



# RAJALAKSHMI ENGINEERING COLLEGE

An AUTONOMOUS Institution  
Affiliated to ANNA UNIVERSITY, Chennai

## **CURRICULUM AND SYLLABUS**

### **DATA SCIENCE (Minor Degree)**

#### **Vision**

- To be a Department of Excellence in Information Technology Education, Research and Development.

#### **Mission**

- To train the students to become highly knowledgeable in the field of Information Technology.
- To promote continuous learning and research in core and emerging areas.
- To develop globally competent students with strong foundations, who will be able to adapt to changing technologies.

## **PROGRAMME OUTCOMES (POs)**

- PO1:** Graduates should be able to learn how to interpret data, extracts meaningful information, and assesses findings.
- PO2:** Graduates should be capable of demonstrating and developing a design of mastery over the key technologies in data science and business analytics such as structured/unstructured data mining, machine learning, visualization techniques, predictive modeling and statistics.
- PO3:** Graduates should be capable of applying ethical principles and responsibilities during Professional practice.
- PO4:** Graduates should be able to function effectively as a team member and to write/ present a substantial technical report / document.
- PO5:** Graduates should independently carry out research / investigation and development work to solve industry and organization-specific problems and challenges using advanced analytics and computational methods.
- PO6:** Graduates should be able to engage in independent and life-long learning in the broadest context of technological change.

## Data Science (Minor Degree)

Now a day's industries, such as social media, healthcare, insurance, e-commerce, transport, government, banking, telecommunications, etc., that are producing massive amounts of data, the so-called –BIG DATA, at a remarkable scale. This has led to converting unprocessed data to useful information to make decisions. In response, the **Department of Information Technology at Rajalakshmi Engineering College** offers Data Science as minor degree program for the other branch of students.

**Data Science (Minor Degree)** course is designed to meet the needs of the Data professionals to grasp the industrial requirements and create data scientists within short time. Data Science program is to create professional who can mine and interpret data and making sense for industrial applications.

**Prerequisite:**

**Probability and Statistics, Programming**

**Eligible Departments: Except Computing and Information Science Disciplines**

### Curriculum

#### SEMESTER V,VI,VII

S. No	COURSE CODE	COURSE TITLE	CATEGORY	CONTACT PERIODS	L	T	P	C
<b>Lab oriented Theory</b>								
1	IT19P65	Python for Data Science	PC	4	2	0	2	3
2	IT19P83	Data Warehousing and Data Mining	PC	3	3	0	0	3
3	IT19P61	Advanced R and python programming	PC	4	2	0	2	3
<b>Theory</b>								
4	IT19P75	Deep Learning Techniques	PC	3	3	0	0	3
5	IT19741	Cloud and big data analytics	PC	4	3	0	2	3
6	OIT1901	Business Intelligence	PC	3	3	0	0	3
<b>Total</b>				<b>21</b>	<b>14</b>	<b>0</b>	<b>8</b>	<b>18</b>

Subject Code	Subject Name	Category	L	T	P	C
IT19P65	PYTHON FOR DATA SCIENCE	PE	2	0	2	3

**Objectives:**

- To understand the basics of Data Science
- To learn the basics of Python
- To learn Sequence data types and associated operations
- To learn and implement Data visualization techniques
- To understand the case studies and implement various machine learning techniques

