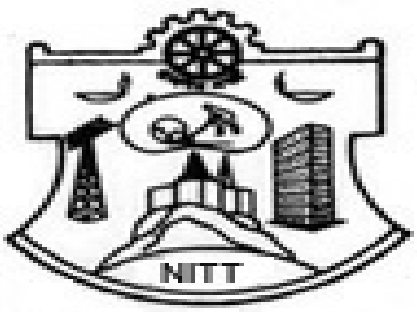
M.Tech. DATA ANALYTICS

Credit Based Flexible Curriculum

(Applicable form 2017-18 onwards)



Department of Computer Applications
National Institute of Technology
Tiruchirappalli– 620 015, Tamilnadu

SYLLABUS

Semester	Subject Code	Subject Name	Credits
I	CA601	Statistical Computing	3
	CA603	Big Data Analytics	3
	CA605	Machine Learning Techniques	3
	*****	Elective -1	3
	*****	Elective-2	3
	*****	Elective-3	3
	CA609	Big Data Management and Data Analytics Lab	2
II	CS618	Real Time Systems	3
	CA602	Next Generation Databases	3
	CA604	High Performance Computing	3
	*****	Elective -4	3
	*****	Elective -5	3
	*****	Elective -6	3
	CA610	Machine Learning Lab	2
III	CA647	Project work-Phase I	12
IV	CA648	Project work-Phase II	12
		Total Credits	64

LIST OF ELECTIVES

Semester	Subject Code	Subject Name	Credits
I	CS655	Digital Forensics	3
	CA611	Cyber Security and Information Assurance	3
	CA612	Natural Language Computing	3
	CA613	Massive Graph Analysis	3
	CA614	Bioinformatics	3
	CA615	Parallel and Distributed Computing	3
	CA616	Data Acquisition and Productization	3
	CA617	Essentials of Human Resource Analytics	3
	CA618	Customer Relationship and Management	3
II	CA619	Principles of Deep Learning	3
	CA620	Image and Video Analytics	3
	CA621	Social Networking and Mining	3
	CA622	Web Intelligence	3
	CA623	Internet of Things	3
	CA624	Health care Data Analytics	3
	CA625	Linked Open Data and Semantic Web	3
	CA626	Financial Risk Analytics and Management	3
	CA627	Logistics and Supply Chain Management	3

SEMESTER - I
CA601 STATISTICAL COMPUTING

Objectives:

- *To learn the probability distributions and density estimations to perform analysis of various kinds of data.*
- *To explore the statistical analysis techniques using Python and R programming languages.*
- *To expand the knowledge in R and Python to use it for further research.*

Probability Theory: Sample Spaces- Events - Axioms – Counting - Conditional Probability and Bayes' Theorem – The Binomial Theorem – Random variable and distributions : Mean and Variance of a Random variable-Binomial-Poisson-Exponential and Normal distributions. Curve Fitting and Principles of Least Squares- Regression and correlation.

Sampling Distributions & Descriptive Statistics: The Central Limit Theorem, distributions of the sample mean and the sample variance for a normal population, Sampling distributions (Chi-Square, t, F, z). Test of Hypothesis- Testing for Attributes – Mean of Normal Population – One-tailed and two-tailed tests, F-test and Chi-Square test - - Analysis of variance ANOVA – One way and two way classifications.

Tabular data- Power and the computation of sample size- Advanced data handling- Multiple regression- Linear models- Logistic regression- Rates and Poisson regression- Nonlinear curve fitting.

Density Estimation- Recursive Partitioning- Smoothers and Generalised Additive Models - Survivals Analysis- Analysing Longitudinal Data- Simultaneous Inference and Multiple Comparisons- Meta-Analysis- Principal Component Analysis- Multidimensional Scaling- Cluster Analysis.

Introduction to R- Packages- Scientific Calculator- Inspecting Variables- Vectors- Matrices and Arrays- Lists and Data Frames- Functions- Strings and Factors- Flow Control and Loops- Advanced Looping- Date and Times. Introduction to Python- Packages- Fundamentals of Python- Inserting and Exporting Data- Data Cleansing- Checking and Filling Missing Data- Merging Data- Operations- Joins.

References:

1. Richard Cotton, "Learning R", O'Reilly, 2013.
2. Dalgaard, Peter, "Introductory statistics with R", Springer Science & Business Media, 2008.
3. Brain S. Everitt, "A Handbook of Statistical Analysis Using R", Second Edition,

LLC, 2014.
4. Samir Madhavan, “Mastering Python for Data Science”, Packt, 2015.
5. Sheldon M. Ross,”Introduction to Probability and Statistics for Engineers and Scientists”, 4th edition, Academic Press; 2009.
6. Paul Teetor, “R Cookbook, O’Reilly, 2011.
7. Mark Lutz ,”Learning Python”, O’Reilly,5th Edition,2013

Outcomes:

Students will be able to:

- *Implement statistical analysis techniques for solving practical problems.*
- *Perform statistical analysis on variety of data.*
- *Perform appropriate statistical tests using R and visualize the outcome.*

CA603 BIG DATA ANALYTICS

Objectives:

- *To optimize business decisions and create competitive advantage with Big Data analytics*
- *To explore the fundamental concepts of big data analytics.*
- *To learn to analyze the big data using intelligent techniques.*
- *To understand the various search methods and visualization techniques.*
- *To learn to use various techniques for mining data stream.*
- *To understand the applications using Map Reduce Concepts.*
- *To introduce programming tools PIG & HIVE in Hadoop echo system.*

Introduction to big data : Introduction to Big Data Platform – Challenges of Conventional Systems - Intelligent data analysis – Nature of Data - Analytic Processes and Tools - Analysis vs Reporting.

Mining data streams : Introduction To Streams Concepts – Stream Data Model and Architecture - Stream Computing - Sampling Data in a Stream – Filtering Streams – Counting Distinct Elements in a Stream – Estimating Moments – Counting Oneness in a Window – Decaying Window - Real time Analytics Platform(RTAP) Applications - Case Studies - Real Time Sentiment Analysis- Stock Market Predictions.

Hadoop: History of Hadoop- the Hadoop Distributed File System – Components of Hadoop Analysing the Data with Hadoop- Scaling Out- Hadoop Streaming- Design of HDFS-Java interfaces to HDFS Basics- Developing a Map Reduce Application-How Map Reduce Works-Anatomy of a Map Reduce Job run-Failures-Job Scheduling-Shuffle and Sort – Task execution - Map Reduce Types and Formats- Map Reduce Features-Hadoop environment.

Frameworks: Applications on Big Data Using Pig and Hive – Data processing operators in Pig – Hive services – HiveQL – Querying Data in Hive - fundamentals of HBase and ZooKeeper - IBM InfoSphere BigInsights and Streams.

Predictive Analytics- Simple linear regression- Multiple linear regression- Interpretation

of regression coefficients. Visualizations - Visual data analysis techniques- interaction techniques - Systems and applications.

References:

1. Michael Berthold, David J. Hand, "Intelligent Data Analysis", Springer, 2007.
2. Tom White "Hadoop: The Definitive Guide" Third Edition, O'reilly Media, 2012.
3. Chris Eaton, Dirk DeRoos, Tom Deutsch, George Lapis, Paul Zikopoulos, "Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data", McGrawHill Publishing, 2012.
4. Anand Rajaraman and Jeffrey David Ullman, "Mining of Massive Datasets", CUP, 2012.
5. Bill Franks, "Taming the Big Data Tidal Wave: Finding Opportunities in Huge Data Streams with Advanced Analytics", John Wiley& sons, 2012.
6. Glenn J. Myatt, "Making Sense of Data", John Wiley & Sons, 2007.
7. Pete Warden, "Big Data Glossary", O'Reilly, 2011.
8. Jiawei Han, Micheline Kamber "Data Mining Concepts and Techniques", 2nd Edition, Elsevier, Reprinted 2008.
9. Da Ruan, Guoquing Chen, Etienne E.Kerre, Geert Wets, "Intelligent Data Mining", Springer, 2007.
10. Paul Zikopoulos, Dirkde Roos, Krishnan Parasuraman, Thomas Deutsch, James Giles , David Corrigan, "Harness the Power of Big Data The IBM Big Data Platform", Tata McGraw Hill Publications, 2012.
11. Arshdeep Bahga, Vijay Madisetti, "Big Data Science & Analytics: A Hands-On Approach ",VPT, 2016
12. Bart Baesens "Analytics in a Big Data World: The Essential Guide to Data Science and its Applications (WILEY Big Data Series)", John Wiley & Sons,2014

Outcomes:

Students will be able to:

- *Work with big data platform and explore the big data analytics techniques business applications.*
- *Design efficient algorithms for mining the data from large volumes.*
- *Analyze the HADOOP and Map Reduce technologies associated with big data analytics.*
- *Explore on Big Data applications Using Pig and Hive.*
- *Understand the fundamentals of various big data analytics techniques.*
- *Build a complete business data analytics solution*

CA605 MACHINE LEARNING TECHNIQUES

Objectives:

- *To introduce the basic concepts and techniques of Machine Learning.*
- *To develop the skills in using recent machine learning software for solving practical problems.*
- *To be familiar with a set of well-known supervised, semi-supervised and unsupervised learning algorithms*

Introduction- overview of machine learning- Different forms of learning- Generative

learning- Gaussian parameter estimation- maximum likelihood estimation- MAP estimation- Bayesian estimation- bias and variance of estimators- missing and noisy features- nonparametric density estimation- applications- software tools.

