YASSER GONZALEZ

Website – http://yassergonzalez.com
Email – contact@yassergonzalez.com

SUMMARY

- Passionate about the design and implementation of algorithms to solve challenging problems that require building scalable predictive models and drawing conclusions from large datasets.
- Keen interest in machine learning & data science, and very enthusiastic about working with open-source technologies.
- Bachelor's degree in Computer Science and master's degree in Information Systems & Technology with emphasis on applied mathematics.

EXPERIENCE

2017–present **Senior Data Scientist.** Canopy Labs, Canada.

2015–2017 **Data Scientist.** Canopy Labs, Canada.

- Developed large-scale recommender systems for diverse application domains – working from the definition of the requirements and business metrics; to the design, implementation and evaluation of the systems in production.
- Employed technologies such as Apache Spark with Python and Scala running on Mesos and YARN, Amazon EMR & S3, and MongoDB.

2013–2015 **Research & Teaching Assistant.** York University, Canada.

- Designed and implemented different methods for optimizing the user interaction in a configuration process.
- Developed techniques to improve the performance of search heuristics on multimodal optimization problems.
- Implemented scientific software in Python (using NumPy, pandas, and SciPy), R, and MATLAB/Octave.

2011–2013 **Research Assistant**. Institute of Cybernetics, Mathematics and Physics, Cuba.

- Developed new estimation of distribution algorithms (EDAs) using copulas and vines to model the probability distributions.
- Implemented a group of R packages available on CRAN, and a C library for dependence modeling using vines.

EDUCATION

2013–2015 **Master's degree, Information Systems & Technology.**
York University, Canada.

- *Thesis:* Efficient Calculation of Optimal Configuration Processes.
- *Selected Courses:* Mining of Massive Datasets, Advanced Information Retrieval Systems, Introduction to Computational Linguistics.

2014–2015 **Data Science Specialization.**
A non-credit series offered by Johns Hopkins University through Coursera.

- *Selected Courses:* R Programming, Getting and Cleaning Data, Exploratory Data Analysis, Statistical Inference, Regression Models, Practical Machine Learning.

2006–2011 **Bachelor's degree, Computer Science.**
University of Havana, Cuba.

- *GPA:* 5.0/5.0 (Summa Cum Laude).
- *Thesis:* Estimation of Distribution Algorithms Based on Copulas and Vines.
- *Selected Courses:* Linear Algebra, Calculus, Probability & Statistics, Design & Analysis of Algorithms, Operating Systems, Computer Networks, Database Systems, Artificial Intelligence, Information Retrieval Systems.

SELECTED OPEN-SOURCE SOFTWARE

- **configurator** – Python package providing different methods for optimizing the user interaction in a configuration process. https://github.com/yasserglez/configurator.
- **copulaedas** – R package for implementing and studying estimation of distribution algorithms (EDAs) based on copulas. https://github.com/yasserglez/copulaedas.
- **vines** – R implementation of the vine graphical models for representing high-dimensional probability distributions. https://github.com/yasserglez/vines.
- **ngram_profile** – Python library for text classification based on character n-grams. https://github.com/yasserglez/ngram_profile.
- **dml** – C library for dependence modeling using C-vines, D-vines and R-vines. https://github.com/yasserglez/dml.

For more information, please see http://yassergonzalez.com/software.

SELECTED PUBLICATIONS

- Y. Gonzalez-Fernandez, S. Chen. (2015). Leaders and Followers – A New Metaheuristic to Avoid the Bias of Accumulated Information. In *IEEE Congress on Evolutionary Computation*, 776–783. IEEE. http://dx.doi.org/10.1109/CEC.2015.7256970.
- Y. Gonzalez-Fernandez, M. Soto. (2014). copulaedas: An R Package for Estimation of Distribution Algorithms Based on Copulas. *Journal of Statistical Software*, 58(9), 1–34. http://www.jstatsoft.org/v58/i09.
- Y. Gonzalez-Fernandez, S. Chen. (2014). Identifying and Exploiting the Scale of a Search Space in Particle Swarm Optimization. In *Conference on Genetic and Evolutionary Computation*, 17–24. ACM. http://doi.acm.org/10.1145/2576768.2598280.
- M. Soto, A. Ochoa, Y. Gonzalez-Fernandez, Y. Milanés, A. Álvarez, D. Carrera, and E. Moreno. (2012). Vine Estimation of Distribution Algorithms with Application to Molecular Docking. In S. Shakya and R. Santana (eds.), *Markov Networks in Evolutionary Computation*, 209–225. Springer. http://link.springer.com/chapter/10.1007/978-3-642-28900-2_13.
- Y. Gonzalez-Fernandez, D. Carrera, M. Soto, A. Ochoa. (2012). Vine Estimation of Distribution Algorithms. In *VIII Congreso Español sobre Metaheurísticas, Algoritmos Evolutivos y Bioinspirados*, 1–7. http://simd.albacete.org/maeb2012/papers/paper_99.pdf.

For more information, please see http://yassergonzalez.com/publications.