<b>UNIT-I</b>	<b>INTRODUCTION TO DATA SCIENCE</b>	<b>6</b>
Introduction - Data Science, Descriptive Statistics, Exploratory Data Analysis, Visualization Before Analysis, Dirty Data, visualizing a Single Variable, Examining Multiple Variables, Data Exploration Versus Presentation- Statistical Methods for Evaluation, Hypothesis Testing, Difference of Means, ANOVA		
<b>UNIT-II</b>	<b>BASICS OF PYTHON SPYDER (TOOL/ANACONDA TOOL)</b>	<b>6</b>
Introduction Spyder/Anaconda - Setting working Directory - Creating and saving a script file - File execution, clearing console, removing variables from environment, clearing environment - Commenting script files - Variable creation – Arithmetic and logical operators - Data types and associated operations		
<b>UNIT-III</b>	<b>SEQUENCE DATA TYPES AND ASSOCIATED OPERATIONS</b>	<b>6</b>
Sequence Data types - Strings, Lists, Arrays, Tuples , Dictionary , Sets , Range - NumPy – and Array		
<b>UNIT-IV</b>	<b>DATA VISUALIZATION</b>	<b>6</b>
Pandas dataframe and dataframe related operations - Reading files, Exploratory data analysis, Data preparation and preprocessing- Data visualization - Scatter plot , Line plot , Bar plot , Histogram , Box plot , Pair plot - Control structures- if-else family , for loop , for loop with if break , while loop – Functions		
<b>UNIT-V</b>	<b>CASE STUDY</b>	<b>6</b>
Case studies - Regression – Classification – Clustering		

**Contact Hours**                      **30**    **3**  
**0**

**List of Experiments**

- 1 Implement basic Python programs for reading input from console
- 2 Apply Python built-in data types: Strings, List, Tuples, Dictionary, Set and their methods to solve any given problem
- 3 Handle numerical operations using math and random number functions
- 4 Import a CSV file and perform various Statistical and Comparison operations on rows/columns.
- 5 Create NumPy arrays from Python Data Structures, Intrinsic NumPy objects and Random Functions.
- 6 Import any CSV file to Pandas DataFrame and perform the following:
  - (a) Visualize the first and last 10 records
  - (b) Get the shape, index and column details
  - (c) Select/Delete the records(rows)/columns based on conditions.
  - (d) Perform ranking and sorting operations.
  - (e) Do required statistical operations on the given columns.
  - (f) Find the count and uniqueness of the given categorical values.

- 7 (g) Rename single/multiple columns.
- 7 Load an image file and do crop and flip operation using NumPy Indexing.
- 8 Implement Data visualization techniques and visualize data using any plotting framework
- (a) Scatter plot (b) Line plot (c) Bar plot (d) Histogram (e) Box plot (f) Pair plot
- 9 Implementation of Linear Regression and Logistic Regression.
- 10 Implementation of classification techniques – SVM and Decision tree.
- 11 Implementation of clustering techniques – K-Mean and Hierarchical methods.

**Course Outcomes:**

On completion of course students will be able to

- CO1: Gain knowledge on the basics of Data science
- CO2: Apply programming knowledge of Python in Data Science
- CO3: Apply the sequence data types and do the associated operations
- CO4: Analyze the data using various Visualization techniques
- CO5: Build analytical models using Machine learning techniques

**Reference Books(s):**

1. Wes McKinney, "Python for Data Analysis", O'Reilly Media.2012
2. Sebastian Raschka, "Python Machine Learning", Packpub.com,2015
3. <https://www.datacamp.com/courses/statistical-thinking-in-python-part-1>

CO/PO	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7	PO 8	PO 9	PO 10	PO 11	PO 12	PS O1	PS O2	PS O3	PS O4
CO1	2	1	3	3	3	1	3	-	-	-	-	1	3	3	1	1
CO2	2	1	3	3	3	1	3	-	-	-	-	1	3	3	1	1
CO3	2	1	3	3	3	1	3	-	-	-	-	1	3	3	1	1
CO4	2	1	3	3	3	1	3	-	-	-	-	1	3	3	1	1
CO5	2	1	3	3	3	1	3	-	-	-	-	1	3	3	1	1
CO (Avg.)	2.00	1.00	3.00	3.00	3.00	1.00	3.00	-	-	-	-	1.00	3.00	3.00	1.00	1.00

Correlation levels 1, 2 or 3 are as defined below:

- 1: Slight (Low)
- 2: Moderate (Medium)
- 3: Substantial (High)
- Nocorrelation : “-”

Subject Code	Subject Name (Theory Course)	Category	L	T	P	C
IT19P83	DATA WAREHOUSING AND DATA MINING	PE	3	0	0	3

**Objectives:** Broad objective of this course is to

- ★Learn the concepts of Data Warehousing and Business Analysis.
- ★Familiar with the concepts of Data Mining.
- ★Understand the concepts of Association and Correlations Algorithms.
- ★Understand the concepts of Classification Algorithms.
- ★Understand the concepts of Clustering and outlier Analysis.