Classification Methods-Nearest neighbour- Decision trees- Linear Discriminant Analysis - Logistic regression-Perceptrons- large margin classification- Kernel methods- Support Vector Machines. Classification and Regression Trees.

Graphical and sequential models- Bayesian networks- conditional independence- Markov random fields- inference in graphical models- Belief propagation- Markov models- Hidden Markov models- decoding states from observations- learning HMM parameters.

Clustering Methods-Partitioned based Clustering - K-means- K-medoids; Hierarchical Clustering - Agglomerative- Divisive- Distance measures; Density based Clustering - DBScan; Spectral clustering.

Neural networks- the perceptron algorithm- multilayer perceptron's- back propagation- nonlinear regression- multiclass discrimination- training procedures- localized network structure- dimensionality reduction interpretation.

References:

1. T. Hastie, R. Tibshirani and J. Friedman, "Elements of Statistical Learning", Springer, 2009.
2. E. Alpaydin, "Machine Learning", MIT Press, 2010.
3. K. Murphy, "Machine Learning: A Probabilistic Perspective", MIT Press, 2012.
4. C. Bishop, "Pattern Recognition and Machine Learning, Springer", 2006.
5. Shai Shalev-Shwartz, Shai Ben-David, "Understanding Machine Learning: From Theory to Algorithms", Cambridge University Press, 2014.
6. John Mueller and Luca Massaron, "Machine Learning For Dummies", John Wiley & Sons, 2016.

Outcomes:

Students will be able to:

- *Select real-world applications that needs machine learning based solutions.*
- *Implement and apply machine learning algorithms.*
- *Select appropriate algorithms for solving a particular group of real-world problems.*
- *Recognize the characteristics of machine learning techniques that are useful to solve real-world problems.*

CA609 BIG DATA MANAGEMENT AND DATA ANALYTICS LAB

Objectives:

- *Optimize business decisions and create competitive advantage with Big Data analytics*
- *Imparting the architectural concepts of Hadoop and introducing map reduce paradigm*
- *Introducing Java concepts required for developing map reduce programs*

- *Derive business benefit from unstructured data*
- *Introduce programming tools PIG & HIVE in Hadoop echo system.*
- *Developing Big Data applications for streaming data using Apache Spark*

Lab Exercises:

1. (i)Perform setting up and Installing Hadoop in its two operating modes:
    - Pseudo distributed,
    - Fully distributed.

   (ii) Use web based tools to monitor your Hadoop setup.
2. (i) Implement the following file management tasks in Hadoop:
    - Adding files and directories
    - Retrieving files
    - Deleting files

   ii) Benchmark and stress test an Apache Hadoop cluster
3. Run a basic Word Count Map Reduce program to understand Map Reduce Paradigm.
    - Find the number of occurrence of each word appearing in the input file(s)
    - Performing a MapReduce Job for word search count (look for specific keywords in a file)
4. Stop word elimination problem:
    - Input:
        - A large textual file containing one sentence per line
        - A small file containing a set of stop words (One stop word per line)
    - Output:
        - A textual file containing the same sentences of the large input file without the words appearing in the small file.
5. Write a Map Reduce program that mines weather data. Weather sensors collecting data every hour at many locations across the globe gather large volume of log data, which is a good candidate for analysis with MapReduce, since it is semi structured and record-oriented. Data available at: https://github.com/tomwhite/hadoop-book/tree/master/input/ncdc/all.
    - Find average, max and min temperature for each year in NCDC data set?
    - Filter the readings of a set based on value of the measurement, Output the line of input files associated with a temperature value greater than 30.0 and store it in a separate file.
6. Purchases.txt Dataset
    - Instead of breaking the sales down by store, give us a sales breakdown by product category across all of our stores
        - What is the value of total sales for the following categories?
            - Toys
            - Consumer Electronics
    - Find the monetary value for the highest individual sale for each separate store

- What are the values for the following stores?
  - Reno
  - Toledo
  - Chandler

- Find the total sales value across all the stores, and the total number of sales.

7. Install and Run Pig then write Pig Latin scripts to sort, group, join, project, and filter your data.
8. Write a Pig Latin scripts for finding TF-IDF value for book dataset (A corpus of eBooks available at: Project Gutenberg)
9. Install and Run Hive then use Hive to create, alter, and drop databases, tables, views, functions, and indexes.
10. Install, Deploy & configure Apache Spark Cluster. Run apache spark applications using Scala.
11. Data analytics using Apache Spark on Amazon food dataset, find all the pairs of items frequently reviewed together.
    - Write a single Spark application that:
      - Transposes the original Amazon food dataset, obtaining a PairRDD of the type: <user_id> → <list of the product_ids reviewed by user_id>
      - Counts the frequencies of all the pairs of products reviewed together;
      - Writes on the output folder all the pairs of products that appear more than once and their frequencies. The pairs of products must be sorted by frequency.

Outcomes:

- *Preparing for data summarization, query, and analysis.*
- *Applying data modelling techniques to large data sets*
- *Creating applications for Big Data analytics*
- *Building a complete business data analytic solution*

SEMESTER - II

CS618 REAL TIME SYSTEMS

Objectives:

- *To study issues related to the design and analysis of systems with real-time constraints.*
- *To learn the features of Real time OS.*
- *To study the various Uniprocessor and Multiprocessor scheduling mechanisms.*
- *To learn about various real time communication protocols.*
- *To study the difference between traditional and real time databases.*

Introduction to real time computing - Concepts; Example of real-time applications – Structure of a real time system – Characterization of real time systems and tasks - Hard and Soft timing constraints - Design Challenges - Performance metrics - Prediction of Execution Time : Source code analysis, Micro-architecture level analysis, Cache and pipeline issues- Programming Languages for Real-Time Systems

Real time OS – Threads and Tasks – Structure of Microkernel – Time services – Scheduling Mechanisms Communication and Synchronization – Event Notification and Software interrupt

Task assignment and Scheduling - Task allocation algorithms - Single-processor and

Multiprocessor task scheduling - Clock-driven and priority-based scheduling algorithms- Fault tolerant scheduling

Real Time Communication -Network topologies and architecture issues – protocols – contention based, token based, polled bus, deadline based protocol, Fault tolerant routing. RTP and RTCP.

Real time Databases – Transaction priorities – Concurrency control issues – Disk scheduling algorithms – Two phase approach to improve predictability.

References:

1. C.M. Krishna, Kang G. Shin – " Real Time Systems", International Edition, McGrawHill Companies, Inc., New York, 1997.
2. Jane W.S. Liu, "Real-Time Systems", Pearson Education India, 2000.
3. Philip A. Laplante and Seppo J. Ovaska, "Real-Time Systems Design and Analysis: Tools for the Practitioner" IV Edition IEEE Press, Wiley, 2013.
4. Sanjoy Baruah, Marko Bertogna, Giorgio Buttazzo, "Multiprocessor Scheduling for Real-Time Systems ", Springer International Publishing, 2015.

Outcomes:

Students will be able to:

- *Gain Knowledge about Schedulability analysis.*
- *Learn about the Real-time programming environments.*
- *Attain knowledge about real time communication and databases.*
- *Develop real time systems.*

CA602 NEXT GENERATION DATABASES

Objectives:

- *To explore the concepts of NoSQL Databases.*
- *To understand and use columnar and distributed database patterns.*
- *To learn to use various Data models for a variety of databases.*

Database Revolutions- System Architecture- Relational Database- Database Design- Data Storage- Transaction Management- Data warehouse and Data Mining- Information Retrieval.

Big Data Revolution- CAP Theorem- Birth of NoSQL- Document Database—XML Databases- JSON Document Databases- Graph Databases.

