



Sri Indu

College of Engineering & Technology

UGC Autonomous Institution

Recognized under 2(f) & 12(B) of UGC Act 1956,

NAAC, Approved by AICTE &

Permanently Affiliated to JNTUH



HANDOUT

IV Year I Semester

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
(ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING)

ACADEMIC YEAR 2025-2026



SRIINDUCOLLEGE OF ENGINEERING & TECHNOLOGY

(An Autonomous Institution under UGC, New Delhi)

(Permanently Affiliated to JNTUH, Approved by AICTE, New Delhi and Accredited by NBA, NAAC) Sheriguda Village, Ibrahimpatnam Mandal, Ranga Reddy Dist. – 501 510

DEPARTMENT OF ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

HANDOUT-INDEX

S. No	Contents
1	Vision, Mission, PEOs, POs, PSOs & Cos
2	Institution Academic Calendar
3	Department Academic Calendar
4	Subjectwise
i)	Syllabus Copy
ii)	Lesson Plan
iii)	Question Bank
iv)	Mid-1 & Mid-2 Questions (Model Paper)
v)	End Examination Questions (Model Paper)



SRIINDUCOLLEGE OF ENGINEERING & TECHNOLOGY

(An Autonomous Institution under UGC, New Delhi)

(Permanently Affiliated to JNTUH, Approved by AICTE, New Delhi and Accredited by NBA,

NAAC) Sheriguda Village, Ibrahimpatnam Mandal, Ranga Reddy Dist. – 501510

VISION OF THE INSTITUTE

To be a premier Institution in Engineering & Technology and Management with competency, values and social consciousness.

MISSION OF THE INSTITUTE

- IM1** Provide high quality academic programs, training activities and research facilities.
- IM2** Promote Continuous Industry-Institute interaction for employability, Entrepreneurship, leadership and research aptitude among stakeholders.
- IM3** Contribute to the economic and technological development of the region, state and nation.

VISION OF THE DEPARTMENT

To develop competent engineers in the domain of Artificial Intelligence & Machine Learning for Note worthy contributions to the society.

MISSION OF THE DEPARTMENT

The Department has following Missions:

- DM1** To strengthen academic collaborations for better exposure.
- DM2** Promote professional environment to imbibe ethical values and entrepreneurial skills.
- DM3** Encourage research development by emphasizing on innovation.
- DM4** Facilitate Industry-Institute collaboration for mutual benefit

PROGRAM EDUCATIONAL OBJECTIVES (PEOs)

- PEO1:** Generate contributions towards advancements in Artificial Intelligence and Machine Learning.
- PEO2:** Promote design, research and implementation of products through strong communication skills, leadership and entrepreneurial skills.
- PEO3:** Apply basic principles and practices of AIML to successfully complete software related projects to meet customer business objectives and/or productively engage in research.

Program Outcomes(POs)

PO1	Engineering Knowledge: Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.
PO2	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.
PO3	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.
PO4	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.
PO5	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.
PO6	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.
PO7	Ethics: Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.
PO8	Individual and team work: Function effectively as an individual, and as a member or leader in diverse teams, and in multi disciplinary settings.
PO9	Communication: Communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write Effective reports and design documentation, make effective presentations, and give and receive clear instructions.
PO10	Project management and finance: Demonstrate knowledge and understanding of the engineering and management principles and apply the set o one's own work, as a member and leader in a team, to manage projects and in multi disciplinary environments.
PO11	Life-long learning: Recognize the need for, and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.

Program specific Outcomes(PSOs)

PSO1	Register mathematical methodology to crack problems using suitable data structures
PSO2	Competence to design and develop software for web based and mobiles androids under real world environment
PSO3	Skill to design the algorithms for machine learning, data compression can be used in different applications.

COURSE OUTCOMES(CO's)

COURSE NAME: (R22INF3245) DATA MINING

At the end of the course, the student will be able to

Course Outcomes(COs)	
C413.1	Ability to understand the types of the data to be mined and present a general classification of tasks and primitives to integrate a data mining system.
C413.2	Apply preprocessing methods for any given raw data.
C413.3	Extract interesting patterns from large amounts of data.
C413.4	Discover the role played by data mining in various fields.
C413.5	Choose and employ suitable data mining algorithms to build analytical applications
C413.6	Evaluate the accuracy of supervised and unsupervised models and algorithms

Outcomes(CO's) withPO's:

PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2
C413.1	2	3	-	-	2	-	-	-	-	-	-	3	3	-
C413.2	3	3	3	-	3	-	2	-	-	-	-	2	-	-
C413.3	2		3	-	2	-	-	-	-	-	-	-	3	-
C413.4	2	2	1	-	-	-	-	-	-	-	-	2	-	2
C413.5	3	-	-	-	-	-	3	-	3	-	-	2	2	2
C413.6	2	-	-	-	1	-	-	-	2	-	-	-	3	-
AVG	2.3	2.7	2.7	-	2.3	-	2.5	-	2.5	-	-	2.3	2.75	2

3:High 2. Medium 1.Low

Mapping of Course Outcomes (CO's) with PSO's:

Cos	PSO1	PSO2
C413.1	3	3
C413.2	3	3
C413.3	3	3
C413.4	3	3
C413.5	3	3
C413.6	3	3
AVG	3	3

SRI INDU COLLEGE OF ENGINEERING AND TECHNOLOGY(AUTONOMOUS)
DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING(AI&ML) DEPARTMENT CALENDAR – 2025-2026(SEMESTER-I)

