

Tools and methods for analysing complex Science, Technology and Innovation (STI) systems: A gentle introduction to Machine Learning (ML), Network Science (NS) and Text Mining (TM)

IRCrES-CNR, Rome – July 1-3, 2019

1. OBJECTIVES OF THE COURSE

Recent years have witnessed an unprecedented availability of information on social, economic, and technological phenomena. Researchers, practitioners, and policymakers have nowadays access to huge datasets (the so-called “Big Data”) on people, companies and institutions, web and mobile devices, satellites, etc., at increasing speed and detail. Relational (big) data are also in a surge, thus documenting an increasing need to shed light on relationships among research and innovation actors. ML, NS, and TM are relatively new techniques able to enlarge our understanding of complex socio-technological systems, either by digging deeply into the data informative power (ML and TM), or by increasing the understanding of the system relational dimension (NS). The training will provide participants with the essential tools for a correct application of some popular ML, NS, and TM methods in various STI contexts. In particular:

- ML techniques proves useful for factor importance detection, as well as for classification purposes in a model-free stance;
- NS techniques are useful to identify and study structure and dynamics of large and complex STI communities;
- TM is a set of tools able to extract relevant information from articulated documents, both for descriptive and policy analysis purposes.

The course foresees three modules (one on ML, one on NS, and one on TM) with the aim of balancing theory and applications. Participants will run some exercises assigned by the instructor under his supervision.

2. GENERAL CONDITIONS FOR PARTICIPATION

Target audience for this course are researchers in research policy, higher education, and innovation studies with a quantitative orientation, who aim to extend their competence on ML, NS, and TM analysis. Basic requisites for admission will be:

- Knowledge on basic principles of statistics;
- Interest in STI studies.

The course is addressed to:

- Senior scientists, early career researchers and PhD students at the last phase of their training;
- Officers from the policy making level;
- Research associations.

Participants will be asked to bring their PC (with RStudio and Stata software installed) when attending the course. It is possible to download the RStudio software [here](#).

3. SUBMISSION OF THE APPLICATIONS

Send an email to eccrome@ircres.cnr.it with a clear indication of your interest for applying and attach an **updated CV**. You will receive the confirmation once the registration has been processed. In case you do not receive any notification from us within one week, please contact the same e-mail address.

Deadline for submission: May 1, 2019.

4. SELECTION

Maximum number of participants for the course is 20. Participants will be selected on the basis of their interests and CV. Notification of acceptance and request of confirmation will be sent after the selection process is completed.

5. FEES AND PAYMENTS

No fees to be paid by EUSPRI members.

Non-EUSPRI members must pay a participation fee of Euro 250 (which includes the social dinner).

Accommodation will be covered only in case of researchers, early researchers, and PhDs coming from European countries.

6. SCHEDULE FOR THE COURSE

The course will take place from July 1st, 2019 to July 3th, 2019.

7. VENUE

IRCrES-CNR – Research Institute on Sustainable Economic Growth of the National Research Council of Italy,
Via dei Taurini 19, Rome, Italy [[MAP](#)]

8. TEACHING STAFF

Dr. Giovanni Cerulli (IRCrES-CNR), Dr. Antonio Zinilli (IRCrES-CNR), Dr. Carlo Drago (Università Cusano)

9. LOCAL ORGANIZING COMMITTEE

Dr. Giovanni Cerulli, Dr. Antonio Zinilli, Dr. Emanuela Reale, Dr. Marco De Biase (IRCrES-CNR)

10. LOCAL CONTACT

Dr. Marco De Biase, IRCrES CNR (eccrome@ircres.cnr.it);

PROGRAMME

Tools and methods for analysing complex Science, Technology and Innovation (STI) systems: A gentle introduction to Machine Learning (ML), Network Science (NS) and Text Mining (TM)

IRCrES-CNR, Rome – July 1-3, 2019

Day 1 – July 1, 2019

9:30

Registration

10:00-10:30

Welcome

10:30-11:30

An introduction to Machine Learning and Data Science for analyzing complex STI systems:
Identification, prediction, trade-offs, and validation approaches

11:30-12:00

Coffee break

12:00 13:00

Monte Carlo, Bootstrap, and model Validation: a primer

13:00-14:00

Lunch

14:00-15:00

Model selection and regularization:
Optimal subset selection; Shrinkage Methods: Lasso, Ridge, and Elastic regression

15:00 – 16:00

Laboratory with Stata and R

16:00 – 16:30

Coffee break

16.30 – 17.30

Applications to real STI datasets

Day 2 – July 2, 2019

10:00-11:00

Basic concepts of Social network analysis

11:00-11:30

Coffee break

11:30-12:30

Main Network models

12:30-13:30

Lunch

13:30-14:30

Application scenarios (some illustrative examples of Network Science on specific datasets of Science Technology and Innovation (STI) systems)

14:30-15:30

Organization of the laboratory assignments
Creation of groups and provision of data

15:30-16:00

Coffee break

16.00-17.00

Young participants' presentation of their assignments

19.30

Social Dinner

Dinner on July 2, 2019

19:30 Meeting at Via dei Taurini, 19
Ristorante Pizzeria Efeso il Barrocciao"

Day 3 – July 3, 2019

10:00-11:00

Concepts and tools for Text Mining analysis

11:00-11:30

Coffee break

11:30-12:30

Text Mining applications with R

12:30-13:30

Lunch
