphs – Eulerian Paths and Circuits – Hamiltonian Paths and Circuits – Planner graph – Graph Coloring –Bipartite Graphs.

Text / Reference Books:

1. Rosen, K.H. and Kamala Krithivasan, Discrete Mathematics and its Applications, McGraw Hill, Eighth edition, 2021
2. Rosen, K.H, Discrete Mathematics and Its Applications with Combinatorics and Graph Theory,McGraw Hill Education, Seventh edition, 2017
3. Tremblay, J.P. and Manohar.R, Discrete Mathematical Structures with Applications to ComputerScience, McGraw Hill Education, First edition, 2017
4. Grimaldi, R.P. , Discrete and Combinatorial Mathematics: An Applied Introduction, Pearson, Fifthedition, 2013
5. Lipschutz, S. and Mark Lipson, Discrete Mathematics (Schaum's Outlines), McGraw Hill Education,Revised Third edition, 2017

c. Mapping of Course Outcomes to Program Outcomes

	PO1	PO2	PO3	PO4	PO5	PO6	PO7
CO1	3	2	-	1	-	2	-
CO2	3	1	2	2	-	-	-
CO3	3	2	2	2	-	-	-
CO4	3	2	2	2	-	2	-
CO5	3	2	2	2	-	1	-

CSC4012	STATISTICS FOR DATA SCIENCE	3 Credits

a. Course Outcome (CO)

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

	Course Outcome	Level
CO1	Learn the fundamental concepts of statistical learning and regression	Understand
CO2	Understand the resampling methods and their effectiveness in evaluating methods	Understand
CO3	Able to apply regularization methods for improving the performance of models	Skill
CO4	Compare classification algorithms using real time data	Analyse
CO5	Able to apply non-linear methods for solving complex problems	Skill

b. Syllabus

Unit – 1	INTRODUCTION TO STATISTICAL LEARNING	12 hrs
Statistical Learning – Trade-Off between Prediction Accuracy and Model Interpretability – Regression vs Classification Problems - Assessing Model Accuracy – Measuring the Quality of Fit – Bias-Variance Trade- Off – Linear Regression – Multiple Linear Regression – Other Factors in Regression Model		
Unit – 2	**RESAMPLING METHODS**	**12 hrs**
Cross Validation – Validation Set Approach – Leave-One-Out Cross Validation – k-fold Cross Validation – Bias-Variance Trade-Off for k-fold Cross Validation – Cross Validation on Classification Problems – Bootstrap.		
Unit – 3	**LINEAR MODEL SELECTION AND REGULARIZATION**	**12 hrs**
Subset Selection – Best Subset Selection – Stepwise Selection – Choosing the Optimum Model - Shrinkage Methods – Ridge Regression – Lasso – Selecting the Tuning Parameter – Dimension Reduction Methods – Principal Components Regression – Partial Least Squares – High Dimensional Data		
Unit – 4	**CLASSIFICATION**	**12 hrs**
Logistic Regression – Estimating Regression Coefficients – Linear Discriminant Analysis - Comparison of Classification Methods – Quadratic Discriminant Analysis		
Unit – 5	**MOVIING BEHIND LINEARITY**	**12 hrs**
Polynomial Regression – Regression Splines – Smoothing Splines – Local Regression – Decision Trees – Bagging – Random Forests - Boosting		

Text / Reference Books:

1. James, G., Witten, D., Hastie, T., Tibshirani, R. An introduction to statistical learning with applications in R. Springer, 2013.
2. Hastie, T., Tibshirani, R., Friedman, J. The Elements of Statistical Learning, 2nd edition. Springer, 2009.
3. Murphy, K. Machine Learning: A Probabilistic Perspective. - MIT Press, 2012.
4. Montgomery, Douglas C., and George C. Runger. Applied statistics and probability for engineers. John Wiley & Sons, 2010

c. Mapping of Course Outcomes to Program Outcomes

	PO1	PO2	PO3	PO4	PO5	PO6	PO7
CO1	3	3	2	2	2	2	2
CO2	3	3	2	2	2	2	2
CO3	3	3	2	2	2	2	2
CO4	3	3	2	2	2	2	2
CO5	3	3	2	2	2	2	2

CSC4013	BIG DATA ANALYTICS	4 Credits

a. Course Outcome (CO)

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

	Course Outcome	Level
CO1	Understand the concepts of Big Data Analytics	Understand
CO2	Learn the components of Map Reduce and Hadoop Eco-System	Understand
CO3	Able to use SPARK	Analyse
CO4	Able to use NOSQL	Analyse
CO5	Learn to plot data	Analyse

b. Syllabus

Unit – 1	INTRODUCTION TO BIG DATA ANALYTICS	12 hrs

Big Data Technology Foundations – Challenges of Conventional Systems - Five Vs - Data analysis – Natureof Data – Big data analytic processes - Ingesting data into the system - Persisting the data in storage - Computing and Analyzing data - Visualizing the results – Processing and Analyzing Big Data batch processing - Stream Processing – overview on big data tools applied in different stages of Big data.

Unit – 2	MAP REDUCE	12 hrs

Introduction – Distributed file system – Map Reduce – Stages in Map Reduce - MapReduce Architecture - Algorithms using map reduce, Matrix-Vector Multiplication by Map Reduce – Hadoop - Understanding the Map Reduce architecture - Writing Hadoop MapReduce Programs - Loading data into HDFS -Executing the Map phase - Shuffling and sorting - Reducing phase execution.

Unit – 3	SPARK	12 hrs

Introduction to data analysis with spark - Programming with RDDs – working with key value pairs –loading and saving your data – Spark SQL – Spark Streaming – Apache spark MLIB – Machine Learning with MLib – Development of real time applications using SPARK.

Unit – 4	NOSQL	12 hrs

NoSQL: Types of Databases – Advantages - No SQL databases: Mongo DB: Introduction – Features – Data types – Mongo DB Query language – CRUD operations – Arrays – Functions: Count – Sort – Limit – Skip –Aggregate – Map Reduce. Cursors – Indexes – Mongo Import – Mongo Export.