SUNDAY				SEPTEMBER				DECEMBER	
MONDAY		JULY		1	IV-I MID I EXAMS			1	IV-I END EXAMS
TUESDAY	1	DEPARTMENT MEETING FOR IV-I & II-I 2025-2026		2	IV-I MID I EXAMS		OCTOBER	2	IV-I END EXAMS
WEDNESDAY	2			3	IV-I MID I EXAMS	1	II-I CRT CLASSES	3	IV-I END EXAMS
THURSDAY	3		AUGUST	4	IV-I SPELL 2 CLASSES	2	GANDHI JAYANTI & DASHARA	4	IV-I END EXAMS 1
FRIDAY	4		II-I SPELL 1 CLASSES COMMENCEMENT	5		3	II-I CRT CLASSES	5	IV-I END EXAMS 2
SATURDAY	5		DEPARTMENT MEETING	6		4	II-I CRT CLASSES	6	IV-I END EXAMS 3
SUNDAY	6			7		5		7	IV-I END EXAMS 4
MONDAY	7	IV-I SPELL 1 CLASSES COMMENCEMENT		8		6	II-I CRT CLASSES	8	IV-I END EXAMS 5
TUESDAY	8			9		7	II-I CRT CLASSES	9	IV-I END EXAMS 6
WEDNESDAY	9			10	IV-I MID I EXAMS SUBMISSION	8	II-I MID I EXAMS & II-I MID I EXAMS	10	IV-I END EXAMS 7
THURSDAY	10			11		9	II-I MID I EXAMS & II-I MID I EXAMS	11	IV-I END EXAMS 8
FRIDAY	11			12	DEPARTMENT MEETING REVIEW ON IV-I SLOW LEARNERS	10	II-I MID I EXAMS & II-I MID I EXAMS	12	IV-I END EXAMS 9
SATURDAY	12	AI CLUB MEETING	WORKSHOP UNDER AI CLUB	13	AI CLUB MEETING	11		13	IV-I END EXAMS 10
SUNDAY	13			14		12		14	IV-I END EXAMS 11
MONDAY	14	II-I SPELL 1 CLASSES COMMENCEMENT		15	II-I CRT CLASSES	13	II-I & II-I SPELL 2 CLASSES	15	II-I & II-I MID II EXAMS 12
TUESDAY	15			16	II-I CRT CLASSES	14		16	II-I & II-I MID II EXAMS 13
WEDNESDAY	16			17	II-I CRT CLASSES	15		17	II-I & II-I MID II EXAMS 14
THURSDAY	17			18	II-I CRT CLASSES	16	II-I & II-I MID I EXAMS SUBMISSION	18	II-I & II-I PREPARATION & PRACTICAL EXAMS 15
FRIDAY	18		15 INDEPENDENCE DAY	19	II-I CRT CLASSES	17	DEPARTMENT MEETING REVIEW ON II-I & II-I SLOW LEARNERS	19	II-I & II-I PREPARATION & PRACTICAL EXAMS 16
SATURDAY	19			20	II-I CRT CLASSES	18	AI CLUB MEETING	20	II-I & II-I PREPARATION & PRACTICAL EXAMS 17
SUNDAY	20			21	BATHUKAMMA	19		21	II-I & II-I PREPARATION & PRACTICAL EXAMS 18
MONDAY	21	BONALI HOLIDAY		22	II-I CRT CLASSES	20	DIWALI	22	II-I & II-I PREPARATION & PRACTICAL EXAMS 19
TUESDAY	22			23	II-I CRT CLASSES	21		23	II-I & II-I MID II MARKS SUBMISSION II-I PREPARATION & PRACTICAL EXAMS 20

WEDNESDAY	23		20	24	III-I CRT CLASSESS	22		19	IV-I MID II EXAMS	24	II-I & III-I PREPARATION & PRACTICAL EXAMS	21
THURSDAY	24		21	25	III-I CRT CLASSESS	23		20	IV-I MID II EXAMS	25	CHRISTMAS DAY	22
FRIDAY	25		22	26	III-I CRT CLASSESS	24		21	IV-I MID II EXAMS	26	II-I & III-I PREPARATION & PRACTICAL EXAMS	23
SATURDAY	26		23	27	GANESH CHATURTHI	25	DEPARTMENT MEETING	22	AI CLUB MEETING	27	II-I & III-I PREPARATION & PRACTICAL EXAMS	24
SUNDAY	27		24	28		26		23		28		25
MONDAY	28	DEPARTMENT MEETING FOR II YR 2025-2026	25	29	III-I CRT CLASSESS	27		24	IV-I PREPARATION & PRACTICAL EXAMS	29	II-I & III-I END EXAMS	26
TUESDAY	29		26	30	III-I CRT CLASSESS	28		25	IV-I PREPARATION & PRACTICAL EXAMS	30	II-I & III-I END EXAMS	27
WEDNESDAY	30		27			29		26	IV-I PREPARATION & PRACTICAL EXAMS	31	II-I & III-I END EXAMS	28
THURSDAY	31		28			30		27	IV-I PREPARATION & PRACTICAL EXAMS			29
FRIDAY			29			31		28	IV-I PREPARATION & PRACTICAL EXAMS			30
SATURDAY			30	AI CLUB MEETING	WOKSHOP UNDER AI LUB			29	IV-I PREPARATION & PRACTICAL EXAMS			31
SUNDAY			31					30				

Faculty In-charge

HOD/_AIML_

Principal

9789350239759.

SRI INDU COLLEGE OF ENGINEERING & TECHNOLOGY

(An Autonomous Institution under UGC, New Delhi)

B.Tech. - IV Year – I Semester

L	T	P	C
3	0	0	3

Professional Elective – III
(R22INF3245) DATA MINING

Course Objectives:

- It presents methods for mining frequent patterns, associations, and correlations.
- It then describes methods for data classification and prediction, and data-clustering approaches.
- It covers mining various types of data stores such as spatial, textual, multimedia, streams.

Course Outcomes:

- Ability to understand the types of the data to be mined and present a general classification of tasks and primitives to integrate a data mining system.
- Apply preprocessing methods for any given raw data.
- Extract interesting patterns from large amounts of data.
- Discover the role played by data mining in various fields.
- Choose and employ suitable data mining algorithms to build analytical applications
- Evaluate the accuracy of supervised and unsupervised models and algorithms.

UNIT - I

Data Mining: Data–Types of Data–, Data Mining Functionalities– Interestingness Patterns– Classification of Data Mining systems– Data mining Task primitives –Integration of Data mining system with a Data warehouse–Major issues in Data Mining–Data Preprocessing.

UNIT - II

Association Rule Mining: Mining Frequent Patterns–Associations and correlations – Mining Methods– Mining Various kinds of Association Rules– Correlation Analysis– Constraint based Association mining. Graph Pattern Mining, SPM.

UNIT - III

Classification: Classification and Prediction – Basic concepts–Decision tree induction–Bayesian classification, Rule–based classification, Lazy learner.

UNIT - IV

Clustering and Applications: Cluster analysis–Types of Data in Cluster Analysis–Categorization of Major Clustering Methods– Partitioning Methods, Hierarchical Methods– Density–Based Methods, Grid–Based Methods, Outlier Analysis.

UNIT - V

Advanced Concepts: Basic concepts in Mining data streams–Mining Time–series data–Mining sequence patterns in Transactional databases– Mining Object– Spatial– Multimedia–Text and Web data – Spatial Data mining– Multimedia Data mining–Text Mining– Mining the World Wide Web.

TEXT BOOKS:

1. Data Mining – Concepts and Techniques – Jiawei Han & Micheline Kamber, 3rd Edition Elsevier.
2. Data Mining Introductory and Advanced topics – Margaret H Dunham, PEA.

REFERENCE BOOK:

1. Ian H. Witten and Eibe Frank, Data Mining: Practical Machine Learning Tools and Techniques (Second Edition), Morgan Kaufmann, 2005.

