Ref No:

Sri Krishna Institute of Technology,
Bangalore

COURSE PLAN

Academic Year 2019-2020

Program:	ISE
Semester :	8
Course Code:	15CS82
Course Title:	Big Data Analytics
Credit / L-T-P:	04/4-0-0
Total Contact Hours:	50
Course Plan Author:	Tejashwini N

Academic Evaluation and Monitoring Cell

Sri Krishna Institute of Technology
#29,Chimney hills,Hesaraghata Main road, Chikkabanavara Post
Bangalore – 560090, Karnataka, INDIA

Phone / Fax :08023721477/28392221/23721315
Web: www.skit.org.in , e-mail: skitprinci@gmail.com

Table of Contents

A. COURSE INFORMATION

1. Course Overview

Degree:	BE	Program:	IS
Semester:	8	Academic Year:	2019-2020
Course Title:	Big Data Analytics	Course Code:	15CS82
Credit / L-T-P:	04/4-0-0	SEE Duration:	180 minutes
Total Contact Hours:	50	SEE Marks:	80 marks
CIA Marks:	20	Assignment	1\Module
Course Plan Author:	Tejashwini N	Sign ..	Dt:
Checked By:		Sign ..	Dt:
CO Targets	CIA Target :	SEE Target:	

Note: Define CIA and SEE % targets based on previous performance.

2. Course Content

Content / Syllabus of the course as prescribed by University or designed by institute.

Module	Content	Teaching Hours	Blooms Learning Levels
1	Hadoop Distributed File System Basics, Running Example Programs and Benchmarks, Hadoop MapReduce Framework, MapReduce Programming	10	L3
2	Essential Hadoop Tools, Hadoop YARN Applications, Managing Hadoop with Apache Ambari, Basic Hadoop Administration Procedures	10	L3
3	Business Intelligence Concepts and Application, Data Warehousing, Data Mining, Data Visualization	10	L3
4	Decision Trees, Regression, Artificial Neural Networks, Cluster Analysis, Association Rule Mining	10	L3
5	Text Mining, Naïve-Bayes Analysis, Support Vector Machines, Web Mining, Social Network Analysis	10	L3
-	**Total**	**50**	

3. Course Material

Books & other material as recommended by university (A, B) and additional resources used by course teacher (C).

1. Understanding: Concept simulation / video ; one per concept ; to understand the concepts ; 15 – 30 minutes
2. Design: Simulation and design tools used – software tools used ; Free / open source
3. Research: Recent developments on the concepts – publications in journals; conferences etc.

Modules	Details	Chapters in book	Availability
A	**Text books (Title, Authors, Edition, Publisher, Year.)**	**-**	**-**
	Douglas Eadline,"Hadoop 2 Quick-Start Guide: Learn the Essentials of Big Data Computing in the Apache Hadoop 2 Ecosystem", 1stEdition, Pearson Education, 2016. ISBN-13: 978-9332570351		
	Anil Maheshwari, "Data Analytics", 1st Edition, McGraw Hill Education, 2017. ISBN-13: 978-9352604180		
B	**Reference books (Title, Authors, Edition, Publisher, Year.)**	**-**	**-**
	Tom White, "Hadoop: The Definitive Guide", 4th Edition, O'Reilly Media, 2015.ISBN-13: 978-9352130672		
	Boris Lublinsky, Kevin T.Smith, Alexey Yakubovich,"Professional Hadoop Solutions", 1stEdition, Wrox Press, 2014ISBN-13: 978-8126551071		

	Eric Sammer,"Hadoop Operations: A Guide for Developers and Administrators",1 stEdition, O'Reilly Media, 2012.ISBN-13: 978-9350239261		
C	**Concept Videos or Simulation for Understanding**	**-**	**-**
C1			
C2			
C3			
C4			
C5			
D	**Software Tools for Design**	**-**	**-**
	HADOOP		
E	**Recent Developments for Research**	**-**	**-**
F	**Others (Web, Video, Simulation, Notes etc.)**	**-**	**-**
1			

4. Course Prerequisites

Refer to GL01. If prerequisites are not taught earlier, GAP in curriculum needs to be addressed. Include in Remarks and implement in B.5.
Students must have learnt the following Courses / Topics with described Content . . .

