PVP SIDDHARTHA INSTITUTE OF TEHNOLOGY, KANURU, VIJAYAWADA
(AUTONOMOUS)

INFORMATION TECHNOLOGY

DATAWAREHOUSING AND DATA MINING

Course Code	19IT4602E	**Year**	III	**Semester**	II
Course Category	PE	**Branch**	IT	**Course Type**	Theory
Credits	3	**L-T-P**	3-0-0	**Prerequisites**	DBMS
Continuous Internal Evaluation :	30	**Semester End Evaluation:**	70	**Total Marks:**	100

Course Outcomes		**Blooms Taxonomy Level**
Upon successful completion of the course, the student will be able to		
CO1	Understand the basic principles of Data Mining and data preprocessing.	L2
CO2	Differentiate the concepts of data warehousing and OLTP.	L3
CO3	Relate the learned algorithms in association and pattern mining to the practical issues.	L3
CO4	Describe and utilize a range of techniques for classifying the data and accuracy improvements.	L3
CO5	Analyze the data and develop some clustering and outlier methods.	L3

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

	PO 1	**PO 2**	**PO 3**	**PO 4**	**PO 5**	**PO 6**	**PO 7**	**PO 8**	**PO 9**	**PO 10**	**PO 11**	**PO 12**	**PSO 1**	**PSO 2**
CO 1	3	2			3								2	2
CO 2		2		2									2	2
CO 3	3	2		3	3								2	2
CO 4		3			3								2	2
CO 5		3			3								2	2

Syllabus

Unit No	**Contents**	**Mapped CO**
I	**Data warehousing** - Data warehouse: Basic Concepts, Data warehouse Modeling: Data Cube and OLAP. Data Mining-What is	**CO1**

	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?	
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.	**CO2**
III	**Mining frequent patterns, Associations and Correlations** - Basic Concepts, Frequent item set Mining methods- Apriori Algorithm, Generating association rules from frequent item sets, improving the efficiency of Apriori, A pattern growth approach for mining frequent item sets. Which patterns are interesting- pattern evaluation methods	**CO3**
IV	**Classification**: Basic Concepts – Basic concepts, Decision Tree Induction, Baye's Classification Methods, Rule based Classification, Classification by Back propagation, Support Vector Machines, Classification using frequent patterns, Lazy Learners, Other classification methods, Model evaluation and Selection, Techniques to improve Classification Accuracy	**CO4**
V	**Cluster Analysis** - Basic Concepts and Methods- Cluster Analysis, partitioning methods, Hierarchical Methods, Density-based methods, Grid-based methods and evaluation of Clustering	**CO5**

Learning Recourses
Text Books
1. Jiawei Han, MichelineKamber and Jian Pei, Data Mining Concepts and Techniques, 3/e, Morgan Kaufmann Publishers, Third edition, 2011.
References
1. Michael Steinbach, Vipin Kumar, Pang-Ning Tan, Introduction to data mining, 1/e, Addison Wesley, 2006 2. Margaret H. Dunham, Data Mining Introductory and Advanced Topics, 1/e, Pearson Publishers, **2006**
e-Resources & other digital material
NPTEL VIDEO LECTURES