	SRI INDU COLLEGE OF ENGINEERING & TECHNOLOGY			Prepared on March 1st
	LESSON PLAN			
	(Regulation :R22)			2025
	Department of Artificial Intelligence & Data Science			
	Sub. Code & Title		(R22INF3245) DATA MINING	
Academic Year: 2024-25		Year/Sem./Section	IV-I/CSE-(AI&ML)(A&B)	
Faculty Name & Designation		Dr.M.Naresh Associate Professor		

Unit/ Item No.	Topic (s)	Book Referenc e	Page (s)		Teaching Methodology	Proposed No. of Periods	Actual Date of Handled	CO/RB T
			From	To				
UNIT-I					12			
1.1	Data–Types of Data	T1	1	1	Black board	1		CO1
1.2	Data Mining Functionalities	T1	2	3	Black board	1		CO1
1.3	Interestingness Patterns	T1	4	6	Black board	1		CO1
1.4	Classification of Data mining Systems	T1	7	7	Black board	1		CO1
1.5	Data mining Task primitives	T1	8	9	Black board	1		CO1
1.6	Integration of Data mining system With a Data warehouse	T1	10	11	Presentation	2		CO1
1.7	Major issues in Data Mining	T1	12	14	Presentation	2		CO1
1.8	Data Preprocessing	T1	15	17	Presentation	2		CO1
1.9	Data warehouse	T1	18	20	Black board	1		CO1
UNIT-II					11			
2.1	Mining Frequent Patterns	T1	22	23	Black board	1		CO2
2.2	Associations and correlations	T1	24	24	Black board	1		CO2
2.3	Mining Methods	T1	25	25	Black board	1		CO2
2.4	Mining Various kinds of	T1	26	28	Black board	2		CO2

	Association Rules							
2.5	Correlation Analysis	T1	29	31	Black board	1		CO2
2.6	Constraint based Association mining	T1	31	33	Black board	1		CO2
2.7	Graph Pattern Mining	T1	35	35	Black board	2		CO2
2.8	SPM	T1	36	37	Black board	2		CO2
UNIT-III					11			
3.1	Classification and Prediction	T1	37	39	Black board	1		CO3
3.2	Basic concepts	T1	39	41	Black board	2		CO3
3.3	Decision tree induction	T1	41	41	Black board	2		CO4
3.4	Bayesian classification	T1	42	44	Black board	2		CO4
3.5	Rule-based classification	T1	44	45	Black board	2		CO4
3.6	Lazy learner	T1	45	46	Black board	2		CO4
UNIT-IV					12			
4.1	Cluster analysis	T1	57	61	Black board	1		CO5
4.2	Types of Data in Cluster Analysis	T1	61	62	Black board	1 2		CO5
4.3	Categorization of Major Clustering Methods	T1	70	71	Black board			CO5
4.4	Partitioning Methods	T1	71	71	Black board	1		CO5
4.5	Hierarchical Methods	T1	72	73	Black board	2		CO5
4.6	Density	T1	73	74	Black board	2		CO5
4.7	Based Methods	T1	74	75	Black board	2		CO5
4.8	Grid-Based Methods	T1	75	76	Black board	1		CO5
4.9	Basic concepts in Miningdata streams	T1	75	76	Black board	1		CO5
UNIT-V					12			
5.1	Mining Time-series data	T1	77	78	Black board	1		CO6
5.2	Mining sequence patterns in Transactional databases	T1	79	81	Black board	1		CO6
5.3	Mining Object	T1	82	84	Black board	2		CO6
5.4	Spatial	T1	85	87	Black board	1		CO6
5.5	Multimedia	T1	88	90	Black board	2		CO6
5.6	Text and Webdata	T1	77	78	Black board	2		CO6

5.7	Spatial Datamining	T1	79	81	Black board	1		CO6
5.8	Multimedia Datamining	T1	82	84	Black board	2		CO6

TEXTBOOKS:

1. Data Mining – Concepts and Techniques – Jiawei Han & Micheline Kamber, 3rd Edition Elsevier.
2. Data Mining Introductory and Advanced topics – Margaret H Dunham, PEA.

REFERENCEBOOK:

Ian H. Witten and Eibe Frank, Data Mining: Practical Machine Learning Tools and Techniques (Second Edition), Morgan Kaufmann, 2005.

WEB REFERENCES

- W1: <https://www.geeksforgeeks.org/data-preprocessing-in-data-mining/>
- W2: <https://www.javatpoint.com/apriori-algorithm>
- W3: https://www.tutorialspoint.com/data_mining/dm_dti.htm
- W4: <https://www.geeksforgeeks.org/data-mining-cluster-analysis/>
- W5: <https://www.geeksforgeeks.org/text-mining-in-data-mining/>

	SRI INDU COLLEGE OF ENGG & TECH QUESTION BANK (Regulation :R22) DEPARTMENT OF CSM			(Regulation :R22) Page: 13 of 7
	Sub.Code & Title	(R22INF3245) DATA MINING		
	Academic Year: 2025-2026	Year/Sem: IV-I	CSE(AI&ML)-(A&B)	
	Faculty Name & Designation	DR.M.NARESH ASSOCIATE PROFESSOR		

QUESTION BANK WITH BLOOMS TAXONOMY LEVEL (BTL)

(1.Remembering 2.Understanding 3. Applying 4.Analyzing
5.Evaluating 6.Creating)

UNIT-I: Data Mining			
	Multiple Choice Questions	BT Level	Course Outcome
1C-1	Smoothing techniques are _____ [] A. binning B. aggregation C. Normalization D. None	I	CO1
1C-2	 is an essential process where intelligent methods are applied to extract data patterns. [] A) Data warehousing B) Data mining C) Text mining D) Data selection	I	CO1
1C-3	Data mining can also apply to other forms such as [] i) Data streams ii) Sequence data iii) Networked data iv) Text data v) Spatial data A) i, ii, iii and v only B) ii, iii, iv and v only C) i, iii, iv and v only D) All i, ii, iii, iv and v	I	CO1
1C-4	Which of the following is not a data mining functionality? [] A) Characterization and Discrimination B) Classification and regression C) Selection and interpretation D) Clustering and Analysis	I	CO1

1C-5	 is a summarization of the general characteristics or features of a target class of data. [] A) Data Characterization B) Data Classification C) Data discrimination D) Data selection	I	CO1
1C-6	 is a comparison of the general features of the target class data objects against the general features of objects from one or multiple contrasting classes. [] A) Data Characterization B) Data Classification C) Data discrimination D) Data selection	I	CO1
1C-7	 is the process of finding a model that describes and distinguishes data classes or concepts. [] A) Data Characterization B) Data Classification C) Data discrimination D) Data selection	I	CO1
1C-8	The various aspects of data mining methodologies is/are [] i) Mining various and new kinds of knowledge ii) Mining knowledge in multidimensional space iii) Pattern evaluation and pattern or constraint-guided mining. iv) Handling uncertainty, noise, or incompleteness of data A) i, ii and iv only B) ii, iii and iv only C) i, ii and iii only D) All i, ii, iii and iv	I	CO1
1C-9	The full form of KDD is [] A) Knowledge Database B) Knowledge Discovery Database C) Knowledge Data House D) Knowledge Data Definition	I	CO1
1C-10	The out put of KDD is [] A) Data B) Information C) Query D) Useful information	I	CO1