Modules	Course Code	Course Name	Topic / Description	Sem	Remarks	Blooms Level
1	15CS 562	Artificial Intelligence	Artificial Neural networks	5		

5. Content for Placement, Profession, HE and GATE

The content is not included in this course, but required to meet industry & profession requirements and help students for Placement, GATE, Higher Education, Entrepreneurship, etc. Identifying Area / Content requires experts consultation in the area.
Topics included are like, a. Advanced Topics, b. Recent Developments, c. Certificate Courses, d. Course Projects, e. New Software Tools, f. GATE Topics, g. NPTEL Videos, h. Swayam videos etc.

Modules	Topic / Description	Area	Remarks	Blooms Level

B. OBE PARAMETERS

1. Course Outcomes

Expected learning outcomes of the course, which will be mapped to POs.

Modules	Course Code.#	Course Outcome **At the end of the course, student should be able to . . .**	Teach. Hours	Instr Method	Assessment Method	Blooms' Level
1	15CS82.1	Apply hadoop Distributed File System Basics to MapReduce Programming	10	Lecture	Slip Test	L3 Apply
2	15CS82.2	Apply essential Hadoop Tools based on Basic Hadoop Administration Procedures	10	Lecture	Assignment	L3 Apply

3	15CS82.3	Apply Business Intelligence Concepts and Application based data	10	Lecture	Assignment	L3 Apply
4	15CS82.4	Apply Decision Trees Concepts and Application based data	10	Lecture / PPT	Assignment	L3 Apply
5	15CS82.5	Apply various mining based on available text	10	Lecture	Slip test	L3 Apply
-	**-**	**Total**	**50**	**-**	**-**	**L3**

2. Course Applications

Write 1 or 2 applications per CO.
Students should be able to employ / apply the course learnings to . . .

Modules	Application Area Compiled from Module Applications.	CO	Level
1	Managing traffic on streets.	CO1	L3
2	Distributed shell	CO2	L3
3	Government education, retailer servicesImage	CO3	L3
4	Processing and Character recognition	CO4	L3
5	Text classification/ Spam Filtering/ Sentiment Analysis	CO5	L3

3. Articulation Matrix

CO – PO Mapping with mapping level for each CO-PO pair, with course average attainment.

-	-	Course Outcomes	Program Outcomes														-
Modules	CO.#	**At the end of the course student should be able to . . .**	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	Level
1	15CS82.1	Apply hadoop Distributed File System Basics to MapReduce Programming	3	3	3		1	1		1				1		3	3	L3
2	15CS82.2	Apply essential Hadoop Tools based on Basic Hadoop Administration Procedures	3	3	3		1	1		1				1		3	3	L3
3	15CS82.3	Apply Business Intelligence Concepts and Application based on data	3	3	3		1	1		1				1		3	3	L3
4	15CS82.4	Apply Decision Trees Concepts and Application based on data	3	3	3		1	1		1				1		3	3	L3
5	15CS82.5	Apply various mining concepts based on available text	3	3	3		1	1		1				1		3	3	L3
-	**15CS82**	Average																**-**
-	*PO, PSO*	*1.Engineering Knowledge; 2.Problem Analysis; 3.Design / Development of Solutions; 4.Conduct Investigations of Complex Problems; 5.Modern Tool Usage; 6.The Engineer and Society; 7.Environment and Sustainability; 8.Ethics; 9.Individual and Teamwork; 10.Communication; 11.Project Management and Finance; 12.Life-long Learning; S1.Software Engineering; S2.Data Base Management; S3.Web Design*																

4. Curricular Gap and Content

Topics & contents not covered (from A.4), but essential for the course to address POs and PSOs.

Modules	Gap Topic	Actions Planned	Schedule Planned	Resources Person	PO Mapping
1		Seminar	2nd week / date	Dr XYZ, Inst	List from B4 above
2		Seminar	3rd Week		

C. COURSE ASSESSMENT

1. Course Coverage

Assessment of learning outcomes for Internal and end semester evaluation.

