Course Name: Data Mining & Knowledge Discovery					
Course Code: CSEN3132					
Contact Hours per week:	**L**	**T**	**P**	**Total**	**Credit points**
	3	**0**	**0**	**3**	**3**

1. Course Outcomes

After completion of the course, students will be able to:
CSEN3132.1. Learn and understand basic knowledge of data mining and related models.
CSEN3132.2. Understand and describe data mining algorithms.
CSEN3132.3. Understand and apply Data mining algorithms.
CSEN3132.4. Suggest appropriate solutions to data mining problems.
CSEN3132.5. Analyze data mining algorithms and techniques.
CSEN3132.6. Perform experiments in Data mining and knowledge discovery using real-world data.

2. Detailed Syllabus

Module 1 [9L]

Introduction and Rule-based Classification: What is Data Mining? Why do we need data mining? Data Mining System - Architecture and Processes. Challenges in Data Mining.
Decision Tree: General approach for solving a classification problem, Decision Tree Induction, Overfitting Pruning.
Rule-based Classification: How a rule-based classifier works, rule-ordering schemes, how to build a rule-based classifier, direct and indirect methods for rule extraction.

Module 2 [9L]

Advanced Classification Techniques: Bayes' Classifier: Bayes' theorem, Naïve Bayes classifier.
Support Vector Machines (SVM): Maximum margin hyperplanes, Linear SVM: separable case, non-separable case, Non-linear SVM.

Module 3 [9L]

Ensemble Methods, Association Rule Mining: Ensemble Methods: Bagging, Boosting, Random Forests
Association Rule Mining: Introduction, Frequent itemset generation, (Apriori principle, candidate generation and pruning), Rule generation, Compact representation of frequent item sets, FP-growth algorithm, Sub-graph mining.

Module 4 [9L]

Cluster Analysis: Introduction: Motivations, objectives and applications of clustering. Different types of clustering.
Partitional Clustering: K-means, Bisecting K-means, PAM.
Hierarchical Clustering: Agglomerative, Divisive, MIN, MAX, dendrogram representation.
Density-based Clustering: DBSCAN. Cluster evaluation, further reading – OPTICS, DENCLUE, CHAMELEON, BIRCH, CURE, ROCK.

3. Textbooks

1. Data Mining Concepts and Techniques, 3rd, Edition, J. Han and M. Kamber, Morgan Kaufmann Publishers, July 2011.

4. Reference Books

1. Introduction to Data Mining, P. N. Tan, M. Steinbach and V. Kumar, Pearson Publishers.
2. Pattern Recognition and Machine Learning, First Edition, C. Bishop, Springer, 2006.
3. Neural Networks and Learning Machines, Third Edition, S. Haykin, PHI Learning, 2009.
4. Pattern Classification, Second Edition, R. Duda, P. Hart and D. Stock, Wiley-Interscience, 2000.