**UNIT-I DATA WAREHOUSING 9**

Data Warehouse: Basic Concepts, A Multitiered Architecture, Data Warehouse Models, Metadata Repository- Data Warehouse Modelling: Data Cube and OLAP, Data Cube: A Multidimensional Data Model: Schemas-Concept Hierarchies-OLAP Operations. (TB1-CH: 4)

**UNIT-II DATA MINING AND VISUALIZATION 9**

Introduction: Kinds of Data, Kinds of Patterns- Data Objects and Attribute Types- Data Visualization: Pixel-Oriented, Geometric Projection, Icon-Based, Hierarchical, Visualizing Complex Data and Relations- Data Pre-processing. (TB1-CH: 1&2)

**UNIT-III ASSOCIATIONS AND CORRELATIONS 9**

Basic Concepts: Frequent Itemsets , Closed Itemsets, and Association Rules - Frequent Item set Mining Methods: Finding Frequent Itemsets by Confined Candidate Generation , Growth Approach for Mining Frequent Itemsets , Mining Frequent Itemsets Using Vertical Data Format , Mining Closed and Max Patterns - Interesting Patterns: Pattern Evaluation Methods. (TB1-CH: 6)

**UNIT-IV CLASSIFICATION 9**

Basic Concepts- Decision Tree Induction: ID3- Bayes Classification Methods: Bayes' Theorem, Naive Bayesian Classification- Classification by Back propagation- Support Vector Machines-Techniques to improve classification accuracy-Prediction. (TB1-CH: 8)

**UNIT-V CLUSTER ANALYSIS AND DATA MINING APPLICATIONS 9**

Cluster Analysis- Partitioning Methods- Hierarchical Methods: Agglomerative versus Divisive Hierarchical Clustering- Density-Based Methods: DBSCAN- Grid-Based Methods: STING: Statistical Information Grid-Outlier Detection-Data Mining Applications: Science and Engineering-Data Mining Tools: Weka & R -Web Mining-Emerging Trends in Data Mining. (TB1-CH: 10)

**Contact Hours : 45 Periods**

**Course Outcomes:**

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

- CO1 Apply the Data Warehousing and Business Analytics concepts.
- CO2 Apply the concepts of Data Mining to large data sets.
- CO3 Make use of Association and Correlations Algorithms.
- CO4 Compare and Contrast the various classifiers.
- CO5 Apply Clustering and outlier Analysis and to solve Data Mining Case Studies.

**Text Books:**

- 1 Jiawei Han and Micheline Kamber, "Data Mining Concepts and Techniques", Third Edition, Elsevier, 2012.

**Reference Books:**

- 1 Pang-Ning Tan, Michael Steinbach and Vipin Kumar, "Introduction to Data Mining", Person Education, 2007.
- 2 K.P. Soman, Shyam Diwakar and V. Aja, "Insight into Data Mining Theory and Practice", Eastern Economy Edition, Prentice Hall of India, 2006.
- 3 G. K. Gupta, "Introduction to Data Mining with Case Studies", Eastern Economy Edition, Prentice Hall of India, 2006.
- 4 Daniel T.Larose, "Data Mining Methods and Models", Wiley-Interscience, 2006.
- 5 Alex Berson and Stephen J.Smith, "Data Warehousing, Data Mining and OLAP", Tata McGraw – Hill Edition, Thirteenth Reprint 2008.

