FR. Conceicao Rodrigues College Of Engineering

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

Department of Information Technology

T.E. (IT) (semester V) (2020-2021)

Lesson Plan

Subject: Data Mining and Business Intelligence (ITC602)

Credits-4

Detailed syllabus:

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? DataCleaning; 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 TreeInduction: Attribute Selection Measures, Tree pruning. 2. Bayesian Classification: NaïveBayes‟ 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	Market Basket Analysis, Frequent Itemsets, Closed Itemsets, and	10	CO4

	Mining	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 Associatio nMining.		CO5
VI	Business Intelligenc e	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:

1. Han, Kamber, "Data Mining Concepts and Techniques", Morgan Kaufmann 3nd Edition.
2. P. N. Tan, M. Steinbach, Vipin Kumar, “Introduction to Data Mining”, Pearson Education.
3. Business Intelligence: Data Mining and Optimization for Decision Making by Carlo Vercellis, Wiley India Publications.
4. 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.

References:

1. Michael Berry and Gordon Linoff “Data Mining Techniques”, 2nd Edition Wiley Publications.
2. Michael Berry and Gordon Linoff “Mastering Data Mining- Art & science of

CRM", WileyStudent Edition.

3. Vikram Pudi & Radha Krishna, "Data Mining", Oxford Higher Education.

Assessment:

Internal Assessment for 20 marks:

Consisting of Two Compulsory Class Tests

Approximately 40% to 50% of syllabus content must be covered in First test and remaining 40% to 50% of syllabus contents must be covered in second test.

CO-Statements:

Sr.No.	Course Outcome Statement
ITC602.1	Demonstrate an understanding of the importance of data mining and the principles of business intelligence.
ITC602.2	Organize and Prepare the data needed for data mining using pre preprocessing techniques.
ITC602.3	Perform exploratory analysis of the data to be used for mining
ITC602.4	Implement the appropriate data mining methods like classification, clustering or Frequent Pattern mining on large data sets.
ITC602.5	Define and apply metrics to measure the performance of various data mining algorithms.
ITC602.6	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-PSO Mapping

Course Name	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2
ITC602.1	1													1
ITC602.2	2	1												
ITC602.3		3	2		2								3	2
ITC602.4		2	3										2	
ITC602.5	2	1	1										2	1
ITC602.6	2	3	3										2	1

CO Assessment Tools

	Test1	Assig1	Lab Work	Tutorial 1	Tutorial 2	Test2	Assig2	University Theory Exam	University Oral Exam	Course Exit Survey
CO1	30%	30%						20%	20%	100%
CO2	30%	20%			10%			20%	20%	100%
CO3	30%	20%		10%				20%	20%	100%
CO4			30%			30%		20%	20%	100%
CO5			10%			30%		20%	20%	100%
CO6			20%			20%	20%	20%	20%	100%

Curriculum Gap/

Content beyond syllabus

Lecture Plan:

No of classes available:	23	No of Classes taken:	23	
Sr. No.	**Topic Planned**	**Planned Date**	**Actual Date**	**Delivery Mechanisms**
1	Introduction- Association Rule Mining	28/01	28/01/2021	Online mode using PPT
2	Association Rules; Frequent Pattern Mining, Efficient and Scalable Frequent Itemset Mining Methods, examples on support, appriori principle	29/01	29/01/2021	Online mode using PPT
3	The Apriori Algorithm for finding Frequent Itemsets Using Candidate Generation, problems on Apriori	04/02	03/02/2021	Online mode using PPT
4	More problems on Apriori Algorithm... Generating Association Rules from Frequent Itemsets	05/02	04/02/2021	Online mode using PPT
5	Improving the Efficiency of Apriori, Class Test on problems on Apriori Algorithm	11/02	11/02/0201	Online mode using PPT

6	A pattern growth approach for mining Frequent Itemsets: FP Growth Algorithm	12/02	12/02/2021	Online mode using PPT
7.	A pattern growth approach for mining Frequent Itemsets: FP Growth Algorithm	18/02	18/02/2021	Online mode using PPT
8.	Mining Frequent itemsets using vertical data formats	25/02	02/03/2021	Online mode using PPT
9	Introduction to Mining Multilevel Association Rules and Multidimensional Association Rules	26/02	05/03/2021	Online mode using PPT
10	FromAssociation Mining to Correlation Analysis, lift	04/03	12/03/2021	Online mode using PPT
11	Pattern evaluation measures	05/03	16/03/2021	Online mode using PPT
12	Cluster Basics, Cluster analysis- Partitioning Methods: K-means	12/03	30/03/2021	Online mode using PPT
13	problem on K-means, K-mediods algorithm	18/03	09/04/2021	Online mode using PPT
14	Hierarchical Methods: Agglomerative, Divisive, BIRCH	19/03	15/04/2021	Online mode using PPT
15	Density based clustering algorithm: DBSCAN	25/03	16/04/2021	Online mode using PPT
16	University Question paper Problems on clustering: Single -linkage clustering and K-Means clustering	26/03	29/04/2021	Discussion on online platform
17	What is BI? and Architecture of BI	01/04	30/04/2021	Online mode using PPT
18	BI Presentations	08/04	5/5/2021	Online mode using PPT
19	BI Presentations	09/04	6/5/2021	Online mode using PPT
20	BI Presentations	15/04	7/5/2021	Online mode using PPT

21	BI Presentations	16/04	10/5/2021	Online mode using PPT
22	BI Presentations	22/04	11/5/2021	Online mode using PPT
23	BI Presentations	23/04	12/5/2021	Online mode using PPT

Lab Plan: Business Intelligence Lab (ITL602)

Lab Outcomes: Students will be able to:

1. Identify sources of Data for mining and perform data exploration
2. Organize and prepare the data needed for data mining algorithms in terms of attributes and class inputs, training, validating, and testing files.
3. Implement the appropriate data mining methods like classification, clustering or association mining on large data sets using open source tools like WEKA
4. Implement various data mining algorithms from scratch using languages like Python/ Java etc.
5. Evaluate and compare performance of some available BI packages
6. 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

Lab Plan: Business Intelligence Lab

Sr. No.	Experiment Title	Week No.	Lab Outcome
1	Tutorial 1 on Data exploration	Week 1	LO 1 and LO 2
2	Tutorial 2 on Data pre-processing	Week 1	LO 1 and LO 2
3	Implementation of any one classifier using python	Week 2	LO4
4	Implementation of classifiers -Decision tree, Naive Bayes and zero R using WEKA	Week 2	LO3
5	Implementation of any one clustering algorithm using Java	Week 3	LO4
6	Implementation of clustering- K means, DBSCAN and OPTICS using WEKA	Week 3	LO3
7	Implementation of Apriori algorithm using Java	Week 4	LO4
8	Use WEKA to implement association mining	Week 4	LO3
9	Implementation of data mining operations using Rapid miner (Project)	Week 5	LO6

Assignment Plan:

Assig No.	Date	Questions	CO/LO
1	**14/05/2021**	**MCQ Questions**	**CO1, CO2, CO3, CO4, CO5**
2	**06/05/2021**	**Presentation on BI Case Study**	**CO6**

Term Work:

Term Work shall consist of at least 10 to 12 practical's based on the below list. Also Term work Journal must include at least 2 assignments. Term Work Marks: 25 Marks (Total marks) = 15 Marks (Experiment) + 5 Marks (Assignments) + 5 Marks (Attendance)

Oral Exam: An Oral exam will be held based on the above syllabus