Department of Computer Science and Engineering
National Institute of Technology Tiruchirappalli

1.Course Outline			
Course Title	Data Mining Techniques		
Course Code	CA721		
Department	CA	**No. of Credits**	3
Pre-requisites Course Code	Data Base Management Systems	**Faculty Name**	Dr. S. Nickolas
PAC Chairman	Dr. S.Domnic		
E-mail	nickolas@nitt.edu	**Telephone No.**	+91-431-2503739
Course Type	Core Course		

Syllabus

Data mining – Motivation – Importance - DM Vs KDD - DM Architecture - Data Types – DM Tasks – DM System Classification - Primitives of DM - Data Mining Query Language - DM Metrics - DM Applications - DM Issues – Social Implications of DM

Data Preprocessing: Summarization - Data cleaning - Data Integration and Transformation - Data Reduction - Discretization and Concept Hierarchy Generation

Mining Frequent Patterns – Frequent Item set Mining Methods. Classification: Classification by Decision Tree Induction – Bayesian Classification – Rule based Classification - Prediction– Accuracy and Error Measures

Cluster Analysis – Types of Data in Cluster Analysis – Categorization of clustering Methods – Partition Methods - Outlier Analysis – Mining Data Streams – Social Network Analysis – Mining the World Wide Web

Data Warehousing: OLTP Vs OLAP - Multidimensional Data Model -DW Architecture Efficient Processing of OLAP queries - Metadata repository – DWH Implementation – OLAM

REFERENCES:

1. JiaweiHan, Micheline amber, "Data Mining: Concepts and Techniques", 3rd Edition, Elsevier India Private Limited, 2012.
2. Margaret H. Dunham, "Data Mining: Introductory and Advanced Topics", Pearson Education, 2012.

3. K.P.Soman, ShyamDiwakar, V.Ajay, "Insight into Data Mining Theory & Practice, Prentice Hall India, 2012

5. G.H.Gupta, "Introduction to Data Mining with Case Studies", 2nd Edition, PHI. 6. Ralph Kimball, Margy Ross "The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling", 3rd Edition ,wiley , Jul 2013

2. Course Overview

Data Mining is the study on process involved in the extraction of useful patterns/knowledge available in the historical databases. It involves the design and implementation of algorithms to find patterns and regularities in databases, perform prediction and forecasting, and generally improve their performance through interaction with data. It is one of the core activities in the more general process called Knowledge Discovery from Data Bases, that deals with extracting useful knowledge from raw data. The knowledge discovery process includes data cleaning, selection, cleaning, transformation, mining using different statistical and machine learning techniques, and visualization of the knowledge extracted after evaluation. The course will cover all these concepts and will illustrate the whole process by examples. Special emphasis will be given to the Machine Learning algorithms as they provide the functionalities for knowledge discovery. Data Warehousing which is built upon Multidimensional Data Modelling for On-line Analytical Processing (OLAP) will also be discussed as part of advanced technology for Data Mining.

3. Course Objectives

- Ability to understand Data Mining techniques and usage of data mining and data warehousing tools for analysis of data.

4. Course Outcomes (CO)

Student will be able to:

- Describe the the basic concepts and techniques of Data Mining.
- Solve the practical problems using recent data mining softwares.
- Doing independent study and research using the experience gained during the course.

5. Course Outcome (CO)	Aligned Programme Outcome (PO)											
	PO-1	PO-2	PO-3	PO-4	PO-5	PO-6	PO-7	PO-8	PO-9	PO-10	PO-11	PO-12
Describe the the basic concepts and techniques of Data Mining.		L	H					M				
Solve the practical problems using recent data mining softwares.	L			M		H						
Doing independent study and research using the experience gained during the course.		H			L				M			

