



PRESIDENCY UNIVERSITY

(Established under the Presidency University Act, 2013 of the Karnataka Act 41 of 2013)

[2022-23 EVEN/ WINTER SEMESTER]

COURSE HAND OUT [Revision 02 -Jan 2022]

SCHOOL: SOCSE	DEPT: CBD	DATE OF ISSUE: 27 Jan 2023
NAME OF THE PROGRAM	: B.Tech	
P.R.C. APPROVAL REF.	: PU/AC-XX.X/CSE17/CBD/2020-2024	
SEMESTER/YEAR	: VI / III	
COURSE TITLE & CODE	: Data Mining & CSE2021	
COURSE CREDIT STRUCTURE	: 3-0-3	
CONTACT HOURS	: 3 sessions per week	
COURSE IC	: S. Thabassum Khan	
COURSE INSTRUCTOR(S)	: Dr.Thivakaran T.K, Mr.Amogh Pramod Kulkarni, Dr.Chinnaiyan R, Mr.Jamil Ahmed, Dr.Charulatha B.S.	
COURSE URL	: https://camu.in/	
PROGRAM OUTCOMES	:	

PO-1: Engineering knowledge: Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.

PO-2: Problem analysis: Identify, formulate, review research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.

PO-3: Design/development of solutions: Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations.

PO-4: Conduct investigations of complex problems: Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.

PO-5: Modern tool usage: Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modeling to complex engineering activities with an understanding of the limitations

PO-6: The engineer and society: Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilities relevant to the professional engineering practice.

PO-7: Environment and sustainability: Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of and need for sustainable development.

[illegible]

COURSE CONTENT (SYLLABUS):

Module:1: Introduction to Data Mining

[6 Sessions] [Blooms 'level selected: Knowledge]

Topics:Introduction to Data mining: Definition, KDD, Challenges, Data Mining Tasks - Data Mining Goals– Stages of the Data Mining Process–Data Mining Techniques– Applications – Major Issues in Data mining.

Module: 2: Data Preprocessing

[7 Sessions] [Blooms 'level selected: Comprehension]

Topics:Types of data – Data Quality – Data Pre-processing Techniques – Similarity and Dissimilarity measures.

Module: 3: Data Mining – Frequent Patterns

[7 Sessions] [Blooms 'level selected: Application]

Motivation and terminology: Basic idea - Item sets – Generating frequent item sets and rules efficiently – Apriori Algorithm – FP Growth.

Assignment: Apply the Apriori algorithms for finding the frequent Item set in the given TDB.

Module: 4: Classification

[8 Sessions] [Blooms 'level selected: Application]

Basic concepts – Decision tree Induction – Bayes classification methods – Rule based classification – Classification by Back Propagation – Lazy learners.

Assignment:

- 1) Find the Gini Index value of the attributes.
- 2) Classify the given model using Decision tree algorithm.

Module: 5: Cluster AnalysisMethods and Pattern Mining

[8 Sessions] [Blooms 'level selected: Application]

Cluster Analysis-Partitioning methods – Hierarchical methods – Basics of Density based method – Pattern mining: A Road Map

Assignment:

- 1) Cluster the objects using Cluster algorithms.
- 2) Problem for Cluster validation.
- 3) Apply theProcess of data mining in the Employee database.

DELIVERY PROCEDURE (PEDAGOGY):

TABLE 3: SPECIAL DELIVERY METHOD/ PEDAGOGY PLANNED WITH TOPICS				
S. No	Lecture Number	Subtopic as per lesson Plan	Pedagogy title/ short explanation of adopted pedagogy	** At end of semester please update whether activity was done
1	L7	Major Issues in Data mining.	Think Pair Share	
2	L12	Data Pre-processing Techniques	Self-learning	
3	L23	Apriori Algorithm	Participative Learning	
4	L37	Applying cluster and classifier algorithm on real –time applications	Participative Learning	

REFERENCE MATERIALS:

Text Books:

T1. Jiawei Han, Micheline Kamber and Jian Pei, “Data Mining: Concepts and Techniques”,Morgan Kaufmann Publishers, Third Edition, 2012.