Modules	Title	Teach. Hours	No. of question in Exam						CO	Levels
			CIA-1	CIA-2	CIA-3	Asg	Extra Asg	SEE		
1	Hadoop Distributed File System Basics	10	2	-	-	1	1	1	CO1	L2,L3
2	Essential Hadoop Tools	10	2	-	-	1	1	1	CO2	L3
3	Basic Intelligence Concepts and Applications	10	-	2	-	1	1	1	CO3	L3
4	Decision Trees	10	-	2	-	1	1	1	CO4	L3
5	Text Mining	10	-	-	4	1	1	1	CO5	L3
-	**Total**	**50**	**4**	**4**	**4**	**5**	**5**	**5**	**-**	**-**

2. Continuous Internal Assessment (CIA)

Assessment of learning outcomes for Internal exams. Blooms Level in last column shall match with A.2.

Modules	Evaluation	Weightage in Marks	CO	Levels
1, 2	CIA Exam – 1	15	CO1, CO2,	L2,L3
3, 4	CIA Exam – 2	15	CO3, CO4	L3
5	CIA Exam – 3	15	CO5	L3
1, 2	Assignment - 1	15	CO1, CO2,	L2,L3
3, 4	Assignment - 2	15	CO3, CO4	L3
5	Assignment - 3	15	CO5	L3
1, 2	Seminar - 1	15	CO1, CO2,	L2,L3
3, 4	Seminar - 2	15	CO3, CO4	L3
5	Seminar - 3	15	CO5	L3
1, 2	Quiz - 1		-	-
3, 4	Quiz - 2		-	-
5	Quiz - 3		-	-
1 - 5	Other Activities – Mini Project		CO1 to CO5	L3
	Final CIA Marks	**20**	**-**	

D1. TEACHING PLAN - 1

Module - 1

Title:	Hadoop Distributed File System Basics	Appr Time:	10 Hrs
a	***Course Outcomes***	**CO**	**Blooms**
-	The student should be able to:	-	**Level**
1	Apply hadoop Distributed File System Basics to MapReduce Programming	CO1	L3
b	***Course Schedule***	-	-
Class No	**Portion covered per hour**	**-**	**-**
1	Hadoop Distributed File System Basics,	CO1	L3
2	Hadoop Distributed File System Basics,	CO1	L3
3	Hadoop Distributed File System Basics,	CO1	L3
4	Hadoop Distributed File System Basics,	CO1	L3
5	Running Example Programs and Benchmarks,,	CO1	L3

6	Running Example Programs and Benchmarks,,	CO1	L3
7	Running Example Programs and Benchmarks,,	CO1	L3
8	Hadoop MapReduce Framework	CO1	L3
9	MapReduce Programming	CO1	L3
10	MapReduce Programming	CO1	L3
c	**Application Areas**		
-	Students should be able employ / apply the Module learnings to . . .		
1	Managing traffic on streets.	CO1	L3
2			
d	**Review Questions**		
-			
1	What are the 3 Vs of Big Data?	CO1	L2
2	How does Big Data impact the business models?	CO1	L2
3	What is Hadoop?	CO1	L2
4	How does Map-Reduce algorithm work?	CO1	L3
5	What are the key issues in managing Big Data?	CO1	L2
6	What is Hadoop? Name the Main Components of a Hadoop Application.	CO1	L3
7	What do you understand by “Rack Awareness”?	CO1	L3
8	What is Speculative Execution?	CO1	L3
9	State some of the important features of Hadoop.	CO1	L3
10	How can you differentiate RDBMS and Hadoop?	CO1	L3
11	What are active and passive NameNodes?	CO1	L3
12	What are the Components of Apache HBase?	CO1	L3
13	How is the DataNode failure handled by NameNode?	CO1	L3
14	Explain the NameNode recovery process.	CO1	L3
15	What are the basic steps involved in map reduce data flow?	CO1	L3
e	**Experiences**	-	-
1		CO1	L2
2			

