INTRODUCTION TO DATA MINING

(Open Elective - II)

Course Code	20IT2601A	**Year**	III	**Semester**	II
Course Category	OE - 2	**Branch**	Offered by IT	**Course Type**	Theory
Credits	3	**L-T-P**	3-0-0	**Prerequisites**	
Continuous Internal Evaluation:	30	**Semester End Evaluation:**	70	**Total Marks:**	100

Course Outcomes		**Blooms Taxonomy Level**
Upon Successful completion of course, the student will be able to		
CO1	Understand the basic principles, process and techniques of data mining.	L2
CO2	Use pre-processing techniques on different datasets.	L3
CO3	Apply techniques and algorithms for Mining frequent patterns, classifying and clustering data.	L3
CO4	Analyze the data for mining frequent patterns, associations, classification and outlier detection in a real scenario.	L4

Contribution of Course Outcomes towards achievement of Program Outcomes & Strength of correlations(3:Substantial,2:Moderate,1:Slight)

	PO1	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PO8**	**PO9**	**PO10**	**PO11**	**PO12**	**PSO1**	**PSO2**
CO1	3												3	
CO2	3			3									3	
CO3	3			3									3	3
CO4	3	3											3	3

Syllabus

Unit No	**Contents**	**Mapped CO**
I	**Introduction:** What is data mining? What kinds of data can be mined? What kinds of pattern can be mined? Which technologies are used? Which kinds of applications are targeted?, Major Issues in Data Mining.	CO1

II	**Getting to Know Your Data:** Data objects and Attribute Types, Basic statistical descriptions of data, Measuring Data Similarity and Dissimilarity. Data Preprocessing: An overview, Data Cleaning, Data integration, Data Reduction, Data Transformation and Discretization.	CO1 CO2
III	**Mining frequent patterns, Associations and Correlations**- Basic Concepts, Frequent itemset Mining methods- Apriori Algorithm, Generating association rules from frequent itemsets, improving the efficiency of Apriori.	CO1 CO3 CO4
IV	**Classification:** Basic Concepts – Basic concepts, Decision Tree Induction, Rule Based Classification, Model evaluation and Selection.	CO1 CO3 CO4
V	**Cluster Analysis:** Basic Concepts and Methods- Cluster Analysis, partitioning methods, Hierarchical Methods and evaluation of Clustering	CO1 CO3 CO4

Learning Resources

Text Books

1. Jiawei Han and Micheline Kamber, "Data Mining Concepts and Techniques" Third Edition, Elsevier, 2012.

References

1. Michael Steinbach, Vipin Kumar, Pang-Ning Tan, Introduction to data mining, First Edition, Addison Wesley, 2006
2. Margaret H. Dunham, Data Mining Introductory and Advanced Topics, 1/e, Pearson Publishers, 2006

e-Resources& other digital material

1. https://www.coursera.org/lecture/code-free-data-science/introduction-to-data-mining-hbb2V
2. https://onlinecourses.swayam2.ac.in/cec19_cs01/preview material