YASSER GONZALEZ

Website – https://yassergonzalez.com
Email – yasser@yassergonzalez.com

SUMMARY

- Passionate about solving practical problems using machine learning.
- Interested in applications of AI and machine learning related to ranking and information retrieval, natural language processing, recommender systems, and mathematical optimization.

EXPERIENCE

2018–present **Research Scientist.** Amazon, Canada.

- Working on natural language understanding in Alexa.
- Applying deep learning and representation learning to build ranking models to score the relevancy of entities mentioned in dialogs.
- Using technologies such Spark for data generation and TensorFlow for model development and deployment.

2017–2018 **Senior Data Scientist.** Canopy Labs, Canada.

- Led the development of recommender systems and customer propensity models using Spark, PyTorch, TensorFlow, scikit-learn, and xgboost.
- Built deep learning models for predictive customer analytics based on the customer's historical sequence of actions.

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

- Developed large-scale recommender systems for multiple industries such as online retail and travel. Participated in the definition of the business metrics as well as the design, implementation and evaluation of the systems in production.
- Used technologies such as Spark (with Python and Scala), Mesos, YARN, Amazon EMR, S3, and MongoDB.

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

- Designed and implemented reinforcement learning techniques to optimize the user interaction in configuration processes.
- Proposed methodologies to improve 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.** 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.

- *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 PUBLICATIONS

- Y. Gonzalez-Fernandez, S. Hamidi, S. Chen, S. Liaskos. (2019). Efficient Elicitation of Software Configurations Using Crowd Preferences and Domain Knowledge. *Automated Software Engineering*, 26(1), 87–123. https://link.springer.com/article/10.1007/s10515-018-0247-4
- 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. https://dx.doi.org/10.1109/CEC.2015.7256970.
- 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. https://doi.acm.org/10.1145/2576768.2598280.
- 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. https://www.jstatsoft.org/v58/i09.

SELECTED OPEN-SOURCE SOFTWARE

- Code used in Kaggle competitions. https://github.com/yasserglez/kaggle.
- Python package with reinforcement learning methods for optimizing the user interaction in a configuration process. https://github.com/yasserglez/configurator.
- R package for developing estimation of distribution algorithms based on copulas. https://github.com/yasserglez/copulaedas.
- R implementation of the vine graphical models for high-dimensional probability distributions. https://github.com/yasserglez/vines.