Unit – 5	CASSANDRA AND DATA VISUALIZATION	12 hrs

Cassandra: Introduction – Features – Data types – CQLSH – Key spaces – CRUD operations – Collections – Counter – TTL – Alter commands – Import and Export – Querying System tables. Big Data Visualization - Basics - Trends in Data Visualization - Interactive Graphics - Visualization Designers – Use of Data Visualization - Visualization Process Into Practice - Visualization Software - TABLEAU DESKTOP - Elastic Stack, Apache Solr and Banana.

Text / Reference Books:

1. Raj Kamal, Preeti Saxena, Big Data Analytics, Introduction to Hadoop, Spark, and Machine-Learning,McGraw-Hill India, 2019
2. Subhashini Chellappan, Seema Acharya, Big Data and Analytics, Wiley, 2ed, 2019,
3. Big Data Fundamentals by Thomas Erl, PEARSON INDIA, 2016
4. Jure Leskovec, Anand Rajaraman, Jeffrey D. Ullman, "Mining of Massive Datasets", CambridgeUniversity Press, 2014.
5. Boris lublinsky, Kevin t. Smith, Alexey Yakubovich, "Professional Hadoop Solutions", Wiley, 2015.
6. Chris Eaton, Dirk deroos et al. , "Understanding Big data ", McGraw Hill, 2012.
7. Tom White, "HADOOP: The definitive Guide" , O Reilly 2012.
8. Spark - The Definitive Guide, Bill Chambers and Matei Zaharia, 2018, O'Reilly Media, Inc,USA,
9. Holden Karau, Andy Konwinski, Patrick Wendell & Matei Zaharia, Learning Spark Lightning Fast Data Analysis, O'Reilly, 2015
10. Josh Wills, Sandy Ryza, Sean Owen, and Uri Laserson, Advanced Analytics with Spark:
11. Patterns for Learning from Data at Scale 2nd Edition, O'Reilly, 2016
12. Sosulski kristen, Data visualization made simple insights into becoming visual, 2019, Routledge
13. Kieran Healy , Data Visualization: A Practical Introduction Paperback, 2018

c. Mapping of Course Outcomes to Program Outcomes

	PO1	PO2	PO3	PO4	PO5	PO6	PO7
CO1	3	2	2	2	2	2	2
CO2	3	2	2	2	2	2	2
CO3	3	2	2	2	2	2	2
CO4	3	2	2	2	2	2	2
CO5	3	2	2	2	2	2	2

CSC4014	MACHINE LEARNING	3 Credits

a. Course Outcome (CO)

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

	Course Outcome	Level
CO1	Understand the fundamentals of machine learning, model preparation and evaluation	Understand
CO2	Learn probability theory and its application in data understanding	Apply
CO3	Apply supervised learning algorithms and unsupervised learning algorithms on real data	Analyze
CO4	Apply ensemble classifiers on data and compare the model performance	Apply
CO5	Explore data repositories and apply Machine Learning Algorithms on case studies	Skill

b. Syllabus

Unit – 1	INTRODUCTION – MODEL PREPARATION – EVALUATION	10 hrs

Introduction to Machine Learning: Human Learning and its Types; Machine Learning and its types; Well-Posed Learning Problem; Applications of Machine Learning; Issues in Machine Learning.

Preparing To Model: Basic Data Types; Exploring Numerical Data; Exploring Categorical Data; Exploring Relationship Between Variables; Data Issues and Remediation; Data Pre-Processing.

Modelling And Evaluation: Selecting A Model; Training Model – Holdout, *K*-Fold Cross-Validation, Bootstrap Sampling; Model Representation and Interpretability – Under-Fitting, Over-Fitting, Bias-Variance Tradeoff – Model Performance Evaluation – Classification, Regression, Clustering – Performance Improvement.

Unit – 2	FEATURE ENGINEERING – REVIEW OF PROBABILITY	6 hrs

Feature Engineering: Feature Construction – Feature Extraction – Feature Selection.

Brief Review ofProbability: Basic Concept Of Probability – Random Variables – Discrete Distributions – Binomial, Poisson – Bernoulli - Continuous Distribution – Uniform, Normal, Laplace; Central Theorem; Monte CarloApproximation.

Unit – 3	CONCEPT LEARNING – SUPERVISED AND UNSUPERVISED LEARNING	12 hrs

Bayesian Concept Learning: Bayes Theorem – Prior and Posterior Probability, Likelihood; Concept Learning; Bayesian Belief Network. Supervised Learning – Regression: Simple Linear Regression; Other Regression Techniques. Supervised Learning – Classification: Basics of Supervised Learning – Classification; Logistic Regression – k-Nearest neighbor – Decision Tree – Support Vector Machine. Unsupervised Learning: Basics of Unsupervised Learning – Clustering Techniques; Association Rules.

Unit – 4	ENSEMBLE LEARNING	10 hrs

Concept of Ensemble Learning – Bagging and Boosting and its impact on bias and variance – RandomForest – Adaboost Classifier – Gradient Boosting Machines – XG Boost

Unit – 5	PUBLIC DATASETS AND CASE STUDIES	7 hrs

Exploring UCI Machine Learning Repository – Kaggle Data Sets – Analyzing Data Sets – Case Studies

Text / Reference Books:

1. Saikat Dutt, Subramanian Chandramouli, Amit Kumar Das, Machine Learning, Pearson EducationIndia, January 2018
2. Tom Mitchell, Machine Learning, McGraw Hill Education, First Edition, 2017
3. Christopher M. Bishop, Pattern Recognition and Machine Learning, Springer, First edition, 2010.
4. Ethem Alpaydin,"Introduction to Machine Learningll, MIT Press, Prentice Hall of India, Third Edition2015.

c. Mapping of Course Outcomes to Program Outcomes

	PO1	PO2	PO3	PO4	PO5	PO6	PO7
CO1	3	3	3	2	2	2	2
CO2	3	3	3	2	2	2	2
CO3	3	3	3	2	2	2	2
CO4	3	3	3	2	2	2	2
CO5	3	3	3	2	2	2	2

CSC4015	INTRODUCTION TO PROGRAMMING USING PYTHON	3 Credits

a. Course Outcome (CO)

