

## List of New Courses Introduced in the Academic Year 2017-2018

### EVEN SEM

S.No	Course Code	Name of the Course
1	SIT1301	Data Mining and Data Warehousing
2	SIT1303	Data Analytics
3	SIT1304	Cloud Computing
4	SIT1619	Internet of Things
5	SCS1619	R Programming
6	SIT4302	Cloud Computing Lab

SIT1301	DATA MINING AND DATA WAREHOUSING	L	T	P	Credits	Total Marks
		3	0	0	3	100

### UNIT 1 DATA WAREHOUSING

9 Hrs.

Data warehousing components - Building a data warehouse - Multi Dimensional Data Model - OLAP Operation in the Multi-Dimensional Model - Three Tier Data Warehouse Architecture - Schemas for Multi-dimensional data Model - Online Analytical Processing (OLAP) - OLAP Vs OLTP Integrated OLAM and OLAP Architecture.

### UNIT 2 DATA MINING

9 Hrs.

Introduction - Data - Types of data - Steps in KDD - System Architecture - Data mining functionalities - Classification of data mining systems - Data mining task primitives - Integration of a data mining system with a data warehouse - Issues - Data pre-processing - Data Mining Application

### UNIT 3 ASSOCIATION RULE MINING

9 Hrs.

Mining frequent patterns - Associations and correlations - Mining methods - Finding Frequent itemset using Candidate Generation - Generating Association Rules from Frequent Itemsets - Mining Frequent itemset without Candidate Generation - Mining various kinds of association rules - Mining Multi-Level Association Rule-Mining Multi-Dimensional Association Rule-Mining Correlation analysis - Constraint based association mining.

### UNIT 4 CLASSIFICATION AND PREDICTION

9 Hrs.

Classification and prediction - Issues Regarding Classification and Prediction - Classification by Decision Tree Induction - Bayesian classification - Baye's Theorem - Naïve Bayesian Classification - Bayesian Belief Network - Rule based classification - Classification by Backpropagation - Support vector machines - Prediction - Linear Regression - Non Linear Regression.

### UNIT 5 CLUSTERING, APPLICATIONS AND TRENDS IN DATA MINING

9 Hrs.

Cluster analysis - Types of data in Cluster Analysis - Categorization of major clustering methods -- Partitioning methods - K Means - K Medoids - Hierarchical methods - Density-based methods - Grid-based methods - Model based clustering methods -- Constraint Based cluster analysis - Outlier analysis - Data mining applications.

Max. 45 Hours

### COURSE OUTCOMES

On completion of the course, student will be able to

- CO1 : Comprehend knowledge about the transactional processing & analytical processing concepts
- CO2 : Analyze the characteristics of data and pre-process to make it ready for analysis.
- CO3 : Build a model for analysis by identifying the correlation between the attributes.
- CO4 : Estimate the future trends by relating the past trends and present scenario..
- CO5 : Evaluate various supervised learning algorithms.
- CO6 : Compare and contrast Supervised and unsupervised learning approaches.

### TEXT / REFERENCE BOOKS

1. Jiawei Han and Micheline Kamber, "Data Mining Concepts and Techniques", 2nd Edition, Elsevier, 2007
2. Alex Berson and Stephen J. Smith, "Data Warehousing, Data Mining & OLAP", Tata McGraw Hill, 2007.
3. Pang-Ning Tan, Michael Steinbach and Vipin Kumar, "Introduction To Data Mining", Person Education, 2007.
4. K.P. Soman, Shyam Diwakar and V. Ajay, "Insight into Data mining Theory and Practice", Easter Economy Edition, Prentice Hall of India, 2006.
5. G. K. Gupta, "Introduction to Data Mining with Case Studies", Easter Economy Edition, Prentice Hall of India, 2006.
6. Daniel T.Larose, "Data Mining Methods and Models", Wile-Interscience, 2006.

### END SEMESTER EXAM QUESTION PAPER PATTERN

Max. Marks : 100

Exam Duration : 3 Hrs.