Reference Books:

- R1. Tan P. N, Steinbach M and Kumar V, “Introduction to Data Mining”, Pearson Education, 2016.
- R2. G K Gupta, “Introduction to Data Mining with Case Studies”, Third Edition, PHI, 2014.
- R3. Alex Berson and Stephen J. Smith, “Data Warehousing, Data Mining and OLAP”, Tata McGraw Hill

Web Based Resources and E-books:**Digital Learning Resources (Library Resources)**

- W1. : https://onlinecourses.swayam2.ac.in/cec20_cs12/preview
Text book of Data Mining: Concepts and Techniques, Jiawei Han, Micheline Kamber and Jian Pei, Morgan Kaufmann Publishers, 2012.
- W2. <https://mitmecsept.files.wordpress.com/2017/04/data-mining-concepts-and-techniques-2nd-edition-impressao.pdf>
- W3. e-book (https://hanj.cs.illinois.edu/bk3/bk3_slidesindex.htm)
- W4. . https://puniversity.informaticsglobal.com:2284/ehost/detail/detail?vid=7&sid=e2d7362a-fd3049a98f0393e963521dbd%40redis&bdata=JnNpdGU9ZWVhc3QtYGl2ZQ%3d%3d#AN=377411_&db=nlebk [OR]
http://182.72.188.195/cgi-bin/koha/opac-detail.pl?biblionumber=4001&query_desc=ti%2Cwrdl%3A%20Data%20Mining%3A%20Concepts%20and%20Techniques

SPECIFIC GUIDELINES TO STUDENTS:

Nil

COURSE SCHEDULE FOR THEORY COMPONENT

TABLE 4: COURSE BROAD SCHEDULE				
Sl. No.	ACTIVITY	PLANNED STARTING DATE	PLANNED CONCLUDING DATE	TOTAL NUMBER OF PERIODS
1	Over View of the course	16-02-2023	16-02-2023	1
2	Module : 01	17-02-2023	03-03-2023	6
3	Module: 02	06-03-2023	24-03-2023	7
4	Mid Term Examination	12-04-2023	15-04-2023	
5	Module:03	25-03-2023	21-04-2023	7
6	Assignment 1	-	-	-
7	Module:04	24-04-2023	15-05-2023	8
8	Article Review -1	-	-	-
9	Module :05	16-05-2023	02-06-2023	8
10	Quiz -1	-	-	-
11	End Term Examination	07-06-2023	24-06-2023	3 Hrs

DETAILED SCHEDULE OF INSTRUCTION:**Main pedagogy: PPT + Chalk Board and Lecture****TABLE 5: DETAILED COURSE SCHEDULE/ LESSON PLAN**

Session no	TOPIC	SUB TOPIC	CO Number	Reference
1	Module 1 : Introduction to Data Mining	Course Integration	-	-
2	Module 1 : Introduction to Data Mining	Definition, KDD, Challenges	CO1	R1
3	Module 1 : Introduction to Data Mining	Data Mining Tasks, Data Mining Goals	CO1	R1
4	Module 1 : Introduction to Data Mining	Stages of the Data Mining Process , Data Mining Techniques	CO1	R1
5	Module 1 : Introduction to Data Mining	Applications	CO1	R1
6	Module 1 : Introduction to Data Mining	Major Issues in Data mining.	CO1	R1
7	Module Integration	Module Integration	CO1	R1
8	Module 2: Data Preprocessing	Introduction to preprocessing	CO2	R1
9	Module 2: Data Preprocessing	Types of data	CO2	R1
10	Module 2: Data Preprocessing	Data Quality	CO2	R1
11	Module 2: Data Preprocessing	Data Pre-processing Techniques	CO2	R1
12	Module 2: Data Preprocessing	Data Pre-processing Techniques	CO2	R1
13	Module 2: Data Preprocessing	Similarity and Dissimilarity measures.	CO2	R1
14	Module 2: Data Preprocessing	Similarity and Dissimilarity measures.	CO2	R1
15	Module Integration	Module Integration	CO2	R1
16	Mid Term	Mid Term	CO2	R1
17	Mid Term	Mid Term	CO2	R1
18	Mid Term	Mid Term	CO2	R1
19	Module 3: Data Mining – Frequent Patterns	Basic idea –Frequent Patterns –Item sets	CO3	R1
20	Module 3: Data Mining – Frequent Patterns	Generating frequent item sets and rules efficiently	CO3	R1
21	Module 3: Data Mining – Frequent Patterns	Generating frequent item sets and rules efficiently	CO3	R1