CO/PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3	PSO4
CO1	3	2	-	2	-	2	1	-	-	-	2	2	-	-	1	2
CO2	3	2	-	2	-	2	1	-	-	-	2	1	-	-	1	2
CO3	2	2	-	2	-	2	2	-	-	-	3	2	-	2	2	2
CO4	1	1	-	2	2	2	1	-	-	-	2	2	-	2	2	2
CO5	-	-	-	2	2	2	-	-	-	-	2	2	-	-	2	2
CO (Avg)	2.25	1.75	-	2.00	2.00	2.00	1.25	-	-	-	2.20	1.80	-	2.00	1.60	2.00

Correlation levels 1, 2 or 3 are as defined below:

1: Slight (Low) 2: Moderate (Medium) 3: Substantial(High)

No correlation : "-"

Subject Code	Subject Name	Category	L	T	P	C
IT19P61	ADVANCED R AND PYTHON PROGRAMMING	PE	2	0	2	3

**Objectives:** Broad objective of this course is

- ★ To know the basics of python programming
- ★ To learn the concept of python modules and frameworks.
- ★ To understand the different types of database connectivity using python.
- ★ To get the basic idea of R programming.
- ★ To implement machine learning algorithm using python data science package.

**UNIT-I INTRODUCTION TO PYTHON PROGRAMMING 6**

Basics of Python-Starting with python interpreter-Control flow statements-Functions-List,tuple,Dictionary-Files-Modules –Packages - Regular expressions – Exception handling- Python OOPS Concepts – Classes and Objects-Constructor-Inheritance

**UNIT-II PYTHON MODULE AND FRAMEWORKS 6**

Python Modules- Collection Module- Math Module –OS Module –Random Module –Sys Module- Statistics Module – JSON Module- -Numpy,Scipy and Pandas modules- Python Frameworks:Django,Flask,Pytorch,Web2PY,Pygame.

**UNIT-III PYTHON DATA BASE CONNECTIVITY 6**

Python MySQL Database Connectivity :Environment Setup-Create Data Base Connection-Create Database-Creating Tables-Insert,select,update,drop,join operations-Performing Transactions-Introduction to Python and MongoDB connectivity-Introduction to Python SQLite Connectivity.

**UNIT-IV INTRODUCTION TO R 6**

OverviewDataTypes,Variables,Operator,DecisionMaking,Loops,Functions,Strings,Vector,List, Matrices, Arrays,Factors,Data Frames,Packages-Data Visualization – Data Interfaces- Reading and writing of CSV files - Exploratory Data Analysis using R- Statistical Methods for Evaluation using R.

**UNIT-V DATA SCIENCE USING R 6**

Association –Classification – Clustering – Time series Analysis-Text Analysis–Prediction Algorithm-Image Analytics – Video Analytics – Data base analytics

**Total Contact Hours : 30**

**List of Experiments**

- 1 Implement python programs using interactive and script mode.
- 2 Implement various control statements in python.
- 3 Implement python programs using packages.
- 4 Implement python OOPs concepts using programs
- 5 Implement python modules, Collection Module, Math Module, Pandas module using simple programs
- 6 Develop programs using Python frameworks such as Django, Flask etc.
- 7 Implement Python MySQL Database Connectivity
- 8 Implement MangoDB connectivity
- 9 Implement Python SQLite connectivity
- 10 Implementation of vector data objects operations
- 11 Study and implementation of various control structures in R
- 12 Study and implementation of Data Visualization with ggplot2
- 13 CLASSIFICATION MODEL
  - a. Install relevant package for classification.
  - b. Choose classifier for classification problem.
  - c. Evaluate the performance of classifier

**CLUSTERING MODEL**

- 14
  - a. Clustering algorithms for unsupervised classification.
  - b. Plot the cluster data using R visualizations.



**Contact Hours** : 15  
**Total Contact Hours** : 45

**Course Outcomes:**

On completion of course students will be able to

CO1 Learn the major concept of python programming and oops principles.

CO2 Understand the concept of Python modules and frameworks

CO3 Connect the different types of database with Python

CO4 Learn the basic idea of R programming.

CO5 Implement the supervised and unsupervised algorithm concept using R data science library.

