Course Title	**Data Mining with Big Data Analytics**				
Course Code	WSS552				
Course Type	Specialization (Elective)				
Level	Master (2nd Cycle)				
Year / Semester	1 / 2 (Spring)				
Teacher's Name	Harris Papadopoulos, PhD and Christos Markides, PhD				
ECTS	10	Lectures / week	3	Laboratories / week	0
Course Purpose and Objectives	The aim of the this course is to provide students with a holistic approach to Big Data, the data model for Big Data, and examine the nature and requirements of a Big Data components, as well as Big Data as a platform. The course will introduce students to Data Mining and focus on intelligent data analysis. The major principles, terminology, problem types and research topics of Data Mining are addressed. The course will examine the main ideas behind some of the leading Machine Learning techniques being used in practical Data Mining and the issues that should be considered in their application. In addition, the course will consist of hands-on approach on Big data and experimentation on real life problems and data to promote the development of critical thinking on the proper application of Data Mining techniques.				
Learning Outcomes	Upon successful completion of this course, students should be able to: • Assess the state of Big Data adoption across a number of industry sectors and apply the key concepts of Data Analytics Lifecycle to tackle Big Data problem. • Describe the Hadoop Architecture, and use the Hadoop file system and components and define and describe the major characteristics of NoSQL databases. • Deploy and manage NoSQL databases using queries. • Define and explain the major principles, terminology and problem types of Data Mining. • Describe and discuss the main Supervised Machine Learning techniques used in practical Data Mining and their theoretical basis and evaluate their strengths and weaknesses. • Explain and propose ways of dealing with the issues involved in the application of Supervised Learning techniques to practical problems • Apply Supervised Learning techniques to a practical problem both in an exploratory or a targeted manner and analyse and evaluate its performance. • Define, explain and demonstrate the main concepts, issues and approaches for designing a Recommender System.				
Prerequisites	None		Required	None	

Course Content	This module consists of the following units: • **Unit 1** introduces the fundamentals of Big Data and the main concepts of Data Mining. • **Unit 2** builds on the fundamentals of Big Data and explains the key aspects of Big Data • **Unit 3** explains the architecture of the Hadoop Framework and foundation core components • **Unit 4** examines the architecture and characteristics of NoSQL databases • **Unit 5** expands on the theory and application of Data Analytics and the Data Analytics Lifecycle • **Unit 6** introduces the main ideas behind some of the leading techniques that are used in practical Data Mining • **Unit 7** examines some advanced Machine Learning techniques • **Unit 8** discusses Web Mining and Recommender Systems
Teaching Methodology	The methodology used to conduct the course is structured around lectures and laboratory exercises, so that students gain theoretical knowledge as well as practical skills. The taught part of the course is delivered to the students with the help of computer presentations. Lecture notes and presentations are available through the web (e-learning platform) for students to use in combination with the textbooks. Furthermore, theoretical principles are explained by means of specific examples and for solving specific problems using practical examples. Lectures are supplemented with supervised and unsupervised computer laboratory. Laboratories will include demonstrations of taught concepts and experimentation with related technologies. Additionally, during laboratory sessions, students apply their gained knowledge and identify the principles taught in the lecture sessions by means of working on different tasks and problems. Students are also allocated exercises for homework, assignments and/or group project to improve both their individual skills and team work.
Bibliography	EMC Education Services (2015), Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data, John Wiley & Sons, 1st ed. Krishnan K. (2013), Data Warehousing in the Age of Big Data, Morgan Kaufmann, 1st ed. Ian H. Witten, Eibe Frank and Mark A. Hall (2011), Data Mining: Practical Machine Learning Tools and Techniques, Morgan Kaufman, 3rd ed. Jure Leskovec, Anand Rajaraman and Jeff Ullman (2014), Mining of Massive Datasets, Cambridge University Press, 2nd ed. (Freely available online) Bing Liu (2011), Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data, Springer-Verlag, 2nd ed. Various relevant academic papers
Assessment	The formal assessment of this course consists of

	• **Participation Activities** (10% of total marks for module) • **Two marked assignments** (50% of total marks for module) • **One closed-book, 3-hour exam** (40% of total marks for module)
Language	English