DeCAIR Training Course: Introduction to Data Science

Prerequisites: basic theoretical knowledge of linear regression, Bayesian classification methods, artificial neural networks, support vector machines, principal component analysis, and clustering.

Theory (8 h)

T1. **Introduction** to data science (1 h) – J. Casillas

T2. Exploratory data analysis: **visualization** (1 h) – J. Casillas

T3. Fundamentals of **classification**: testing and evaluation, decision trees, lazy, ensemble learning (2,5 h) – A. Fernández

T4. **Clustering** and **association** rules (1,5 h) – J. Casillas

T5. Basic data **preprocessing**: selection and processing of instances and features, noise processing (2 h) – S. García

Practical (16 h)

P1. Visualization tools* (3 h): exploratory data analysis by different tools – J. Casillas

P2. **BigML*** (5 h): basic decision trees, ensemble learning, clustering (k-means), association rules (Apriori), model's visualization – J. Casillas

P3. **KNIME**** (8 h): supervised learning by classification and regression, unsupervised learning by clustering and association rules, basic data preprocessing – A. Fernández, S. García

    * on-line freely-available software   ** multiplatform open-source software

Applications: many different application examples will be afforded, such as Twitter analysis on politics, prediction of telemarketing campaigns, sales analysis, patient profiling, disease diagnosis, relational analysis of social questionnaires, migration flows, etc.

Agenda

Day 1: Monday, **March 14**, 2022; 3pm~6:30pm (CET)
    T1 (1 h) / **T2** (1 h) / *Break* (30 m) / **P1** (1 h)

Day 2: Tuesday, **March 15**, 2022; 3pm~6:30pm (CET)
    P1 (2 h) / *Break* (30 m) / **T3** (1 h)

Day 3: Monday, **March 21**, 2022; 3pm~6:30pm (CET)
    T3 (1,5 h) / *Break* (30 m) / **T4** (1,5 h)

Day 4: Tuesday, **March 22**, 2022; 3pm~6:30pm (CET)
    P2 (2 h) / *Break* (30 m) / **P2** (1 h)

Day 5: Monday, **March 28**, 2022; 4pm~7:30pm (CET)
    T5 (1 h) / Break (30 m) / **P2** (2 h)

Day 6: Wednesday, **March 30**, 2022; 4pm~7:30pm (CET)
    T5 (1 h) / Break (30 m) / **P3** (2 h)

Day 7: Thursday, **March 31**, 2022; 4pm~7:30pm (CET)
    P3 (2 h) / Break (30 m) / **P3** (1 h)

Day 8: Wednesday, **April 6**, 2022; 4pm~7:30pm (CET)
    P3 (2 h) / *Break* (30 m) / **P3** (1 h)

Place

Fully on-line through **Google Meet** (access provided by UGR).

Teachers

Jorge Casillas received the PhD graduate degree in Computer Science in 2001 from the University of Granada, Spain. He is a Full Professor with the Department of Computer Science and Artificial Intelligence, University of Granada. He was Visiting Research Fellow at the University of the West of England (Bristol, UK) and the University of Birmingham (Birmingham, UK).

He teaches on artificial intelligence, data science, big data, machine learning, and intelligent control and robotics. He is author of 40+ refereed journal articles and 60+ conference papers. His publications accumulate 4,400+ citations in Google Scholar (h-index 31).

His works in early 2000 in the pursuit of interpretable fuzzy models were pioneers. Since 2007, he has been working on the design of machine learning algorithms to deal with data streams in real-time. In recent years, his interest in a fairer, more equitable and minority-sensitive artificial intelligence has grown. Currently, he is leading a national research project on the use of responsible machine learning to characterize people with chronic diseases.

URL: https://ccia.ugr.es/~casillas, e-mail: casillas@decsai.ugr.es

Alberto Fernández received the M.Sc. and Ph.D. degrees in computer science from the University of Granada, Granada, Spain, in 2005 and 2010, respectively. He is an Associate Professor with the Department of Computer Science and Artificial Intelligence, University of Granada, Spain.

He has published more than 100 papers in highly rated JCR journals and international conferences. In 2021 he was bestowed with the IEEE CIM Outstanding Paper Award in IEEE Computational Intelligence Magazine. In 2013, 2014, and 2017 Dr. Fernández received the University of Granada Prize for Scientific Excellence Works in the field of Engineering. He has also been awarded in 2011 with the Lofti A. Zadeh Best Paper prize (IFSA Association). He has been recently selected as a Highly Cited Researcher (in the field of Computer Science, 2017 Clarivate Analytics). He is member of editorial board of several JCR journals such as Applied Intelligence, Plos-One, Electronics, and Mathematics.

His research interests include eXplainable Artificial Intelligence, Fairness-Accountability-Transparency-Ethics, classification in imbalanced domains, data science in big data applications and Bioinformatics.

E-mail: alberto@decsai.ugr.es
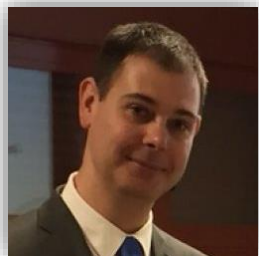
Salvador García received the B.S. and Ph.D. degrees in Computer Science from the University of Granada, Granada, Spain, in 2004 and 2008, respectively. He is a Full Professor in the Department of Computer Science and Artificial Intelligence, University of Granada, Granada, Spain.

Dr. García has published more than 100 papers in international journals (more than 75 in Q1), h-index 59. As edited activities, he is an associate editor in chief of "Information Fusion" (Elsevier), and an associate editor of "Swarm and Evolutionary Computation" (Elsevier) and "AI Communications" (IOS Press) journals. He is a co-author of the books entitled "Data Preprocessing in Data Mining", "Learning from Imbalanced Data Sets" and "Big Data Preprocessing: Enabling Smart Data" published by Springer.

His research interests include data science, data preprocessing, Big Data, evolutionary learning, Deep Learning, metaheuristics and biometrics. He belongs to the list of the Highly Cited Researchers in the area of Computer Sciences (2014-2020): http://highlycited.com/ (Clarivate Analytics).

E-mail: salvagl@decsai.ugr.es