PCS 213 DATA MINING AND KNOWLEDGE MANAGEMENT

	L	T	P	Cr
	3	0	2	4.0

Course Objectives: To have an advanced level of knowledge to extract latent, potentially useful information from stored data, display it to the final user in a comprehensible manner and incorporate it into an intelligent decision-making system.

Introduction to Big Data Analytics: Big Data Overview, State of the Practice in Analytics, the Data Scientist, Big Data Analytics in Industry Verticals, Data Analytics Lifecycle.

Review of the Basic Data Analytic Methods using R: Introduction to R – look at the data, Analyzing and Exploring the Data, Statistics for Model Building and Evaluation.

Data mining: Introduction, association rules mining, Naive algorithm, Apriori algorithm, direct hashing and pruning (DHP), Dynamic Item set counting (DIC), Mining frequent pattern without candidate generation(FP, growth), performance evaluation of algorithms, Mining Customer values: From Association rule to direct mining: A case study.

Classification: Introduction, decision tree, tree induction algorithm – split algorithm based on information theory, split algorithm based on Gini index; naïve Bayes method; estimating predictive accuracy of classification method; classification software, software for association rule mining; case study; KDD Insurance Risk Assessment: A Case study.

Cluster analysis: Introduction, partitional methods, hierrarchical methods, density based methods, dealing with large databases, cluster software; Efficient Clustering of Very Large Document Collections: A case study.

Web data mining: WebTerminology and Characterstics, Locality and Hirarchy in the web, Web Content Mining, Web Usage Mining, Web Structure Mining, Web mining Software**.**

Search engines: Characterstics of Search engines, Search Engine Functionality, Search Engine Architecture, Ranking of web pages, The search engine history, Enterprise Search, Enterprise Search Engine Software.

Data warehousing: Introduction, Operational data stores, ETL, Data warehouses – design guidelines for data warehouse implementation ,Data warehouse metadata; OLAP –introduction, Characteristics, Multidimensional view and data cube, Data cube operations**,** Data Warehouse Governance: Best Practices at Blue Cross and Blue Shield of North Carolina: A Case Study**.**

Laboratory Work: The laboratory will cover the most important data mining techniques-classification, clustering, and association rule mining and exploring advanced data mining tools.

Recommended Books:

1. Han J., Kamber M. and Pei J., Data mining concepts and techniques, Morgan Kaufmann Publishers (2011) 3rd ed.
2. Pudi V., Krishana P.R., Data Mining, Oxford University press, (2009) 1st ed.
3. Adriaans P., Zantinge D., Data mining, Pearson education press (1996), 1st ed.
4. Pooniah P., Data Warehousing Fundamentals,Willey interscience Publication, (2001), 1st ed.