

BIG DATA TECHNIQUES AND APPLICATIONS IN THE AGRI-FOOD SYSTEM
Zaragoza (Spain), 17-21 June 2019

PROGRAMME

- 1. Introduction to Big Data (2 hours) (G. Anzaldi)**
 - 1.1. What is Big Data? Where is the data? Where 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) (M. Solanki)**
 - 2.1. Accessing: from CSV to API, streaming and static data (1 h)
 - 2.2. Organising: SQL, NoSQL, Semantic Web/Linked data (1 h)
 - 2.3. Handling: Hadoop, map-reduce (1 h)
 - 2.4. Infrastructure: Big Data architectures and analytical libraries (1 h)
 - 2.4.1. Popular stacks
 - 2.4.2. Data warehouse versus data lake
 - 2.5. Hybrid architectures (IoT, Big Data, and classical approaches) (1 h)
- 3. Analysing Big Data (9 hours)**
 - 3.1. Statistical and machine learning processes phases (2 h) (F. van Eeuwijk)
 - 3.1.1. Problem formalisation
 - 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? (2 h) (F. van Eeuwijk)
 - 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 (1 h) (S. Coleman)
 - 3.3.1. Regression: livestock growth prediction
 - 3.3.2. Classification: livestock fertility prediction
 - 3.4. Unsupervised learning (1 h) (S. Coleman)
 - 3.4.1. Clustering: customer segmentation
 - 3.4.2. Dimension reduction: Principal Components Analysis (PCA)
 - 3.5. Other learning techniques (1 h) (X. Domingo)
 - 3.5.1. Reinforcement learning
 - 3.5.2. Nearest Neighbour techniques: k-NN
 - 3.5.3. Neural Networks and Deep Learning techniques
 - 3.6. Optimisation techniques (Genetic Algorithms, Integer Linear Programming, Particle Swarm) (1 h) (X. Domingo)
 - 3.7. Combining models in ensembles (0.5 h) (X. Domingo)
 - 3.8. Business intelligence tools (e.g. Kibana, PowerBI, Zeppelin) (0.5) (X. Domingo)
- 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. (1 h) (L. Echeverría)
- 4.2. Digitalisation of agri-food systems: horizontal and vertical integration, digital products and services, and digital business models and customer access (1 h) (L. Echeverría)
- 4.3. ICARDA's activities on Big Data (1 h) (C. Biradar)
- 4.4. Big data applications in the agri-food sector revisited
 - 4.4.1. Case study Bayer (1 h) (J. Betrán)
 - 4.4.2. Case study Carrefour (1 h) (expert to be nominated by Carrefour)
 - 4.4.3. Case study John Deere (1 h) (D. Arrobas)

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) (5.1, 5.2.1, 5.2.2 and 5.2.3: 3 h) (M. Solanki, X. Domingo, L. Echeverría)
- 5.2. Hands on
 - 5.2.1. Accessing data from external sources (M. Solanki, X. Domingo, L. Echeverría)
 - 5.2.2. Data integration (M. Solanki, X. Domingo, L. Echeverría)
 - 5.2.3. Data exploration (M. Solanki, X. Domingo, L. Echeverría)
 - 5.2.4. Consumer segmentation for targeted marketing (S. Coleman, X. Domingo, F. van Eeuwijk, L. Echeverría)
 - 5.2.5. Predict plant growth from DNA and environmental conditions (F. van Eeuwijk, X. Domingo, L. Echeverría)
 - 5.2.6. Image recognition (X. Domingo, C. Biradar, L. Echeverría)
 - 5.2.7. Open data and other public data sources (X. Domingo, F. van Eeuwijk, S. Coleman, C. Biradar, L. Echeverría)