1C-11	Choose the options below that a data warehouse can include. [] A . Flat files B . Online data C . Database tables D . All the above	I	CO1
1C-12	Which of the following is not a data mining task? [] A . Classification B . Clustering C . Regression D . Linear Programming	I	CO1
1C-13	DMQL stands for? [] A . Data Mining Query Language B . Dataset Mining Query Language C . DBMiner Query Language D . Data Marts Query Language	I	CO1
1C-14	_____ may be defined as the data objects that do not comply with the general behavior or model of the data available. [] A . Outlier Analysis B . Evolution Analysis C . Prediction D . Classification	I	CO1
1C-15	In which step of Knowledge Discovery, multiple data sources are combined? [] A . Data Cleaning B . Data Integration C . Data Selection D . Data Transformation	I	CO1

Fill in the blanks

1F-1	Extracting knowledge from multiple, heterogeneous and external sources refers to_____	I	CO1
1F-2	_____perform a linear transformation on the original data?	I	CO1
1F-3	Many people treat data mining as synonym for another popularly used term----- --	I	CO1
1F-4	The discrete wavelet transformation is closely related to the ----transform	I	CO1
1F-5	_____ is task of discovering interesting patterns from large amounts of data.	I	CO1

1F-6	_____ data marts are sources directly from enterprise data warehouse.	I	CO1
1F-7	_____ contains a subset of corporate –wide data that is of value to a specific group of users.	I	CO1
1F-8	_____ converts data from legacy or host format to warehouse format. Data transformation	I	CO1
1F-9	_____ is the estimate of the strength of the implication of the rule.	I	CO1
1F-10	Binning Method is used for-----	I	CO1
1F-11	In _____ the Data Mining system will not utilize any function of data warehouse system or database.	I	CO1
1F-12	_____ is one of the data mining functionalities that help the analyst find the missing numeric values.	I	CO1
1F-13	Removing duplicate records is a process called _____	I	CO1
1F-14	_____ is a process of finding potentially useful patterns and valuable information from huge amount of data.	I	CO1
1F-15	_____ is defined as the process where data relevant to the analysis is decided and retrieved from the data collection.	I	CO1
Match the following			
1M-1	Match the following. a) Classification 1) Grouping data into similar objects b) Clustering 2) continuous data into categorical c) Data Mining 3) categorical labels based on input data d) Discretization 4) knowledge to be discovered	II	CO1
1M - 2	Match the following. a Data Visualization 1) Filling in missing values and correcting inconsistent data b. Data Integration 2) specifically focused on in the context of Data Mining c. Data Cleaning 3) Reducing volume but producing similar analysis results d. Data Reduction 4) Combining data from multiple sources into a coherent	II	CO1
1M - 3	Match the following. a) Histogram 1. Displays relationship between two continuous variables b) Scatter Plot 2. Visualizes the distribution of a single continuous variable	II	CO1

	c) Heat map d) Box Plot	3. Shows data distribution and identifies outliers using quartiles 4. Uses colours to represent values in a matrix		
1M - 4	Match the following. a) Data Validation b) Missing Data Imputation c) Outlier Detection d) Data Deduplication	1) Replacing null values with averages 2) Removing repeated customer records 3) Checking if age is between 0 and 120 4) Identifying data points far from others	II	CO1
1M - 5	Match the following. a) Association Rule Mining b) Bar Chart c) non consistency Correction d) Validation	1) Compares quantities across different categories 2) Ensuring data falls 3) Discovering interesting correlations 4) Resolving conflicting data entries	II	CO1
5 - MARKS QUESTIONS				
1D-1	Discuss the types of the data can be Mined.		II I	CO1
1D-2.	Describe the various phases in knowledge discovery process with a neat diagram.		II	CO1
1D-3	Explain about Data Mining Functionalities.			
1D-4.	Discuss about Interestingness Patterns		II II	CO1
1D-5.	List out the applications of data mining.		IV II	CO1
1D-6.	Explain about Classification of Data mining Systems.		II II	CO1
1D-7.	Describe about data mining task primitives.		I,II	CO1
1D-8	Discuss about subset selection.		II	CO1
1D-9	Explain how to integrate data mining system with a data warehouse.		V II	CO1
1D-10	Discuss the major issues in data mining.		II,I	CO1
1D-11	Explain the need of data preprocessing and various forms of preprocessing.		II	CO1
1D-12	Differentiate between data mining and data warehouse.		I II	CO1
1D-13	Explain about the Data Cleaning techniques in detail.		V I	CO1

1D-14	Describe Data Normalization methods in brief	II, II	CO1
1D-15	Explain briefly about Dimensionality Reduction	II, II	CO1

UNIT-II: Association Rule Mining			
	Multiple Choice Questions	BT Level	Course Outcome
2C-1	What does Apriori algorithm do? [] a. It mines all frequent patterns through pruning rules with lesser support b. It mines all frequent patterns through pruning rules with higher support c. Both a and b d. None of the above	I	CO2
2C-2	What does FP growth algorithm do? [] a. It mines all frequent patterns through pruning rules with lesser support b. It mines all frequent patterns through pruning rules with higher support c. It mines all frequent patterns by constructing a FP tree d. All of the above	I	CO2
2C-3	What techniques can be used to improve the efficiency of apriori algorithm? [] a. Hash-based techniques b. Transaction Reduction c. Partitioning d. All of the above	I	CO2
2C-5	What do you mean by support(A)? [] a. Total number of transactions containing A b. Total Number of transactions not containing A c. Number of transactions containing A / Total number of transactions d. Number of transactions not containing A / Total number of transactions	I	CO2
2C-6	How do you calculate Confidence(A -> B)? [] a) Support(A B) / Support(A) b)Support(A B) / Support (A) c) Support(A B) / Support (B) d) none	I	CO2

2C-7	Which of the following is direct application of frequent itemset mining? [] a. Social Network Analysis b. Market Basket Analysis c. Outlier Detection d. Intrusion Detection	I	CO2
2C-8	What is not true about FP growth algorithms? [] a)It mines frequent itemsets without candidate generation. b. There are chances that FP trees may not fit in the memory c. FP trees are very expensive to build d. It expands the original database to build FP trees.	I	CO2
2C-9	When do you consider an association rule interesting? [] a. If it only satisfies min_support b. If it only satisfies min_confidence c. If it satisfies both min_support and min_confidence d. There are other measures to check so	I	CO2
2C-10	What is the difference between absolute and relative support? [] a. Absolute - Minimum support count threshold and Relative - Minimum support threshold b. Absolute - Minimum support threshold and Relative - Minimum support count threshold c. Both mean same	I	CO2
2C-11	What is the relation between candidate and frequent itemsets? [] a. A candidate itemset is always a frequent itemset b. A frequent itemset must be a candidate itemset c. No relation between the two d. Both are same	I	CO2
2C-12	Which algorithm is used for frequent itemset mining? [] A . Decision tree algorithm B . K-nearest neighbors algorithm C . Apriori algorithm D . Naive Bayes algorithm	I	CO2
2C-13	Which algorithm is used for frequent itemset mining? [] A . Decision tree algorithm B . K-nearest neighbors algorithm C . Apriori algorithm D . Naive Bayes algorithm	I	CO2