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

	Course Outcome	Level
CO1	Understand the fundamental concepts of python and its main components	Understand
CO2	Develop (Read and Write) python programs using variables, assignments, and conditionalstatements using functions	Apply
CO3	Illustrate and implement different data structures	Analyse
CO4	Demonstrate Object-oriented concepts and file handling	Apply
CO5	Analyze and plot data using python visualization libraries	Skill

b. Syllabus

Unit – 1	INTRODUCTION TO PYTHON	9 hrs
Introduction to a programming language – History of Python- Python environment setup – Python 2 vs. Python 3 - Comments and documentation in Python- Keywords and Identifiers -Programming Errors - Writing and Running python programs.		
Unit – 2	**VARIABLES AND CONDITIONALS IN PYTHON**	**9 hrs**
Variables – Constants- Strings - Assignment statements – Expressions-Operators – Type Conversions - Control Flow statements and Loops- Functions.		
Unit – 3	**DATA STRUCTURES IN PYTHON**	**9 hrs**
List Basics- List Indexing and Slicing-Appending-Sorting and Ranging-Tuples-Creation-Deletion-Converting tuple to list- Assignment- Dictionaries-Adding-Modifying and Retrieving Values-Traversing all keys in the dictionary-Operations and methods-Sets- manipulating and accessing sets.		
Unit – 4	**EXCEPTION HANDLING AND FILE I/O**	**9 hrs**
Exception handling - Catching and Handling Exceptions-Object Orientated Concepts (Basics) – Creating python class and Objects - Object properties and methods- Inheritance- Operator overloading - Polymorphism- File handling - Opening, Reading, Writing and Deleting files.		
Unit – 5	**GRAPH PLOTTING**	**9 hrs**
Introduction to plotting python libraries -Plots and Graphs- Applied Visualizations – Seaborn – Matplotlib		

Text / Reference Books:

1. Eric Matthes, Python Crash Course, No Starch Press, Second edition, 2019
2. Martin C. Brown, Python: The Complete Reference, McGraw Hill Education, Fourth edition, 2018
3. Ryan Marvin Mark Ng'Ang'A Amos Omondi, Python Fundamentals, Packet Publishing, Second edition, 2018
4. Zed Shaw, Learn Python 3 the Hard Way: A Very Simple Introduction to the Terrifyingly Beautiful World of Computers and Code, Addison-Wesley Professional, First edition, 2017

c. Mapping of Course Outcomes to Program Outcomes

	PO1	PO2	PO3	PO4	PO5	PO6	PO7
CO1	3	3	3	3	3	1	2
CO2	3	3	2	3	3	1	2
CO3	3	3	2	3	3	1	2
CO4	3	3	3	3	3	1	2
CO5	3	3	3	2	3	2	2

CSC4016	BIG DATA ANALYTICS LAB	2 Credits

a. Course Outcome (CO)

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

	Course Outcome	Level
CO1	Get familiar with Hadoop distributions, configuring Hadoop and performing File management tasks	Understand
CO2	Experiment Map Reduce in Hadoop frameworks	Create
CO3	Implement Map Reduce programs in variety application	Analyze
CO4	Explore Map Reduce support for debugging	Analyze
CO5	Understand different approaches for building Hadoop Map Reduce programs for real-time applications	Analyze

b. Syllabus

Programs should include but not limited to:

I. Install, configure and run Hadoop and HDFS
II. Implement NoSQL database operations: CRUD operations, arrays using MongoDB
III. Implement basic sql functions using MongoDB
IV. Implement word count / frequency programs using map reduce
V. Implement clustering techniques using spark
VI. Implement a map reduce program that processes a dataset

c. Mapping of Course Outcomes to Program Outcomes

	PO1	PO2	PO3	PO4	PO5	PO6	PO7
CO1	3	2	2	2	2	2	2
CO2	3	2	2	2	2	2	2
CO3	3	2	2	2	2	2	2
CO4	3	2	2	2	2	2	2
CO5	3	2	2	2	2	2	2

CSC4017	MACHINE LEARNING LAB	2 Credits

a. Course Outcome (CO)

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

	Course Outcome	Level
CO1	Design and develop solutions for informed and uninformed search problems in AI	Apply
CO2	Explore advance packages to solve AI problems	Skill
CO3	Implement classical supervised learning classification algorithms	Apply
CO4	Implement various clustering algorithms and analyze clusters formed	Apply
CO5	Implement various ensemble classifiers and analyze the results	Skill

b. Syllabus

Programs should include but not limited to:

i. Data Analysis and Data Visualization
ii. Linear Regression and Multiple Regression
iii. Logistic Regression
iv. Decision Trees
v. k-Nearest Neighbor
vi. Multi layer perceptron
vii. Support vector machines
viii. Random Forest
ix. Adaboost and XG Boost

c. Mapping of Course Outcomes to Program Outcomes

	PO1	PO2	PO3	PO4	PO5	PO6	PO7
CO1	3	3	3	2	2	2	2
CO2	3	3	3	2	2	2	2
CO3	3	3	3	2	2	2	2
CO4	3	3	3	2	2	2	2
CO5	3	3	3	2	2	2	2

CSC4018	PYTHON PROGRAMMING LAB	2 Credits

a. Course Outcome (CO)

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

	Course Outcome	Level
CO1	Understand the fundamental concepts of python and its main components	Understand
CO2	Develop (Read and Write) python programs using variables, assignments, and conditionalstatements using functions	Create
CO3	Illustrate and implement different data structures	Analyze
CO4	Demonstrate Object-oriented concepts and file handling	Analyze
CO5	Analyze and plot data using python visualization libraries	Analyze

b. Syllabus

Programs should include but not limited to:

I. Test and Debug simple Python programs
II. Different datatypes in python (variables constants and strings)
III. Programs on different operators
IV. Control statements and Loops
V. Working on Functions
VI. Data structures in python (List, Tuple, Dictionary and Set)
VII. Objects and Classes manipulation using python
VIII. Open, Read and write data from/to files in Python
IX. Different plots using Matplotlib
X. Visualization of data using seaborn

c. Mapping of Course Outcomes to Program Outcomes

	PO1	PO2	PO3	PO4	PO5	PO6	PO7
CO1	3	2	2	2	2	2	2
CO2	3	2	2	2	2	2	2
CO3	3	2	2	2	2	2	2
CO4	3	2	2	2	2	2	2
CO5	3	2	2	2	2	2	2

SEMESTER – II

CSC4021	PRINCIPLES OF DEEP LEARNING	4 Credits

a. Course Outcome (CO)

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

	Course Outcome	Level
CO1	Recognize characteristics of deep learning models to solve real world problems	Understand
CO2	Understand the concepts of convolutional neural networks and its usage in classification problem	Apply
CO3	Learn deep learning supporting environments	Analyze
CO4	Understand the concept of recurrent neural networks and apply them in solving real world problems	Apply
CO5	Generate auto encoders and generative models for real time applications	Skill

b. Syllabus

Unit – 1	INTRODUCTION TO DEEP LEARNING & ARCHITECTURES	12 hrs

Machine Learning Vs. Deep Learning – Representation Learning – Width Vs. Depth of Neural Networks -Activation Functions: RELU – LRELU – ERELU –Unsupervised Training of Neural Networks – Regularization Dropout - Drop connect – Optimization methods for neural networks – Adagrad – Adadelta – Rmsprop Adam – NAG.

Unit – 2	CONVOLUTIONAL NEURAL NETWORKS & TRANSFER LEARNING	12 hrs

Architectural Overview – Motivation - Layers – Filters – Parameter sharing – Regularization - Popular CNN Architectures: LeNet – ResNet – Vggnet – AlexNet. Transfer learning Techniques –DenseNet – PixelNet.

Unit – 3	TRAINING NEURAL NETWORKS	12 hrs

Data Preprocessing – Data Augmentation – Batch Normalization – Transfer Learning – Deep Transfer Learning Strategies – Update Rules - Hyperparameter Tuning – Learning RateScheduling – Variants of CNN – ResNet – GoogleNet – Xception etc

Unit – 4	RECURRENT NEURAL NETWORKS	12 hrs

Recurrent Neural Networks – Bidirectional RNNs –Long Short Term Memory Networks.

Unit – 5	AUTO ENCODERS AND DEEP GENERATIVE MODELS	12 hrs

Auto encoders – Size and Depth of Auto encoders – Variational Auto Encoders -Deep Belief networks – Boltzmann Machines – Deep Boltzmann Machine – Generative Adversarial Networks.

Text / Reference Books:

1. Ian Goodfellow, Yoshua Bengio and Aaron Courville, Deep Learning (Adaptive Computation and Machine Learning series), First edition, The MIT Press, 2016
2. Josh Patterson, Adam Gibson, Deep Learning: A Practitioner's Approach, First edition, O'Reilly Media, 2017
3. Giancarlo Zaccone, Md. Rezaul Karim, Ahmed Menshawy, Deep Learning with TensorFlow: Explore neural networks with Python, First edition, Packt Publisher, 2017

c. Mapping of Course Outcomes to Program Outcomes

	PO1	PO2	PO3	PO4	PO5	PO6	PO7
CO1	3	3	3	2	2	2	2
CO2	3	3	3	2	2	2	2
CO3	3	3	3	2	2	2	2
CO4	3	3	3	2	2	2	2
CO5	3	3	3	2	2	2	2

CSC4022	DATA VISUALIZATION	4 Credits

a. Course Outcome (CO)

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

	Course Outcome	Level
CO1	Learn data visualization, visualization designers , use of data visualization and data visualization into practice	Apply
CO2	Learn visualization tools, Comparisons, Distributions and Chart Interface	Analyse
CO3	Understand of data, explore the data and design standards for the visual data	Analyse
CO4	Learn Information Dashboard categorization, issues and visual perception	Analyse
CO5	Learn Information dashboard design, organization, testing with case studies	Skill

b. Syllabus

Unit – 1	FUNDAMENTALS OF DATA VISUALIZATION	12 hrs

Basics of Data visualization - Trends in Data Visualization - Interactive Graphics - Visualization Designers –Use of Data Visualization - Visualization Process Into Practice - Basic Productivity Applications -MICROSOFT EXCEL – iWORK - GOOGLE CHARTS - Visualization Software - TABLEAU DESKTOP.

Unit – 2	VISUALIZATION TOOLS, DISTRIBUTIONS AND CHART INTERFACE	12 hrs

Business Intelligence Tools - Programming Packages - Selecting Tools to Build Data Graphics - Comparisons of Categories and Time – Distributions – Proportions – Relationships – Locations - Trends—Showing Comparisons Over Time or Composition Over Time - Word Frequency and Sentiment - Connections and Networks - Chart Interface

Unit – 3	DESIGN STANDARDS FOR THE VISUAL DATA	12 hrs

Understand your data - Review the Data Dictionary – different kinds of data – Preparing the data - Survey Data - IDENTIFY MISSING VALUES - Compute Descriptive Statistics - Explore the Data Visually - Devise the Problem, Challenge, and/or Questions - Design Standards for visual data

Unit – 4	INFORMATION DASHBOARD CATEGORIZATION AND VISUAL PERCEPTION	12 hrs

Information dashboard - categorizing dashboards - typical dashboard data - dashboard design issues andbest practices - visual perception - principles of visual perception for dashboard design

Unit – 5	INFORMATION DASHBOARD DESIGN WITH CASE STUDIES	12 hrs

INFORMATION DASHBOARD DESIGN II: Characteristics of dashboards - key goals in visual design process - dashboard display media - designing dashboards for usability- meaningful organization – maintaining consistency.