Module – 2

Title:	Essencial Hadoop Tools	Appr Time:	10 Hrs
a	***Course Outcomes***	**CO**	**Blooms**
-		-	**Level**
1	Apply essential Hadoop Tools based on Basic Hadoop Administration Procedures	CO2	L3
b	***Course Schedule***	-	-
Class No	**Portion covered per hour**	**-**	**-**
11	Essential Hadoop Tools	CO2	L2
12	Essential Hadoop Tools	CO2	L3
13	Hadoop YARN Applications	CO2	L3
14	Hadoop YARN Applications	CO2	L3
15	Hadoop YARN Applications	CO2	L3
16	Managing Hadoop with Apache Ambari	CO2	L3
17	Managing Hadoop with Apache Ambari	CO2	L3
18	Basic Hadoop Administration Procedures	CO2	L3
19	Basic Hadoop Administration Procedures	CO2	L3
20	Basic Hadoop Administration Procedures	CO2	L3

c	**Application Areas**	**-**	**-**
-	Students should be able employ / apply the Module learnings to . . .	-	-
1	Distributed shell	CO2	L3
2			
d	**Review Questions**	-	-
-			
16	What are the main components of Job flow in YARN architecture ?	CO2	L2
17	What is the role of Application Master in YARN architecture ?	CO2	L2
18	Write the structure of YARN applications.	CO2	L2
19	Write a Apache Ambari dashboard view of hadoop cluster.	CO2	L3
20	What are the different views of Apache Ambari.	CO2	L3
21	Write the basicHDFS administration.	CO2	L3
22	Explain capacity scheduler background.	CO2	L3
e	**Experiences**	-	-
1		CO3	L2
2			

E1. CIA EXAM – 1

a. Model Question Paper - 1

Crs Code:			Sem:	I	Marks:		Time:	
Course:								
-	**-**	**Note: Answer all questions, each carry equal marks. Module : 1, 2**				**Marks**	**CO**	**Level**
1	a	What are the key issues in managing Big Data?				5	CO1	L2
	b	What is Hadoop? Name the Main Components of a Hadoop Application.				5	CO1	L3
	c	State some of the important features of Hadoop.				5	CO1	L3
	d							
2	a	What do you understand by “Rack Awareness”?				8	CO1	L3
	b	What is Speculative Execution?				9	CO1	L3
	c							
	d							
3	a	What is the role of Application Master in YARN architecture ?				10	CO2	L2
	b	What are the main components of Job flow in YARN architecture ?				5	CO2	L2
	c							
	d							
4	a	Explain capacity scheduler background.				7	CO2	L2
	b	Write a Apache Ambari dashboard view of hadoop cluster.				8	CO2	L3
	c							
	d							

b. Assignment -1

Model Assignment Questions							
Crs Code:	15CS82	Sem:8		Marks:5		Time:90-120	

Course:	Big Data Analytics		

SNo	Assignment Description	Marks	CO	Level
1	State some of the important features of Hadoop.	5	CO1	L3
2	How does Big Data impact the business models?	5	CO1	L3
3	What are active and passive NameNodes?	4	CO1	L3
4	How does Map-Reduce algorithm work?	5	CO1	L3
5	What are the key issues in managing Big Data?	8	CO1	L3
6	How does Map-Reduce algorithm work?	9	CO1	L3
7	What are the key issues in managing Big Data?	6	CO1	L2
8	What is Hadoop? Name the Main Components of a Hadoop Application.	9	CO1	L3
9	What do you understand by "Rack Awareness"?	8	CO1	L3
10	What is Speculative Execution?	6	CO1	L3
11	How is the DataNode failure handled by NameNode?	9	CO1	L3
12	What is the role of Application Master in YARN architecture ?	10	CO2	L2
13	Write the structure of YARN applications.	7	CO2	L2
14	Write a Apache Ambari dashboard view of hadoop cluster.	8	CO2	L3
15	What are the main components of Job flow in YARN architecture ?	8	CO2	L2

