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

<table>
<thead>
<tr>
<th>Semester</th>
<th>Subject Code</th>
<th>Subject Name</th>
<th>Credits</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>CA601</td>
<td>Statistical Computing</td>
<td>3</td>
</tr>
<tr>
<td>I</td>
<td>CA603</td>
<td>Big Data Analytics</td>
<td>3</td>
</tr>
<tr>
<td>I</td>
<td>CA605</td>
<td>Machine Learning Techniques</td>
<td>3</td>
</tr>
<tr>
<td>I</td>
<td>*****</td>
<td>Elective -1</td>
<td>3</td>
</tr>
<tr>
<td>I</td>
<td>*****</td>
<td>Elective-2</td>
<td>3</td>
</tr>
<tr>
<td>I</td>
<td>*****</td>
<td>Elective-3</td>
<td>3</td>
</tr>
<tr>
<td>I</td>
<td>CA609</td>
<td>Big Data Management and Data Analytics Lab</td>
<td>2</td>
</tr>
<tr>
<td>II</td>
<td>CS618</td>
<td>Real Time Systems</td>
<td>3</td>
</tr>
<tr>
<td>II</td>
<td>CA602</td>
<td>Next Generation Databases</td>
<td>3</td>
</tr>
<tr>
<td>II</td>
<td>CA604</td>
<td>High Performance Computing</td>
<td>3</td>
</tr>
<tr>
<td>II</td>
<td>*****</td>
<td>Elective -4</td>
<td>3</td>
</tr>
<tr>
<td>II</td>
<td>*****</td>
<td>Elective -5</td>
<td>3</td>
</tr>
<tr>
<td>II</td>
<td>*****</td>
<td>Elective -6</td>
<td>3</td>
</tr>
<tr>
<td>II</td>
<td>CA610</td>
<td>Machine Learning Lab</td>
<td>2</td>
</tr>
<tr>
<td>III</td>
<td>CA647</td>
<td>Project work-Phase I</td>
<td>12</td>
</tr>
<tr>
<td>IV</td>
<td>CA648</td>
<td>Project work-Phase II</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td><strong>Total Credits</strong></td>
<td></td>
<td><strong>64</strong></td>
</tr>
</tbody>
</table>
## List of Electives

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


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.


References:

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.


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

References:

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.


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


References:

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:
     o A large textual file containing one sentence per line
     o A small file containing a set of stop words (One stop word per line)
   • Output:
     o 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
     o 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
o 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:
    o Transposes the original Amazon food dataset, obtaining a PairRDD of the type: <user_id> → <list of the product_ids reviewed by user_id>
    o Counts the frequencies of all the pairs of products reviewed together;
    o 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:

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.


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


References:

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.


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:
2. Ananth Grama and George Karypis, “Introduction to parallel computing”, Addison
-Wesley, 2009.
5. Michael Quinn, “Parallel Programming in C with MPI and OpenMP”, Indian edition,

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:


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


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.


References:

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.


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.


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

References:

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:
3. Tanja Falkowski, "Community Analysis in Dynamic Social Networks", (Dissertation), University Magdeburg, 2009.

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 Bioinformatics from computing perspective.
• To comprehend bioinformatics databases, file formats and its applications.
• To understand the applications of Bioinformatics


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:
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 bioinformatics databases.

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.


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

References:

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.


Data Productization


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:

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.


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.


References:

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.

Challenges.


References:


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.


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.


Reference:

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.


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:


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.


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

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.


References:

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.


References:

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.

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:

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

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.


References:

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.


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.


References:

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:

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.

References

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.