Text / Reference Books:

1. Sosulski kristen, Data visualization made simple insights into becoming visual, 2019, Routledge
2. Kieran Healy , Data Visualization: A Practical Introduction Paperback, 2018
3. Cole Nussbaumer Knaflic, Storytelling with Data: A Data Visualization Guide for Business Professionals, 2015
4. Chun-houh Chen, Wolfgang Härdle, Antony Unwin, Handbook of Data Visualization, Springer, 2008
5. Stephen Few, Information Dashboard Design: Displaying Data for At-a-Glance Monitoring, 2013

c. Mapping of Course Outcomes to Program Outcomes

	PO1	PO2	PO3	PO4	PO5	PO6	PO7
CO1	3	3	3	2	2	2	2
CO2	3	3	3	2	2	2	2
CO3	3	3	3	2	2	2	2
CO4	3	3	3	2	2	2	2
CO5	3	3	3	2	2	2	2

CSC4023	DEEP LEARNING LAB	2 Credits

a. Course Outcome (CO)

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

	Course Outcome	Level
CO1	Design and develop solutions for problems of voluminous data	Apply
CO2	Explore keras – tensor flow and theano frame work	Skill
CO3	Implement various convolutional neural networks for classification	Apply
CO4	Implement encoding networks for real time applications	Apply
CO5	Implement generative adversarial networks	Skill

b. Syllabus

Programs should include but not limited to:

i. Keras API
ii. Tensorflow API
iii. Theano framework
iv. Alex net
v. Resnet
vi. VGG nets
vii. Auto Encoder
viii. VAE
ix. GANs

c. Mapping of Course Outcomes to Program Outcomes

	PO1	PO2	PO3	PO4	PO5	PO6	PO7
CO1	3	3	3	2	2	2	2
CO2	3	3	3	2	2	2	2
CO3	3	3	3	2	2	2	2
CO4	3	3	3	2	2	2	2
CO5	3	3	3	2	2	2	2

CSC4024	DATA VISUALIZATION LAB	2 Credits

a. Course Outcome (CO)

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

	Course Outcome	Level
CO1	Understand the importance of data visualization for business intelligence and decision making	Understand
CO2	Know various approaches to understand visual perception	Understand
CO3	Learn about categories of visualization and applications	Apply
CO4	Familiarize with data visualization tools	Skill
CO5	Gain knowledge of effective data visuals to solve workplace problems	Apply

b. Syllabus

Programs should include but not limited to:

i. Exploration of various data visualization tools
ii. Basic Visualization in python
iii. Working in Tableau
iv. Connecting to data and prepare data for visualization in tableau
v. Data Aggregation and Statistical functions in tableau
vi. Data Visualization in tableau
vii. Basic dashboards in tableau

c. Mapping of Course Outcomes to Program Outcomes

	PO1	PO2	PO3	PO4	PO5	PO6	PO7
CO1	3	2	2	2	2	2	2
CO2	3	2	2	2	2	2	2
CO3	3	2	2	2	2	2	2
CO4	3	2	2	2	2	2	2
CO5	3	2	2	2	2	2	2

CSC4025	PROJECT / DISSERTATION	12 Credits

a. Course Outcome (CO)

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

	Course Outcome	Level
CO1	Analyze and understand the real life problem and apply their knowledge to determine the solution	Analyze
CO2	Engage in the creative design process through the integration and application of diverse technical knowledge and expertise to meet the requirements	Skill
CO3	Use various tools and techniques, coding practices for developing real time solutions to the problem	Apply
CO4	Test and deploy the solution determined through various test cases.	Skill
CO5	Documentation of the work done in a report format	Create

b. Description

The project done will be evaluated by the faculty members in the project review meetings. Two midterm reviews and one final review will be conducted where the student has to present the work done in the project followed by viva-voce. The complete work carried out has to be represented in the form of a project report. The report should also contain the plagiarism check report (max 10%) certified by the project guide.

The report should strictly follow the specifications of the report. Four copies should be submitted by the student. One copy will be returned to the student after evaluation.

c. Mapping of Course Outcomes to Program Outcomes

	PO1	PO2	PO3	PO4	PO5	PO6	PO7
CO1	3	3	3	2	2	2	2
CO2	3	3	3	2	2	2	2
CO3	3	3	3	2	2	2	2
CO4	3	3	3	2	2	2	2
CO5	3	3	3	2	2	2	2

DISCIPLINE SPECIFIC ELECTIVES

CSCEC04	DATA WAREHOUSING AND DATA MINING	4 Credits

a. Course Outcome (CO)

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

	Course Outcome	Level
CO1	Understand the principles of Data warehousing and OLAP	Understand
CO2	Familiar with the Data mining processing methods	Understand
CO3	Perform classification and prediction of data	Apply
CO4	Perform clustering of data and analyze patterns	Apply
CO5	Analyse spatial and multimedia data	Analyze

b. Syllabus

Unit – 1	DATA WAREHOUSING AND BUSINESS ANALYSIS	12 hrs

Data warehousing Components –Building a Data warehouse –Data Warehouse Architecture – DBMS Schemas for Decision Support – Data Extraction, Cleanup, and Transformation Tools –Metadata – reporting – Query tools and Applications – Online Analytical Processing (OLAP) – OLAP and Multidimensional Data Analysis.

Unit – 2	DATA MINING & ASSOCIATION RULE MINING	12 hrs

Data Preprocessing – Data Cleaning – Data Integration and Transformation – Data Reduction – Data Discretization and Concept Hierarchy Generation- Architecture Of A Typical Data Mining Systems- Classification Of Data Mining Systems- Efficient and Scalable Frequent Item set Mining Methods – MiningVarious Kinds of Association Rules – Association Mining to Correlation Analysis – Constraint-Based Association Mining.

Unit – 3	CLASSIFICATION AND PREDICTION	12 hrs

Issues Regarding Classification and Prediction – Classification by Decision Tree Introduction – Bayesian Classification – Rule Based Classification – Classification by Back propagation – Support Vector Machines – Associative Classification – Lazy Learners – Other Classification Methods – Prediction – Accuracy and Error Measures – Evaluating the Accuracy of a Classifier or Predictor – Ensemble Methods – Model Selection.