2C-14	What does FP growth algorithm do? [] A . It mines all frequent patterns through pruning rules with lesser support B . It mines all frequent patterns through pruning rules with higher support C . It mines all frequent patterns by constructing a FP tree D . It mines all frequent patterns by constructing an itemset	I	CO2
2C-15	Which of the following is the direct application of frequent itemset mining? [] A . Social Network Analysis B . Market Basket Analysis C . Outlier Detection D . Intrusion Detection	I	CO2

Fill in the blanks

2F-1	The value that says that transactions in D that support X also support Y is called _____.	I	CO2
2F-2	If T consist of 500000 transactions, 20000 transaction contain bread, 30000 transaction contain jam, 10000 transaction contain both bread and jam. Then the support of bread and jam is _____.	I	CO2
2F-3	The left hand side of an association rule is called_____	I	CO2
2F-4	The right hand side of an association rule is called_____	I	CO2
2F-5	All set of items whose support is greater than the user-specified minimum support are called as_____.	I,II	CO2
2F-6	If a set is a frequent set and no superset of this set is a frequent set, then it is called _____.	I	CO2
2F-7	A priori algorithm is otherwise called as _____.	I	CO2
2F-8	A rule concerns associations between the presence or absence of items, then it referred as -----	I	CO2
2F-9	-----association rule consist of atleast two dimensions.	I	CO2
2F-10	----- algorithm is used to mine frequent item sets using candidate generation.	I	CO2

2F-11	Correlation Analysis is a data mining technique used to identify _____	I	CO2
2F-12	The FP-tree (Frequent Pattern tree) is a data structure used in the FP Growth algorithm that stores the _____ and _____	I	CO2
2F-13	Multilevel association rules can be mined efficiently using _____.	I	CO2
2F-14	Steps in Apriori algorithm are _____ and _____	I	CO2
2F-15	If there is a pair of items, X and Y, which are frequently bought together then association rule is represented as _____	I	CO2
Match the following			
2M-1	Match the following 1) Association Rule a) Probability that Y is purchased given X 2) Confidence b) Measures how much more likely 3) Support c) An implication of the form $X \rightarrow Y$ 4) Lift d) Frequency of the itemset in the dataset	II	CO2
2M - 2	Match the following 1) Item set a) Item sets that occur more than once 2) Frequent Item set b) Application of association rule mining 3) Market Basket Analysis c) Right-hand side of the rule (Y in $X \rightarrow Y$) 4) Consequent d) collection of one or more items	II	CO2
2M - 3	Match the following 1) A priori Algorithm a) Frequency of the item set in the dataset 2) Antecedent b) Right-hand side of the rule (Y in $X \rightarrow Y$) 3) Support c) Left-hand side of the rule (X in $X \rightarrow Y$) 4) Consequent d) Classic algorithm to find frequent item sets	II	CO2
2M - 4	Match the following. 1) Association Rule Mining a) Compares quantities across different categories 2) Bar Chart b) Ensuring data falls 3) non consistency Correction c) Discovering interesting correlations 4) Validation d) Resolving conflicting data entries	II	CO2
5 - MARKS QUESTIONS			
2D-1.	Explain the measures of association rule mining?	I	CO2
2D-2.	Explain in detail about frequent pattern mining in data mining.	II	CO2
2D-3.	Give an overview of correlation analysis.	II	CO2

2D-4.	Compute all the frequent item sets using Apriori algorithm for the given data where min-sup = 2. Confidance =50%. <table><tr><td>TID List of ItemIDs</td><td>TID List of ItemIDs</td></tr><tr><td>1</td><td>I1, I2, I4, I5</td></tr><tr><td>2</td><td>I2, I4, I7</td></tr><tr><td>3</td><td>I2,I3,I4,I5</td></tr><tr><td>4</td><td>I1,I3,I4,I7</td></tr><tr><td>5</td><td>I1,I2,I3,I4,I5</td></tr><tr><td>6</td><td>I3,I4,I5,I6</td></tr></table>	TID List of ItemIDs	TID List of ItemIDs	1	I1, I2, I4, I5	2	I2, I4, I7	3	I2,I3,I4,I5	4	I1,I3,I4,I7	5	I1,I2,I3,I4,I5	6	I3,I4,I5,I6	II	CO2
TID List of ItemIDs	TID List of ItemIDs																
1	I1, I2, I4, I5																
2	I2, I4, I7																
3	I2,I3,I4,I5																
4	I1,I3,I4,I7																
5	I1,I2,I3,I4,I5																
6	I3,I4,I5,I6																
2D-5.	Explain in detail about multilevel association rules.	II	CO2														
2D-6.	Explain the constraint based association mining.	II I	CO2														
2D-7.	Write the FP-growth algorithm.	II	CO2														
2D-8	Explain about the SPM	VI	CO2														
2D-9	Explain about the identification of sub graphs in a graph	IV II	CO2														
2D-10	Quote an example for quantitative association rule.	II	CO2														
2D-11	Describe various types of association rules.	II	CO2														
2D-12	Give a note on Closed Frequent Item	IV I	CO2														
2D-13	Define Association rule mining two step process	III	CO2														
2D-14	Show how to compute confidence of an association rule? Give example	III	CO2														
2D-15	Compute all the frequent item sets using Apriori algorithm for the given data where min-sup = 50% Confidance =70%. <table><tr><td>TID List of ItemIDs</td><td>TID List of ItemIDs</td></tr><tr><td>1</td><td>I1, I2, I4, I5</td></tr><tr><td>2</td><td>I2, I4, I7</td></tr><tr><td>3</td><td>I2,I3,I4,I5</td></tr><tr><td>4</td><td>I1,I3,I4,I7</td></tr><tr><td>5</td><td>I1,I2,I3,I4,I5</td></tr><tr><td>6</td><td>I3,I4,I5,I6</td></tr></table>	TID List of ItemIDs	TID List of ItemIDs	1	I1, I2, I4, I5	2	I2, I4, I7	3	I2,I3,I4,I5	4	I1,I3,I4,I7	5	I1,I2,I3,I4,I5	6	I3,I4,I5,I6	II	CO2
TID List of ItemIDs	TID List of ItemIDs																
1	I1, I2, I4, I5																
2	I2, I4, I7																
3	I2,I3,I4,I5																
4	I1,I3,I4,I7																
5	I1,I2,I3,I4,I5																
6	I3,I4,I5,I6																

