FR. Conceicao Rodrigues College Of Engineering

Father Agnel Ashram, Bandstand, Bandra-west, Mumbai-50

Department of Computer Engineering

T.E. (IT) (semester VI) (2018-2019)

Subject: Data Mining & Business Intelligence (DMBI-ITC602)

Credits-4

Syllabus

Course Code	Course Name	Theory	Practical	Tutorial	Theory	Oral & Practical	Tutorial	Total
ITC602	Data Mining and Business Intelligence	04		--	04	--	--	04

Course Code	Course Name	Examination Scheme						
		Theory Marks				Term Work	Oral & Practical	Total
		Internal assessment			End Sem. Exam			
		Test1	Test2	Avg. of two Tests				
ITC602	Data Mining and Business Intelligence	20	20	20	80	--	--	100

Sr. No.	Module	Detailed Content	Hours	CO Mapping
0	Prerequisites	Knowledge of databases, and Date warehousing, OLAP	02	--
I	Introduction to Data Mining	What is Data Mining; Kind of patterns to be mined; Technologies used; Major issues in Data Mining	03	CO1
II	Data Exploration and Data Preprocessing	Types of Attributes; Statistical Description of Data; Data Visualization; Measuring similarity and dissimilarity. Why Preprocessing? Data Cleaning; Data Integration; Data Reduction: Attribute subset selection, Histograms, Clustering and Sampling; Data Transformation & Data Discretization: Normalization, Binning, Histogram Analysis and Concept hierarchy generation.	09	CO2 CO3
III	Classification	Basic Concepts; Classification methods: 1. Decision Tree Induction: Attribute Selection Measures, Tree pruning. 2. Bayesian Classification: Naïve Bayes‟ Classifier. Prediction: Structure of regression models; Simple linear regression, Multiple linear regression. Accuracy and Error measures, Precision, Recall, Holdout, Random Sampling, Cross Validation.	09	CO4 CO5
IV	Clustering	Cluster Analysis: Basic Concepts; Partitioning Methods: K-Means, K-Mediods; Hierarchical Methods: Agglomerative, Divisive, BIRCH; Density-Based Methods: DBSCAN What are outliers? Types, Challenges; Outlier Detection Methods: Supervised, Semi Supervised, Unsupervised, Proximity based, Clustering Based.	10	CO4 CO5

V	Frequent Pattern Mining	Market Basket analysis, Frequent itemsets, closed itemsets and Association Rules; Frequent Pattern Mining, Efficient and Scalable Frequent Itemset Mining Methods, The Apriori Algorithm for finding Frequent Itemsets Using Candidate Generation, Generating Association Rules from Frequent Itemsets, Improving the Efficiency of Apriori, A pattern growth approach for mining Frequent Itemsets; Mining Frequent itemsets using vertical data formats; Introduction to Mining Multilevel Association Rules and Multidimensional Association Rules; From Association Mining to Correlation Analysis, lift, ; Introduction to Constraint-Based Association Mining.	10	CO4 CO5
VI	Business Intelligence	What is BI? Business intelligence architectures; Definition of decision support system; Development of a business intelligence system using Data Mining for business Applications like Fraud Detection, Clickstream Mining, Market Segmentation, retail industry, telecommunications industry, banking & finance CRM etc.	09	CO6

Text Books:

Han, Kamber, "Data Mining Concepts and Techniques", Morgan Kaufmann 3nd Edition.

P. N. Tan, M. Steinbach, Vipin Kumar, “Introduction to Data Mining”, Pearson Education.

Business Intelligence: Data Mining and Optimization for Decision Making by Carlo Vercellis ,Wiley India Publications.

G. Shmueli, N.R. Patel, P.C. Bruce, “Data Mining for Business Intelligence: Concepts, Techniques, and Applications in Microsoft Office Excel with XLMiner”, 2nd Edition, Wiley India.

Course Outcomes

CO1: Demonstrate an understanding of the importance of data mining and the principles of business intelligence.

CO2: Organize and Prepare the data needed for data mining using pre preprocessing techniques.

CO3: Perform exploratory analysis of the data to be used for mining.

CO4: Implement the appropriate data mining methods like classification, clustering or Frequent Pattern mining on large data sets.

CO5: Define and apply metrics to measure the performance of various data mining algorithms.

CO6: Apply BI to solve practical problems : Analyze the problem domain, use the data collected in enterprise apply the appropriate data mining technique, interpret and visualize the results and provide decision support.