PART A : 10 questions of 2 marks each - No choice

20 Marks

PART B : 2 questions from each unit of internal choice, each carrying 16 marks

80 Marks

SIT1303	DATA ANALYTICS	L	T	P	Credits	Total Marks
		3	0	0	3	100

**UNIT 1 INTRODUCTION TO DATA ANALYTICS**

**10 Hrs.**

Introduction-Data Analytics, Data mining and Knowledge discovery-Data and Relations-Dissimilarity measures - similarity measures - sequence relations - sampling and quantization - Analysis vs Reporting - Modern data analytic tools - Statistical Concepts - Probability - Sampling and Sampling distribution - statistical Inference - Prediction and prediction error – Re sampling.

**UNIT 2 DATA ANALYSIS**

**10 Hrs.**

Linear Correlation - Regression Modeling - Multivariate analysis - Bayesian modeling - Inference and Bayesian networks - Support vector and Kernel Methods - Analysis of time series: Linear system analysis, nonlinear dynamics - Rule induction - Basics of Fuzzy and Neural networks.

**UNIT 3 CLASSIFICATION AND CLUSTERING METHODS**

**10 Hrs.**

Classification criteria - Naive Bayes classifier - Linear Discriminant Analysis - Support Vector Machine - Nearest neighbor classifier - Learning vector Quantization - Decision Trees - Cluster Partitions - Sequential clustering - Prototype Based Clustering - Fuzzy Clustering - Relational Clustering - Self Organizing Map.

**UNIT 4 DATA VISUALIZATION**

**8 Hrs.**

Introduction - Classification of Visual data analysis techniques - Data type to be visualized - Visualization techniques - Interaction techniques - Principle Component Analysis, Multi dimensional Scaling, Sammon Sampling, Histograms - Spectral analysis

**UNIT 5 FRAMEWORKS**

**7 Hrs.**

Distributed File system - Physical organization of Computer nodes - Large scale file system organization - Map Reduce - The map task - Grouping and Aggregation - The reduce task - Combiners - Details of map reduce Execution - Coping with Node failures.

**Max. 45 Hours**

**COURSE OUTCOMES:**

On completion of the course, student will be able to

- CO1:Comprehend the fundamental concepts of data analytics
- CO2:Analyze classification algorithms and choose the suitable algorithm for application under consideration
- CO3:Explore different clustering techniques
- CO4:Acquire an appreciation for the data analytics and present them using visualization techniques
- CO5:Develop frame works for data analytics
- CO6:Generate new ideas and innovations in Data analysis.

**TEXT / REFERENCE BOOKS**

1. Michael Berthold, David J. Hand, Intelligent Data Analysis, Springer, 2007
2. Thomas A. Runkler, Data Analytics, Models and Algorithms for Intelligent data Analysis, Springer 2012
3. Anand Rajaraman and Jeffery David Ullman, Mining of massive data sets, Cambridge University press, 2012
4. Bill Franks, Taming the Big data tidal wave: Finding opportunities in Huge Data Streams with advanced analytics, John Wiley & sons, 2012
5. Thomas H. Davenport, Jinho Kim Keeping up with the Quants: Your Guide to understanding and using Analytics, Harvard Business Press books, 2013

**END SEMESTER EXAM QUESTION PAPER PATTERN**

**Max. Marks : 100**

**Exam Duration: 3 Hrs.**

**PART A:** 2 Questions from each unit, each carrying 2 marks

20 Marks

**PARTB:** 2 Questions from each unit with internal choice, each carrying 16 marks

80 Marks

<b>SIT1304</b>	<b>CLOUD COMPUTING</b>	<b>L</b>	<b>T</b>	<b>P</b>	<b>Credits</b>	<b>Total Marks</b>
		<b>3</b>	<b>0</b>	<b>0</b>	<b>3</b>	<b>100</b>

#### **UNIT 1 UNDERSTANDING CLOUD COMPUTING**

**9 Hrs.**

History of Cloud computing - Cloud Computing Architectural Framework - Types of Clouds - pros and cons of cloud computing - difference between web 2.0 and cloud - key challenges in cloud computing - Major Cloud players - Cloud Deployment Models - Virtualization in Cloud Computing - types of virtualization - Parallelization in Cloud Computing - cloud resource management - dynamic resource allocation - Optimal allocation of cloud models.

#### **UNIT 2 CLOUD SERVICE MODELS**

**9Hrs.**

Software as a Service (SaaS) - Infrastructure as a Service (IaaS)- Platform as a Service (PaaS)- Service Oriented Architecture (SoA) - Elastic Computing - On Demand Computing.