Unit – 4	CLUSTER ANALYSIS	12 hrs

Types of Data in Cluster Analysis – A Categorization of Major Clustering Methods – Partitioning Methods – Hierarchical methods – Density-Based Methods – Grid-Based Methods – Model-Based ClusteringMethods – Clustering High-Dimensional Data – Constraint-Based Cluster Analysis – Outlier Analysis.

Unit – 5	MINING OBJECT, SPATIAL, MULTIMEDIA, TEXT AND WEB DATA	12 hrs

Multidimensional Analysis and Descriptive Mining of Complex Data Objects – Spatial Data Mining – Multimedia Data Mining – Text Mining – Mining the World Wide Web.

Text / Reference Books:

1. Jiawei Han, Micheline Kamber and Jian Pei, Data Mining Concepts and Techniques, Elsevier, Thirdedition, 2011.
2. Alex Berson and Stephen J. Smith, Data Warehousing, Data Mining & OLAP, Tata McGraw – HillEdition, Tenth Reprint 2007.
3. K.P. Soman, Shyam Diwakar and V. Ajay, Insight into Data mining Theory and Practice, Easter Economy Edition, Prentice Hall of India, 2006.
4. G. K. Gupta, Introduction to Data Mining with Case Studies, Easter Economy Edition, Prentice Hall ofIndia, 2006.
5. Pang-Ning Tan, Michael Steinbach and Vipin Kumar, Introduction to Data Mining, Pearson Education,2007.

c. Mapping of Course Outcomes to Program Outcomes

	PO1	PO2	PO3	PO4	PO5	PO6	PO7
CO1	3	1	2	2	3	1	1
CO2	2	3	3	3	3	2	2
CO3	2	2	2	2	3	2	2
CO4	2	2	1	2	3	2	2
CO5	3	2	2	3	3	1	2

CSCEC05	PATTERN RECOGNITION	4 Credits

a. Course Outcome (CO)

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

	Course Outcome	Level
CO1	Understand the concepts of Bayesian Decision Theory	Understand
CO2	Apply Maximum Likelihood and Bayesian Estimation in solving problems	Apply
CO3	Compare linear and non-linear classifiers	Analyze
CO4	Learn the principles of structural pattern recognition with examples	Understand
CO5	Ability to extract and select features from the given data	Skill

b. Syllabus

Unit – 1	INTRODUCTION TO PATTERN RECOGNITION AND BAYESIAN DECISION THEORY	12 hrs

Machine Perception – Sub-problems of Pattern Classification – Learning and Adaptation – BayesianDecision Theory – Continuous Features – Minimum Error Rate Classification – Classifiers – Discriminants – Decision Surfaces – Normal Density – Discriminant Functions for Normal Density – Bayesian DecisionTheory for Discrete Features

Unit – 2	MAXIMUM LIKELIHOOD AND BAYESIAN ESTIMATION	12 hrs

Maximum Likelihood Estimation – Bayesian Estimation – Bayesian Parameter Estimation: General Theory and Gaussian Case – Problems on Dimensionality – Expectation-Maximization – Bayesian Belief Networks – Hidden Markov Models

Unit – 3	LINEAR AND NON-LINEAR CLASSIFIERS	12 hrs

Linear Discriminant Functions and Decision Hyperplanes – Perceptron – Least Squares Methods – MeanSquare Estimation – XOR Problem – SVM for XOR Problem – Polynomial Classifiers – Radial Basis FunctionNetworks

Unit – 4	STRUCTURAL PATTERN RECOGNITION	12 hrs

Elements Of Formal Grammars – String Generation as Pattern Description – Recognition of SyntacticDescription – Parsing – Stochastic Grammars and Applications - Graph based Structural Representation.

Unit – 5	FEATURE SELECTION AND GENERATION	12 hrs

Feature Selection: Pre-processing – Hypothesis Testing – ROC Curve – Feature Subset Selection. FeatureGeneration: Basis Vectors and Images – KL Transform – Singular Value Decomposition – Independent Component Analysis.

Text / Reference Books:

1. Duda R.O., and Hart.P.E., David G Stork, Pattern Recognition, Second Edition, Wiley, May 2021.
2. Duda R.O., Hart P.E., Pattern Classification, Wiley, Second edition, January 2012.
3. Sergios Theodoridis and K. Koutroumbas , Pattern Recognition, Fourth edition, Elsevier, 2008.
4. Robert J.Schalkoff, Pattern Recognition:, Statistical, Structural and Neural Approaches,John Wiley &Sons Inc., New York, 2007.
5. Trevor H, Robert T,Jerome Friedman, The Elements of Statistical Learning, SpringerSeries,2017
6. C.M. Bishop, Pattern Recognition and Machine Learning, Springer First edition, August 2016

c. Mapping of Course Outcomes to Program Outcomes

	PO1	PO2	PO3	PO4	PO5	PO6	PO7
CO1	3	3	3	2	2	2	2
CO2	3	3	3	2	2	2	2
CO3	3	3	3	2	2	2	2
CO4	3	3	3	2	2	2	2
CO5	3	3	3	2	2	2	2

CSCEC06	NATURAL LANGUAGE PROCESSING	4 Credits

a. Course Outcome (CO)

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

	Course Outcome	Level
CO1	Learn the fundamentals of Natural Language Processing	Understand
CO2	Ability to perform text processing operations	Apply
CO3	Ability to estimate parameters and evaluate language models	Apply
CO4	Understand the syntax and semantics of text data	Understand
CO5	Ability to apply NLP concepts in related real time applications	Analyze

b. Syllabus

Unit – 1	INTRODUCTION TO NLP	12 hrs

Introduction to various levels of natural language processing, Ambiguities and computational challengesin processing various natural languages. Introduction to Real life applications of NLP such as spell and grammar checkers, information extraction, question answering, and machine translation.

Unit – 2	TEXT PROCESSING	12 hrs

Character Encoding, Word Segmentation, Sentence Segmentation, Introduction to Corpora, Corpora Analysis.

