



GUJARAT TECHNOLOGICAL UNIVERSITY

BACHELOR OF ENGINEERING SYLLABUS

Subject Code : 3164205

Subject Name : Predictive Analytics

WEF Academic Year :	2020-21
Semester :	6
Category of the Course :	Professional Elective

Prerequisite :

- Linear Algebra, Probability, programming language.

Rationale :

In the age of data, comprehending and analyzing data is simple. Today's industries are data-driven. Understanding different types of data, modelling them, and assessing those models are all made possible by this course. Objective of this course is to learn, how to develop models to predict categorical and continuous outcomes and be aware of how to choose models automatically using the binary and multiclass classifier and numerical predictor models.

Course Scheme :

Teaching Scheme			Total Credits	Assessment Pattern and Marks				Total Marks
L	T	P	C	Theory		Tutorial/ Practical		
				University exams (ESE)	Progressive Assessment (PA)	External Practical /viva Exam(ESE)	Internal evaluation Practical /viva Exam(PA)	
3	0	2	4	70	30	30	20	150

Course Content :

Unit No.	Content	No. of Hours	Weightage (%)
1	Regression : Overview of supervised and unsupervised machine learning, Predictive model, Descriptive model, Correlation, Simple linear regression and least squares, Multiple linear regression, Regression model evaluation, Ridge regression, Lasso regression	08	22 %



GUJARAT TECHNOLOGICAL UNIVERSITY

BACHELOR OF ENGINEERING SYLLABUS

Subject Code : 3164205

Subject Name : Predictive Analytics

2	Classification : Logistic regression, Naïve Bayes classifier, SVM for classification, SVM for regression, K-nearest neighbor classifiers	08	22 %
3	Evaluating ML models : Bias, Variance and model complexity, Bias-variance tradeoff, Cross validation: Holdout method, K-fold cross validation, Bootstrapping	06	17 %
4	Additive Models, Trees, and Boosting : Generalized additive models, Regression and classification trees , Boosting methods-exponential loss and AdaBoost, Numerical Optimization via gradient boosting	08	22 %
5	Unsupervised Learning and Random forests : Association rules, Cluster analysis, Principal Components, Random forests and analysis.	06	17 %
Total Hours :		36	100 %

Textbook :

1. An Introduction to Statistical Learning with Applications in R, Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani, (Springer, 2015)
2. The Elements of Statistical Learning-Data Mining, Inference, and Prediction, Trevor Hastie, Robert Tibshirani, Jerome Friedman, Second Edition, Springer Verlag, 2009.
3. Machine Learning, Saikat Dull, S. Chjandramouli, Das, Pearson
4. C. M. Bishop – Pattern Recognition and Machine Learning, Springer, 2006

Reference Books :

1. Michael Bowles, Machine Learning in Python, Wiley
2. Python Machine Learning By Example, Yuxi Liu, Packt
3. Machine Learning: A Probabilistic Perspective, Kevin Murphy, MIT Press



GUJARAT TECHNOLOGICAL UNIVERSITY

BACHELOR OF ENGINEERING SYLLABUS

Subject Code : 3164205

Subject Name : Predictive Analytics

Course Outcomes :

No.	Course Outcomes	RBT Level*
1	Be able to apply regression to predict continuous variable.	RM, UN, CR
2	Be able to apply classification methods to predict categorical variable.	UN, CR
3	Understand model complexity and overfitting of model.	EL, AP
4	Understand the metric used for evaluating models.	EL
5	Understand the use of ML models for unsupervised learning.	CR

*RM: Remember, UN: Understand, AP: Apply, AN: Analyze, EL: Evaluate, CR: Create

Suggested Course Practical List :

Below is the suggested practical list. Extra practical can be added as per the requirements.

1	Implement simple linear regression model.
2	Implement multiple linear regression model.
3	WAP to test performance of simple linear regression model.
4	WAP to demonstrate use of Logistic regression on IRIS dataset.
5	WAP to demonstrate use of Naive Bayes on IRIS dataset.
6	WAP to demonstrate use of SVM on IRIS dataset.
7	WAP to demonstrate use of KNN on IRIS dataset.
8	WAP to demonstrate various model evaluation parameters.
9	WAP to demonstrate various validation methods.
10	Implement various ensemble models on IRIS dataset and evaluate its performance.

List of Laboratory/Learning Resources Required :

1. Software: Python (<https://www.python.org/>)
2. Software: Anaconda Distribution (<https://www.anaconda.com/>)
3. <https://www.ibm.com/in-en/topics/predictive-analytics>
4. <https://www.investopedia.com/terms/p/predictive-analytics.asp>

Other Resources/MOOCs :

https://onlinecourses.nptel.ac.in/noc23_ma46/preview
https://onlinecourses.swayam2.ac.in/imb20_mg1