D2. TEACHING PLAN - 2

Module – 3

Title:		Appr Time:	10 Hrs
a	***Course Outcomes***	**CO**	**Blooms**
-	At the end of the topic the student should be able to . . .	-	**Level**
1	Apply Business Intelligence Concepts and Application based on data	CO3	L3
b	***Course Schedule***		
Class No	**Portion covered per hour**	**-**	**-**
21	Business Intelligence Concepts and Application	CO3	L2
22	Business Intelligence Concepts and Application	CO3	L3
23	Business Intelligence Concepts and Application	CO3	L3
24	Business Intelligence Concepts and Application	CO3	L3
25	Data Warehousing	CO3	L3
26	Data Warehousing	CO3	L3
27	Data Mining	CO3	L3
28	Data Mining	CO3	L3
29	Data Visualization	CO3	L3
30	Data Visualization	CO3	L3
c	**Application Areas**	**-**	**-**
-	Students should be able employ / apply the Module learnings to . . .	-	-
1	Government education, retailer services	CO3	L3
d	**Review Questions**	-	-
-	The attainment of the module learning assessed through following questions	-	-

23	Describe the Business Intelligence and Data Mining cycle.	CO3	L2
24	Describe the data processing chain.	CO3	L3
25	What are the similarities between diamond mining and data mining?	CO3	L2
26	What are the different data mining techniques? Which of these would be relevant in your current work?	CO3	L3
27	What is a dashboard? How does it help?	CO3	L2
28	Create a visual to show the weather pattern in your city. Could you show together temperature, humidity, wind, and rain/snow over a period of time.	CO3	L3
29	Why should organizations invest in business intelligence solutions? Are these more important than IT security solutions? Why or why not?	CO3	L3
30	List 3 business intelligence applications in the hospitality industry.	CO3	L3
31	Describe 2 BI tools used in your organization.	CO3	L2
32	Businesses need a 'two-second advantage' to succeed. What does that mean to you?	CO3	L3
33	What is the purpose of a data warehouse?	CO3	L2
34	What are the key elements of a data warehouse? Describe each one.	CO3	L3
35	What are the sources and types of data for a data warehouse?	CO3	L3
36	How will data warehousing evolve in the age of social media?	CO3	L3
37	What is data mining? What are supervised and unsupervised learning techniques?	CO3	L2
38	Describe the key steps in the data mining process. Why is it important to follow these processes?	CO3	L3
39	What are the major mistakes to be avoided when doing data mining?	CO3	L3
40	What are the key requirements for a skilled data analyst?	CO3	L2
41	What are some of the most popular data mining techniques?	CO3	L3
42	What is a confusion matrix?	CO3	L2
43	Why is data preparation so important and time consuming?	CO3	L3
44	What is data visualization?	CO3	L3
45	What are some key requirements for good visualization.	CO3	L2
46	Describe some key steps in data visualization.	CO3	L3
47	What are the data visualization techniques? When would you use table or graphs?	CO3	L3
48	How would you judge the quality of data visualizations?	CO3	L3
e	**Experiences**	-	-
1			
2			
3			
4			
5			

Module – 4

Title:	Data Transmission and Telemetry Measurement of Non – Electrical Quantities	Appr Time:	10 Hrs
a	***Course Outcomes***	**CO**	**Blooms**
-	At the end of the topic the student should be able to . . .	-	**Level**
1	Apply Decision Trees Concepts and Application based on data	CO4	L3
b	***Course Schedule***		
Class No	**Portion covered per hour**	**-**	**-**
31	Decision Trees	CO4	L2
32	Decision Trees	CO4	L3
33	Regression	CO4	L3
34	Regression	CO4	L3
35	Artificial Neural Networks	CO4	L3

36	Artificial Neural Networks	CO4	L3
37	Cluster Analysis	CO4	L3
38	Cluster Analysis	CO4	L3
39	Association Rule Mining	CO4	L3
40	Association Rule Mining	CO4	L3
c	**Application Areas**	**-**	**-**
-	Students should be able employ / apply the Module learnings to . . .	-	-
1	Image Processing and Character recognition	CO4	L4
d	**Review Questions**	-	-
-	The attainment of the module learning assessed through following questions	-	-
49	What is a decision tree? Why are decision trees the most popular classification technique?	CO4	L2
50	What is a splitting variable? Describe three criteria for choosing splitting variable.	CO4	L3
51	What is pruning? What are pre-pruning and post-pruning? Why choose one over the other?	CO4	L3
52	What are gini's coefficient, and information gain?	CO4	L3
53	What is a regression model?	CO4	L2
54	What is a scatter plot? How does it help?	CO4	L3
55	Compare and contrast decision trees with regression models?	CO4	L2
56	What is a neural network? How does it work?	CO4	L3
57	Compare a neural network with a decision tree.	CO4	L2
58	What makes a neural network versatile enough for supervised as well as non-supervised learning tasks?	CO4	L3
59	Examine the steps in developing a neural network for predicting stock prices.	CO4	L3
60	What is unsupervised learning? When is it used? Describe three business applications in your industry where cluster analysis will be useful.	CO4	L3
61	What are association rules? How do they help?	CO4	L3
62	How many association rules should be used?	CO4	L2
e	**Experiences**	-	-
1			
2			