#### **UNIT 3 CLOUD DEPLOYMENT MODELS**

**9Hrs.**

Deployment of applications on the cloud - Hypervisor - Case studies - Xen, VMware, Eucalyptus, Amazon EC2, KVM, Virtual Box, Hyper-V.

#### **UNIT 4 CLOUD COMPUTING FOR EVERYONE**

**9 Hrs.**

Cloud data centers - Energy efficiency in data centre - Mobile cloud computing service models - Collaboration with services and applications: CRM management - Project management - Email - on line database - calendar - schedules - Word Processing - Presentation - Spreadsheet - Databases - Desktop - Social Networks and Groupware.

#### **UNIT 5 CLOUD SECURITY**

**9 Hrs.**

Cloud security - Security threats and solutions in clouds - Auditing protocols - dynamic auditing - storage security - Privacy preserving -- Fully Homomorphic Encryption - big data security - Cloud availability - DoS attacks - Fault tolerance management in cloud computing - Cloud computing in India.

**Max. 45 Hours**

#### **COURSE OUTCOMES**

On completion of the course, student will be able to

CO1 : Comprehend the type, architecture and features of a given Cloud setup with its suitability for a given application.

CO2 : Compare the types of cloud services and connect to the service requirements of an application.

CO3 : Deploy an application on suitable cloud and make use of the suitable virtualization tool to create cloud infrastructure.

CO4 : Evaluate a cloud data centre and measure its energy efficiency.

CO5 : Appraise the Cloud security and the associated risks to protect the online and cloud infrastructure from external threats and third party attacks.

CO6 : Manage and collaborate cloud services with dynamic auditing.

#### **TEXT / REFERENCE BOOKS**

1. Cloud computing a practical approach - Anthony T.Velte, Toby J. Velte Robert Elsenpeter TATA McGraw - Hill, New Delhi - 2010
2. Cloud Computing: Web-Based Applications That Change the Way You Work and Collaborate Online - Michael Miller - Que 2008
3. Cloud Security: A Comprehensive Guide to Secure Cloud Computing, Ronald L. Krutz, Russell Dean Vines, Wiley-India, 2010

#### **END SEMESTER EXAM QUESTION PAPER PATTERN**

**Max. Marks : 100**

**Exam Duration : 3 Hrs.**

**PART A** : 10 questions of 2 marks each - No choice

20 Marks

**PART B** : 2 questions from each unit of internal choice, each carrying 16 marks

80 Marks

<b>SIT1619</b>	<b>INTERNET OF THINGS</b>	<b>L</b>	<b>T</b>	<b>P</b>	<b>Credits</b>	<b>Total Marks</b>
		<b>3</b>	<b>0</b>	<b>0</b>	<b>3</b>	<b>100</b>

**UNIT 1**

**9 Hrs.**

Introduction to IoT, Current technological trends and future prospects, - Evolution of IoT - IoT and related terms – Business Scope, Relation with embedded system, Introduction to Arduino and Raspberrypi, Overview of Elements of IoT - Basic Architecture of an IoT.

**UNIT 2**

**9 Hrs.**

Application Sensors & Actuators - Edge Networking (WSN) – Gateways - IoT Communication Model – WPAN & LPWA, IoT platform for available applications, Hardware Devices: Arduino, Raspberry pi and Smartwifti, etc, Wearable Development Boards, Softwares, Programs and Stacks available for building IoT applications, Installation of various packages necessary for project and list of tools.

**UNIT 3**

**9 Hrs.**

IoT Communication Model, Cloud computing in IoT, IoT in cloud architecture, Logging on to cloud, Selecting and Creating cloud service ,Managing Cloud Account Credentials and Generating Key Pair, Creating IoT Cloud Service.

**UNIT 4**

**9 Hrs.**

Big Data Analytics – Data Exploration, Data Visualization, Visualization tools for IoT.

**UNIT 5**

**9 Hrs.**

DIY Kits – Soil moisture monitoring, Weather monitoring, Air quality Monitoring, Movement Detection, etc, IFTTT and other apps.

