

DEPARTMENT OF STATISTICS				CLASS: I M.Sc. Statistics				
Sem	Course Type	Course Code	Course Title	Credits	Contact Hours/week	CIA	Ext	Total
I	C Elective – 2(B)	21P1SME1 (B)	Data Mining	4	5	25	75	100

Nature of Course			
Knowledge and skill	✓		Employability oriented
Skill oriented			Entrepreneurship oriented

### Course Objectives:

1. To teach principles, concepts and applications of data warehousing and data mining
2. To introduce the task of data mining as an important phase of knowledge recovery process
3. Characterize the kinds of patterns that can be discovered by association rule mining

Unit	Description	Hours	K-level	CLO(s)
I	Introduction: Data mining - Kinds of data – Data mining Functionalities - Classification of Data mining Systems - Major Issues on Data mining - Introduction to OLAP - OLAP technology for Data Mining - Data warehousing - Data warehousing to Data mining - Optimizing Data for mining - Data preprocessing	16	K3	1
II	Data Mining Primitives: Data mining Query language - Association Rules in large - Data mining - KDD Process - Fuzzy sets and logic - Classification and Prediction: Information retrieval - Dimensional Modeling of Data - Pattern Matching - Estimation Error- EM and MLE.	15	K2	2
III	Models based on Summarization: Bayes Theorem - Chi squared Statistics - Regression - Decision Tree - Neural Networks - Genetic Algorithms - Cluster Analysis – Outlier - Cluster Vs Classification - Clustering Issues - Impact of Outliers on clustering- Clustering problems - Clustering Approaches.	17	K3	3
IV	Clustering Algorithms: Hierarchical algorithm – Single Link- MST Single Link - Complete Link - Average Link- Dendrogram - Partitional Algorithm – MST - Squared Error - K-Means - Nearest Neighbor – PAM – BEA – GA - Categorical algorithm - Large Database.	12	K4	4
V	Web Mining: Introduction - Web data - Web Knowledge Mining Taxonomy - Web Content mining - Web Usage Mining Research - Ontology based web mining Research - Web mining Applications.	15	K4	5

### **Books for Reference:**

1. Berry, J.A. and Linoff, G.S. (2011) Data Mining Techniques (3/e), Wiley, New York.
2. Chattamvelli, R. (2009) Data mining Methods, Alpha Science International, Oxford.
3. Dunham, M.H. (2006) Data mining: Introductory and Advanced Topics, Pearson, New Delhi .
4. Gorunescu, F. (2010) Data mining Concepts, Models and Techniques. Springer, New York.
5. Han, J. and Kamber, M (2001) Data Mining Concepts and Techniques, Morgan Kaufmann Publications, Massachusetts.
6. Hand, D. Mannila, H. and Smyth, P (2001) Principles of Data mining, MIT Press, Cambridge.
7. Larose, D.T. (2005) Discovering Knowledge in Data: An Introduction to Data Mining, Wiley, Toronto.
8. Pujari, A.K. (2001) Data Mining Techniques, Universities Press, Hyderabad.
9. Sivanandam S.N. and Sumathi, S (2006) Data Mining Concepts, Tasks and Techniques, Springer, New Delhi.
10. Brieman, L. Friedman, J.H. Olshen, R.A. and Stone, C.J. (1984) Classification and regression trees, Wadsworth & Brooks, California.

### **Web References:**

1. Types of Sources of Data in Data Mining  
<https://www.geeksforgeeks.org/types-of-sources-of-data-in-data-mining/>
2. Introduction to OLAP - OLAP technology for Data Mining  
[https://www.tutorialspoint.com/dwh/dwh\\_olap.htm](https://www.tutorialspoint.com/dwh/dwh_olap.htm)
3. Data mining Primitives  
[http://www.brainkart.com/article/Data-mining-primitives\\_8311/](http://www.brainkart.com/article/Data-mining-primitives_8311/)
4. Neural Network <https://www.getsmarter.com>
5. Web Mining  
<https://www.techopedia.com/definition/15634/web-mining>

### **Rationale for Nature of the course**

Analyzes massive volumes of data to discover business intelligence that helps companies solve problems, mitigate risks, and seize new opportunities.

### **Activities having direct bearing on Skill development / Employability / Entrepreneurship**

Seminars / Assignments on different algorithms in current scenario

### **Pedagogy**

Chalk and Talk, PPT, Seminar, Interaction, Problem solving.

## Lecture Schedule

Unit	Topics	Hours	Mode
I	Introduction, Data mining, Kinds of data, Data mining Functionalities	3	PPT, Chalk and Talk and Assignments
	Classification of Data mining Systems, Major Issues on Data mining	3	
	Introduction to OLAP - OLAP technology for Data Mining	3	
	Data warehousing, Data warehousing to Data mining	4	
	Optimizing Data for mining, Data preprocessing	3	
II	Data mining Query language , Association Rules in large	3	PPT, Chalk and Talk and Assignments
	Data mining, KDD Process, Fuzzy sets and logic	3	
	Classification and Prediction, Information retrieval	3	
	Dimensional Modeling of Data, Pattern Matching	3	
	Estimation Error, EM and MLE	3	
III	Bayes Theorem , Chi squared Statistics	3	PPT, Chalk and Talk, Assignments and seminar
	Regression, Decision Tree, Neural Networks	3	
	Genetic Algorithms , Cluster Analysis , Outlier	3	
	Cluster Vs Classification , Clustering Issues	3	
	Impact of Outliers on clustering	2	
	Clustering problems , Clustering Approaches	3	
IV	Clustering Algorithms, Hierarchical algorithm, single link	2	PPT, Chalk and Talk, Assignments and seminar
	MST Single Link - Complete Link - Average Link	2	
	Dendrogram - Partitional Algorithm	2	
	MST - Squared Error - K-Means	2	
	- Nearest Neighbor – PAM – BEA – GA -	2	
	Categorical algorithm - Large Database	2	
V	Web Mining Introduction , Web data	3	PPT, Chalk and Talk, Assignments and seminar
	Web Knowledge Mining Taxonomy	3	
	Web Content mining	3	
	Web Usage Mining Research	3	
	Ontology based web mining Research, Applications	3	

