Master's in Data Science

Course

Data Processes

Description

More and more applications in industry, academia, and everyday life are or should be based on careful analysis of data. But data has not been gathered to be directly analyzed and consequently processes are required to transform raw data to datasets usable for knowledge extraction. This presents many opportunities, and the right tools and methods at each steps need to be known in order to be able to learn from the data without drowning.

The goal of this course is to go through the complete data science process. We will learn: i) how to understand the goals of a project and translate it to knowledge discovery goals; ii) to explore data to understand it, iii) to prepare data sets; iv) to model, validate and evaluate models, v) communicate results and vi) deploy. Students will learn concepts, techniques and tools they need to deal with various facets of data science practice, including data collection and integration, exploratory data analysis, predictive modeling, descriptive modeling, data product creation, evaluation, and effective communication. The focus in the treatment of these topics will be on breadth, rather than depth, and emphasis will be placed on integration and synthesis of concepts and their application to solving real life problems. Through real-world examples of wide interest, we introduce methods for:

- Business understanding to be able to prepare a data science project plan
- Data munging/scraping/sampling/cleaning in order to get an informative, manageable data set
- Exploratory data analysis to generate hypotheses and intuition about the data
- Modelling based on tools such as regression, classification, clustering, association
- Evaluation of models
- Communication of results and deployment

Expected Learning Outcomes

After successful completion of this course, students will be able to…

- Produce a project plan for a data science project
- Be able to understand and prepare datasets for knowledge extraction
- Use Python/R and other tools to scrape, clean, and process data
- Apply the correct technique to make predictions/descriptions based on data
- Effectively communicate the outcome of data analysis

- Deploy the models produced