E2. CIA EXAM – 2

a. Model Question Paper - 2

Crs Code:	15CS82	Sem:	VIII	Marks:	30	Time:	75 minutes
Course:	Big Data Analytics						

-	-	**Note: Answer all questions, each carry equal marks. Module : 3, 4**	**Marks**	**CO**	**Level**
1	a	List 3 business intelligence applications in the hospitality industry.	5	CO3	L2
	b	What is the purpose of a data warehouse?	4	CO3	L2
	c	What are the sources and types of data for a data warehouse?	6	CO3	L3
	d				
2	a	How will data warehousing evolve in the age of social media?	8	CO3	L3
	b	What is data mining? What are supervised and unsupervised learning	7	CO3	L3

		techniques?			
	c				
	d				
3	a	What is a decision tree? Why are decision trees the most popular classification technique?	7	CO4	L3
	b	What is a splitting variable? Describe three criteria for choosing splitting variable.	8	CO4	L3
	c				
	d				
4	a	What is a regression model?	5	CO4	L2
	b	What is a scatter plot? How does it help?	5	CO4	L3
	c	What is a neural network? How does it work?	5	CO4	L3
	d				

b. Assignment – 2

Model Assignment Questions							
Crs Code:	15CS82	Sem:	VIII	Marks:	5	Time:	90 – 120 minutes
Course:	Big Data Analytics						

SNo	Assignment Description	Marks	CO	Level
1	List 3 business intelligence applications in the hospitality industry.	5	CO3	L3
2	Describe 2 BI tools used in your organization.	8	CO3	L2
3	Businesses need a 'two-second advantage' to succeed. What does that mean to you?	8	CO3	L3
4	What is the purpose of a data warehouse?	4	CO3	L2
5	What are the key elements of a data warehouse? Describe each one.	8	CO3	L3
6	What are the sources and types of data for a data warehouse?	6	CO3	L3
7	How will data warehousing evolve in the age of social media?	9	CO3	L3
8	What is data mining? What are supervised and unsupervised learning techniques?	8	CO3	L2
9	What is a decision tree? Why are decision trees the most popular classification technique?	7	CO4	L2
10	What is a splitting variable? Describe three criteria for choosing splitting variable.	8	CO4	L3
11	What is pruning? What are pre-pruning and post-pruning? Why choose one over the other?	8	CO4	L3
12	What are gini's coefficient, and information gain?	8	CO4	L4
13	What is a regression model?	5	CO4	L2
14	What is a scatter plot? How does it help?	5	CO4	L3
15	Compare and contrast decision trees with regression models?	8	CO4	L2
16	What is a neural network? How does it work?	7	CO4	L3
17	Compare a neural network with a decision tree.	7	CO4	L2
18	What makes a neural network versatile enough for supervised as well as non-supervised learning tasks?	8	CO4	L3