Column Databases— Data Warehousing Schemes- Columnar Alternative- Sybase IQ- C-Store and Vertica- Column Database Architectures- SSD and In-Memory Databases— In-Memory Databases- Berkeley Analytics Data Stack and Spark.

Distributed Database Patterns— Distributed Relational Databases- Non-relational

Distributed Databases- MongoDB - Sharing and Replication- HBase- Cassandra- Consistency Models— Types of Consistency- Consistency MongoDB- HBase Consistency- Cassandra Consistency.

Data Models and Storage- SQL- NoSQL APIs- Return SQL- Advance Databases— PostgreSQL- Riak- CouchDB- NEO4J- Redis- Future Databases— Revolution Revisited- Counter revolutionaries- Oracle HQ- Other Convergent Databases- Disruptive Database Technologies.

References:

1. Abraham Silberschatz, Henry F. Korth, S. Sudarshan, "Database System Concepts", Sixth Edition, McGrawHill.
2. Guy Harrison, "Next Generation Databases", Apress, 2015.
3. Eric Redmond, Jim R Wilson, "Seven Databases in Seven Weeks", LLC. 2012.
4. Dan Sullivan, "NoSQL for Mere Mortals", Addison-Wesley, 2015.
5. Adam Fowler, "NoSQL for Dummies ", John Wiley & Sons, 2015.

Outcomes:

Students will be able to:

- *Explore the relationship between Big Data and NoSQL databases*
- *Work with NoSQL databases to analyze the big data for useful business applications.*
- *Work with different data models to suit various data representation and storage needs.*

CA604 HIGH PERFORMANCE COMPUTING

Objectives:

- *To know how modern high performance processors are organized their strengths and weaknesses.*
- *To study about the architecture of parallel systems.*
- *To gain knowledge about the analytical parallel algorithms.*

Principles of Parallel Algorithms- Graph Algorithms- Minimum Spanning Tree- Prim's Algorithm - Single-Source Shortest Paths-Dijkstra's Algorithm - All-Pairs Shortest Paths - -. Algorithms for Sparse Graphs - Search Algorithms for Discrete Optimization Problems - Sequential Search Algorithms - - Parallel Depth-First Search - Parallel Breadth-First Search - Dynamic Programming - Serial Monadic DP Formulations –No serial Monadic DP Formulations - Serial Polyadic DP Formulations.

Shared-memory parallel programming with OpenMP- Introduction to OpenMP - Parallel

execution - Data scoping -OpenMP work sharing for loops - Synchronization Reductions - Loop scheduling - Miscellaneous - Case study-OpenMP-parallel Jacobi algorithm - Advanced OpenMP-Wavefront parallelization - Efficient OpenMP programming - Profiling OpenMP programs.

Distributed-memory parallel programming with MPI- Message passing- MPI – example - Messages and point-to-point communication - Collective communication - Non blocking point-to-point communication - Virtual topologies - Example- MPI parallelization of Jacobi solver - Communication parameters -Synchronization-serialization- contention - Implicit serialization and synchronization - Contention - Reducing communication overhead - Optimal domain decomposition - Aggregating messages - Non blocking vs. asynchronous communication.

Hybrid parallelization with MPI and OpenMP- Basic MPI/OpenMP programming models - Vector mode implementation - Task mode implementation - Case study- Hybrid Jacobi solver - MPI taxonomy of thread interoperability - Hybrid decomposition and mapping - Potential benefits and drawbacks of hybrid programming.

NVidia – GPU Computing – CUDA – Case studies.

References:

1. Georg Hager and Gerhard Wellein, “Introduction to High Performance Computing for Scientists and Engineers”, Chapman & Hall, 2010.
2. Ananth Grama and George Karypis, “Introduction to parallel computing”, Addison -Wesley, 2009.
3. John Levesque and Gene Wagenbreth, “High Performance Computing: Programming and Applications”, Chapman & Hall, 2010.
4. John L. Hennessy and David Patterson, “Computer Architecture- A Quantitative Approach”, Elsevier, 2012.
5. Michael Quinn, “Parallel Programming in C with MPI and OpenMP”, Indian edition, McGraw Hill Education, 2017.

Outcomes:

Students will be able to:

- *Investigate modern design structures of pipelined and multiprocessors systems.*
- *Write algorithms using parallel programming principle.*
- *Design the architecture of parallel systems.*

CA610 MACHINE LEARNING LAB

Objectives:

- *To introduce basic machine learning techniques.*
- *To develop the skills in using recent machine learning software for solving practical problems in high-performance computing environment.*
- *To develop the skills in applying appropriate supervised, semi-supervised or unsupervised learning algorithms for solving practical problems.*

1. Exercises to solve the real-world problems using the following machine learning methods:

- Linear Regression
- Logistic Regression
- Multi-Class Classification
- Neural Networks
- Support Vector Machines
- K-Means Clustering & PCA

2. Develop programs to implement Anomaly Detection & Recommendation Systems.

3. Implement GPU computing models to solving some of the problems mentioned in Problem 1.

Outcomes:
Students will be able to:

- *Implement and apply machine learning algorithms to solve problems.*
- *Select appropriate algorithms for solving a of real-world problems.*
- *Use machine learning techniques in high-performance computing environment to solve real-world problems.*

CA647 PROJECT WORK –Phase I

Internal project work of 6 Months duration to be extended in phase II.

CA648 PROJECT WORK –Phase II

Internal project work of 6 Months duration with submission of thesis and viva-voce examination

SEMESTER 1: ELECTIVES

CS655 DIGITAL FORENSICS

Objectives:

- *To understand the basics of digital forensics and the techniques for conducting the forensic examination on different digital devices.*
- *To understand how to examine digital evidences such as the data acquisition, identification analysis.*
- *To understand the various categories of tools and procedures used in the digital forensic process.*

Computer forensics fundamentals, Benefits of forensics, computer crimes, computer forensics evidence and courts, legal concerns and private issues.

Understanding Computing Investigations – Procedure for corporate High-Tech investigations, understanding data recovery work station and software, conducting and investigations.

Data acquisition- understanding storage formats and digital evidence, determining the best acquisition method, acquisition tools, validating data acquisitions, performing RAID data acquisitions, remote network acquisition tools, other forensics acquisition tools.

Processing crimes and incident scenes, securing a computer incident or crime, seizing digital evidence at scene, storing digital evidence, obtaining digital hash, reviewing case.

Current computer forensics tools- software, hardware tools, validating and testing forensic software, addressing data-hiding techniques, performing remote acquisitions, E-Mail investigations- investigating email crime and violations, understanding E-Mail servers, specialized E-Mail forensics tool.

References:

1. Warren G. Kruse II and Jay G. Heiser, "Computer Forensics: Incident Response Essentials", Addison Wesley, 2002.
2. Bill Nelson, Amelia Phillips, F.Enfinger and Christopher Stuart, "Guide to Computer Forensics and Investigations", 2nd edition. Thomson Course Technology, 2006.
3. John R. Vacca, Computer Forensics, "Computer Crime Scene Investigation", 2ndEd, Charles River Media, 2005.
4. Bill Nelson, Amelia Phillips, F.Enfinger and Christopher Stuart, "Guide to Computer Forensics and Investigations, 4th ed., Thomson Course Technology, 2010.
5. Anthony T. S. Ho and Shujun Li, "Handbook of Digital Forensics of Multimedia Data and Devices", IEEE Press, John Wiley & Sons, 2015.

Outcomes:

Students will be able to:

- *Know how to apply forensic analysis tools to recover important evidence for identifying computer crime.*
- *Become well-trained cyber-crime investigators.*
- *Apply various techniques of digital forensics for the systematic crime investigation.*

CA611 CYBER SECURITY AND INFORMATION ASSURANCE

Objectives:

- *To understand and apply the models of Information Security.*
- *To study the Information assurance tools and methods.*
- *To study and analyze cryptographic and forensic methods.*
- *To analyze and simulate the network and application security.*
- *To explore the nature and logic behind the cyber security threats as an ethical*

hacker.

Critical characteristics of Information - NSTISSC Security Model -Components of information System –SDLC – Information assurance - Security Threats and vulnerabilities - Overview of Security threats-– Security Standards .

Classical Cryptography - Symmetric Cryptography- Asymmetric Cryptography - Modern Cryptography – Access Control - DRM – Steganography – Biometrics.

Network security - Intrusion Prevention, detection and Management - Firewall – E-commerce Security - Computer Forensics - Security for VPN and Next Generation Networks.

Host and Application security -Control hijacking, Software architecture and a simple buffer overflow - Common exploitable application bugs, shellcode - Buffer Overflow - Side-channel attacks - Timing attacks, power analysis, cold-boot attacks, defenses – Malware - Viruses and worms, spyware, key loggers, and botnets; defenses auditing, policy - Defending weak applications - Isolation, sandboxing, virtual machines.