CO-PO and CO-PSO Mapping

Course Name	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7	PO 8	PO 9	PO10	PO 11	PO 12	PSO 1	PSO 2
CO1	1													1
CO2	2	1												
CO3		3	2										3	2
CO4		2	3		2								2	
CO5	2	1	1										2	1
CO6	2	3	3		1								2	1

CO Assessment Tools

	Direct Methods							Indirect Methods
	Test 1	**Assig1**	**Lab Work/ Tutorial**	**Test2**	**Assig2**	**University Theory Exam**	**University Practical Exam**	**Course Exit Survey**
CO1	25%	20%	10%			20%	25%	100%
CO2	25%	20%	10%			20%	25%	100%
CO3	25%	20%	10%			20%	25%	100%
CO4			10%	25%	20%	20%	25%	100%
CO5			10%	25%	20%	20%	25%	100%
CO6			10%	25%	20%	20%	25%	100%

Lecture Plan:

No of classes available:	**40**	**No of Classes taken:**	**41**	
Sr. No.	**Topic Planned**	**Planned Date**	**Actual Date**	**Delivery Mechanisms**
1	What is Data Mining; Kind of patterns to be mined	9/1	9/1	Chalk and board
2	Technologies used, Major issues in data mining	10/1	9/1	Chalk and board
3	Types of attributes, statistical distribution	11/1	10/1	Chalk and board
4	Data visualization, statistical distribution	15/1	11/1	Chalk and board
5	Measuring dissimilarity	16/1	14/1,15/1	Chalk and board
6	Measuring similarity, dissimilarity	17/1	16/1	Chalk and board
7	Preprocessing, Data cleaning, data integration	18/1	17/1,22/1	Chalk and board
8	Data reduction, attribute subset selection, sampling	22/1	23/1	Chalk and board
9	histogram, clustering, data transformation	23/1	24/1	Chalk and board
10	Discretization, normalization, binning, concept hierarchy generation	24/1	24/1	Chalk and board
11	Classification, Basic concepts, Decision tree induction	25/1	25/1	Chalk and board
12	Attribute selection measures, tree pruning	29/1	29/1	Chalk and board
13	Bayesian classification, Naïve bayesian	30/1	30/1	Chalk and board

14	Prediction: structure of regression model, Simple linear regression	1/2	7/2	Chalk and board
15	multiple regression ,Accuracy , error measures	12/2	8/2	Chalk and board
16	holdout, random sampling, cross validation	20/2	8/2	Chalk and board
17	Bootstrap, Roc curves, Bagging	21/2	20/2	Chalk and board
18	Boosting, Random forests	22/2	20/2	Chalk and board
19	Cluster analysis: basic	26/2	21/2	Chalk and board
20	Partiitoning: kmeans	27/2	21/2	Chalk and board
21	kmediods	28/2	22/2	Chalk and board
22	Hierarchical: agglomerative	1/3	27/2, 28/2	Chalk and board
23	Divisive, BIRCH	5/3	28/2	Chalk and board
24	Density based: DBSCAN	6/3	1/3, 5/3	Chalk and board
25	Optics	7/3	6/3	Chalk and board
26	What are outliers, types, challenges	8/3	6/3	Chalk and board
27	Outlier detection methods: supervised, semisupervised,Proximity based, clustering	12/3	7/3	Chalk and board
28	Market basket analysis, frequent itemsets, association rules	13/3	7/3	Chalk and board
29	Frequent pattern mining methods, Apriori using candidate generation	14/3	12/3	Chalk and board
30	Generating association rules from frequent items, efficicency, using vertical data formats	19/3	13/3	Chalk and board
31	Closed mining, maximal patterns	20/3	14/3	Chalk and board
32	Multilevel association rules, multidimensional	22/3	18/3	Chalk and board,
33	Correlation analysis	26/3	20/3	Chalk and board,
34	Pattern evaluation measures	27/3	20/3	Chalk and board,
35	Constraint based association mining	27/3	22/3	Chalk and board
36	What is BI, Data information , knowledge	28/3	26/3	Chalk and board

37	Mathematical models, BI architectures	28/3	26/3	Chalk and board,
38	Enabling factors in BI project, Development of BI system	29/3	27/3	Chalk and board,
39	Representation of decision making process, evolution of information systems	29/3	27/3	Chalk and board,
40	Decision support systems	02/4	28/3	Chalk and board,
41	Development of decision support systems	02/4	28/3	Chalk and board
42	Data mining for business application like fraud detection	03/4	01/4	Chalk and board,
43	Click stream mining	03/4	01/4	Chalk and board
44	Market segmentation, CRM	04/4	02/4	Chalk and board
45	Retail industry, banking	05/4	02/4	Chalk and board