Advanced Course

BIG DATA TECHNIQUES AND APPLICATIONS IN THE AGRI-FOOD SYSTEM

Zaragoza (Spain), 17-21 June 2019

1. Objective of the course

The volume of available data has doubled in the last five years. This is creating new opportunities for business and science alike but Big Data applications are not yet commonplace in the Agri-Food system. In this course we will give an overview of the methods and computational architectures that are being used to give answers to previously unanswerable questions using Big Data analytics.

The course will involve lectures and practical sessions. It will introduce participants to web services for accessing agri-food data, supervised and unsupervised learning, segmentation methods and deep learning. Participants will also gain experience in organising data using traditional methods, such as SQL, as well as more modern approaches including NoSQL and the Semantic Web.

The focus of the course is on the whole agri-food system from primary production to the food consumer. Experts from the agri-food system will contribute to the course with illustrative examples of business and science success stories.

At the end of the course the participants will:

– Know what is meant by big data and how it is affecting the way that we do business.
– Be familiar with a range of computational architectures suitable for big data.
– Understand a range of analytical techniques that can be applied to big data including traditional statistical methods, machine learning and artificial intelligence.
– Be aware of the issues that arise when transferring big data analytics to a production environment.
– Have experienced a number of real world examples of business applications of big data.
– Be capable to write simple scripts that use publicly available libraries for analysing big data.

2. Organization

The course is jointly organized by the International Centre for Advanced Mediterranean Agronomic Studies (CIHEAM), through the Mediterranean Agronomic Institute of Zaragoza (IAMZ), and the International Center for Agricultural Research in the Dry Areas (ICARDA). The course will take place at the Mediterranean Agronomic Institute of Zaragoza and will be given by well qualified lecturers from international organizations and from universities, technological centres and private companies in different countries.

The course will be held over a period of one week, from 17 to 21 June 2019, in morning and afternoon sessions.

3. Admission

The course is designed for 25 professionals with a university degree. It is intended for different actors in the agri-food system, such as: data managers, professionals involved in precision agriculture and livestock farming, food processors and distributors, marketing and sales managers, members from the administration involved with data advising decision makers and final users, specialists from consultancy services, and R+D experts involved in data analytics.

Participants are expected to have a quantitative background, and some familiarity with scripting languages (e.g. R, Python).

All lectures and class material will be given in English.

4. Registration

Candidates must apply online at the following address: http://www.admission.iamz.ciheam.org/en/

Applications must include the curriculum vitae and copy of the supporting documents most related to the subject of the course.

The deadline for the submission of applications is 8 April 2019.

Applications from those requiring authorization to attend the course, may be accepted provisionally.

Registration fees for the course amount to 500 euro. This sum covers tuition fees only.

5. Scholarships

Candidates from CIHEAM member countries (Albania, Algeria, Egypt, France, Greece, Italy, Lebanon, Malta, Morocco, Portugal, Spain, Tunisia and Turkey) and from ICARDA Middle East and...
North Africa (MENA) partners may apply for scholarships covering registration fees, and for scholarships covering the cost of travel and full board accommodation. Candidates from other countries who require financial support should apply directly to other national or international institutions.

6. Insurance

It is compulsory for participants to have medical insurance valid for Spain. Proof of insurance cover must be given at the beginning of the course. Those who so wish may participate in a collective insurance policy taken out by the Organization, upon payment of the stipulated sum.

7. Teaching organization

The course requires personal work and interaction among participants and with lecturers. The international characteristics of the course favour the exchange of experiences and points of view. The course has an applied approach. Formal lectures are complemented with examples and case studies on the use of big data at different points of the agri-food chain. Furthermore, during the course participants will be introduced to the use of different data sources and management tools, and will carry out guided exercises including big data analysis and interpretation of results.