	7	I1,I2,I3,I4,I5		
	8	I3,I4,I5		
	9	I2, I4, I7		

UNIT-III: Classification				
	Multiple Choice Questions			Course Outcome
			BT Level	
3C-1	The basic idea of the apriori algorithm is to generate_____ item sets of a particular size & scans the database. [] A. candidate. B. primary. C. secondary. D. superkey.			CO3
3C-2	_____ is the most well known association rule algorithm and is used in most commercial products. [] A. Apriori algorithm. B. Partition algorithm. C. Distributed algorithm. D. Pincer-search algorithm.			CO3
3C-3	An algorithm called _____ is used to generate the candidate item sets for each pass after the first. [] A. apriori. B. apriori-gen. C. sampling. D. partition.			CO3
3C-4	The basic partition algorithm reduces the number of database scans to _____ & divides it into partitions. [] A. one. B. two. C. three. D. four.			CO3
3C-5	_____ and prediction may be viewed as types of classification. [] A. Decision. B. Verification. C. Estimation. D. Illustration.			CO3

3C-6	_____ can be thought of as classifying an attribute value into one of a set of possible classes. [] A. Estimation. B. Prediction. C. Identification. D. Clarification.	I	CO3
3C-7	Prediction can be viewed as forecasting a _____ value. [] A. non-continuous. B. constant. C. continuous. D. variable.	I	CO3
3C-8	_____ data consists of sample input data as well as the classification assignment for the data. [] A. Missing. B. Measuring. C. Non-training. D. Training.	I	CO4
3C-9	Rule based classification algorithms generate _____ rule to perform the classification [] A. if-then. B. while. C. do while. D. switch.	II	CO4
3C-10	A _____ is a flowchart-like tree structure [] A. Navie Bayesian B. Rule Based C. Decision tree D. Binary tree.	I	CO4
3C-11	To remove sub-nodes of a decision node, _____ is used. [] A . Clustering B . Classification C . Regression D . Pruning	I	CO4
3C-12	ID3 use _____ attribute selection measure. [] A . Information Gain B . Gain Ratio C . Gini Index D . Tree pruning.	I	CO4
3C-13	Decision tree builds _____ models in the form of a tree structure[] A . classification B . regression C . both D . none	I	CO4

3C-14	Naive Bayes classifiers are a collection of classification algorithms based on____[] A . Naive concept B . Bayes theorem C . Duch rule D . None	I	CO4
3C-15	A classification model can overfit the test data due to presence of_____.[] A . Data B . Information C . Light D . Noise	I	CO4
Fill in the blanks			
3F-1	-----is the process of scaling data such that it falls with in a specified range.	II	CO3
3F-2	. A-----represent the test on an attribute	II	CO3
3F-3	-----represents the outcome of the test	II	CO3
3F-4	. The ----- represents the class label	II	CO3
3F-5	-----gives the impurity of the data partition	II	CO3
3F-6	Baye's Theorem____	II	CO3
3F-7	_____ are the methods to remove model over fitting	II	CO3
3F-8	A categorical variable is generalized form of -----with more than two states.	II	CO4
3F-9	classification is _____	III	CO4
3F-10	over fitting means _____	III	CO4
3F-11	Is discrimination between spam and ham emails a classification task? (True/False) _____.	III	CO4
3F-12	Data classification is a two-step process, consisting of a _____ and _____.	III	CO4
3F-13	_____results in a simple tree and produces, the least impurity in the partition.	III	CO4
3F-14	he accuracy of classification model is evaluated using the formula_____	III	CO4

3F-15	_____ of a decision tree has only one incoming edge and no outgoing.	III	CO4
Match the following			
3M-1	Match the following 1) Binary Classification a) Classifying handwritten digits (0–9 2) Multi-class Classification b) A movie tagged as [Action, Drama, Thriller] 3) Multi-label Classification c) Fraud detection where fraud cases are rare 4) Imbalanced Classification d) Classifying email as spam or not spam	II	CO3
3M - 2	Match the following 1) Hierarchical Classification a) Predicting discrete labels or categories 2) Classification b) Learning from labeled training data 3) Prediction c) Categorizing animals: Animal 4) Supervised Learning d) Forecasting temperature for tomorrow	II	CO3
3M - 3	Match the following 1) Supervised Learning a) A model used to assign labels 2) Classification b) Classifying email as spam or not spam 3) Classifier c) Learning from labeled training data 4) Binary Classification d) Suspend a thread for a period of time	II	CO4
3M-4	Match the following 1) Hierarchical Classification a) Predicting discrete labels or categories 2) Classification b) Learning from labeled training data 3) Prediction c) Categorizing animals: Animal 4) Supervised Learning d) Forecasting temperature for tomorrow		
5 - MARKS QUESTIONS			
3D-1.	Compare the methods of classification and prediction	II	CO3
3D-2.	Explain classification as a two step process.	III	CO3
3D-3.	Explain the way of evaluating performance of classification model?	II	CO3
3D-4.	Discuss decision tree over fitting and pruning techniques.	VI	CO3
3D-5.	Explain about the measures for selecting the best split.	II	CO3
3D-6.	List the characteristics of k-nearest neighbor algorithm.	I	CO4
3D-7.	Discuss in detail about rule-based classification.	III	CO4
3D-8	Explain the merits and de-merits of the lazy learning method.	I	CO4
3D-9	Define Pre pruning and post pruning.	I	CO4

3D-10	Describe about Lazy Learners for classification.	II	CO4
3D-11	Explain the way of solving a classification problem using k-nearest neighbor algorithm?	IV II	CO4
3D-12	Discuss about attribute selection measure Information Gain.	II II	CO4
3D-13	Describe Bayesian Belief Network with an example.	II II	CO4
3D-14	Explain the need for tree pruning in decision tree induction?	II	CO4
D-15	Define the significance of information gain?	II	CO4

UNIT-IV: Clustering and Applications			
	Multiple Choice Questions	BT Level	Course Outcome
4C-1	_____ Clustering Technique Start With As Many Clusters As There Are Records, With Each Cluster Having Only One Record. [] A) Agglomerative. B. Divisive. C. Partition. D. Numeric.	I	CO5
4C-2	In _____ Algorithm Each Cluster Is Represented By The Center Of Gravity Of The Cluster. [] A. K-Medoid. B. K-Means. C. STIRR. D. ROCK.	I	CO5
4C-3	In _____ Each Cluster Is Represented By One Of The Objects Of The Cluster Located Near Thecenter. [] A. K-Medoid. B. K-Means. C. Stirr. D. Rock.	I	CO5
4C-4	Pick Out A Hierarchical Clustering Algorithm. [] A. Dbscan B. Birch. C. Pam. D. Cure.	I	CO5

