

Course Title	Data Mining and Machine Learning				
Course Code	CSC664				
Course Type	Elective				
Level	Master (2 <sup>nd</sup> Cycle)				
Year / Semester	2 <sup>nd</sup> Year / 1 <sup>st</sup> Semester				
Teacher's Name	TBA				
ECTS	10	Lectures / week	3 Hours/ 14 weeks	Laboratories / week	N/A
Course Purpose and Objectives	The objective of this course is to introduce basic concepts, tasks, methods, and techniques in data mining and machine learning. The emphasis is on various data mining problems and their solutions with applications real-world problems. Students will develop an understanding of the data mining process and issues, learn various Machine Learning (ML) techniques for data mining, and apply the techniques in solving data mining problems using data mining tools and systems.				
Learning Outcomes	<p>Upon successful completion of the course, students will be able to:</p> <ul style="list-style-type: none"> <li>• Apply basic ML tasks and types of analysis, such as supervised learning, unsupervised learning, re-inforcement learning, classification and regression, and feature selection.</li> <li>• Discuss the inner workings of standard ML classification and feature selection algorithms.</li> <li>• Solve the problem of selecting algorithms, tuning their hyper-parameters, and estimating the performance of the final predictive model.</li> <li>• Perform and apply Data Mining pipelines to real-world problems, dealing with problems such as representing the problems as a Data Mining task, representing appropriately the data, applying and tuning a Data Mining pipeline, and interpreting results.</li> <li>• Discuss, compare, apply and synthesize with key statistical estimation and hypothesis testing concepts, with a focus on the ones that are routinely employed within ML algorithms.</li> </ul>				
Prerequisites	CSC635	Co-requisites	None		
Course Content	<p>Introduction. What is Data Mining and Machine Learning, What kind of data can be mined, What kind of patterns can be mined.</p> <p>Data preprocessing. Descriptive Data characterization, Data cleaning, Data integration and transformation, Data Reduction</p>				

	<p>Probability theory and concepts for ML, axioms of probability, conditional probability, Bayes theorem, maximum likelihood estimation, maximum a posteriori estimation</p> <p>Logistic Regression and fitting with gradient descent</p> <p>Hypothesis testing, and permutation-based hypothesis testing</p> <p>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</p> <p>Classification: Alternative Techniques: Rule-Based Classifier, Nearest-Neighbor Classifiers, Bayesian Classifiers, Artificial Neural Network (ANN), Support Vector Machine (SVM), Ensemble Methods, Class Imbalance Problem</p> <p>Cluster Analysis: Basic Concepts and Algorithms: Overview, K-means, Agglomerative Hierarchical Clustering, DBSCAN, Cluster Evaluation, Characteristics of Data, Clusters, and Clustering Algorithms, Prototype-Based Clustering, Density-Based Clustering, Scalable Clustering Algorithms</p> <p>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</p> <p>Graph Mining, Co-Citation and Bibliographic Coupling, PageRank, HITS Algorithm, Graph Clustering.</p> <p>Metrics of performance, Receiver Operating Characteristic Curves (ROC), and Area Under the ROC curve</p> <p>Estimation of performance and hyper-parameter tuning using cross validation techniques</p> <p>Basics of optimization and constrained optimization</p>
Teaching Methodology	Face- to- face
Bibliography	<p>Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data, Bing Liu, Springer, Latest Edition.</p> <p>Introduction to Data Mining: International Edition Pang-Ning Tan, Michael Steinbach, Vipin Kumar, Pearson, Latest Edition.</p> <p>Mining of Massive Datasets, Anand Rajaraman and Jeffrey Ullman, Cambridge University Press, Latest Edition.</p> <p>Machine Learning, Tom Mitchell, McGraw Hill, Latest Edition.</p>

	<p>Pattern Recognition and Machine Learning, Christopher Bishop, Springer, Latest Edition.</p> <p>The Elements of Statistical Learning, Jerome H. Friedman, Robert Tibshirani, and Trevor Hastie, 2nd Edition, Springer, Latest Edition.</p> <p>An Introduction to Statistical Learning, with Applications in R, Latest Edition.</p> <p>Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani, Springer, Latest Edition.</p>										
Assessment	<table border="1"> <tr> <td>Mid – Term Examination</td> <td>25%</td> </tr> <tr> <td>Final Examination</td> <td>40%</td> </tr> <tr> <td>Assignments/Lab</td> <td>25%</td> </tr> <tr> <td>Class Participation and Attendance</td> <td>10%</td> </tr> <tr> <td></td> <td>100%</td> </tr> </table>	Mid – Term Examination	25%	Final Examination	40%	Assignments/Lab	25%	Class Participation and Attendance	10%		100%
Mid – Term Examination	25%										
Final Examination	40%										
Assignments/Lab	25%										
Class Participation and Attendance	10%										
	100%										
Language	English										