Mobile, GSM and Wireless LAN security - Protection measures - Business risk analysis – Information Warfare and Surveillance – Case study on Attack prevention, detection and response.

References:

1. William Stallings, "Cryptography and Network Security: Principles and Practice", 6th Edition, PHI, 2014.
2. Michael E. Whitman and Herbert J Mattord, "Principles of Information Security", 6th edition, Vikas Publishing House, 2017.
3. Bill Nelson, Amelia Phillips, F.Enfinger and Christopher Stuart, "Guide to Computer Forensics and Investigations, 4th ed., Thomson Course Technology, 2010.
4. Matt Bishop, "Computer Security: Art and Science", 1st edition, Addison-Wesley Professional, 2015.

Outcomes:

Students will be able to:

- *Identify the information security models and their characteristics.*
- *Analyze the different types of cryptographic and forensic methods.*
- *Study the network security issues.*
- *Discover the layers of application security.*
- *Identify and solve different cyber security threats.*

CA612 NATURAL LANGUAGE COMPUTING

Objectives:

- *To get introduced to language processing technologies for processing the text*

data.

- *To understand the role of Information Retrieval and Information Extraction in Text Analytics.*
- *To acquire knowledge on text data analytics using language models.*

Natural Language Processing – Linguistic Background -– Mathematical Foundations - Morphological Analysis-Tokenization- Stemming-Lemmatization - Boundary Determination.

Reading unstructured data - Representing text data - Part of speech tagging - Syntactic representation - Text similarity - WordNet based similarity- Shallow parsing -Semantic representation.

Information retrieval and Information extraction - Named Entity Recognition - Relation Identification-Template filling.

Language model - Probabilistic Models - n-gram language models- Hidden Markov Model- Topic Modelling - Graph Models -Feature Selection and classifiers -Rule-based Classifiers - Maximum entropy classifier – Clustering-Word and Phrase-based Clustering.

Tools – Natural Language Tool kit, Apache OpenNLP. Applications of Text Analytics – Applications in Social media - Life science - Legal Text–Visualization -Case studies.

References:

1. Christopher D. Manning and Hinrich Schutze, "Foundations of Statistical Natural Language Processing", MIT Press, 1999.
2. Steven Struhl, "Practical Text Analytics: Interpreting Text and Unstructured Data for Business Intelligence", Kogan Page, 2015.
3. Matthew A. Russell, "Mining the Social Web", O'Reilly Media, 2013.
4. Steven Bird, Ewan Klein and Edward Loper, "Natural Language Processing with Python", 1st Edition, O'Reilly Media, 2009.

Outcomes:

Students will be able to:

- *Process the text data at syntactic and semantic level.*
- *Extract the ¬key information from Text data.*
- *Analyze the text content to provide predictions related to a specific domain using language models.*

CA613 MASSIVE GRAPH ANALYSIS

Objectives:

- *To explore the concept of Graphs and related algorithms.*
- *To learn new ways to model, store, retrieve and analyze graph-structured data.*
- *To be aware of advanced concepts in graph analytic techniques and its applications.*

Introduction and Application of Large-scale Graph, Characteristics, Complex Data Sources - Social Networks, Simulations, Bioinformatics; Categories- Social, Endorsement, Location, Co-occurrence graphs; Graph Data structures, Parallel, Multicore, & Multithreaded Architectural Support for Graph Processing, Mapping Graph Algorithms to Architectures.

Basic and Advanced Large-scale Graph Analysis- Parallel Prefix & List Ranking, Link Analysis, Page Ranking Algorithms; Parallel BFS, Spanning Tree, Connected Components, Minimum Spanning Tree Matroid Algorithms, Social Networking Algorithms, Parallel Betweenness Centrality.

Dynamic Parallel Algorithms - Streaming Data Analysis -Data Structures for Streaming Data - Tracking Clustering Coefficients - Tracking Connected Components -Anomaly Detection, Massive-Graphs in Computational Biology, Genome Assembly.

Distributed Computation for Massive Data Sets- Spectral, Modularity-based Clustering, Random Walks; Large Graph Representation and Implementation- V-Graph Representation, Map Reduce, Surfer, Graph Lab.

Advanced Topics- Power Law Distribution, Game-Theoretic Approach, Rank Aggregation and Voting Theory, Recommendation Systems, Social network analysis: case study -Facebook, LinkedIn, Google+, and Twitter.

References:

1. Matthew O. Jackson, “Social and Economic Networks”, Princeton University Press, 2010.
2. Stanley Wasserman, Katherine Faust, "Social Network Analysis: Methods and Applications”, (Structural Analysis in the Social Sciences), Cambridge University Press, 1995.
3. Tanja Falkowski, "Community Analysis in Dynamic Social Networks", (Dissertation), University Magdeburg, 2009.
4. Ladislav Novak, Alan Gibbons, "Hybrid Graph Theory and Network Analysis", Cambridge Tracts in Theoretical Computer Science, 2009.
5. Eric D. Kolaczyk, "Statistical Analysis of Network Data Methods and Models", Springer Series in Statistics, 2009.
6. Akihito Hora, Nobuaki Obata, “Quantum Probability and Spectral Analysis of Graphs”, Springer, 2007.
7. Richard Brath, David Jonker, “Graph Analysis and Visualization: Discovering Business Opportunity in Linked Data”, John Wiley & Sons, 2015.

Outcomes:

Students will be able to:

- *Explore the graph analytic techniques and its applications.*
- *Model a problem into a graph database and perform analytical tasks over the*

graph in a scalable manner.

- *Apply Graph theoretical techniques in massive networks.*

CA614 BIOINFORMATICS

Objectives:

- *To understand Bio informatics from computing perspective.*
- *To comprehend bio informatics databases, file formats and its applications.*
- *To understand the applications of Bio informatics*

History of bioinformatics-History of Bioinformatics-role of Bioinformatics in biological sciences- scope of bioinformatics -introduction to internet-WWW- network basics- LAN & WAN standards-network topologies and protocols- FTP- HTTP - division of Bioinformatics- Bioinformatics and internet-challenges in Bioinformatics.

Databases in bioinformatics-Databases in Bioinformatics- Genbank- NCBI- EMBL- DDBJ -UniGene- SGD- EMI Genomes- -protein databases-PIR- SWISSPROT-TrEMBL-Prosite-PRINTS -structural databases-PDB- SCOP- CATH- PDB_SELECT- PDBSUM- DSSP- FSSP-DALI- PRODOM- protein families & pattern databases-Pfam- KEGG - sequence storage sequence accuracy-EST-STS- sequence retrieval systems- Entrez-SRS- sequence query refinement using Boolean operators- limits- preview- history and index.

Sequence submission-Sequence submission tools-BANKIT-SEQUIN-WEBIN-SAKURA-literature databases-PubMed and medline. Data mining and its techniques- data warehousing- Sequence annotation- principles of genome annotation- annotation tools & resources.

Applications of bioinformatics-Applications of Bioinformatics-phylogenetic analysis-steps in phylogenetic analysis-microarrays-DNA and protein microarrays-Bioinformatics in pharmaceutical industry- informatics & drug- discovery – pharma informatics resources drug discovery and designing-SNP.

File formats-File formats-raw/plain format-NCBI-Genbank flat file format-ASN.1- GCG-FASTA- EMBL- NBRF- PIR-swissprot sequence formats- PDB format-Introduction to structure prediction methods.

References:

1. Attwood T.K, Parry-Smith, “Introduction to Bioinformatics”, Addison Wesley Longman, 1999.
2. David W Mount, “Bioinformatics: Sequence and Genome Analysis”, 2nd edition, CBS Publishers, 2004.
3. Arun Jagota, “Data Analysis and Classification for Bioinformatics”, Pine Press, 2001.
4. Des Higgins and Willie Taylor, “Bioinformatics Sequence, Structures and Databanks”, Oxford University Press, 2000.
5. Jason T.L.Wang, Mohammed J. Zaki, Hannu T.T. Toivonene and Dennis Shasha, “Data Mining in Bioinformatics”, Springer International Edition, 2005.
6. K. Erciyes, “Distributed and Sequential Algorithms for Bioinformatics”, Springer, 2015.