8. Programme

1. Introduction to Big Data (2 hours)
   1.1. What is Big Data? Where is the data? Where do they come from?
   1.2. What new kind of questions can be answered?
   1.3. Digital transformation and digital services [Software/Platform/Infrastructure as a Service (SaaS, PaaS, IaaS)]: sensors, cyber physical system (CPS), Internet of Things (IoT), cybersecurity, cloud computing, etc.
   1.4. Organizational transformation: new roles (Chief Data Officer and Data Scientists), new digital products and services, new relationship with providers and customers
   1.5. Developing your businesses technological roadmap
   1.6. Socio-economic aspects of Big Data: data ownership, stakeholder involvement, legal and ethical implications
   1.7. From the data we have, to the knowledge we need

2. Accessing, organising and handling Big Data (5 hours)
   2.1. Accessing: from CSV to API, streaming and static data
   2.2. Organising: SQL, NoSQL, Semantic Web/Linked data
   2.3. Handling: Hadoop, map-reduce
   2.4. Infrastructure: Big Data architectures and analytical libraries
      2.4.1. Popular stacks
      2.4.2. Data warehouse versus data lake
   2.5. Hybrid architectures (IoT, Big Data, and classical approaches)

3. Analysing Big Data (9 hours)
   3.1. Statistical and machine learning processes phases
      3.1.1. Problem formalization
      3.1.2. Exploratory analysis
      3.1.3. Data cleaning and feature engineering
      3.1.4. Learning models
      3.1.5. Presentation of results
   3.2. How do we improve our models?
      3.2.1. Addition of new data (cases and features)
      3.2.2. Cleaning improvement
      3.2.3. Training models
      3.2.4. Score analysis and model validation
   3.3. Supervised learning
      3.3.1. Regression: livestock growth prediction
      3.3.2. Classification: livestock fertility prediction
   3.4. Unsupervised learning
      3.4.1. Clustering: customer segmentation
      3.4.2. Dimension reduction: Principal Components Analysis (PCA)
   3.5. Other learning techniques
      3.5.1. Reinforcement learning
      3.5.2. Nearest Neighbour techniques: k-NN
      3.5.3. Neural Networks and Deep Learning techniques
   3.6. Optimization techniques (Genetic Algorithms, Integer Linear Programming, Particle Swarm)
   3.7. Combining models in ensembles
   3.8. Business intelligence tools (e.g. Kibana, PowerBI, Zeppelin)

4. Robust implementation (6 hours)
   4.1. Architecture needed to exploit the results: model generation versus exploitation, real time versus batch analytics, pipeline processes, publish/subscribe architectures, etc.
   4.2. Digitalization of agri-food systems: horizontal and vertical integration, digital products and services, and digital business models and customer access
   4.3. ICARDA’s activities on Big Data
   4.4. Big data applications in the agri-food sector revisited
      4.4.1. Case study Bayer
      4.4.2. Case study Carrefour
      4.4.3. Case study John Deere

5. Practical work (14 hours)
   5.1. Understand a relevant state of the art Big Data architecture and tools with a simple example (e.g. elastic stack: elastic + logstash + kibana)
   5.2. Hands on
      5.2.1. Accessing data from external sources
      5.2.2. Data integration
      5.2.3. Data exploration
      5.2.4. Consumer segmentation for targeted marketing
      5.2.5. Predict plant growth from DNA and environmental conditions
      5.2.6. Image recognition
      5.2.7. Open data and other public data sources

---

**GUEST LECTURERS**

G. ANZALDI, Eurecat, Lleida (Spain)
D. ARROBAS, John Deere, Madrid (Spain)
J. BETRÁN, Bayer, Toulouse (France)
C. BIRADAR, ICARDA, Cairo (Egypt)
S. COLEMAN, Newcastle University (United Kingdom)
X. DOMINGO, Eurecat, Lleida (Spain)
L. ECHEVERRÍA, Eurecat, Lleida (Spain)
M. SOLANKI, Agrimetrics, Reading (United Kingdom)
F. VAN EEUWIJK, Wageningen UR (The Netherlands)
Expert from Carrefour