### Course Learning Outcomes

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

CLO's	Course Learning Outcomes	Knowledge Level
CLO-1	Explain the concepts of data warehouse and Analyze OLAP tool Analyze the basic functions of data warehouse and data mining.	Up to K3
CLO-2	Compare and evaluate different data mining techniques like classification, prediction and association rule mining	Up to K2
CLO-3	Categorize and carefully differentiate between situations for applying different data-mining techniques: Chi square, Regression, cluster, and outlier analysis.	Up to K3
CLO-4	Evaluate the performance of different data-mining algorithms.	Up to K4
CLO-5	Describe complex data types with respect to web mining	Up to K4

### MAPPING CLOs WITH PSOs

#	PSO1	PSO2	PSO3	PSO4	PSO5	PSO6	PSO7
CLO-1				3			2
CLO-2		3	1				
CLO-3		1				2	
CLO-4	3	2	3				
CLO-5						1	3

### CIA I – Blue Print

Units	CLOs	K- Level	Section A		Section B		Section C	
			Short Answers		(Either/or Choice)		(Open Choice)	
			No. of Questions	K- Level	No. of Questions	K- Level	No. of Questions	K- Level
1	CLO 1	Up to K3	3	K2, K2,K3	2	K3,K3	2	K3, K3
2	CLO 2	Up to K2	2	K1, K1	2	K2,K2	1	K2
No. of Questions to be asked			5		4		3	
No. of Questions to be answered			5		2		2	
Marks for each question			2		5		10	
Total Marks for each section			10		10		20	

### CIA-I :: Distribution of Section-wise Marks with K levels

K Levels	Section A (No choice)	Section B (Either/ or)	Section C (Open choice)	Total marks	% of marks without choice	Consolidated
K1	4	-	-	4	6.67	46.67%
K2	4	10	10	24	40.00	
K3	2	10	20	32	53.33	53.33%
K4	-	-	-	-	-	-
K5	-	-	-	-	-	-
Total Marks	10	20	30	60	100	100%

**CIA II – Blue Print**

Units	CLOs	K- Level	Section A		Section B		Section C	
			Short Answers		(Either/or Choice)		(Open Choice)	
			No. of Questions	K- Level	No. of Questions	K- Level	No. of Questions	K- Level
1	CLO 3	Up to K3	3	K1,K1, K2	2	K3, K3	2	K3, K3
2	CLO 4	Up to K4	2	K2, K3	2	K4, K4	1	K4
No. of Questions to be asked			5		4		3	
No. of Questions to be answered			5		2		2	
Marks for each question			2		5		10	
Total Marks for each section			10		10		20	

**CIA-II :: Distribution of Section-wise Marks with K levels**

K Levels	Section A (No choice)	Section B (Either/ or)	Section C (Open choice)	Total marks	% of marks without choice	Consolidated
K1	4	-	-	4	6.67	13.34%
K2	4	-	-	4	6.67	
K3	2	10	20	32	53.33	53.33%
K4	-	10	10	20	33.33	33.33%
K5	-	-	-	-	-	-
Total Marks	10	20	30	60	100	100%

### Summative Examination – Blue Print

Units	CLOs	K Level	Section A		Section B		Section C (Either/or Choice)	Section D (Open Choice)
			MCQs		Short Answers			
			No. of questions	K Level	No. of questions	K level		
1	CLO -1	Up to K3	2	K2 & K3	1	K1	2(K2 & K2)	1(K3)
2	CLO -2	Up to K2	2	K1 & K1	1	K1	2(K1 & K1)	1(K2)
3	CLO -3	Up to K3	2	K2 & K3	1	K2	2(K3 & K3)	1(K3)
4	CLO -4	Up to K4	2	K3 & K4	1	K3	2(K4 & K4)	1(K4)
5	CLO -5	Up to K4	2	K3 & K4	1	K2	2(K4 & K4)	1(K4)
No. of Questions to be asked			10		5		10	5
No. of Questions to be answered			10		5		5	3
Marks for each question			1		2		5	10
Total Marks for each Section			10		10		25	30

#### Distribution of section wise marks with K levels for Summative Examination

K Levels	Section A MCQs	Section B (Short Answers)	Section C (Either/ or)	Section D (Open choice)	Total marks	% of marks without choice	Consolidated
K1	2	4	10	-	16	13.33	35%
K2	2	4	10	10	26	21.67	
K3	4	2	10	20	36	30	30%
K4	2	-	20	20	42	35	35%
K5	-	-	-	-	-	-	-
Total marks	10	10	50	50	120	100	100%

#### Course Designers:

1. Dr. M. Venkateswaran
2. Dr. A. Saberunnisa