Unit – 3	LEXICAL SYNTAX AND LANGUAGE MODELLING	12 hrs

Introduction to word types, POS Tagging, Maximum Entropy Models for POS tagging, Multi-word Expressions. The role of language models. Simple N-gram models. Estimating parameters and smoothing. Evaluating language models.

Unit – 4	SYNTAX & SEMANTICS	12 hrs

Introduction to phrases, clauses and sentence structure, Shallow Parsing and Chunking, Shallow Parsing with Conditional Random Fields (CRF), Lexical Semantics, Word Sense Disambiguation, WordNet, Thematic Roles, Semantic Role Labelling with CRFs.

Unit – 5	APPLICATIONS OF NLP	12 hrs

NL Interfaces, Text Summarization, Sentiment Analysis, Machine Translation, Question answering.

Text / Reference Books:

1. Chris Manning and Hinrich Schütze, Foundations of Statistical Natural Language Processing, MITPress Cambridge, MA, Second edition, 2003.
2. NitinIndurkhya, Fred J. Damerau, Handbook of Natural Language Processing, CRC Press, Secondedition,2010.
3. James Allen, Natural Language Understanding, Pearson Publication, Eighth edition. 2012.

c. Mapping of Course Outcomes to Program Outcomes

	PO1	PO2	PO3	PO4	PO5	PO6	PO7
CO1	3	2	2	2	2	2	2
CO2	3	2	2	2	2	2	2
CO3	3	2	2	2	2	2	2
CO4	3	2	2	2	2	2	2
CO5	3	2	2	2	2	2	2

CSCEC38	WEB ANALYTICS	4 Credits

a. Course Outcome (CO)

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

	Course Outcome	Level
CO1	Learn the foundations of Web analytics	Understand
CO2	Ability to compare and contrast the clickstream data collection techniques, their impact on metrics, and their inherent limitations	Apply
CO3	Apply web analytics techniques to effectively use the resulting insights to support website design decisions, campaign optimization, search analytics, etc.	Apply
CO4	Understand the basics of software tools, techniques, and reports that are relevant to web analytics apply them to solve problems	Understand
CO5	Analyse and interpret web channel data and understand the difficulties and issues involved in it.	Analyse

b. Syllabus

Unit – 1	INTRODUCTION	12 hrs
Understanding web analytics – The foundations of Web analytics: Techniques and Technologies – Present and Future of Web analytics		
Unit – 2	**DATA COLLECTION**	**12 hrs**
Importance and Options –Web server log files: Click stream data – User submitted information – Web server performance data – Page tags –First and third party tracking.		
Unit – 3	**WEB ANALYSIS STRATEGY**	**12 hrs**
Key performance indicators – Web analytics process – Heuristics evaluations – Site visits – Surveys – Measuring reach – Measuring acquisition – Measuring conversion – Measuring retention – Security and privacy implications of Web analytics.		
Unit – 4	**WEB ANALYTICS TOOLS**	**12 hrs**
Content organization tools – Process measurement tools – Visitor segmentation tools – Campaign analysis tools – Commerce measurement tools – Google analytics – Omniture – Web trends – Yahoo! Web analytics.		
Unit – 5	**GOOGLE ANALYTICS**	**12 hrs**
Key features and capabilities – Quantitative and qualitative data - Working of Google analytics – Privacy - Tracking visitor clicks, Outbound links and Non HTML files.		

Text / Reference Books:

1. Bernard J. Jansen, ―Understanding User-Web Interactions via Web analytics‖, Morgan and Claypool, 2009.
2. Avinash Kaushik, ―Web Analytics2.0‖, John Wiley and Sons, 2010.
3. Brian Clifton, ―Advanced web metrics with Google analytics‖, John Wiley and Sons, 2012.
4. Justin Cutroni, ―Google Analytics‖, O"Reilly, 2015.
5. Jerri L. Ledford, Joe Teixeira and Mary E. Tyler, ―Google Analytics‖, John Wiley and Sons, 2013.

c. Mapping of Course Outcomes to Program Outcomes

	PO1	PO2	PO3	PO4	PO5	PO6	PO7
CO1	3	2	2	2	2	2	2
CO2	3	2	2	2	2	2	2
CO3	3	2	2	2	2	2	2
CO4	3	2	2	2	2	2	2
CO5	3	2	2	2	2	2	2

CSCEC39	MULTIVARIATE DATA ANALYSIS	4 Credits

a. Course Outcome (CO)

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

	Course Outcome	Level
CO1	Learn the basic statistics methods and its applications	Understand
CO2	Able to apply univariate statistical principles for solving problems	Apply
CO3	Able to apply multivariate statistical principles on real time data	Apply
CO4	Develop multivariate models for data analysis	Understand
CO5	Learn the concepts of factor analysis and its applications	Analyse

b. Syllabus

Unit – 1	INTRODUCTION	12 hrs

Basic multivariate statistics–mean, variance, Covariance, correlation, linear combination of variables, geometric concepts, distances

Unit – 2	UNIVARIATE STATISTICS	12 hrs

Univariate Descriptive Statistics – Sampling Distribution – Estimation – Hypothesis Testing

Unit – 3	MULTIVARIATE STATISTICS	12 hrs

Multivariate Descriptive Statistics – Multivariate Normal Distribution – Multivariate Inferential Statistics – Analysis of Variance (ANOVA) – Multivariate Analysis of Variance (MANOVA)

Unit – 4	MULTIVARIATE MODELS	12 hrs

Multiple Linear Regression – Sampling Distribution of Regression Coefficients – Model Adequacy Tests – Test of Assumptions – Model Diagnostics

Unit – 5	FACTOR ANALYSIS	12 hrs

Factor Analysis – Estimation and Model Adequacy Testing – Rotation Factor Scores – Structural Equation Modelling – Correspondence Analysis.