D3. TEACHING PLAN - 3

Module – 5

Title:	Loop and Horn Antenna and Antenna Types	Appr Time:	10 Hrs
a	***Course Outcomes***	**CO**	**Blooms**
-	At the end of the topic the student should be able to . . .	-	**Level**
1	Apply various mining concepts based on available text	CO5	L3
b	***Course Schedule***	-	-
Class No	**Portion covered per hour**	**-**	**-**
41	Text Mining	CO5	L2
42	Text Mining	CO5	L3
43	Naïve-Bayes Analysis	CO5	L3
44	Naïve-Bayes Analysis	CO5	L3
45	Support Vector Machines	CO5	L3
46	Support Vector Machines	CO5	L3
47	Web Mining	CO5	L2
48	Web Mining	CO5	L3
49	Social Network Analysis	CO5	L3
50	Social Network Analysis	CO5	L3
c	**Application Areas**	**-**	**-**
-	Students should be able employ / apply the Module learnings to . . .	-	-
1	Text classification/ Spam Filtering/ Sentiment Analysis	CO5	L3
d	**Review Questions**	-	-
-	The attainment of the module learning assessed through following questions	-	-
63	Why is text mining useful in the age of social media?	CO5	L3
64	What kinds of problems can be addressed using text mining?	CO5	L3
65	What kinds of sentiments can be found in the text?	CO5	L3
66	What are the three types of web mining?	CO5	L3
67	What are the two major ways that a website can become popular?	CO5	L3
68	What are the privacy issues in web mining?	CO5	L3
69	A user spends 60 minutes on the web, visiting 10 web pages in all. Given the click stream data, what kind of an analysis would you do?	CO5	L3
70	What is click stream analysis?	CO5	L3

e	**Experiences**	-	-
1		CO10	L2
2		CO9	

E3. CIA EXAM – 3

a. Model Question Paper - 3

Crs Code:	15CS82	Sem:	VIII	Marks:	30	Time:	75 minutes
Course:	Big Data Analytics						

-	-	**Note: Answer all questions, each carry equal marks. Module : 5**	**Marks**	**CO**	**Level**
1	a	Why is text mining useful in the age of social media?	8	CO5	L3
	b	What kinds of problems can be addressed using text mining?	7	CO5	L3
	c				
	d				
2	a	Briefly explain support vector machine.	8	CO5	L3
	b	Briefly explain naive-Bayes analysis	9	CO5	L3
	c				
	d				
3	a	What are the three types of web mining?	6	CO5	L3
	b	What kinds of sentiments can be found in the text?	5	CO5	L3
	c				
	d				
4	a	What are the two major ways that a website can become popular?	8	CO5	L3
	b	What are the privacy issues in web mining?	5	CO5	L3
	c	What is click stream analysis?	2	CO5	L2
	d				

b. Assignment – 3

Model Assignment Questions							
Crs Code:	15CS82	Sem:	VIII	Marks:	5	Time:	90 – 120 minutes
Course:	Big Data Analytics						

SNo	**Assignment Description**	**Marks**	**CO**	**Level**
1	Why is text mining useful in the age of social media?	8	CO5	L3
2	What kinds of problems can be addressed using text mining?	7	CO5	L3
3	What are the two major ways that a website can become popular?	8	CO5	L3
4	What are the privacy issues in web mining?	5	CO5	L3
5	What is click stream analysis?	2	CO5	L2
6	Briefly explain support vector machine.	8	CO5	L3
7	Briefly explain naive-Bayes analysis	9	CO5	L3
8	What are the three types of web mining?	6	CO5	L3
9	What kinds of sentiments can be found in the text?	5	CO5	L3

F. EXAM PREPARATION

1. University Model Question Paper

Course:		Big Data Analytics						Month / Year		May /2018	
Crs Code:		15CS82	Sem:	VIII	Marks:	80		Time:		180 minutes	
Module		Answer all FIVE full questions. All questions carry equal marks.							**Marks**	**CO**	**Level**
1	a	What are active and passive NameNodes?							4	CO1	L3
	b	How does Map-Reduce algorithm work?							5	CO1	L3
	c	What are the key issues in managing Big Data?							6	CO1	L3
	d										
		OR									
-	a	What are the key issues in managing Big Data?							4	CO1	L2
	b	What is Hadoop? Name the Main Components of a Hadoop Application.							7	CO1	L3
	c	What do you understand by "Rack Awareness"?							5	CO1	L3
	d										
2	a	What are the main components of Job flow in YARN architecture ?							6	CO2	L2
	b	What is the role of Application Master in YARN architecture ?							6	CO2	L2
	c	Write the structure of YARN applications.							4	CO2	L2
	d										
		OR									
-	a	Write a Apache Ambari dashboard view of hadoop cluster.							7	CO2	L3
	b	What are the different views of Apache Ambari.							3	CO2	L3
	c	Write the basic HDFS administration.							6	CO2	L3
	d										
3	a	Describe the Business Intelligence and Data Mining cycle.							7	CO3	L2
	b	What are the different data mining techniques? Which of these would be relevant in your current work?							6	CO3	L3
	c	What are the similarities between diamond mining and data mining?							3	CO3	L2
	d										
		OR									
-	a	What is a dashboard? How does it help?							4	CO3	L2
	b	Create a visual to show the weather pattern in your city. Could you show together temperature, humidity, wind, and rain/snow over a period of time.							8	CO3	L3
	c	Why should organizations invest in business intelligence solutions? Are these more important than IT security solutions? Why or why not?							4	CO3	L3
	d										
4	a	What is a decision tree? Why are decision trees the most popular classification technique?							8	CO4	L2
	b	What is a splitting variable? Describe three criteria for choosing splitting variable.							8	CO4	L3
	c										
	d										
		OR									
-	a	What is a regression model?							2	CO4	L2