Outcomes:

Students will be able to:

- *Explore bioinformatics from computing perspective.*
- *Apply data mining techniques to provide better health care services.*
- *Explore and extract hidden information from bio informatics databases.*

CA615 PARALLEL AND DISTRIBUTED COMPUTING

Objectives:

- *To learn core ideas behind parallel and distributed computing.*
- *To explore the methodologies adopted for concurrent and distributed environment.*
- *To understand the networking aspects of parallel and distributed computing.*
- *To provide an overview of the computational aspects of parallel and distributed computing.*
- *To learn parallel and distributed computing models.*

Parallel and Distributed Computing— Introduction- Benefits and Needs- Parallel and Distributed Systems- Programming Environment- Theoretical Foundations- Parallel Algorithms— Introduction- Parallel Models and Algorithms- Sorting- Matrix Multiplication- Convex Hull- Pointer Based Data Structures.

Synchronization- Process Parallel Languages- Architecture of Parallel and Distributed Systems- Consistency and Replication- Security- Parallel Operating Systems.

Management of Resources in Parallel Systems- Tools for Parallel Computing- Parallel Database Systems and Multimedia Object Servers.

Networking Aspects of Distributed and Parallel Computing- Process- Parallel and Distributed Scientific Computing.

High-Performance Computing in Molecular Sciences- Communication- Multimedia Applications for Parallel and Distributed Systems- Distributed File Systems.

References:

1. Jacek Błażewicz, et al., “Handbook on parallel and distributed processing”, Springer Science & Business Media, 2013.
2. Andrew S. Tanenbaum, and Maarten Van Steen, “Distributed Systems: Principles and Paradigms”. Prentice-Hall, 2007.
3. George F.Coulouris, Jean Dollimore, and Tim Kindberg, “Distributed systems: concepts and design”, Pearson Education, 2005.
4. Gregor Kosec and Roman Trobec, “Parallel Scientific Computing: Theory, Algorithms, and Applications of Mesh Based and Meshless Methods”, Springer, 2015.

Outcomes:

Students will be able to:

- *Explore the methodologies adopted for concurrent and distributed environment.*
- *Analyse the networking aspects of Distributed and Parallel Computing.*

- *Explore the different performance issues and tasks in parallel and distributed computing.*
- *Develop parallel algorithms for solving real–world problems.*

CA616 DATA ACQUISITION AND PRODUCTIZATION

Objectives:

- *To explore the fundamental concepts of data pre-processing, extraction, cleaning, annotation, integration.*
- *To understand the various information visualization techniques.*
- *To understand data productization using Internet of things.*

Data Acquisition

Introduction to Data Warehouse- OLTP and OLAP concepts- Introduction to Data Mining- Data Objects and Attribute Types-Basic Statistical Descriptions of Data- Exploratory Data analysis- Measuring Data Similarity and Dissimilarity- Graphical representation of data.

Introduction to Data Acquisition – Applications –Process- Data Extraction- Data Cleaning and Annotation- Data Integration -Data Reduction- Data Transformation –Data Discretization and Concept Hierarchy Generation.

Visualization-Introduction -Terminology- Basic Charts and Plots- Multivariate Data Visualization- Data Visualization Techniques– Pixel-Oriented Visualization Techniques- Geometric Projection Visualization Techniques- Icon-Based Visualization Techniques- Hierarchical Visualization Techniques- Visualizing Complex Data and Relations- Data Visualization Tools– Rank Analysis Tools- Trend Analysis Tools- Multivariate Analysis Tools- Distribution Analysis Tools- Correlation Analysis Tools- Geographical Analysis Tools.

Data Productization

IoT Overview- IoT Design methodology- Semantic Web Infrastructure- Intelligence Applications- Programming Framework for IoT- Distributed Data Analysis for IoT- Security and Privacy in IoT- Applied IoT- Cloud Based Smart Facilities Management.

Virtualization on Embedded Boards IoT- Stream Processing in IoT-Internet of Vehicles and Applications - Case study on Data Acquisition using Dashboards, Android and iOS apps.

References:

1. Han, Jiawei, Jian Pei, and Micheline Kamber, “Data mining: concepts and techniques”, 3rd Edition, Elsevier, 2011.
2. Margaret H. Dunham, "Data Mining: Introductory and Advanced Topics", Pearson Education, 2012.
3. Arshdeep Bahga, Vijay Madisetti, “Internet of Things -A hands-on approach”, Universities Press, 2015.
4. Manoel Carlos Ramon, “Intel Galileo and Intel Galileo Gen 2: API Features and Arduino Projects for Linux Programmers”, Apress, 2014.
5. Karl Pover, “Learning Qlikview Data Visualization”, Packt, 2013.
6. Rajkumar Buyya, Amir Vahid Dastjerdi, “Internet of Things: Principles and Paradigms”, Elsevier, 2016.

Outcomes:

Students will be able to:

- *Apply of data pre-processing, extraction, cleaning, annotation, integration on data.*
- *Apply the suitable visualization techniques to output analytical results.*
- *Explore on applications using Internet of things.*

CA617 ESSENTIALS OF HUMAN RESOURCE ANALYTICS

Objective:

- *To provide the knowledge and necessary skills to accomplish personnel roles in the domain of HR.*
- *To provide knowledge to carry out analytics within the context of business objectives and outcomes.*

Introduction to HR Analytics: Overview of HR Process, HR as an expense, the analytics and prediction Strategic Human capital measures, business analysis and rational action. Benefits of Analytics in Improving HR Process, Intersection of people and profits. Technology Used, SWOT Analysis of HR analytics.

Employee Engagement Measurement Process: Attracting, motivating and retaining people Organization Gap and Alignment Analytics. Process to assess and prioritize organization gaps and identify alignment opportunities. HR Alignment Inventory. Developing performance metrics/Predicting future 'performance' Developing metrics to capture the fallouts of HR Policies.

Organization-Wide Alignment Audits: Assessing the individual components and the holistic system and the best practices, clarifying organizational practices and gaps, Recruitment Analytics and On Boarding Analytics Staffing Analytics Performance & Skill Gap Analytics Compensation & Benefit Analytics Training &Learning Analytics Promotion and Succession Planning Analytics Compliance Analytics Attrition& Retention Analytics, Identification of Key Business Objectives Conducting HR Practice Audits Conducting On-Site Visits Performing Data Triangulation.

Approach to HR Solutions: Identifying job responsibilities, tasks, and employee attributes needed on different jobs to assure mission-critical goals. Assessing competencies as a consistent foundation for organization/job design, succession and compensation. HR Dashboards Advanced Data Analytics (Forecasting, Predicting and Segmentation etc.) & Business Insights High End Consulting, KPI Catalogue Creation.

Program Evaluation and Return-On-Investment Analysis: Applied research for data-driven organization change and improvement using sophisticated HR analysis and metrics. Clarifying HR issues and drive focused, systematic organization change. Evaluating critical HR initiatives and/or business objectives. Make program improvements and shifts.

References:

1. Jac Fitz-enz, "The New HR Analytics: Predicting the Economic Value of Your Company's Human Capital Investments", American Management Association, 2010.
2. John W. Boudreau," Beyond HR: The New Science of Human Capital", Harvard Business School Press, 2007.

Outcomes:

Students will be able to:

- *Identify necessary skills to carry out the personnel roles in the domain of HR.*
- *Identify and develop metrics to improve employer-employee relationship and improve employee retention.*
- *Identify skilled personnel and job tasks to achieve mission-critical goals.*
- *Align organization's mission and goals with key metrics and benchmarks.*
- *Apply HR analytics to improve organizational performance by providing better insights on human resources data.*

CA618 CUSTOMER RELATIONSHIP AND MANAGEMENT

Objective:

- *To train the participants in the concepts of customer relationship management with industry case studies and strategies for implementing them in any organization.*
- *To better understand customer needs and to maintain long-term customer relationships.*
- *Be able to pursue a strategy of Relationship Marketing.*

Introduction to Customer Relationship Management: Concept- Evolution of Customer Relationships: Customers as strangers- acquaintances- friends and partners. Objectives- Benefits of CRM to Customers and Organizations- Customer Profitability Segments- Components of CRM: Information- Process- Technology and People- Barriers to CRM. Relationship Marketing and CRM: Relationship Development Strategies: Organizational Pervasive Approach- Managing Customer Emotions- Brand Building through Relationship Marketing- Service Level Agreements- Relationship

Challenges.

CRM Marketing Initiatives- Customer Service and Data Management :CRM Marketing Initiatives: Cross-Selling and Up-Selling- Customer Retention- Behaviour Prediction- Customer Profitability and Value Modeling- Channel Optimization- Personalization and Event-Based Marketing. CRM and Customer Service: Call Center and Customer Care: Call Routing- Contact Center Sales-Support- Web Based Self Service- Customer Satisfaction Measurement- Call-Scripting- Cyber Agents and Workforce Management. CRM and Data Management: Types of Data: Reference Data- Transactional Data- Warehouse Data and Business View Data- Identifying Data Quality Issues- Planning and Getting Information Quality- Using Tools to Manage Data- Types of Data Analysis: Online Analytical Processing (OLAP) - Clickstream Analysis- Personalization and Collaborative Filtering- Data Reporting.

