COURSE TITLE	**DATA MINING AND INFORMATION RETRIEVAL**
COURSE CODE	**01CE0515**
COURSE CREDITS	**3**

Objective:

1 The goal of the course is to learn how to understand and apply different data mining and information retrieval techniques. It also emphasizes on statistical analysis techniques and algorithms that are used to identify trends in data.

2 The goal of the course is to learn how to understand and apply different data mining and information retrieval techniques. It also emphasizes on statistical analysis techniques and algorithms that are used to identify trends in data

Course Outcomes: After completion of this course, student will be able to:

1 Understand various data warehousing techniques used in industry. (Analysis)

2 Apply different pre-processing techniques on data set. (Apply)

3 Apply the association rules mining over various data sets. (Apply)

4 Apply and demonstrate different classification and Clustering methods on data set. (Apply)

5 Understand and correlate the concepts of various components of Information Retrieval Systems (Understand)

Pre-requisite of course:NA

Teaching and Examination Scheme

Theory Hours	Tutorial Hours	Practical Hours	ESE	IA	CSE	Viva	Term Work
2	0	2	50	30	20	25	25

Contents : Unit	Topics	Contact Hours
1	**Introduction to data mining and Data warehousing:** Motivation for Data Mining, Definition and Functionalities, Classification of DM Systems, DM task primitives, Integration of a Data Mining system with a Database or a Data Warehouse, Issues in DM, KDD Process, OLAP/OLTP, Dimensional Modeling, Cube , Schema, defining Schema's Star Schema, Snow-flakes Schema and Fact Constellation, ETL Process	6
2	**Data Pre-processing** Data summarization, , data cleaning, data integration and transformation, , data reduction, data discretization and concept hierarchy generation, feature extraction, feature transformation, feature selection, introduction to Dimensionality Reduction	8

DR. KRUNAL NARUMAL VAGHELA
Digitally signed by (Name of HOD)

DR. RAJENDRASINH BAHADURSINH JADEJA
Digitally signed by (Name of Dean/ Principal)

Contents : Unit	Topics	Contact Hours
3	**Mining Frequent Patterns, Associations and Correlations** Basic concept, Market Basket Analysis, frequent item-set mining methods, mining various kind of association rules, from association mining to correlation analysis, Advanced Association Rule Techniques, Measuring the Quality of Rules.	7
4	**Classification, Prediction and Clustering** Classification: Introduction, Model Construction and Usage, Rule Based Algorithm, Tree based Algorithm, Naïve Bayes Algorithm, Prediction: Linear Regression, Non-Linear Regression, Neural Network, accuracy and error measures, evaluation of the accuracy of a classifier, Clustering: Introduction to Clustering, Types of Data in Cluster Analysis, Overview of Major Clustering Methods, K-Means Clustering, Outlier Analysis	8
5	**Introduction to Information Retrieval:** Concept end architecture of Information Retrieval, Boolean Retrieval, the term vocabulary and posting list, text processing – tokenization, Stop words Removal, Stemming, Posting Lists Intersection via Skip Pointers, Positional List and Phrase Queries	6
Total Hours		**35**

Suggested List of Experiments:

Contents : Unit	Topics	Contact Hours
1	**Practical 1** Explore data mining tools.	2
2	**Practical 2** Explore Weka modules.	2
3	**Practical 3** Preparing and understanding the dataset.	2
4	**Practical 4** Applying Preprocessing techniques on dataset.	2
5	**Practical 5** Applying Preprocessing techniques on dataset.	2
6	**Practical 6** Applying APRIORI algorithm on dataset.	2
7	**Practical 7** Apply APRIORI algorithm on dataset.	2
8	**Practical 8** Applying classification method on dataset	2
9	**Practical 9** Applying classification method on dataset	2
10	**Practical 10** Applying classification method on dataset.	2
11	**Practical 11** Applying clustering method on dataset.	2

DR. KRUNAL NARUMAL VAGHELA
Digitally signed by (Name of HOD)

DR. RAJENDRASINH BAHADURSINH JADEJA
Digitally signed by (Name of Dean/ Principal)

Suggested List of Experiments:

Contents : Unit	Topics	Contact Hours
12	**Practical 12** Applying clustering method on dataset.	2
13	**Practical 13** Demonstration of IR technique.	2
14	**Practical 14** Case study on Data Mining and IR techniques	2
Total Hours		**28**

Textbook :

1 Data mining concepts and techniques, Jiawei Han, Jian Pei, Hanghang Tong, Elsevier Science, 2022

References:

1 Introduction to information retrieval, Introduction to information retrieval, Christopher D. Manning, Prabhakar Raghavan, Hinrich Schütze, Cambridge University Press, 2008

2 Data Mining : Introductory and Advance Topics, Data Mining : Introductory and Advance Topics, Margaret H Dunham, Pearson Education., 2006

3 Information retrieval systems: Characteristics, testing and evaluation., Information retrieval systems: Characteristics, testing and evaluation., Frederick Wilfrid Lancaster, Wiley, 1979

Suggested Theory Distribution:

The suggested theory distribution as per Bloom's taxonomy is as follows. This distribution serves as guidelines for teachers and students to achieve effective teaching-learning process

Distribution of Theory for course delivery					
Remember / Knowledge	**Understand**	**Apply**	**Analyze**	**Evaluate**	**Higher order Thinking / Creative**
10.00	35.00	35.00	20.00	0.00	0.00

Instructional Method:

1 The course delivery method will depend upon the requirement of content and need of students. The teacher in addition to conventional teaching method by black board, may also use any of tools such as demonstration, role play, Quiz, brainstorming, MOOCs etc.

2 The internal evaluation will be done on the basis of continuous evaluation of students in the laboratory and class-room.

3 Practical examination will be conducted at the end of semester for evaluation of performance of students in laboratory.

4 Students will use supplementary resources such as online videos, NPTEL videos, e courses, Virtual Laboratory as suggested by subject faculty.

DR. KRUNAL NARUMAL VAGHELA
Digitally signed by (Name of HOD)

DR. RAJENDRASINH BAHADURSINH JADEJA
Digitally signed by (Name of Dean/ Principal)

Supplementary Resources:

1 https://www.cs.waikato.ac.nz/ml/weka/
2 https://rapidminer.com/
3 https://orangedatamining.com/
4 https://www.sas.com/en_us/insights/analytics/data-mining.html
5 https://www.webcrawler.com/

DR. KRUNAL NARUMAL VAGHELA
Digitally signed by (Name of HOD)

DR. RAJENDRASINH BAHADURSINH JADEJA
Digitally signed by (Name of Dean/ Principal)