**Max. 45 Hrs.**

**COURSE OUTCOMES:**

On completion of the course, student will be able to

- CO1 : Illustrate the technology and current trends in Internet of things.
- CO2 : Understand the various elements of IoT system and hardware devices.
- CO3 : Learn the programming languages and platforms for building IoT applications
- CO4 : Demonstrate cloud computing and its relevance for developing IoT applications.
- CO5 : Implement IoT applications that manage big data with data analytics and visualization tools.
- CO6 : Implement hands-on projects using an appropriate software and hardware devices in various applications.

**TEXT / REFERENCE BOOKS**

1. The Internet of Things: Applications and Protocols, Wiley publications. Author(s): Oliver Hersent, David Boswarthick, Omar Elloumi.
2. Architecting the Internet of Things, Springer publications. Author(s):Dieter Uckelmann, Mark Harrison, Florian Michahelles.
3. Internet of Things with Arduino Cookbook, Packt Publications. Author(s): Marco Schwatz.
4. Internet of Things and Data Analytics, Wiley Publications.

**END SEMESTER EXAM QUESTION PAPER PATTERN**

**Max. Marks: 100**

**Exam Duration: 3 Hrs.**

**PART A:** 2 Questions from each unit, each carrying 2 marks

20 Marks

**PART B:** 2 Questions from each unit with internal choice, each carrying 16 marks

80 Marks

<b>SCS1621</b>	<b>R PROGRAMMING</b>	<b>L</b>	<b>T</b>	<b>P</b>	<b>Credits</b>	<b>Total Marks</b>
		<b>3</b>	<b>0</b>	<b>0</b>	<b>3</b>	<b>100</b>

**UNIT 1**

**9 Hrs.**

Introduction to R - History and fundamentals of R, Installation and use of R / R Studio / R Shiny, Installing R packages, R – Nuts and Bolts -Getting Data In and Out - Control Structures and Functions- Loop Functions- Data Manipulation- String Operations- Matrix Operations.

**UNIT 2**

**9 Hrs.**

R Data interfaces - CSV Files, XML files, Web Data- Data Preprocessing: Missing Values, Principle Component Analysis - Data Visualization – Charts & Graphs-Pie Chart, Bar Chart, Box plot, Histogram, Line graph, Scatter Plot.

**UNIT 3**

**9 Hrs.**

Statistical Modeling in R - Descriptive statistics-R Packages: Regression (MASS package) - Distribution (STATS package) - ANOVA - Time Series Analysis.

**UNIT 4**

**9 Hrs.**

Machine Learning in R - Classification: Decision Trees, Random Forest, SVM – Clustering - Association Rule Mining - Outlier Detection.

**UNIT 5**

**9 Hrs.**

Overview of R Shiny - R Hadoop - Case Study - Hypothesis Generation, Importing Data set and Basic Data Exploration, Feature Engineering, Model Building..

**Max. 45 Hrs.**

**COURSE OUTCOMES**

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

- CO1 : Install R Packages and develop applications.
- CO2 : Demonstrate Data Preprocessing Algorithms and Visualization methods on datasets in R.
- CO3 : Apply the Statistical Models in R.
- CO4 : Demonstrate Machine Learning Algorithms in R.
- CO5 : Create applications in R Frameworks.
- CO6 : Develop Engineering Models for Hypothesis Generation and Data Extraction.

**TEXT / REFERENCE BOOKS**

1. Hands-On Programming with R: Write Your Own Functions and Simulations By Garrett Golemund, O'Reilly Media, Inc., 2014.
2. R for Data Science, Hadley Wickham, Garrett Golemund," O'Reilly Media, Inc.2016.
3. Introduction to Statistics and Data Analysis - With Exercises, Solutions and Applications in R By Christian Heumann, Michael Schomaker and Shalabh, Springer, 2016
4. The R Software-Fundamentals of Programming and Statistical Analysis -Pierre Lafaye de Micheaux, Rémy Drouilhet, Benoit Liquet, Springer 2013
5. A Beginner's Guide to R (Use R) By Alain F. Zuur, Elena N. Ieno, Erik H.W.G. Meesters, Springer 2009

**END SEMESTER EXAM QUESTION PAPER PATTERN**

Max. Marks : 100

PART A: 2 Questions from each unit, each carrying 2 marks

PARTB: 2 Questions from each unit with internal choice, each carrying 16 marks

Exam Duration: 3 Hrs.

20 Marks

80 Marks