CRM Strategy- Planning: Understanding Customers: Customer Value- Customer Care- Company Profit Chain: Satisfaction- Loyalty- Retention and Profits. Objectives of CRM Strategy- The CRM Strategy Cycle: Acquisition- Retention and Win Back- Complexities of CRM Strategy.

CRM Implementation and Evaluation: Planning and Implementation of CRM: Business to Business CRM- Sales and CRM- Sales Force Automation- Sales Process/ Activity Management- Sales Territory Management- Contact Management- Lead Management- Configuration Support- Knowledge Management CRM Implementation: Steps- Business Planning- Architecture and Design- Technology Selection- Development- Delivery and Measurement.

CRM Evaluation: Basic Measures: Service Quality- Customer Satisfaction and Loyalty- Company 3E Measures: Efficiency- Effectiveness and Employee Change.

CRM New Horizons: e-CRM: Concept- Different Levels of E- CRM- Privacy in E-CRM - Software App for Customer Service:# Activity Management- Agent Management- Case Assignment- Contract Management- Customer Self Service- Email Response Management- Escalation- Inbound Communication Management- Invoicing- Outbound Communication Management- Queuing and Routing- Scheduling - Social Networking and CRM - Mobile-CRM - CRM Trends- Challenges and Opportunities - Ethical Issues in CRM.

References:

1. Anderrson Kristin and Carol Kerr,"Customer Relationship Management", Tata McGraw-Hill, 2002.
2. Ed Peelen, "Customer Relationship Management", Prentice Hall, 2005.
3. Bhasin Jaspreet Kaur, "Customer Relationship Management", Dreamtech Press, 2012
4. Valarie A Zeithmal, Mary Jo Bitner, Dwayne D Gremler and Ajay Pandit, "Services Marketing Integrating Customer Focus Across the Firm", Tata McGraw Hill, 2010.
5. Urvashi Makkar and Harinder Kumar Makkar, "CRM Customer Relationship Management", McGraw Hill Education, 2013.

Outcomes:

Students will be able to:

- *Explore the concepts of customer relationship management with industry case studies.*
- *Develop metrics for customer retention.*
- *Apply data mining concepts to implement CRM in real world applications.*
- *Devise strategies to implement CRM in any organization.*

SEMESTER 2: ELECTIVES

CA619 PRINCIPLES OF DEEP LEARNING

Objectives:

- *To acquire knowledge on the basics of neural networks.*
- *To implement neural networks using computational tools for variety of problems.*
- *To explore various deep learning algorithms.*

Basics of Deep leaning- Deep learning architectures: Convolutional Neural Networks : Neurons in Human Vision-The Shortcomings of Feature Selection-Vanilla Deep Neural Networks Don't Scale-Filters and Feature Maps-Full Description of the Convolutional Layer-Max Pooling-Full Architectural Description of Convolution Networks-Closing the Loop on MNIST with Convolutional Networks-Image Preprocessing Pipelines Enable More Robust Models-Accelerating Training with Batch Normalization-Building a Convolutional Network for CIFAR-10-Visualizing Learning in Convolutional Networks-Leveraging Convolutional Filters to Replicate Artistic Styles-Learning Convolutional Filters for Other Problem Domains-Training algorithms.

Memory Augmented Neural Networks : Neural Turing Machines-Attention-Based Memory Access-NTM Memory Addressing Mechanisms-Differentiable Neural Computers-Interference-Free Writing in DNCs-DNC Memory Reuse-Temporal Linking of DNC Writes-Understanding the DNC Read Head-The DNC Controller Network-Visualizing the DNC in Action-Implementing the DNC in TensorFlow-Teaching a DNC to Read and Comprehend.

Deep Reinforcement Learning: Deep Reinforcement Learning Masters Atari Games-What Is Reinforcement Learning?-Markov Decision Processes (MDP)-Explore Versus Exploit-Policy versus Value Learning-Pole-Cart with Policy Gradients-Q-Learning and Deep Q-Networks-Improving and Moving Beyond DQN.

Implementing Neural Networks in TensorFlow : What Is TensorFlow?-How Does TensorFlow Compare to Alternatives?-Installing TensorFlow-Creating and Manipulating TensorFlow Variables-TensorFlow Operations-Placeholder Tensors-Sessions in TensorFlow-Navigating Variable Scopes and Sharing Variables-Managing Models over the CPU and GPU-Specifying the Logistic Regression Model in TensorFlow-Logging and Training the Logistic Regression Model-Leveraging TensorBoard to Visualize

Computation Graphs and Learning-Building a Multilayer Model for MNIST in TensorFlow.

Applications: Deep learning for computer vision, Deep Learning Applications at the Enterprise Scale, Deep Learning Models for Healthcare Applications.

Reference:

1. Nikhil Buduma, Nicholas Locascio, "Fundamentals of Deep Learning: Designing Next-Generation Machine Intelligence Algorithms", O'Reilly Media, 2017.
2. Ian Goodfellow, Yoshua Bengio, Aaron Courville, "Deep Learning (Adaptive Computation and Machine Learning series", MIT Press, 2017.

Outcomes:

Students will be able to:

- *Develop algorithms simulating human brain.*
- *Implement Neural Networks in Tensor Flow for solving problems.*
- *Explore the essentials of Deep Learning and Deep Network architectures.*
- *Define, train and use a Deep Neural Network for solving real world problems that require artificial Intelligence based solutions.*

CA620 IMAGE AND VIDEO ANALYTICS

Objectives:

- *To teach the fundamentals of digital image processing, image and video analysis.*
- *To understand the real time use of image and video analytics.*
- *To demonstrate real time image and video analytics applications and others.*

Digital image representation- Visual Perception- Sampling and Quantization- Basic Relations between Pixels- Mathematical Tools Used in Digital Image Processing: Fundamental Operations –Vector and Matric Operations- Image Transforms (DFT, DCT, DWT, Hadamard).

Fundamentals of spatial filtering: spatial correlation and convolution-smoothing-blurring- sharpening- edge detection - Basics of filtering in the frequency domain: smoothing-blurring- sharpening--Histograms and basic statistical models of image.

Colour models and Transformations – Image and Video segmentation-Image and video demonising- Image and Video enhancement- Image and Video compression.

Object detection and recognition in image and video-Texture models Image and Video

classification models- Object tracking in Video.

Applications and Case studies- Industrial- Retail- Transportation & Travel- Remote sensing-Video Analytics in WSN: IoT Video Analytics Architectures.

References:

1. R.C. Gonzalez and R.E. Woods." Digital Image Processing". 3rd Edition. Addison Wesley, 2007.
2. W. Härdle, M. Müller, S. Sperlich, A. Werwatz, "Nonparametric and Semi parametric Models", Springer, 2004.
3. Rick Szelisk, "Computer Vision: Algorithms and Applications", Springer 2011.
4. Jean-Yves Dufour, "Intelligent Video Surveillance Systems", Wiley, 2013.
5. Caifeng Shan, Fatih Porikli, Tao Xiang, Shaogang Gong, "Video Analytics for Business Intelligence", Springer, 2012.
6. AsierPerallos, Unai Hernandez-Jayo, Enrique Onieva, Ignacio Julio García Zuazola, "Intelligent Transport Systems: Technologies and Applications", Wiley, 2015.
7. Basudeb Bhatta, "Analysis of Urban Growth and Sprawl from Remote Sensing Data", Springer, 2010

Outcomes:

Students will be able to:

- *Describe the fundamental principles of image and video analysis and have an idea of their application.*
- *Apply image and video analysis in real world problems.*

CA621 SOCIAL NETWORKING AND MINING

Objectives:

- *To understand the components of the social network.*
- *To model and visualize the social network.*
- *To mine the users in the social network.*
- *To understand the evolution of the social network.*
- *To mine the interest of the user.*

Introduction- Introduction to Web - Limitations of current Web – Development of Semantic Web – Emergence of the Social Web – Statistical Properties of Social Networks -Network analysis - Development of Social Network Analysis - Key concepts and measures in network analysis - Discussion networks - Blogs and online communities - Web-based networks.

Modeling And Visualization- Visualizing Online Social Networks - A Taxonomy of

Visualizations - Graph Representation - Centrality- Clustering - Node-Edge Diagrams - Visualizing Social Networks with Matrix-Based Representations- Node-Link Diagrams - Hybrid Representations - Modelling and aggregating social network data – Random Walks and their Applications –Use of Hadoop and Map Reduce - Ontological representation of social individuals and relationships.