**Reference Books(s):**

- 1 MARK LUTZ, David Ascher,, “Learning Python”, 2nd Edition, O’REILLY Media Inc, 2003.
- 2 Sebastian Raschka,”Python Machine Learning”,2nd Edition, 2017 Packt Publishing.
- 3 Andreas C. Müller, Sarah Guido,” Introduction to Machine Learning with Python: A Guide for Data Scientists”, 2018 Release O’REILLY Media Inc.
- 4 Garrett Golemund,”Hands on Programming with R”, 2nd Edition, O’REILLY Media Inc
- 5 Hadley Wickham, Garrett Golemund,”R for Data Science”, O’REILLY Media Inc.

CO/PO	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7	PO 8	PO 9	PO1 0	PO1 1	PO1 2	PSO 1	PSO 2	PSO 3	PSO 4
CO1	2	3	3	3	2	1	-	-	-	-	1	2	2	1	2	2
CO2	2	3	3	3	3	2	-	-	-	-	1	2	2	1	2	2
CO3	2	3	3	3	3	2	-	2	2	-	1	2	2	1	2	2
CO4	2	3	3	3	3	2	-	-	-	-	1	2	2	1	2	2
CO5	2	3	3	3	3	2	-	2	2	-	3	2	2	1	2	2
CO(Avg)	2.00	3.00	3.00	3.00	2.80	1.80	-	2.00	2.00	-	1.40	2.00	2.00	1.00	2.00	2.00

Correlation levels 1, 2 or 3 are as defined below:

1: Slight (Low) 2: Moderate (Medium) 3: Substantial(High)

No correlation : “-”

Subject Code	Subject Name (Theory course)	Category	L	T	P	C
IT19P75	DEEP LEARNING TECHNIQUES	PE	3	0	0	3

**Objectives:**

- ★ To get familiar with basic concepts and theory
- ★ To understand the basics of Neural Networks
- ★ To be familiar with CNN
- ★ To have deep knowledge in optimizers
- ★ To know about advanced deep architectures

**UNIT-I INTRODUCTION 9**

Introduction – Feature Descriptor, Bayesian Learning, Discriminant Function, Linear Classifier, Support Vector Machine, Linear Machine, Multiclass Support Vector Machine, Optimization – Optimization Techniques in Machine Learning, Non-Linear Functions.

**UNIT-II BASICS OF NEURAL NETWORKS 9**

Introduction to Neural Networks -Multilayer Perceptron, Backpropagation Learning, Loss Function, Backpropagation Learning Examples, Autoencoders – Autoencoders Training, Autoencoders Variants

**UNIT-III CONVOLUTIONAL NEURAL NETWORKS 9**

Convolutional – Cross Correlation, CNN Architecture, MLP versus CNN, LeNet, AlexNet, VGG16 and Transfer Learning, Vanishing and Exploding Gradient

**UNIT-IV OPTIMIZATION TECHNIQUES 9**

GoogleNet, ResNet, Optimizers – Momentum Optimizer, Momentum and Nesterov Accelerated Gradient (NAG) Optimizer, Adagrad Optimizer, RMSProp, AdaDelta and Adam Optimizer. Normalization – Batch Normalization, Group Normalization, Training Tricks – Regularization, Early Stopping, Dropouts

**UNIT-V ADVANCED DEEP ARCHITECTURES 9**

Recurrent Neural Networks (RNN), Advanced RNNs – LSTM, GRU Generative Adversarial Networks (GAN), Advanced GANs, Case Study: Face Recognition, Image Denoising and Image Classification.

**Total Contact Hours : 45**

**Course Outcomes:**

On completion of course students will be able to

- CO1 Know the basic concepts and theory
- CO2 Have knowledge in Neural Networks
- CO3 Implement CNN
- CO4 Have more knowledge in deep learning optimizers
- CO5 Know the advanced deep architectures

**Text Book(s):**

- 1 Goodfellow, I., Bengio, Y., and Courville, A., Deep Learning, MIT Press, 2016
- 2 <https://nptel.ac.in/courses/106/105/106105215/>

**Reference Books(s):**

- 1 Bishop, C. ,M., Pattern Recognition and Machine Learning, Springer, 2006.
- 2 Yegnanarayana, B., Artificial Neural Networks PHI Learning Pvt. Ltd, 2009.
- 3 Golub, G.,H., and Van Loan,C.,F., Matrix Computations, JHU Press,2013.
- 4 Satish Kumar, Neural Networks: A Classroom Approach, Tata McGraw-Hill Education, 2004.