L-Low M-Medium H-High

6. Course Teaching and Learning Activities			
Week	**No. of Classes**	**Topic Covered**	**Mode of Delivery**
1	Class-I	Data mining – Motivation – Importance	Chalk and Talk
	Class-II	DM Vs KDD - DM Architecture	Chalk and Talk , Power Point Presentation
	Class-III	Data Types – DM Functionalities	Chalk and Talk , Power Point Presentation
2	Class-I	DM System Classification - Primitives of DM - Data Mining Query Language	Chalk and Talk , Power Point Presentation
	Class-II	Interestingness of Patterns - DM Applications	Chalk and Talk , Power Point Presentation
	Class-III	DM Issues – Social Implications of DM	Chalk and Talk , Power Point Presentation
3	Class-I	Data Objects and Attribute Types- Statistical Descriptions of Data – Similarity and Dissimilariy Measures	Chalk and Talk , Power Point Presentation
	Class-II	Graphic Displays and Visualization Techniques	Power Point Presentation
	Class-III	Data Preprocessing: An Overview	Chalk and Talk ,
4	Class-I	Data cleaning and Integration	Chalk and Talk , Power Point Presentation
	Class-II	Data Integration and Transformation	Chalk and Talk , Power Point Presentation
	Class-III	Data Reduction and Transformation	Chalk and Talk , Power Point Presentation
5	Class-I	Discretization and Concept Hierarchy Generation	Chalk and Talk , Power Point Presentation
	Class-II	Mining Frequent Patterns – Frequent Itemset Mining Methods.	Chalk and Talk
	Class-III	Apriori Algorithm for mining Frequent Itemsets	Chalk and Talk , Power Point Presentation
6	Class-I	Tutorial for Finding Frequent Items	Assignment
	Class-II	FP-Tree Mining	Chalk and Talk , Power Point Presentation
	Class-III	Pattern Evaluation Methods	Chalk and Talk , Power Point Presentation
7	Class-I	Advanced Pattern Mining – Frequent Patterns	Chalk and Talk , Power Point Presentation
	Class-II	Data Warehousing: OLTP Vs OLAP	Chalk and Talk , Power Point Presentation
	Class-III	Multidimensional Data Model	Chalk and Talk , Power Point Presentation
8	Class-I	DW Architecture Efficient Processing of OLAP queries	Chalk and Talk , Power Point Presentation
	Class-II	Metadata repository	Chalk and Talk , Power Point Presentation
	Class-III	DWH Implementation - OLAM	Chalk and Talk , Power Point Presentation
9	Class-I	Classification-Basic Concepts - Classification by Decision Tree Induction	Chalk and Talk , Power Point Presentation
	Class-II	Bayesian Classification	Chalk and Talk , PPT

	Class-III	Rule based Classification – Model Evaluation and Selection	Chalk and Talk , Power Point Presentation
10	Class-I	Prediction	Chalk and Talk , Power Point Presentation
	Class-II	Accuracy and Error Measures	Chalk and Talk , Power Point Presentation
	Class-III	Cluster Analysis – Types of Data in Cluster Analysis – Categorization of clustering Methods	Chalk and Talk , Power Point Presentation
11	Class-I	Partition Methods – K-Means	Chalk and Talk , Power Point Presentation
	Class-II	Partition Methods – K-Medoids	Chalk and Talk , Power Point Presentation
	Class-III	Hierarchical Methods	Chalk and Talk , Power Point Presentation
12	Class-I	Outlier Analysis - Outlier Detection Methods	Chalk and Talk , Power Point Presentation
	Class-II	Mining Data Streams - Social Network Analysis	Chalk and Talk , Power Point Presentation
	Class-III	Mining the Web	Chalk and Talk , Power Point Presentation

7. Course Assessment Methods – Theorey

Sl. No.	Mode of Assessment	Week/Date	Duration	Weightage(%)
1.	Cycle Test –1	6th week	60 mins	20
2.	Cycle Test –2	12th week	60 mins	20
3.	Assignment	7thand 10thweek	7 days	10
4.	End Semester Exam	-	180 mins	50
			Total	100

8. Essential Readings (Textbooks, Reference books, Websites, Journals, etc.)

REFERENCES:

1. JiaweiHan ,MichelineKamber, "Data Mining: Concepts and Techniques", 3rd Edition, Elsevier India Private Limited,2012.
2. Margaret H. Dunham, "Data Mining: Introductory and Advanced Topics", Pearson Education, 2012.
3. K.P.Soman, ShyamDiwakar, V.Ajay, "Insight into Data Mining Theory & Practice, Prentice Hall India,2012,
4. G.H.Gupta, "Introduction to Data Mining with Case Studies", 2nd Edition, PHI.

9. Course Exit Survey (mention the ways by which the feedback about the course is assessed and indicate the attainment level)

1. The students through the class rep may give their feedback at any time to the course co-ordinator which will be duly addressed.
2. The students may also give their feedback during Class Committee meeting.
3. 'Course Outcome Survey' form will be distributed on the last working day to all the students and the feedback on various rubrics will be analyzed.
4. The COs will be computed after arriving at the final marks.

10. Course Policy (including plagiarism, academic honesty, attendance, etc.)

- **Plagiarism**
  The students are expected to come out with their original code for problems given in assignments during the class work, and tests/examinations. If found to copy from internet/other students, zero marks will be assigned and action will be taken.
- **Attendance**
  100% is a must. However, relaxation will be given for leave on emergency requirements (medical, death, etc.) and representing institute events. Minimum 75% is required.
- **Academic Honesty**
  i. Possession of any electronic device, if any, found during the test or exam, the student will be debarred for 3 years from appearing for the exam and this will be printed in the Grade statement/Transcript.
  ii. Tampering of MIS records, if any, found, then the results of the student will be with held and the student will not be allowed to appear for the Placement interviews conducted by the Office of Training & Placement, besides (i).

11. Additional Course Information

- The students can get their doubts clarified at any time with their faculty member with prior appointment.

For Senate's Consideration

Dr. S. Nickolas	**Dr.S. Domnic**	**Dr. S.R.Balasundaram**
Course Faculty	**Class Committee Chairperson**	**HOD**