



Introduction to Big Data for Economics

Module Code: ECU33143

Module Title: Introduction to Big Data for Economics

- **ECTS Weighting:** 5
- **Semester/Term Taught:** Semester 1
- **Contact Hours:** 20 hours lectures + 5 hours tutorials
- **Module Personnel:** Lecturer - Jian Cao

Module Learning Aims:

This module introduces the field of Big Data, examining its main features, such as real-time availability, scale, types of variables and structure. It focuses on showing how different types of Big Data, including administrative datasets, high-frequency information, web-scraped datasets, can be used to answer questions in economics.

This module will introduce students to the main topics that have been arising around the concept of “Big Data”. With increasing computing and storage capacity, new opportunities and challenges arise for researchers. The goal of this module is to provide an overview of different datasets, techniques and applications, with an emphasis on practical implementation.

Module Learning Outcomes:

On successful completion of the module students should be able to:

- Explain the challenges and opportunities arising from the “Big Data”.
- Distinguish which type of data to collect based on the typology of empirical questions to tackle and how to efficiently collect them.
- Apply the latest technique to analyze and summarize large amount of data.
- Apply Big Data methodologies such as Data Mining and High-Dimensional Data Analysis in different research settings.



Module Content:

Topics covered in this module will include:

- 1) Understanding Big Data in Economic Contexts: Key characteristics of "big data"
- 2) Challenges and opportunities specific to "Big data" such as:
 - a. Privacy & confidentiality
 - b. Data management practices
 - c. Scalable approach
- 3) Data Collection: Strategies & Challenges
- 4) Data Analysis: Techniques & Tools for Big Data
- 5) Applying Big Data: Case Studies in Economics

Recommended Reading List:

Taddy, Matt. *Business Data Science: Combining Machine Learning and Economics to Optimize, Automate, and Accelerate Business Decisions*. McGraw Hill, 2019.

Gareth James, Daniela Witten, Trevor Hastie & Robert Tibshirani. *An Introduction to Statistical Learning*. Springer, 2021.

Module Pre-Requisite:

ECU22011/ ECU22012 and ECU22031/ ECU22032

Assessment Details:

Assignments (20%)

Final Exam (80%)

Module Website:

Blackboard