Mining Communities- Aggregating and reasoning with social network data- Advanced Representations - Extracting evolution of Web Community from a Series of Web Archive - Detecting Communities in Social Networks - Evaluating Communities – Core Methods for Community Detection & Mining - Applications of Community Mining Algorithms - Node Classification in Social Networks.

Text and Opinion Mining- Text Mining in Social Networks -Opinion extraction – Sentiment classification and clustering - Temporal sentiment analysis - Irony detection in opinion mining - Wish analysis - Product review mining – Review Classification – Tracking sentiments towards topics over time.

Tools for Social Network Analysis- UCINET – PAJEK – ETDRAW – StOCNET – Splus – R – NodeXL – SIENA and RSIENA – Real world Social Networks (Facebook- Twitter- etc.)

References:

1. Charu C. Aggarwal, "Social Network Data Analytics", Springer; 2011.
2. Peter Mika, "Social Networks and the Semantic Web", 1st edition, Springer, 2007.
3. BorkoFurht, "Handbook of Social Network Technologies and Applications", 1st edition, Springer, 2010.
4. GuandongXu, Yanchun Zhang and Lin Li, "Web Mining and Social Networking – Techniques and applications", 1st edition, Springer, 2011.
5. Giles, Mark Smith, John Yen, "Advances in Social Network Mining and Analysis", Springer, 2010.
6. Ajith Abraham, Aboul Ella Hassanien, VáclavSnáel, "Computational Social Network Analysis: Trends, Tools and Research Advances", Springer, 2009.
7. Toby Segaran, "Programming Collective Intelligence", O'Reilly, 2012.
8. Sule Gündüz-Ogüdücü, A. Şima Etaner-Uyar, "Social Networks: Analysis and Case Studies", Springer, 2014.

Outcomes:

Students will be able to:

- *Work on the internal components of the social network.*
- *Model and visualize the social network.*
- *Mine the behavior of the users in the social network.*
- *Predict the possible next outcome of the social network.*
- *Mine the opinion of the user.*

CA622 WEB INTELLIGENCE

Objectives:

- *To know the importance of qualitative data, get insights and techniques.*
- *To develop customer-centric approach in dealing with data.*
- *To know the principles, tools and methods of web intelligence.*
- *To apply analytics for business situations.*

Web Analytics – Basics – Traditional Ways – Expectations – Data Collection – Clickstream Data – Weblogs – Beacons – JavaScript Tags – Packet Sniffing – Outcomes data – Competitive data – Search Engine Data.

Qualitative Analysis – Customer Centricity – Site Visits – Surveys – Questionnaires – Website Surveys – Post visits – Creating and Running- Benefits of surveys – Critical components of successful strategy.

Web Analytic concepts – URLS – Cookies – Time on site – Page views – Understand standard reports – Website content quality – Navigation reports (top pages, top destinations, site overlay). – Search Analytics – Internal search, SEO and PPC – Measuring Email and Multichannel Marketing - Competitive intelligence and Web 2.0 Analytics – Segmentation – Connectable reports.

Google Analytics: Analytics - Cookies - Accounts vs Property - Tracking Code - Tracking Unique Visitors - Demographics - Page Views & Bounce Rate Acquisitions - Custom Reporting.

Goals & Funnels – Filters - Ecommerce Tracking - Real Time Reports - Customer Data Alert - Adwords Linking - Adsense Linking -Attribution Modeling - Segmentation - Campaign Tracking - Multi-Channel Attribution.

References:

1. Avinash Kaushik, "Web Analytics 2.0: The Art of Online Accountability and Science Of Customer Centricity ", 1st edition, Sybex, 2009.
2. Michael Beasley, "Practical Web Analytics for User Experience: How Analytics can help you Understand your Users", Morgan Kaufmann, 2013.
3. Magy Seif El-Nasr, Anders Drachen, Alessandro Canossa, eds., "Game Analytics: Maximizing the Value of Player Data", Springer, 2013.
4. Bing Liu, "Web Data Mining: Exploring Hyperlinks, Content, and Usage Data", 2nd Edition, Springer, 2011.
5. Justin Cutroni, "Google Analytics", O'Reilly, 2010.
6. Eric Fettman, Shiraz Asif, Feras Alhlou , "Google Analytics Breakthrough", John Wiley & sons, 2016.

Outcomes:

Students will be able to:

- *Know the concepts and terminologies related to web analytics.*
- *Explore various parameters used for web analytics and their impact.*
- *Explore the use of tools and techniques of web analytics.*
- *Get experience on websites, web data insights and conversions.*

CA623 INTERNET OF THINGS

Objectives:

- *To understand the fundamentals of internet of things.*
- *To provide knowledge about IoT devices, applications and examples.*

- *To acquire skills to program the embedded devices and connecting them to the web and cloud.*

Internet of things: Overview, technology of the internet of things, enchanted objects, Design principles for connected devices, Privacy, Web thinking for connected devices

Writing Code: building a program and deploying to a device, writing to Actuators, Blinking Led, Reading from Sensors, Light Switch, Voltage Reader, Device as HTTP Client, HTTP, Push Versus Pull

Pachube, Netduino, Sending HTTP Requests—the Simple Way, Sending HTTP Requests—the Efficient Way

HTTP: Device as HTTP Server, Relaying Messages to and from the Netduino, Request Handlers, Web Html, Handling Sensor Requests, Handling Actuator Requests

Going Parallel: Multithreading, Parallel Blinker, prototyping online components, using an API, from prototypes to reality, business models, ethics, privacy, disrupting control, crowdsourcing

References:

1. Adrian McEwen and Hakim Cassimally, "Designing the Internet of Things", John Wiley & Sons, 2013.
2. Cuno Pfister, "Getting Started with the Internet of Things: Connecting Sensors and Microcontrollers to the Cloud", Maker Media, 2011.
3. Rob Barton, Gonzalo Salgueiro, David Hanes, "IoT Fundamentals: Networking Technologies, Protocols, and Use Cases for the Internet of Things", Cisco Press, 2017.
4. Radomir Mihajlovic, Muthu Ramachandran, Reinhold Behringer, Petar Kocovic "Emerging Trends and Applications of the Internet of Things", IGI Global, 2017.
5. Hwaiyu Geng, "Internet of Things and Data Analytics Handbook", John Wiley & Sons, 2017.
6. Marco Schwartz, "Internet of Things with Arduino Cookbook", Packt Publishing, 2016.

Outcomes:

Students will be able to:

- *Program embedded devices.*
- *Program simple actuators and sensors.*
- *Build client programs that push sensor readings from a device to a web service.*

CA624 HEALTHCARE DATA ANALYTICS

Objectives:

- *To explore the various forms of electronic health care information.*
- *To learn the techniques adopted to analyse health care data.*
- *To understand the predictive models for clinical data*

Introduction: Introduction to Healthcare Data Analytics- Electronic Health Records– Components of EHR- Coding Systems- Benefits of EHR- Barrier to Adopting EHR- Challenges- Phenotyping Algorithms.

Analysis: Biomedical Image Analysis- Mining of Sensor Data in Healthcare- Biomedical Signal Analysis- Genomic Data Analysis for Personalized Medicine.

Analytics: Natural Language Processing and Data Mining for Clinical Text- Mining the Biomedical- Social Media Analytics for Healthcare.

Advanced Data Analytics: Advanced Data Analytics for Healthcare– Review of Clinical Prediction Models- Temporal Data Mining for Healthcare Data- Visual Analytics for Healthcare- Predictive Models for Integrating Clinical and Genomic Data- Information Retrieval for Healthcare- Privacy-Preserving Data Publishing Methods in Healthcare.

Applications: Applications and Practical Systems for Healthcare– Data Analytics for Pervasive Health- Fraud Detection in Healthcare- Data Analytics for Pharmaceutical Discoveries- Clinical Decision Support Systems- Computer-Assisted Medical Image Analysis Systems- Mobile Imaging and Analytics for Biomedical Data.

References:

1. Chandan K. Reddy and Charu C Aggarwal, "Healthcare data analytics", Taylor & Francis, 2015
2. Hui Yang and Eva K. Lee, "Healthcare Analytics: From Data to Knowledge to Healthcare Improvement, Wiley, 2016.