4C-5	CLARANS Stands For _____. [] A. CLARA Net Server. B. Clustering Large Application Range Network Search. C. Clustering Large Applications Based On Randomized Search. D. Clustering Application Randomized Search.	I	CO5
4C-6	Interval Scaled Variables Are----Measurements Of A Linear Scale. [] A) Numeric B)Continuous C)Differentiable D) Non-Continuous	I	CO5
4C-7	Clustering Large Applications Can Be Shortened As ---- [] A) DBSCAN B)OPTICS C)STING D)CLARA	I	CO5
4C-8	Most Of The Partitioning Methods Cluster Objects Are Based On ---- [] A) Number Of Clusters B)Distance Between Objects C)Number Of Objects In Each Class D)Learning Rate	I	CO5
4C-9	DBSCAN Is A -----Clustering Algorithm [] A) Partitioning Methods B)Hierarchical Methods C)Density Based Methods D)Grid Based Methods	I	CO5
4C-10	Which Of The Following Is Required By K-Means Clustering ? [] A) Defined Distance Metric B) Number Of Clusters C) Initial Guess As To Cluster Centroids D) All Of The Mentioned	I	CO5
4C-11	Pick Out A Hierarchical Clustering Algorithm. [] A. DBSCAN B. BIRCH. C. PAM. D. CURE.		

4C-12	The Occurrences Of _____ Is Inevitable When Euclidean Distance Criterion Is Employed. [] A . Outlier B . Empty Cluster C . Centroids D . DBSCAN		
4C-13	_____ Method Works By Grouping Data Objects Into A Tree Of Clusters[] A . Partition Based Clustering B . Grid Based Clustering C . Density Based Clustering D . Hierarchical Based Clustering		
4C-14	In _____ Algorithm, Each Cluster Is Represented By The Mean Value Of The Objects In The Cluster. [] A . K-Means B . K-Medoid C . DBSCAN D . Hierarchal		

Fill in the blanks			
4F-1	Cluster contains a set of_____ data objects	I	CO5
4F-2	_____methods can find arbitrary shaped clusters.	I	CO5
4F-3	Dissimilarity matrix is also called as -----structure.	I	CO5
4F-5	A categorical variable is generalized form of -----with more than two states.	I,II	CO5
4F-6	Data matrix is also called as -----structure.	I	CO5
4F-7	Rock stands for -----	I	CO5
4F-8	-----is a statistical information grid approach	I	CO5
4F-9	Data which are inconsistent with the remaining set of data is called as _____	I	CO5
4F-10	A hierarchical method can be classified as being either _____	I	CO5
4F-11	._____methods quantize the object space into a finite number of cells that form a grid structure.	I	CO5
4F-12	._____ methods hypothesize a model for each of the clusters and find the best fit of the data to the given model.	I	CO5
4F-13	Data matrix is also called as -----structure	I	CO5
4F-14	._____ clustering approach that performs clustering by incorporation of user-specified or application-oriented constraints.	I	CO5
4F-15	_____ is a "top-down" approach in hierarchical clustering m	I	CO5
Match the following			
4M-1	Match the following 1) K-means 2) Hierarchical 3) DBSCAN 4) OPTICS	a) Builds tree of clusters b) finds clusters by density c) orders points to handle varying density d) when the item selected or deselected	II CO5
4M -2	Match the following 1) Partitioning Methods 2) Hierarchical Methods 3) Density-Based Methods 4) Grid-Based Methods	a) Create a tree of clusters b) Form clusters based on regions of high density c) Quantize data space into finite cells to d) Divide data into k clusters	II CO5

4M -3	Match the following 1) Statistical Methods 2) Distance-Based Methods 3) Machine Learning Methods 4) Clustering-Based Methods	a) distance metrics are outliers b) normal data to detect deviations c) identify points with low probability d) small clusters can be outliers.	II	CO5
4M -4	Match the following 1) OPTICS 2) K-means 3) Machine Learning Methods 4) Clustering-Based Methods	a) Partitioning method b) orders points to handle varying density c) small clusters can be outliers d) normal data to detect deviations	II	CO5
5 - MARKS QUESTIONS				
4D-1.	Explain the applications of cluster analysis,		II	CO5
4D-2.	Explain different Clustering Method.		III	CO5
4D-3.	Describe K-Means Additional issues?		II	CO5
4D-4.	List out all partitioning methods for clustering data.		III	CO5
4D-5.	Give a brief note on PAM(K-Medoids) Algorithm with example.		IV	CO5
4D-6.	Discuss the two approaches to improve quality of hierarchical clustering		V	CO5
4D-7.	Explain Hierarchical clustering.		I	CO5
4D-8	Discuss hierarchical methods for clustering and contrast agglomerative and divisive approaches.		II	CO5
4D-9	Discuss the merits and demerits of hierarchical approaches for clustering		II	CO5
4D-10	Compare agglomerative and divisive methods.		II	CO5
4D-11	Explain about Density Based(DBSCAN) method is used for clustering?		II	CO5
4D-12	Discuss about the drawbacks of k-means algorithm? How can we modify the algorithm to diminish that problem?		VI	CO5
4D-13	Differentiate between clustering and classification		VI	CO5
4D-14	Explain about the Grid-based methods.		I	CO5
4D-15	Compare the performance of various outlier detection approaches		II	CO5

UNIT-V: Advanced Concepts			
	Multiple Choice Questions	BT Level	Course Outcome
5C-1	-----sequence data consist of long sequences of numeric data, [] recorded at equal time intervals . a. time-series data b. Symbolic sequence data c. Biological sequences d.sequence data	I	CO6
5C-2	-----consist of long sequences of event or nominal data, which typically are not observed at equal time intervals. [] a. time-series data b. Symbolic sequence data c. Biological sequences d.sequence data		CO6
5C-3	-----sequences include DNA and protein sequences. [] a. time-series data b. Symbolic sequence data c. Biological sequences d.sequence data	I	CO6
5C-4	For effective trend analysis, the data often need to be “deseasonalized” based on a -----computed by autocorrelation [] a. seasonal index b. long-term movements c. Cyclic movements d. Random movements	I	CO6
5C-5	----- are used to represent the probabilities of substitutions of nucleotides or amino acids and probabilities of insertions and deletion [] a. Substitution matrices b. multiple sequence c. pairwise sequence d. a phylogenetic tree	I	CO6
5C-6	----- data are data that relate to both space and time[] a. Time-series b. Spatiotemporal c. Multimedia d. Text and web	I	CO6
5C-7	A----- typically consists of a large number of interacting physical and [] information components a. cyber-physical system b. cyber-crime system	I	CO6

	c. crime-physical system d. cyber-data system		
5C-8	-----mining is an interdisciplinary field that draws on information retrieval, data mining, [] a. Data stream b. Text c. Spatial d. Multimedia	I	CO6
5C-9	-----mining analyzes web content such as text, multimedia data, and structured Data. [] a. Web content b. Spatial c. Multimedia d. Data stream	I	CO6
5C-10	----- are used to represent the probabilities of substitutions of nucleotides or amino acids and probabilities of insertions and deletion [] a. Substitution matrices b. multiple sequence c. pairwise sequence d. a phylogenetic tree	I	CO6
5C-11	-----consist of long sequences of event or nominal data, which typically are not observed at equal time intervals. [] a. time-series data b. Symbolic sequence data c. Biological sequences d. sequence data	I	CO6
5C-12	Which of the following is not type of spatial data? [] A . Line data B . Point data C . Polygon data D . Text data	I	CO6
5C-13	For effective trend analysis, the data often need to be “deseasonalized” based on a -----computed by autocorrelation [] a. seasonal index b. long-term movements c. Cyclic movements d. Random movements	I	CO6
5C-14	----- data are data that relate to both space and time [] a. Time-series b. Spatiotemporal c. Multimedia d. Text and web	I	CO6

