Nirma University
Institute of Technology, School of Technology
M. Tech. Computer Science and Engineering (Data Science)

Semester – II

L	T	P	C
3	0	2	4

Course Code	3CS42D303
Course Name	Predictive Analytics

Course Learning Outcomes (CLOs):

At the end of the course, students will be able to

1. apply statistical and regression analysis methods to identify new trends and patterns, uncover relationships, create forecasts, predict likelihoods, and test predictive hypotheses
2. compare the underlying predictive modeling techniques
3. develop the modeling skills from an industry perspective
4. select appropriate predictive modeling approaches suitable to various tasks

Syllabus:

Teaching Hours

Unit I **4**

Introduction: Introducing CRISP-DM methodology, Need for it in business scenario

Unit II **8**

Modeling Techniques: Overview of Modeling Techniques, Unsupervised learning - Clustering, Supervised Learning - Classification, Linear Discriminant Analysis, Ensemble Learning, Random Forest models.

Unit III **10**

Regression: Concept of Regression, Covariance, Correlation and ANOVA Review, Simple linear regression, multiple linear regression, parameter estimation, logistic regression, Maximum Likelihood Estimation (MLE) of parameters

Unit IV **5**

Model Evaluation: Metrics for Performance Evaluation, Accuracy, confusion matrix, Precision, Recall, ROC Curves

Unit V **10**

Decision Trees and Unstructured data analysis: Introduction to Decision Trees, CHI-Square Automatic Interaction Detectors (CHAID), Classification and Regression Tree (CART), Analysis of Unstructured data, Naive Bayes Classification

Unit VI **8**

Advanced Topics: Forecasting and time series analysis, auto-regressive and moving average models, applications in stock market prediction weather prediction.

Self-Study:

The self-study contents will be declared at the commencement of semester. Around 10% of the questions will be asked from self-study contents.

Laboratory Work:

Laboratory work will be based on above syllabus with minimum 5 experiments to be incorporated.

Suggested Readings^:

1. James, Witten, Hastie and Tibshirani, An Introduction to Statistical Learning: with Applications in R, Springer
2. Dinov, Ivo D., Data Science and Predictive Analytics, Springer
3. Hastie, Trevor, et al., The elements of statistical learning, Vol. 2. No. 1. New York: Springer
4. Montgomery, Douglas C., and George C. Runger., Applied statistics and probability for engineers, John Wiley & Sons
5. Hann, J. and Kamber, M. Data Mining: Concepts and Techniques, Morgan Kaufmann.

L=Lecture, T=Tutorial, P=Practical, C=Credit

^this is not an exhaustive list