

Sergio Ramírez Gallego

ML Engineer & Data Scientist



Born in 1988 in Jaén, Spain



(34) 699 72 40 43



srg00017@gmail.com



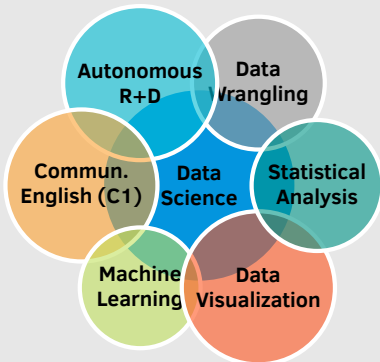
/in/sramirezg



sramirez

Skills

Overview



Programming

Novice —————> Daily

Scala • Spark • Python

Java • Git • Flink • R

C • Docker • Cloud • Linux scripting

Software

Spark-IT-FS* - A distributed feature selection framework for Apache Spark | 111 ★

fast-mRMR - mRMR feature selector for Spark | 57 ★

Spark-MDLP* - A Spark implementation of the MDLP discretizer (presented in Spark Summit '17) | 39 ★

FeuerFreiKiller - A deep learning project to automatically detect wildfire in nature using spectrogram images.

*<https://spark-packages.org/>.

Education

2018 - Now **Self-learning: online courses (see LinkedIn profile)**

2013 - 2018 **Ph. D., Computer Science CUM LAUDE** University of Granada, Spain
🏆 Best Ph. D. Thesis Project award by Spanish Association of Artificial Intelligence (AEPIA)

2012 - 2013 **MSc., Soft Computing and Intelligent Systems (GPA: 8.1/10.0)** University of Granada, Spain

2009 - 2012 **BEng., Computer Engineering (GPA: 8.8/10.0)** University of Jaen, Spain

2006 - 2009 **BEng., Computer Technical Engineering (GPA: 7.9/10.0)** University of Jaen, Spain
🏆 Bachelor extraordinary award

Research

2012 - 2018 **Ph. D. Candidate, Graduate Research and Teaching Assistant** University of Granada

Thesis: Distributed Data Reduction Algorithms for Big Data

- Scalable data reduction (discretization, feature and instance selection) algorithms for Spark. Achievements: fast feature selection on 30M cols (kdd2010), and discretization on 60M rows.
- **Tools:** Spark, Scala, Java, MOA, Flink, \LaTeX , Git, Maven.
- Publications: <https://scholar.google.com/citations?user=cEkfaW4AAAAJ>

Experience

Mar 2019 - Now **Data Scientist & Big Data Developer** NTT Docomo Digital Spain Inc.

- Projects: Default prediction in mobile billing.
- Tasks: Production-ready code for all stages in the ML pipeline. Development of a proper validation schema for uncertain and delayed labeling in imbalanced classification. Production deployment, unit testing, scaling and performance tuning. Product definition. Interviewed junior roles.
- **Toolset:** Spark, Scala, Python, MongoDB, Jenkins, LDA, LightGBM.

Jun 2018 - Mar 2019 **Data Scientist & Big Data Developer** Stratio Big Data Inc.

- Projects: Default prediction in mobile billing. Customer churn prediction. Advertisement optimization in social media with reinforcement learning.
- **Toolset:** Spark, Scala, Python, MongoDB, Jenkins, XGBoost, Kibana.

Dec 2011 - Mar 2018 **Data Scientist and Research Consultant** University of Granada

- Projects:
 - **Indra Systems:** Workload prediction in cloud systems. Development of an extension for big data development in Eclipse IDE.
 - **ADIF and others:** Time-series forecasting of railway deficiencies. Multi-objective optimization for train infrastructures.
 - **LaPalma cooperative:** Forecasting of greenhouse climate conditions. Automatic code generation of analytic plots.
 - **CETAQUA foundation:** Time-series forecasting applied to water consumption in Spanish cities. Data analysis and modeling tasks.
- Toolset: R, ggplot2, forecast, tsExpKit, caret, tidyr, knitr.