

Introduction to Data Science and Machine Learning - SS 2018

Bachelor of Science WI / IS
Faculty of Management, Economics, and Social Sciences
Department of Information Systems for Sustainable Society
University of Cologne
Version - May 30, 2018

| | |
|---|--|
| Instructor Prof. Dr. Wolf Ketter | Term SS 2018 |
| TAs Nastaran Naseri, Karsten Schroer, & Philipp Kienscherf | Class Time Wednesday, 16:00 to 19:30 |
| Website www.is3.uni-koeln.de and ILIAS | Class Location 101 Hörsaal XXV (Gutenberg HS), Universitätsstraße 24, Erdgeschoß |

Welcome to Introduction to Data Science and Machine Learning (DSML). This course and the accompanying reading materials aim to provide you with the fundamental knowledge and skills required for an introductory course on data science and machine learning using information systems that drive business success.

Business analytics is the use of data-driven decision making. Companies, governments, and other organizations now collect and have access to large amounts of data about suppliers, clients, employees, citizens, transactions, etc. Data Mining and predictive analytics provide a powerful toolkit for detecting actionable patterns in data and generating predictions. These methods are used in many industries: Mobile companies use their customer database to predict customer churn or to personalize SMS messages for improving customer service; Financial institutions use past loan data to predict defaulting chance for loan applicants; Charities use data from a campaign in one location to target the right people in another location; Politicians use databases of supporters to segment and best target each audience; Movie rental and eCommerce websites provide recommendations based on users online behavior; Renewable energy providers use weather forecasts to predict their electricity generation to better trade in different types of markets.

In the course we will work with a combination of stylized data, real business problems and real data. We will examine types of questions that data mining can answer and the appropriate data mining tools for answering different questions. The emphasis is on understanding the core concepts and logic behind data science and machine learning techniques and their relation to specific business analytics situations. The course provides the theoretical underpinnings of these techniques and is focused on a business analytics team project that will encapsulate the learnings. In your team project, you will develop the idea, the business justification, and the analytics for a compelling (sustainable) business application.

You will learn about the process of data analytics. You will learn to identify problems, to define the structure of an information system, to evaluate competing solutions for a business problem, and you will start to “speak business and analytics”. You will gain an insight into how important these activities are in creating information systems that are truly aligned with business needs. Throughout the course, you will also learn about selected topics in sustainability, especially in terms of renewable electricity.

Some of the most important thoughts that we will present throughout the lecture are summarized in this reader and the references therein. This material is designed to help you find additional references, and to help you recapture what was taught in the lectures. It is not designed, however, to serve as an exclusive source of exam preparation material. In our lectures and workshops, we will present facts, examples, and teach you skills that are not on the following pages, and we reserve the right to ask you about them during the final exam and in your assignments. Please make sure you take advantage of all modes of learning that we offer in this course!