CO/PO	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7	PO 8	PO 9	PO1 0	PO1 1	PO1 2	PSO 1	PSO 2	PSO 3	PSO 4
CO1	3	2	1	2	2	-	-	-	-	-	-	2	-	1	1	3
CO2	2	3	2	3	2	-	-	-	-	-	-	3	-	1	2	3
CO3	2	3	3	3	3	-	-	-	-	-	-	2	-	1	-	3
CO4	2	2	2	2	1	-	-	-	-	-	-	-	1	1	3	3
CO5	2	2	3	3	2	-	-	-	-	-	-	3	1	1	3	3
CO(Avg)	2.20	2.40	2.20	2.60	2.00	-	-	-	-	-	-	2.50	1.00	1.00	2.25	3.00

Correlation levels 1, 2 or 3 are as defined below:

1: Slight (Low) 2: Moderate (Medium) 3: Substantial (High) No correlation : “-”

Subject Code	Subject Name (Lab oriented Theory Course)	Category	L	T	P	C
IT19741	CLOUD AND BIG DATA ANALYTICS	PC	3	0	4	5

### Objectives:

- To understand and appreciate the evolution of cloud from the existing technologies.
- To be familiar with cloud computing and have knowledge on the various service models.
- To introduce the concepts of Big Data and Hadoop, and implement map reduce.
- To describe the data stream analytics methodologies.
- To visualize and analyze various data visualization techniques.

### UNIT-I CLOUD ENABLING TECHNOLOGIES 9

Technologies for Network-Based Systems - System Models for Distributed and Cloud Computing - Implementation Levels of Virtualization - Virtualization Structures/Tools and Mechanisms - Virtualization of CPU, Memory, and I/O Devices - Virtual Clusters and Resource Management - Virtualization for Data-Center Automation.

### UNIT-II CLOUD ARCHITECTURE AND SERVICES 9

Layered Cloud Architecture Design - NIST Cloud Computing Reference Architecture - Public, Private and Hybrid Clouds - IaaS – PaaS – SaaS - Architectural Design of Compute and Storage Clouds -Public Cloud Platforms: GAE, AWS, and Azure.

### UNIT-III INTRODUCTION TO BIG DATA AND HADOOP 9

Introduction to Big Data, Types of Digital Data, Challenges of conventional systems - Web data, Evolution of analytic processes and tools, Analysis Vs reporting - Big Data Analytics, Introduction to Hadoop - Distributed Computing Challenges - History of Hadoop, Hadoop Eco System - Use case of Hadoop – Hadoop Distributors – HDFS – Processing Data with Hadoop – Map Reduce.

### UNIT-IV MINING DATA STREAMS 9

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 - Realtime Analytics Platform(RTAP) applications - case studies - real time sentiment analysis, stock market predictions.

### UNIT-V DATA VISUALIZATION 9

Input for Visualization: Data and Tasks - Encoding Data with Marks and Channels - Creating a Scatter Plot - Common Visualization Idioms: Bar Chart - Vertical & Horizontal - Pie Chart and Coxcomb Plot -Line Chart - Area Chart - Making Maps - Visualizing Trees and Networks - Data Reduction: Histograms - Hexbin Mapping – Crossfiltering.

**Total Contact Hours : 45**

### List of Experiments

- : Install Virtualbox/VMware Workstation with different flavours of linux or windows OS on top of windows7 or8.
- : Install a C compiler in the virtual machine created using virtual box and execute Simple Programs
- : Install Google App Engine. Create hello world app and other simple web applications using python/java
- : Use GAE launcher to launch the web applications
- : Install Hadoop single node cluster and run simple applications like word count
- : Implement the following Substitution and Transposition Techniques:
  - a) Caesar Cipher
  - b) Playfair Cipher
  - c) Hill Cipher
  - d) Vigenere Cipher
  - e) Rail fence – row & Column Transformation
  - f) Affine Cipher
- : Implement the following algorithms
  - a) DES