5C-15	A -----typically consists of a large number of interacting physical and information components a. cyber-physical system b. cyber-crime system c. crime-physical system d. cyber-data system	I	CO6
-------	---	---	-----

Fill in the blanks			
5F-1	_____ mining is the process of using graph and network mining theory and methods to analyze the nodes and connection structures on the Web.	I	CO6
5F-2	_____ refer to data that flow into a system in vast volumes, change dynamically, are possibly infinite, and contain multidimensional features.	I	CO6
5F-3	“High quality” in text mining usually refers to a combination of _____	I	CO6
5F-4	CPS systems may be interconnected so as to form large _____ cyber-physical networks.	I	CO6
5F-5	Spatial data, in many cases, refer to _____ data stored in geospatial data repositories	I,II	CO6
5F-6	A _____ data set consists of sequences of numeric values obtained over repeated measurements of time.	I	CO6
5F-7	Many time-series similarity queries require _____ matching.	I	CO6
5F-8	An _____ is the process of lining up sequences to achieve a maximal identity	I	CO6
5F-9	The _____ is a synopsis data structure that can be used to approximate the frequency distribution of element values in a data stream.	I	CO6
5F-10	_____ pattern is considered frequent if its count satisfies a _____ support	I	CO6
5F-11	NLP stands for _____	I	CO6

5F-12	Mining multiple types of data is called as _____	II	CO6
5F-13	The evaluation of web page relevancy can be carried out by _____ method.	II	CO6
5F-14	In _____ multiple pages point to a single page	I	CO6
5F-15	The process of reducing words to their morphological route is called _____	I	CO6

Match the following

5M-1	Match the following 1) Tokenization a) Removing common words like 2) Stop Words Removal b) Breaking text into words or meaningful units 3) Stemming c) Reducing words to their dictionary base form using context 4) Lemmatization d) Reducing words to their root form	II	CO6
5M - 2	Match the following 1) Sentiment Analysis a) Discovering abstract topics in a collection of documents 2) TF-IDF b) Identifying names of people, places, organizations 3) Named Entity Recognition c) Determining positive, negative. 4) Topic Modeling d) Weighing terms by their frequency and uniqueness	II	CO6
5M - 3	Match the following 1) Page Rank a) Segmenting user web activity into meaningful sessions 2) Session Identification b) Mining user behavior patterns from web logs and 3) Web Usage Mining c) Analyzing the link structure between web pages 4) Web Structure Mining d) Algorithm to rank web pages based on link popularity	II	CO6
5M - 4	Match the following 1) Sentiment Analysis a) Discovering abstract topics in a collection 2) TF-IDF b) Identifying names of people, places, organizations 3) Named Entity Recognition c) Determining positive, negative. 4) Topic Modeling d) Weighing terms by their frequency and uniqueness	II	CO6

5 - MARKS QUESTIONS

5D-1.	Explain the applications of Web Mining?.	IV	CO6
5D-2.	Explain in detail about Time series Mining.	I	CO6
5D-3.	Define Structured data	II	CO6

		IV	CO6
5D-4.	Describe the Terminology of Web Structure?	II	CO6
5D-5.	Explain the following: i) Spatial data mining. ii) Mining sequence patterns in transactional databases	IV	CO6
5D-6.	Define Text Mining?	V	CO6
		IV	CO6
5D-7.	Explain the methods under Text Transformation.	II	CO6
5D-8	Explain the way of applying the mining techniques to unstructured text database?	II	CO6
5D-9	Give examples for unstructured text	V	CO6
5D-10	Discuss various kinds of patterns to be mined from web/server logs in web usage	I	CO6
5D-11	Discuss in detail about web content mining	II	CO6
5D-12	Explain the Advantages & Disadvantages of Text Mining?	I	CO6
5D-13	Explain about the Multimedia data mining	III	CO6
5D-14	List the applications of web usage mining	II	CO6
5D-15	Explain about the World wide web mining	II	CO6

SRI INDUCOLLEGE OF ENGINEERING& TECHNOLOGY

(An Autonomous Institution under UGC, New Delhi)

Recognized under 2(f) and 12(B) of UGC Act 1956

IV B.Tech- I Semester – End Examinations (Model paper)**DATA MINING****CSE(AIML)****Duration: 3 Hrs****Max Marks: 60****Section – A**

Answer <u>All</u> the following questions		Marks: 10Qx1M = 10M		
		Marks	Course Outcome	BT Level
1	Define Pre pruning and post pruning.	1	CO1	2
2	Define maximal frequent item set.	1	CO1	1
3	List out the applications of data mining.	1	CO2	1
4	Define associative classification?	1	CO2	1
5	List the advantages of using decision trees.	1	CO3	2
6	Define the significance of information gain?	1	CO4	2
7	Quote an example for quantitative association rule.	1	CO5	2
8	Define Clustering?	1	CO5	2
9	Define outlier?	1	CO6	2
10	List out all partitioning methods for clustering data.	1	CO6	2

Section – B

Answer any <u>FIVE</u> questions choosing at least one from each Unit Marks: 5Qx10M = 50M			
	Marks	Course Outcome	BT Level
UNIT-I			
6. Define Data mining? Explain essential steps involved in the process of knowledge discovery with a neat sketch. (OR)	10	CO1	2,1
7. List out data pre-processing techniques and explain about data cleaning methods?			
UNIT-II			
8. Explain in detail about frequent pattern mining in data mining. (OR)	10	CO2	1,3
9. Explain in detail about frequent pattern mining in data mining.			
UNIT – III			
10. Discuss about Decision tree induction algorithm with an example (OR)	10	CO3,CO4	3,2
11. Explain the way of solving a classification problem using k-nearest neighbor algorithm?			
UNIT – IV			
12. Discuss hierarchical methods for clustering and contrast agglomerative and divisive approaches.	10	CO5	5,2

(OR) 13. Discuss about the drawbacks of k-means algorithm? How can we modify the algorithm to diminish that problem?			
UNIT – V 14. Write short notes on a. Spatial Database. b. Time Series Databases c. Multimedia Mining d. World Wide Web (OR) 15 Explain outlier analyses with an example?	10	CO6	2,3