Outcomes:

Students will be able to:

- *Analyse health care data using appropriate analytical techniques.*
- *Apply analytics for decision making in healthcare services.*
- *Apply data mining to integrate health data from multiple sources and develop efficient clinical decision support systems.*

CA625 LINKED OPEN DATA AND SEMANTIC WEB

Objectives:

- *To understand the fundamentals of linked open data, its representation and applications.*
- *To learn the design considerations for linked data and technologies behind in*

publishing and consuming linked data on applications.

- *To understand the fundamental concepts, advantages and limitations of the Semantic Web and its related techniques and tools.*
- *To learn the basics of ontology and use ontology engineering approaches in semantic applications.*

Introduction- Introduction to Linked Data (LD) and the Semantic Web- visions and basic concepts- focusing on linked "entities" on the Web - The Rationale for Linked Data Structure Enables Sophisticated Processing - Hyperlinks Connect Distributed Data - From Data Islands to a Global Data Space - Introducing Big Lynx Productions - Principles of Linked Data - The Principles in a Nutshell - Naming Things with URIs - URIs dereferencable - Providing Useful RDF Information - The RDF Data Model - RDF Serialization Formats - Including Links to other Things - Relationship Links - Identity Links - Vocabulary Links.

The Web of Data- - Bootstrapping the Web of Data - Topology of the Web of Data – Cross-Domain Data - Geographic Data - Media Data - Government Data - Libraries and Education - Life Sciences Data - Retail and Commerce - User Generated Content and Social Media- Linked Data Design Considerations- Using URIs as Names for Things - Describing Things with RDF - Literal Triples and Outgoing Links - Publishing Data about Data - Choosing and Using Vocabularies - Making Links with RDF - Making Links within a Data Set - Making Links with External Data Sources - Setting RDF Links Manually - Auto-generating RDF Links.

Publishing Linked Data- Linked Data Publishing Patterns - The Recipes - Serving Linked Data as Static RDF/XML Files - Serving Linked Data as RDF Embedded in HTML Files - Serving RDF and HTML with Custom Server-Side Scripts - Serving Linked Data from Relational Databases - Serving Linked Data from RDF Triple Stores - Serving Linked Data by Wrapping Existing Application or Web APIs - Linked Data Publishing Checklist- Consuming Linked Data- Deployed Linked Data Applications - Generic Applications - Domain-specific Applications - Developing a Linked Data Mashup - Architecture of Linked Data Applications - Effort Distribution between Publishers- Consumers and Third Parties.

Semantic Web- Introduction to Semantic Data - Semantic modeling- Modeling for Human Communication- Explanation and Prediction- Mediating Variability- Expressivity in Modeling -RDF The basis of the Semantic Web - Semantic Web application architecture- RDF Parser/Serializer- RDF Store - Querying the Semantic Web-SPARQL- SPARQL—Query Language for RDF- Advanced Features of SPARQL - RDF and inferencing.

RDFS and Ontology- RDF schema- The RDF Schema Language - RDFS-Plus - Using RDFS-Plus in the wild - SKOS—managing vocabularies with RDFS-Plus - Introduction to Ontology- OWL-Web Ontology Language- Basic OWL- Examples include BIBO- FOAF- Good Relations- CIDOC-CRM- DPLA - Counting and sets in OWL - Ontologies on the Web—putting it all together - Ontology Mapping - Good and bad modeling practices - Expert modeling in OWL The future of the Semantic Web.

References:

1. Tom Heath and Christian Bizer, "Linked Data: Evolving the Web into a Global Data Space - Synthesis Lectures on the Semantic Web: Theory and Technology", 1st Edition, Morgan & Claypool, 2011.
2. Dean Allemang and James Hendle, "Semantic Web for the Working Ontologist: Effective Modeling in RDFS and OWL", Second Edition, Elsevier, 2011.
3. Bob DuCharme, "Learning SPARQL: Querying and Updating SPARQL 1.1", Second Edition, O'Reilly Media, 2013.
4. Grigoris Antoniou, Paul Groth, Frank van Harmelen and Rinke Hoekstra, "Semantic Web Primer, Third Edition, MIT Press, 2012.
5. Rajendra Akerkar, "Foundations of the Semantic Web", Narosa Publishing House, New Delhi and Alpha Science Intern, 2009.
6. Leslie Sikos, "Mastering Structured Data on the Semantic Web: From HTML5 Microdata to Linked Open Data ", Apress, 2015.
7.

Outcomes:

Students will be able to:

- *Describe the fundamentals of linked open data, its representation and advantages*
- *Explain the design considerations for linked data and technologies behind in publishing and consuming linked data on applications.*
- *Explore fundamental concepts, advantages and limitations of the Semantic Web and its related techniques and use various tools for constricting applications.*
- *Use ontology engineering approaches in developing semantic applications.*

CA626 FINANCIAL RISK ANALYTICS AND MANAGEMENT

Objectives:

- *To identify the different risks involved in Finance arena.*
- *To understand and solve the different risks pertaining to stock market and its instruments.*
- *To analyze the legal issues affecting the business.*

Introduction to Risk -Understanding Risk- Nature of Risk, Source of Risk, Need for risk management, Benefits of Risk Management, Risk Management approaches. Risk Classification- credit risk, market risk, operational risk and other risk

Risk Measurements -Measurement of Risk – credit risk measurement, market risk measurement, interest rate risk measurement, Asset liability management, measurement of operational risk

Risk Management- Risk management- Managing credit risk, managing operational risk, managing market risk, insurance

Risk in Instruments -Tools for risk management – Derivatives, combinations of derivative instruments, Neutral and volatile strategies, credit derivatives, credit ratings, swaps

Regulation and Other Issues: Other issues in risk management – Regulatory framework, Basel committee, legal issues, accounting issues, tax issues, MIS and reporting,

integrated risk management

References:

1. Dun, Bradstreet, “Financial Risk Management”, TMH, 2006.
2. John C Hull, “Risk management and Financial Institutions”, Pearson, 2015.
3. Aswath Damodharan, “Strategic Risk Taking”, Pearson, 2008.

Outcomes:

Students will be able to:

- *Identify and categorize the various risks faced by an organization.*
- *Explore the tools and practices needed to assess and evaluate financial risks.*
- *Explore risk management practices in an industry.*
- *Identify and solve legal issues that impact financial and other risk affecting business.*

CA627 LOGISTICS AND SUPPLY CHAIN MANAGEMENT

Objective:

- *To create an understanding of design and management of supply chain networks.*
- *To analyze an existing supply chain of a company.*
- *To understand and apply the various supply chain management concepts.*

Understanding of Supply Chain :Objectives of a supply chains - decision phases - stages of supply chain-supply chain process view- cycle view of supply chain process - push/pull view of supply chain processes - key issues in SCM - Supply chain drivers and obstacles - - inventory- transportation- facilities and information – Overview of Retail/FMCG industries.

Inventory Management :A framework for structuring drivers in supply chain - supply chain strategies - strategic fit - Obstacles to achieve strategic fit - value of information - Role of cycle inventory- Economics of scale to exploit fixed costs and discounts- cycle time related costs- levels of safety- single stage inventory model- risk pooling- centralized and decentralized systems of planning inventory in supply chain.

Network Planning and supply chain Integration: Network design- warehouse location- service level requirements- integrating inventory positioning and network design- supply chain integration. Push-pull and pull-push type systems- demand driven strategies- Impact of internet on supply chain strategies- Transportation in a supply chain- facilities affecting transportation decision- modes of transportation and their performance characteristics.

Distribution strategies and strategic alliances: Introduction- Centralized vs. decentralized control- direct shipment- cross docking- push based vs. pull based supply chain- third party logistics (3PL) - Retailer-Supplier relationship issues- requirements- success and failures- distributor integration types and issues.

Global logistics and Risk management :Agile supply chains- Introduction to global SCM- risk management- issues in international SCM- regional differences in logistics- design for logistics- supplier integration in to new product development- pricing issues and smart pricing. IT and Business processes in supply chain.

References

1. Christopher Martin, "Logistics And Supply Chain Management", Pearson Education Asia, 2002
2. Kapoor K.K., Kansal Purva, "Marketing Logistics: A Supply Chain Approach", Pearson Education Asia, 2003
3. David Simchi-Levi, Ravi Shankar, "Designing And Managing Supply Chain Concepts- Strategies and Case Studies", 3rd Edition, McGraw Hill Publication, 2008.
4. Sunil Chopra, Peter Meindl, "Supply Chain Management: Strategy, Planning, and Operation", 6 edition, Pearson, 2014

Outcomes:

The students will be able to

- *Configure a supply chain network for an organization from a global perspective.*
- *Analyze an existing supply chain of a company.*
- *Apply various supply chain management concepts.*
- *Improve an existing supply chain and design an efficient supply chain in alignment with the strategic goals of the company.*