GANPAT UNIVERSITY									
FACULTY OF TECHNOLOGY									
Programme	Bachelor of Technology					Branch/Spec.	Computer Science & Engineering (BDA)		
Semester	VII					Version	1.0.1.0		
Effective from Academic Year			2018 – 19			Effective for the batch Admitted in			June 2016
Subject code	2CSE70E14		Subject Name			PREDICTIVE MODELLING			
Teaching scheme						Examination scheme (Marks)			
(Per week)	Lecture(DT)		Practical(Lab.)		Total		CE	SEE	Total
	L	TU	P	TW					
Credit	3	0	2	0	5	Theory	40	60	100
Hours	3	0	4	0	7	Practical	40	60	100

Pre-requisites:

Data Structures, Mathematics

Learning Outcome:

After learning the course the students should be able to

- Define the predictive models using PMML
- Design and analyze appropriate predictive models
- Apply statistical tools for analysis

Theory syllabus

Unit	Content	Hrs
1	**Introduction To Predictive Modeling:** Core ideas in data mining - Supervised and unsupervised learning - Classification vs Prediction - Steps in data mining- SEMMA Approach - Sampling -Pre-processing - Data cleaning - Data Partitioning - Building a model - Statistical models - Statistical models for predictive analytics.	**8**
2	**Predictive Modeling Basics:** Data splitting – Balancing- Overfitting –Oversampling –Multiple Regression - Artificial neural networks (MLP) - Variable importance- Profit/loss/prior probabilities - Model specification - Model selection - Multivariate Analysis.	**8**

3	**Predictive Models:** Association Rules-Clustering Models –Decision Trees- Ruleset Models- K-Nearest Neighbors – Naive Bayes - Neural Network Model – Regression Models– Regression Trees – Classification & Regression Trees (CART) – Logistic Regression – Mulitple Linear Regression Scorecards –Support Vector Machines – Time Series Models - Comparison between models - Lift chart - Assessment of a single model.	**9**
4	**Predictive Modeling Markup Language:** Introduction to PMML – PMML Converter - PMML Structure – Data Manipulation in PMML – PMML Modeling Techniques - Multiple Model Support – Model Verification.	**7**
5	**Tools and Technologies:** Weka – RapidMiner – IBM SPSS Statistics- IBM SPSS Modeler – SAS Enterprise Miner – Apache Mahout – R Programming Language	**7**
6	**Case Studies:** Real time case study with modeling and analysis.	**6**

Practical content

- Practical will be based on Clustering Models
- Practical will be based on Decision Trees
- Practical will be based on K-Nearest Neighbors
- Practical will be based on Naive Bayes
- Practical will be based on Neural Network Model
- Practical will be based on Classification & Regression Trees (CART)
- Practical will be based on Logistic Regression
- Practical will be based on Support Vector Machines
- Study on different algorithm on Weka
- Study on different algorithm RapidMiner
- Study on different algorithm IBM SPSS Statistics- IBM SPSS Modeler
- Study on different algorithm SAS Enterprise Miner
- Study on different algorithm Apache Mahout
- Study on different algorithm R Programming Language

Text Books

1	Kattamuri S. Sarma, “Predictive Modeling with SAS Enterprise Miner: Practical Solutions for Business Applications”, 2nd Edition, SAS Publishing, 2007.

Reference Books

1	Alex Guazzelli, Wen-Ching Lin, Tridivesh Jena, James Taylor, “PMML in Action Unleashing the Power of Open Standards for Data Mining and Predictive Analytics”, 2nd Edition, Create Space Independent Publishing Platform,2012.
2	Ian H. Witten, EibeFrank , “Data Mining: Practical Machine Learning Tools and Techniques”, Morgan Kaufmann Series in Data Management Systems, Morgan Kaufmann, 3rd Edition, 2011.

3	Eric Siegel , “Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die”, 1st Edition, Wiley, 2013.
4	Conrad Carlberg, “Predictive Analytics: Microsoft Excel”, 1st Edition, Que Publishing, 2012.
5	Jeremy Howard, Margit Zwemer, Mike Loukides, “Designing Great Data Products- Inside the Drivetrain Approach, a Four-Step Process for Building Data Products – Ebook”, 1st Edition, O'Reilly Media, March 2012.