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	CA712	**Faculty Name**	Dr. S. Nickolas
PAC Chairman	Dr. R. Eswari		
E-mail	nickolas@nitt.edu	**Telephone No.**	+91-431-2503739
Course Type	Core Course		

2. Course Overview

Data Mining studies algorithms and computational paradigms that allow computers to find patterns and regularities in databases, perform prediction and forecasting. It is currently regarded as the key element of a more general process called Knowledge Discovery that deals with extracting useful knowledge from raw data. The knowledge discovery process includes data selection, cleaning, coding, using different statistical and machine learning techniques, and visualization of the generated structures. The course will cover all these issues and will illustrate the whole process by examples. Special emphasis will be give to the Machine Learning methods as they provide the real knowledge discovery tools. Important related technologies such as data warehousing and on-line analytical processing (OLAP) will be also discussed.

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	PPT, Chalk and Talk
	Class-II	DM Vs KDD - DM Architecture	PPT, Chalk and Talk
	Class-III	Data Types – DM Tasks –DM System Classification - Primitives of DM	PPT, Chalk and Talk
2	Class-I	Data Mining Query Language	PPT, Chalk and Talk
	Class-II	DM Metrics - DM Applications	PPT, Chalk and Talk
	Class-III	DM Issues – Social Implications of DM	PPT, Chalk and Talk
3	Class-I	Know Your Data: Data Objects and Attributes	PPT, Chalk and Talk
	Class-II	Central Tendency and Dispersion Measrues	PPT, Chalk and Talk
	Class-III	Similarity and Disimilarity Metrics	PPT, Chalk and Talk
4	Class-I	Finding Similarity in Documents	PPT, Chalk and Talk
	Class-II	Data Preprocessing-Data Quality	PPT, Chalk and Talk
	Class-III	Data Cleaning	PPT, Chalk and Talk
5	Class-I	Finding Missing Values	PPT, Chalk and Talk
	Class-II	Data Integration : Chi-square Test	PPT, Chalk and Talk
	Class-III	Data Integration : Correlation Analysis	PPT, Chalk and Talk
6	Class-I	Data Reduction	PPT, Chalk and Talk
	Class-II	Discretization and Concept Hierarchy Generation	PPT, Chalk and Talk
	Class-III	Discretization and Concept Hierarchy Generation	PPT, Chalk and Talk
7	Class-I	Mining Frequent Patterns – Frequent Itemset Mining Methods.	PPT, Chalk and Talk
	Class-II	Apriori Algorithm for mining Frequent Itemsets	PPT, Chalk and Talk
	Class-III	Tutorial for Finding Frequent Items	
8	Class-I	FP-Tree Mining	PPT, Chalk and Talk
	Class-II	FP-Tree Mining	PPT, Chalk and Talk

	Class-III	Classification: Classification by Decision Tree Induction	PPT, Chalk and Talk
9	Class-I	Attribute Selection by Information Gain	PPT, Chalk and Talk
	Class-II	Attribute Selection by Gini Index	PPT, Chalk and Talk
	Class-III	Bayesian Classification	PPT, Chalk and Talk
10	Class-I	Rule based Classification	PPT, Chalk and Talk
	Class-II	Prediction	PPT, Chalk and Talk
	Class-III	Accuracy and Error Measures	PPT, Chalk and Talk
11	Class-I	Cluster Analysis – Types of Data in Cluster Analysis – Categorization of clustering Methods	PPT, Chalk and Talk
	Class-II	Partition Methods	PPT, Chalk and Talk
	Class-III	Outlier Analysis	PPT, Chalk and Talk
12	Class-I	Mining Data Streams	PPT, Chalk and Talk
	Class-II	Social Network Analysis	PPT, Chalk and Talk
	Class-III	Mining the World Wide Web	PPT, Chalk and Talk
13	Class-I	Data Warehousing: OLTP Vs OLAP	PPT, Chalk and Talk
	Class-II	Multidimensional Data Model	PPT, Chalk and Talk
	Class-III	DW Architecture Efficient Processing of OLAP queries	PPT, Chalk and Talk
14	Class-I	Metadata repository	PPT, Chalk and Talk
	Class-II	DWH Implementation - OLAM	PPT, Chalk and Talk
	Class-III	OLAP To OLAM	PPT, Chalk and Talk

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	10th week	60 mins	20
3.	Assignment/Seminar	7thand 10thweek	7 days	10
CPA	Compensation Assesment	11	60 mins	20
4.	End Semester Exam	-	120 mins	50
			Total	100

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

REFERENCES:

1. JiaweiHan ,MichelineKamber, "Data Mining: Concepts and Techniques", 2nd Edition, Elsevier India Private Limited,2008.
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 as assignments during the class work, and tests/examinations. If found to be copied 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.R. Eswari	Dr. PJA Alphonse
Course Faculty	PAC Chairperson	HOD