	b	Compare and contrast decision trees with regression models?	6	CO4	L2
	c	What is a neural network? How does it work?	8	CO4	L3
	d				
5	a	Why is text mining useful in the age of social media?	5	CO5	L3
	b	What kinds of problems can be addressed using text mining?	6	CO5	L3
	c	What kinds of sentiments can be found in the text?	5	CO5	L3
	d				
		OR			
	a	What are the three types of web mining?	5	CO5	L3
	b	What are the two major ways that a website can become popular?	6	CO5	L3
	c	What are the privacy issues in web mining?	5	CO5	L3
	d				

2. SEE Important Questions

Course:	Big Data Analytics					Month / Year	May /2018	
Crs Code:	15CS82	Sem:	VIII	Marks:	80	Time:	180 minutes	
	Note	Answer all FIVE full questions. All questions carry equal marks.				-	-	
Module	Qno.	Important Question				**Marks**	**CO**	**Year**
1	1					16 / 20		
	2							
	3							
	4							
	5							
2	1					16 / 20		
	2							
	3							
	4							
	5							
3	1					16 / 20		
	2							
	3							
	4							
	5							
4	1					16 / 20		
	2							
	3							
	4							
	5							
5	1					16 / 20		
	2							
	3							
	4							
	5							

Course Outcome Computation

Academic Year:

Odd / Even semester

INTERNAL TEST					**T1**						**T2**	
Course Outcome	**CO1**		**CO2**		**CO3**		**CO4**		**CO5**		**CO6**	
QUESTION NO	**Q1**	**LV**	**Q2**	**LV**	**Q3**	**LV**	**Q1**	**LV**	**Q2**	**LV**	**Q3**	**LV**
MAX MARKS	10	-	10	-	10	-	10	-	10	-	10	-
USN-1	5	2	10				10	3	9	3	4	1
USN-2	5	2	8	3								
USN-3	7	3	7	3	10	3	8	3	8	3	5	2
USN-4					4	1	10	3	8	3	6	2
USN-5	8	3	6	2	9	3	10	3	8	3		
USN-6							10	3	9	3	4	1
Average CO Attainment		2.5		2.75		2.33		3		3		1.5

LV Threshold : 3:>60%, 2:>=50% and <=60%, 1: <=49%
CO1 Computation :(2+2+2+3)/4 = 10/4=2.5

PO Computation

Program Outcome	PO1		PO3		PO3		PO1		PO12		PO12	
Weight of CO - PO	3		1		3		2		2		3	
Course Outcome	CO1		CO2		CO3		CO4		CO5		CO6	
Test/Quiz/Lab	T1						T2					
QUESTION NO	Q1	LV	Q2	LV	Q3	LV	Q1	LV	Q2	LV	Q3	LV
MAX MARKS	10	-	10	-	10	-	10	-	10	-	10	-
USN-1	5	2	10	3			10	3	9	3	4	1
USN-2	5	2	8	3								
USN-3	7	3	7	3	10	3	8	3	8	3	5	2
USN-4					4	1	10	3	8	3	6	2
USN-5	8	3	6	2	9	3	10	3	8	3		
USN-6							10	3	9	3	4	1
Average CO Attainment		2.5		2.75		2.33		3		3		1.5