- b) RSA Algorithm
  - c) Diffie-Hellman
  - d) MD5
  - e) SHA-1
- 8 Implement the Digital Signature Algorithm (DSA).
- 9 Create a visualization of the dataset you chose for your project using Tableau
- a) Bar Chart, Vertical & Horizontal
  - b) Pie Chart and Coxcomb Plot
  - c) Line Chart
  - d) Area Chart
- 10 Create a histogram or aggregated bar chart of your project dataset. Using Tableau

**Course Outcomes:**

On completion of course students will be able to

CO1 Learn the key and enabling technologies that help in the development of cloud.

CO2 Develop the ability to understand and use the architecture of compute and storage cloud, service and delivery models.

CO3 Understand the usage scenarios of Big Data Analysis and Hadoop framework and apply Mapreduce over HDFS.

CO4 Apply stream data models.

CO5 Use various data visualization techniques.

**Text Book(s):**

- 1 Kai Hwang, Geoffrey C. Fox, Jack G. Dongarra, "Distributed and Cloud Computing, From Parallel Processing to the Internet of Things", Morgan Kaufmann Publishers, 2012.
- 2 Seema Acharya, Subhasini Chellappan, "Big Data Analytics" Wiley India; First Edition , 2015.
- 3 Visualization Analysis and Design, Tamara Munzner, AK Peters Visualization Series, CRC Press, Nov. 2014.
- 4 Anand Rajaraman and Jeffrey David Ullman, "Mining of Massive Datasets", Cambridge University Press, First Edition , 2012.
- 5 Jiawei Han, Micheline Kamber "Data Mining Concepts and Techniques", Second Edition, Elsevier, Reprinted 2008.

**Reference Books(s):**

1 Rittinghouse, John W., and James F. Ransome, —Cloud Computing: Implementation, Management and Security, CRC Press, 2017.

2 Rajkumar Buyya, Christian Vecchiola, S. ThamaraiSelvi, —Mastering Cloud Computing, Tata Mcgraw Hill, 2013.

3 Jay Liebowitz, "Big Data and Business Analytics" Auerbach Publications, CRC press First Edition, 2013.

4 Tom White "Hadoop: The Definitive Guide" Third Edition, O'Reilly Media, 2012.

5 Bill Franks, "Taming the Big Data Tidal Wave: Finding Opportunities in Huge Data Streams with Advanced Analytics", John Wiley & sons, First Edition, 2012.

<i>CO/PO</i>	<i>PO 1</i>	<i>PO 2</i>	<i>PO 3</i>	<i>PO 4</i>	<i>PO 5</i>	<i>PO 6</i>	<i>PO 7</i>	<i>PO 8</i>	<i>PO 9</i>	<i>PO1 0</i>	<i>PO1 1</i>	<i>PO1 2</i>	<i>PSO 1</i>	<i>PSO 2</i>	<i>PSO 3</i>	<i>PSO 4</i>
<b>CO1</b>	-	-	2	3	3	-	2	-	-	-	-	2	2	3	-	2
<b>CO2</b>	-	-	2	2	3	-	2	-	-	-	-	2	2	3	2	2
<b>CO3</b>	-	-	2	2	3	-	2	-	-	-	-	2	1	3	-	2
<b>CO4</b>	-	2	2	3	3	-	2	-	-	-	-	2	1	3	-	2
<b>CO5</b>	1	2	2	3	3	-	2	-	-	-	-	2	1	3	2	2
<b>CO(Avg)</b>	1.00	2.00	2.00	2.60	3.00	-	2.00	-	-	-	-	2.00	1.40	3.00	2.00	2.00

Correlation levels 1, 2 or 3 are as defined below:

1: Slight (Low) 2: Moderate (Medium) 3:

Substantial(High)No correlation : “-”

Subject Code	Subject Name (Theory Course)	Category	L	T	P	C
OIT1901	BUSINESS INTELLIGENCE	OE	3	0	0	3

**Objectives:** Broad objective of this course is

- ★ Be exposed with the basic rudiments of business intelligence system
- ★ understand the modeling aspects behind Business Intelligence
- ★ understand of the business intelligence life cycle and the techniques used in it
- ★ Be exposed with different data analysis tools and techniques
- ★

**UNIT-I BUSINESS INTELLIGENCE 9**

Effective and timely decisions – Data, information and knowledge – Role of mathematical models – Business intelligence architectures: Cycle of a business intelligence analysis – Enabling factors in business intelligence projects – Development of a business intelligence system – Ethics and business intelligence.

