Data Analytics (COMP0047)

**Description**

**Aims:**
The module is aimed at introducing to data analytics providing some fundamental data-science tools. Students will learn statistical tools to individuate regularities, discover patterns and laws in complex datasets together with instruments to analyse, characterize, validate, parameterize and model complex data. Practical issues on business data analysis and statistics will be covered with specific case studies also in collaboration with industry partners.

**Learning outcomes:**
On successful completion of the module, a student will be able to:
1. analyse main statistical features of complex datasets;
2. understand how to analyse, characterize empirically complex data;
3. understand how to compute relevant statistical quantities and quantify their confidence intervals;
4. understand how to build sensible models and how to parameterize and validate these models;
5. understand how to quantify inter-dependency/causality structure between different variables;
6. understand how to use the outcome of data-analytics to develop better tools for forecasting;

**Content:**
Empirical investigation of complex data:
- Essential practical familiarization with complex and big data, and with the most commonly used software packages to analyse them. Typical challenges with real data. Basics on data acquisition, manipulation, cleaning, filtering, representation and plotting; Univariate and multivariate statistics:
  - Marginal probability, joint probability and conditional probability. Empirical estimation of probability distributions. Measures of dependency. Cause and effect, Granger causality. Information theoretic measures: mutual information, transfer entropy. Spurious correlations and

**Key information**

- **Year**: 2019/20
- **Credit value**: 15 (150 study hours)
- **Delivery**: PGT L7, Campus-based
- **Reading List**: [View on UCL website](https://www.ucl.ac.uk/)
- **Tutor**: Prof Tomaso Aste
- **Term**: Term 2
- **Timetable**: [View on UCL website](https://www.ucl.ac.uk/)

**Assessment**

- Coursework: 40%
- Coursework: 20%
- Written examination (departmentally managed): 40%

**Find out more**

For more information about the department, programmes, relevant open days and to browse other modules, visit [ucl.ac.uk](https://www.ucl.ac.uk/)

**Disclaimer:** All information correct as of August 2019. Please note that aspects of the module may be subject to change. UCL will make best efforts to inform applicants of major changes.
regularization. Forecasting and regressions. Hypothesis testing and validation; Modelling and filtering through networks:
-Basics on complex networks: definitions and properties. Construction of networks of interactions form correlation and causality measures. Information filtering though networks; Probabilistic modelling:
-Constructing predictive probabilistic models form data. Test and validate model performances. Select between alternative models; Applications and case-study:
-Application of the studied material and methods to practical cases and real data will be done within the course through case-studies developed in collaboration with industry partners. Some case studies will discussed in class and used as demonstrations of the methodologies covered during the lectures. Other case studies will instead be given as assignments, and will represent the core material for the coursework;

Requisites:
In order to be eligible to select this module, a student must be registered on a programme for which it is a formally-approved option or elective choice AND must have a good knowledge of basic mathematics and statistics.