David Charte

Data scientist

Granada, Spain
+34 697 494 169
david@deivi.ch
deivi.ch
fdavidcl
born 19 oct 1994

Experience

2018 — **Doctoral student**, *Universidad de Granada*, supervisors: Francisco Herrera & Francisco Charte.
Lecturer training contract under Spanish FPU program ref. FPU17/04069, defending thesis soon.

- **Main topic**: *Finding alternative representations for data through deep learning techniques*
- Collaborated with **Repsol** on optimization of refinery processes (machine learning and autoencoders)
- Collaborated with **ArcelorMittal** on semantic segmentation of metallographic microstructures (adaptation and training of semi-supervised fully convolutional models)
- Directed two bachelor's theses on automatic melody synthesis with autoencoders and neural search for COVID-19 detection in chest X-rays, respectively
- Published a 5-part free online course (in Spanish) on linear algebra and dimensionality reduction

2020 — **Online course author and lecturer**, *CampusMVP*, Spain.
Course on Data Science and Machine Learning with both video-based and written lectures

2018 — **Researcher**, *Universidad de Granada*, supervisor: Francisco Herrera.
Research contract with project *BigDaPTOOLS*. Task: *Development of data preprocessing libraries in R*

2016–2018 — **Undergraduate researcher**, *Universidad de Granada*.
Research grant. Topic: *Interpretative analysis of unsupervised deep learning techniques and extraction of multi-view models for supervised learning*

Education

2017–2018 — **M.Sc. in Data Science and Computer Engineering**, *Universidad de Granada*, Granada.
Emphasis in data science

2012–2017 — **B.Sc. in Computer Science**, *Universidad de Granada*, Granada, *9.40/10*.

2012–2017 — **B.Sc. in Mathematics**, *Universidad de Granada*, Granada, *9.04/10*.

2017 — **Intl. Summer School on Deep Learning**, *Universidad de Deusto & Rovira i Virgili University*.

2014 — **A practical approach to Data Science and Big Data**, *Intl. University of Andalusia (UNIA)*.

2013 — **New Trends on Computer Engineering**, *Centro Mediterráneo (UGR)*.

2007–2012 — **Project for detection/stimulus of mathematical talent (ESTALMAT)**, *SAEM-Thales*, Granada.

Skills

Soft skills Learns fast · Loves teaching · Team player · Natural problem solver · Organized and meticulous

Languages Spanish (native) · English (Advanced, CEFR C1) · French (Intermediate, CEFR B2) · Swedish (Basic)

Data Science

Models Autoencoders · (Fully) Convolutional Networks · Standard machine learning

Technologies Tensorflow/Keras · Pytorch · Scikit-Learn · Matplotlib

Development

Programming Python · R · Ruby · C++ · C · Shell · SQL

Web HTML · JavaScript · CSS · RWD · Vue.js

Misc

Systems Linux/UNIX · Docker · NGINX

Tools Git · GitHub · LaTeX · Emacs

Journal Publications

2021 José Daniel Pascual-Triana, David Charte, Marta Andrés Arroyo, Alberto Fernández, and Francisco Herrera (2021). "Revisiting data complexity metrics based on morphology for overlap and imbalance: snapshot, new overlap number of balls metrics and singular problems prospect". In: *Knowledge and Information Systems*, pp. 1–29.

2020 David Charte, Francisco Charte, María J del Jesus, and Francisco Herrera (2020). "An analysis on the use of autoencoders for representation learning: Fundamentals, learning task case studies, explainability and challenges". In: *Neurocomputing* 404, pp. 93–107.

2020 Siham Tabik, Anabel Gómez-Ríos, José Luis Martín-Rodríguez, Iván Sevillano-García, Manuel Rey-Area, David Charte, Emilio Guirado, Juan-Luis Suárez, Julián Luengo, MA Valero-González, et al. (2020). "COVIDGR dataset and COVID-SDNet methodology for predicting COVID-19 based on Chest X-Ray images". In: *IEEE Journal of Biomedical and Health Informatics* 24.12, pp. 3595–3605.

2019 David Charte, Francisco Charte, Salvador García, and Francisco Herrera (2019). "A snapshot on nonstandard supervised learning problems: taxonomy, relationships, problem transformations and algorithm adaptations". In: *Progress in Artificial Intelligence* 8.1, pp. 1–14.

2019 David Charte, Francisco Herrera, and Francisco Charte (2019). "Ruta: Implementations of neural autoencoders in R". In: *Knowledge-Based Systems* 174, pp. 4–8.

2018 David Charte, Francisco Charte, Salvador García, María J. del Jesus, and Francisco Herrera (2018). "A practical tutorial on autoencoders for nonlinear feature fusion: Taxonomy, models, software and guidelines". In: *Information Fusion* 44, pp. 78 –96.

2018 Francisco Charte, Antonio J. Rivera, David Charte, María J. del Jesus, and Francisco Herrera (2018). "Tips, guidelines and tools for managing multi-label datasets: The mldr.datasets R package and the Cometa data repository". In: *Neurocomputing* 289, pp. 68 –85.

2015 Francisco Charte and David Charte (2015). "Working with multilabel datasets in R: the mldr package". In: *The R Journal* 7 (2), pp. 149–162.

Submitted works

David Charte, Francisco Charte and Francisco Herrera. "Reducing Data Complexity using Autoencoders with Class-informed Loss Functions"

Julián Luengo, Raúl Moreno, Iván Sevillano, David Charte, Adrián Peláez-Vegas, et al. "A tutorial on the segmentation of metallographic images: taxonomy, new MetalDAM dataset, deep learning-based ensemble model, experimental analysis and challenges"

Projects

2021 **Slicer (convolutional)**, *Convolutional autoencoder model for complexity reduction*.
Source code: `http://github.com/fdavidcl/slicer-conv`
Tensorflow implementation of a convolutional autoencoder which learns from labels with an SVM loss.

2017 **Cometa**, *The comprehensive multi-label data archive*.
Source code: `https://github.com/fdavidcl/cometa`
Docker container that deploys an automatized web repository to prepare and host multi-label datasets.

2016 **Ruta**, *Software for unsupervised deep architectures*.
Source code: `https://github.com/fdavidcl/ruta`
R package for training unsupervised Deep Learning models.

2014 **mldr**, *R package for analyzing and manipulating multilabel datasets*.
Source code: `https://github.com/fcharte/mldr`
R library for exploratory data analysis of multi-label datasets.

Interests and communities

Interests Data science, free culture, (human and machine) languages, scientific dissemination

2014–2020 **LibreIM**, *Student community dedicated to Mathematics and Computer Science*, Co-founder.

- We organized regular talks for compsci & math students, several of them given by myself

2016–2017 **Interferencias**, *Non-profit group interested in online rights and security*, Participant.