**UNIT-II KNOWLEDGE DELIVERY 9**

The business intelligence user types, Standard reports, Interactive Analysis and Ad Hoc Querying, Parameterized Reports and Self-Service Reporting, dimensional analysis, Alerts/Notifications, Visualization: Charts, Graphs, Widgets, Scorecards and Dashboards, Geographic Visualization, Integrated Analytics, Considerations: Optimizing the Presentation for the Right Message.

**UNIT-III EFFICIENCY 9**

Efficiency measures – The CCR model: Definition of target objectives- Peer groups – Identification of good operating practices; cross efficiency analysis – virtual inputs and outputs – Other models. Pattern matching – cluster analysis, outlier analysis.

**UNIT-IV BUSINESS INTELLIGENCE APPLICATIONS 9**

Marketing models – Logistic and Production models –Telecommunication Industry & Banking Case studies.

**UNIT-V FUTURE OF BUSINESS INTELLIGENCE 9**

Future of business intelligence – Emerging Technologies, Machine Learning, Predicting the Future, BI Search & Text Analytics – Advanced Visualization – Rich Report, Future beyond Technology.

**Contact Hours : 45 Periods**

**Course Outcomes:**

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

- CO1: Explain the fundamentals of business intelligence.
- CO2: Link data mining with business intelligence.
- CO3: Apply various modeling techniques.
- CO4: Explain the data analysis and knowledge delivery stages.
- CO5: Apply business intelligence methods to various situations and decide on appropriate technique.

**Text Books:**

- 1 Efraim Turban, Ramesh Sharda, Dursun Delen, “Decision Support and Business Intelligence Systems”, 9<sup>th</sup> Edition, Pearson 2013.

**Reference Books:**

- 1 Larissa T. Moss, S. Atre, “Business Intelligence Roadmap: The Complete Project Lifecycle of Decision Making”, Addison Wesley, 2003.
- 2 Carlo Vercellis, “Business Intelligence: Data Mining and Optimization for Decision Making”, Wiley Publications, 2009.
- 3 David Loshin Morgan, Kaufman, “Business Intelligence: The Savvy Manager’s Guide”, Second Edition, 2012.
- 4 Cindi Howson, “Successful Business Intelligence: Secrets to Making BI a Killer App”, McGraw- Hill, 2007.
- 5 Ralph Kimball, Margy Ross, Warren Thornthwaite, Joy Mundy, Bob Becker, “The Data Warehouse Lifecycle Toolkit”, Wiley Publication Inc., 2007.
- 6 G. Shmueli, N.R. Patel, P.C. Bruce, “Data Mining for Business Intelligence: Concepts, Techniques, and Applications in Microsoft Office Excel with XLMiner”, 1<sup>st</sup> Edition, Wiley India.

CO/PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3	PSO4
CO1	2	2	2	2	2	1	-	-	-	-	1	1	2	2	2	2
CO2	2	3	3	2	2	1	-	-	-	-	1	1	2	2	2	2
CO3	2	3	3	3	3	1	-	-	-	-	1	1	2	2	2	2
CO4	2	3	3	3	3	1	-	-	-	-	1	1	2	2	2	2
CO5	2	3	3	3	3	1	1	1	-	-	1	3	3	3	3	3
CO (Avg)	2.00	2.80	2.80	2.60	2.60	1.00	1.00	1.00	-	-	1.00	1.40	2.20	2.20	2.20	2.20

Correlation levels 1, 2 or 3 are as defined below:

1: Slight (Low) 2: Moderate  
(Medium) 3: Substantial(High)

No correlation : “-”