Text / Reference Books:

1. Chris Manning and Hinrich Schütze, Foundations of Statistical Natural Language Processing, MITPress Cambridge, MA, Second edition, 2003.
2. NitinIndurkhya, Fred J. Damerau, Handbook of Natural Language Processing, CRC Press, Secondedition,2010.
3. James Allen, Natural Language Understanding, Pearson Publication, Eighth edition. 2012.

c. Mapping of Course Outcomes to Program Outcomes

	PO1	PO2	PO3	PO4	PO5	PO6	PO7
CO1	3	2	2	2	2	2	2
CO2	3	2	2	2	2	2	2
CO3	3	2	2	2	2	2	2
CO4	3	2	2	2	2	2	2
CO5	3	2	2	2	2	2	2

CSCEC40	GRAPH ALGORITHMS AND MINING	4 Credits

a. Course Outcome (CO)

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

	Course Outcome	Level
CO1	Provide the basic concepts and important properties of graphs	Understand
CO2	Learn and explore several methods on algorithms such as graph traversal, shortest paths, minimum spanning tree	Apply
CO3	Introduce students to the field of graph mining and its application in various domains	Apply
CO4	Apply greedy based graph mining algorithms	Apply
CO5	Apply graph mining algorithms on real time data and analyse results	Skill

b. Syllabus

Unit – 1	INTRODUCTION TO GRAPHS	12 hrs

Introduction to graphs: Introduction to graphs and basic terminology, Representations of a graph, types of graphs, basic algorithms for decomposing graphs into parts, connectivity of graphs, matching on graphs.

Unit – 2	GRAPH ALGORITHMS	12 hrs

Graph colouring, graphs on surface, directed graphs, Shortest path algorithms, algorithms to discover minimum spanning tree, Flows in Networks and some important flow algorithms, Searching Graphs and Related algorithms.

Unit – 3	GRAPH MINING	12 hrs

Motivation for Graph Mining, Applications of Graph Mining, Mining Frequent Subgraphs – Transactions, BFS/Apriori Approach (FSG and others), DFS Approach (gSpan and others), Diagonal and Greedy Approaches, Constraint-based mining and new algorithms, Mining Frequent Subgraphs, graph visualizations

Unit – 4	ADVANCED GRAPH MINING ALGORITHMS	12 hrs

Diagonal and Greedy Approaches, Constraint-based mining and new algorithms, Mining Frequent Subgraphs, graph visualizations

Unit – 5	APPLICATIONS OF GRAPH MINING	12 hrs

Web mining, centrality analysis, Link analysis algorithms, graph clustering and community detection, Node classification and Link prediction, Influential spreaders, Influence maximization, Geo-social and location-based networks

Text / Reference Books:

1. Diestel, R. (2010).Graph Theory, 4th ed.Springer-Verlag, Heidelberg

2. J. Han and M. Kamber, Data minining–Concepts and Techniques, 2ndEdition, Morgan Kaufman Publishers, 2006

3. Bing Liu, Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data, Springer publishing, 2009

4. Jure Leskovec, Anand Rajaraman, Jeff Ullman. Mining of Massive Datasets. Book 2nd edition. Cambridge University Press

5. David Easley and Jon Kleinberg. Networks, Crowds, and Markets. Cambridge University Press, 2010.

6. Deepayan Chakrabarti andChristos Faloutsos. Graph Mining: Laws, Tools, and Case Studies. Synthesis Lectures on Data Mining and Knowledge Discovery, Morgan & Claypool Publishers,2012

7. Albert-László Barabási. Network Science. Cambridge University Press, 2016.

c. Mapping of Course Outcomes to Program Outcomes

	PO1	PO2	PO3	PO4	PO5	PO6	PO7
CO1	3	2	2	2	2	2	2
CO2	3	2	2	2	2	2	2
CO3	3	2	2	2	2	2	2
CO4	3	2	2	2	2	2	2
CO5	3	2	2	2	2	2	2

CSCEC41	OPTIMIZATION TECHNIQUES	4 Credits

a. Course Outcome (CO)

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

	Course Outcome	Level
CO1	Learn the fundamentals of linear programming	Understand
CO2	Understand other linear programming techniques	Apply
CO3	Analyze the network related issues	Apply
CO4	Learn the queuing theory concepts	Understand
CO5	Understand the non-linear programming and its applications	Analyse

b. Syllabus

Unit – 1	LINEAR PROGRAMMING	12 hrs

Convex sets, supporting and separating hyperplanes, standard linear programming problem, basic feasible solutions, simplex algorithm and simplex method, geometry of simplex method. Duality in linear programming, duality theorems.

Unit – 2	ADVANCE CONCEPTS OF LINEAR PROGRAMMING	12 hrs

Dual simplex method with justification, post optimality analysis, sensitivity analysis and parametric linear programming. Integer Linear Programming: Introduction, Gomory Cut Method, Branch and Bound Method

Unit – 3	NETWORK ANALYSIS	12 hrs

Definition and formulation, critical path method, Project Evaluation and Review Technique (PERT), Optimal allocation of resources (men-power) through time schedule

Unit – 4	QUEUEING THEORY	12 hrs

Introduction, steady state solution of M/M/c/∞/FIFO and M/M/C/N/FIFO with associated distributions of queue length and waiting time.(c=1 as particular case)

Unit – 5	NON-LINEAR PROGRAMMING	12 hrs

Quadratic Programming, Kuhn-Tucker Conditions, Wolfe-Beal-Fletcher algorithms for solving quadratic programming problems.

Text / Reference Books:

1. Chris Manning and Hinrich Schütze, Foundations of Statistical Natural Language Processing, MIT Press Cambridge, MA, Second edition, 2003.
2. NitinIndurkhya, Fred J. Damerau, Handbook of Natural Language Processing, CRC Press, Second edition,2010.
3. James Allen, Natural Language Understanding, Pearson Publication, Eighth edition. 2012.

c. Mapping of Course Outcomes to Program Outcomes

	PO1	PO2	PO3	PO4	PO5	PO6	PO7
CO1	3	2	2	2	2	2	2
CO2	3	2	2	2	2	2	2
CO3	3	2	2	2	2	2	2
CO4	3	2	2	2	2	2	2
CO5	3	2	2	2	2	2	2