Course Title	Data Mining and Web Mining				
Course Code	CIS405				
Course Type	Compulsory				
Level	Bachelor (1st cycle)				
Year / Semester	4th Year / 1st Semester				
Teacher's Name	TBA				
ECTS	6	Lectures / week	3 hours /14 weeks	Laboratories / week	None
Course Purpose and Objectives	The objective of this course is to introduce basic concepts, tasks, methods, and techniques in data mining. The emphasis is on various data mining problems and their solutions with application on the web. Students will develop an understanding of the data mining process and issues, learn various techniques for data mining, and apply the techniques in solving data mining problems using data mining tools and systems. The course will also introduce major web mining techniques and develop knowledge and skills to discover useful information from data effectively. More specifically, web usage mining techniques for Web site management, user profiling, and personalization, as well as Web content and structure mining techniques, such as Web information retrieval and link analysis, aiming at supporting search engines will be explained and discussed.				
Learning Outcomes	Upon successful completion of this course students should be able to: • Explain what Data Mining is and what problems it can address • Carry out various data preprocessing actions • Explain and exemplify what classification and prediction is in the context of data mining • Classify data using the methods of decision tree induction, Bayesian methods, rule-based classification and using the method of backpropagation • Evaluate the accuracy of a classifier and predictor • Recall what clustering is and apply the various methods for clustering data • Explain how web search engines index, and rank web content. • Explain how web mining can be applied to extract useful information from Web 2.0 media such as social networking web sites, blogs, reviews' sites, etc. • Gain hands-on experience by conducting a group-based term project on designing and developing a data/web mining application, or performing an extensive analysis using data/web mining techniques.				
Prerequisites	CSE310		Co-requisites	None	

Course Content	Introduction to Data Mining, What kind of data can be mined, What kind of patterns can be mined. Data preprocessing. Descriptive Data characterization, Data cleaning, Data integration and transformation, Data Reduction Classification: Basic Concepts, Decision Trees, and Model Evaluation, General Approach to Solving a Classification Problem, Decision Tree Induction, Model Overfitting, Evaluating the Performance of a Classifier Classification: Alternative Techniques: Rule-Based Classifier, Nearest-Neighbor Classifiers, Bayesian Classifiers, Artificial Neural Network (ANN), Support Vector Machine (SVM), Ensemble Methods, Class Imbalance Problem Cluster Analysis: Basic Concepts and Algorithms: Overview, K-means, Agglomerative Hierarchical Clustering, DBSCAN, Cluster Evaluation, Additional Issues and Algorithms, Characteristics of Data, Clusters, and Clustering Algorithms, Prototype-Based Clustering, Density-Based Clustering, Graph-Based Clustering, Scalable Clustering Algorithms Association Analysis: Basic Concepts and Algorithms, Problem Definition, Frequent Itemset Generation, Rule Generation, Compact Representation of Frequent Itemsets, Alternative Methods for Generating Frequent Itemsets, FP-Growth Algorithm, Evaluation of Association Patterns Information Retrieval and Web Search: Basic Concepts of Information Retrieval, Information Retrieval Models: Boolean, Vector Space, Statistical Language Model. Evaluation Measures, Text and Web Page Pre-Processing, Inverted Index and Its Compression, Latent Semantic Indexing, Web Search, Meta-Search, Web Spamming. Social Network Analysis, Co-Citation and Bibliographic Coupling, PageRank, HITS Algorithm. Web Crawling: A Basic Crawler Algorithm, Breadth-First Crawlers, Preferential Crawlers, Universal, topical Crawlers. Structured Data Extraction: Wrapper Generation. Web Usage Mining, Data Collection and Pre-Processing, Sources and Types of Data, Data Modeling for Web Usage Mining, Discovery and Analysis of Web Usage Patterns, Recommender Systems and Collaborative Filtering
Teaching Methodology	Face – to – face
Bibliography	Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data, Bing Liu, Springer, Latest Edition. Introduction to Data Mining: International Edition, Pang-Ning Tan, Michael Steinbach, Vipin Kumar, Pearson, Latest Edition.

	Mining of Massive Datasets, Anand Rajaraman and Jeffrey Ullman, Cambridge University Press, Latest Edition. Data Mining: Concepts and Techniques, Han, J. and Kamber, M., Morgan Kaufmann, Latest Edition Machine Learning, Tom Mitchell, McGraw Hill, Latest Edition.
Assessment	Final Examination 35% Midterm Examination 25% Coursework 30% Class Participation and attendance 10% 100%
Language	English