22	Module 3: Data Mining – Frequent Patterns	Apriori Algorithm	CO3	R1
23	Module 3: Data Mining – Frequent Patterns	Apriori Algorithm	CO3	R1
24	Module 3: Data Mining – Frequent Patterns	FP Growth	CO3	R1
25	Module 3: Data Mining – Frequent Patterns	FP Growth	CO3	R1
26	Module Integration	Module Integration	CO3	R1
27	Module4 : Classification	Basic concepts of Classification	CO4	R1
28	Module4 : Classification	Decision tree Induction	CO4	R1
29	Module4 : Classification	Bayes classification methods	CO4	R1
30	Module4 : Classification	Bayes classification methods	CO4	R1
31	Module4 : Classification	Rule based classification	CO4	R1
32	Module4 : Classification	Classification by Back Propagation	CO4	R1
33	Module4 : Classification	Classification by Back Propagation	CO4	R1
34	Module4 : Classification	Lazy learners	CO4	R1
35	Module Integration	Module Integration	CO4	R1
36	Module5: Cluster Analysis Methods and Pattern Mining	Partitioning methods	CO5	R1
37	Module5: Cluster Analysis Methods and Pattern Mining	Partitioning methods	CO5	R1
38	Module5: Cluster Analysis Methods and Pattern Mining	Hierarchical methods	CO5	R1
39	Module5: Cluster Analysis Methods and Pattern Mining	Hierarchical methods	CO5	R1
40	Module5: Cluster Analysis Methods and Pattern Mining	Basics of Density based method	CO5	R1
41	Module5: Cluster Analysis Methods and Pattern Mining	Basics of Density based method	CO5	R1
42	Module5: Cluster Analysis Methods and Pattern Mining	Pattern mining: A Road Map	CO5	R1
43	Module5: Cluster Analysis Methods and Pattern Mining	Pattern mining: A Road Map	CO5	R1
44	Module Integration	Module Integration	CO5	R1
45	Course Integration	Course Integration	CO 1,2,3,4,5	

ASSESSMENT SCHEDULE:**TABLE 6 ASSESSMENT SCHEDULE**

Sl.no	Assessment type	Contents	Course outcome Number	Duration In Hours	marks	Weightage	Venue, DATE & TIME
1		Module-1	CO 1				
2		Module-2	CO2				
3	Midterm Examination	Module 1, 2	CO 1,2	1.5 hour	60 Marks	30%	12-04-23 to 15-04-23
4	Assignment-1	Module-3	CO3	NA	10 Marks	5%	Last week of April
5	Article-1 Review of digital / e-resources from Pres. Univ. link given in the References Section - (Mandatory to submit screenshot accessing digital resource. Otherwise it will not be evaluated)	Module-4 https://presiuniv.knimbus.com/user#/home	CO4	NA	10 Marks	5%	Last week of May
6	QUIZ-1	Module- 4, 5	CO4, 5	1 Hour	20 Marks	10%	Last week of May
7	End Term	Module1,2,3,4,5	CO1, 2, 3, 4, 5	3 hours	100 Marks	50%	07-06-23 to 24-06-23

COURSE CLEARANCE CRITERIA:

AS PER ACADEMIC REGULATIONS OF THE UNIVERSITY

MAKEUP EXAM POLICY:

AS PER ACADEMIC REGULATIONS OF THE UNIVERSITY

CONTACT TIMINGS IN THE CHAMBER FOR ANY DISCUSSIONS:

WEDNESDAY- 3pm to 4pm

SAMPLE THOUGHT PROVOKING QUESTIONS:

TABLE 7: SAMPLE THOUGHT PROVOKING QUESTIONS				
SL NO	QUESTION	MARKS	COURSE OUTCOME NO.	BLOOM'S LEVEL
1	Suppose that you are employed as a datamining consultant for an Internets Search engine company, describe how data mining can help the company by giving Specific examples of how techniques, such as clustering, classification, Association rule mining, and anomaly detection can be applied.	10	CO3	Application
2	Which of the following quantities are likely to show more temporal autocorrelation: daily rainfall or daily temperature? Why?	10	CO3	Application
3	Consider the following set of frequent3-itemsets: {1, 2, 3}, {1, 2, 4}, {1, 2, 5}, {1, 3, 4}, {1, 3, 5}, {2, 3, 4}, {2, 3, 5}, {3, 4, 5}. Assume that there are only five items in the data set. List all candidate 4-itemsets obtained by a candidate generation procedure using the $F_{k-1} \times F_1$ merging strategy. List all candidate 4- item sets obtained by the candidate generation procedure in Apriori. List all candidate 4-itemsets that survive the candidate • pruning step of the Apriori algorithm.	10	CO4	Application
4	Discuss which algorithm is an influential algorithm for mining frequent item sets for Boolean association rules? Explain with an example?	10	CO4	Application

TARGET SET FOR COURSE OUTCOME ATTAINMENT:

TABLE 8: TARGET SET FOR ATTAINMENT OF EACH CO and ATTAINMENT ANALYSIS AFTER RESULTS						
Sl.no	C.O. No.	Course Outcomes	Threshold Set for the CO	Target set for attainment in percentage	Actual C.O. Attainment In Percentage	Remarks on attainment & Measures to enhance the attainment
					*	*
01	CO1	Explain the basic concepts and issues involved in Data Mining.	50	75%		
02	CO2	Discuss different preprocessing techniques on Data Analysis	45	70%		
03	CO3	Discover frequent item sets by using Association rule algorithms	55	60%		
04	CO4	Apply different Classification and Clustering techniques used in data mining	60	60%		
05	CO5	Describe the advances in data mining for the selected real time application	60	60%		

Signature of the course Instructor **In-Charge (s)**

APPROVAL:

This course has been duly verified Approved by the D.A.C.

Signature of the Chairperson D.A.C.

Name and signature of the **Instructor In-Charge (s)** **AFTER completing entries in Table number 3 and 8 at end of semester;**

Name and signature of the DAC Chairperson **AFTER completing entries in Table number 3 and 8 at end of semester;**

BLOOM'S TAXONOMY SAMPLE VERBS

Learning Outcomes Verbs at Each Bloom Taxonomy Level to be used for writing the course Outcomes.

TABLE 9: REFERENCE SAMPLES OF BLOOMS TAXONOMY VERBS		
Cognitive Level	Illustrative Verbs	Definitions
Knowledge	arrange, define, describe, duplicate, identify, label, list, match, memorize, name, order, outline, recognize, relate, recall, repeat, reproduce, select, state	remembering previously learned information
Comprehension	classify, convert, defend, discuss, distinguish, estimate, explain, express, extend, generalize, give example(s), identify, indicate, infer, locate, paraphrase, predict, recognize, rewrite, report, restate, review, select, summarize, translate	grasping the meaning of information
Application	apply, change, choose, compute, demonstrate, discover, dramatize, employ, illustrate, interpret, manipulate, modify, operate, practice, predict, prepare, produce, relate schedule, show, sketch, solve, use write	applying knowledge to actual situations
Analysis	analyze, appraise, breakdown, calculate, categorize, classify, compare, contrast, criticize, derive, diagram, differentiate, discriminate, distinguish, examine, experiment, identify, illustrate, infer, interpret, model, outline, point out, question, relate, select, separate, subdivide, test	breaking down objects or ideas into simpler parts and seeing how the parts relate and are organized
Synthesis	arrange, assemble, categorize, collect, combine, comply, compose, construct, create, design, develop, devise, explain, formulate, generate, plan, prepare, propose, rearrange, reconstruct, relate, reorganize, revise, rewrite, set up, summarize, synthesize, tell, write	rearranging component ideas into a new whole
Evaluation	appraise, argue, assess, attach, choose, compare, conclude, contrast, defend, describe, discriminate, estimate, evaluate, explain, judge, justify, interpret, relate, predict, rate, select, summarize, support, value	making judgments based on internal evidence or external criteria