

# Sergio Ramírez Gallego

## Machine Learning Engineer



Born in 1988



(34) 699 72 40 43



srg00017@gmail.com



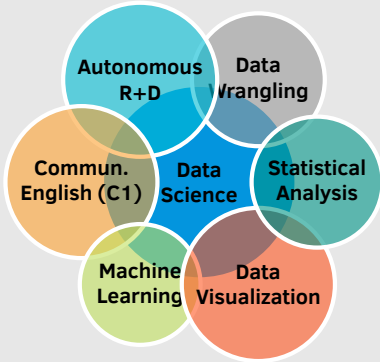
/in/sramirezg



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/>.

## Experience

Jan 2020 - **Machine Learning Engineer**

Pragsis Bidoop, Part of Accenture Applied Intelligence

Now

- Projects: Crowd behaviour analysis from drone view.
- Creation of several computer vision models with Deep Learning. Crowd segmentation with EfficientNet+UNet (CNN). Object detection with YOLOv3 and IOU tracking. Human pose estimation with AlphaPose. Dataset creation and annotation.
- **Stack:** OpenCV2, Keras, PyTorch, Tensorflow, Albumentations, Pillow, UNet, CVAT annotation tool.

Nov 2019 - **Lead Machine Learning Engineer**

Omdena

Now

- Project: Improving the aftermath management of an earthquake with AI.
- Deep Learning model (CNN) to automatically detect buildings footprints from satellite imagery. Data extraction from Mapbox API. Post-processing: creation of a risk heatmap with distance to dangerous buildings. Definition of the final solution.
- **[LINK]** Medium article.
- **Stack:** geopandas, geojson, GDAL, fast.ai, PyTorch, Solaris.

Mar 2019 - **Machine Learning Engineer**

NTT Docomo Digital Spain Inc.

Jan 2020

- Projects: Default prediction in mobile carrier billing.
- Tasks: Scala production-ready code for all stages in the ML pipeline. Data analysis and modeling with XGBoost. Development of a proper model validation schema for uncertain and delayed labeling in imbalanced classification. Production deployment, unit testing, scaling and performance tuning. Product definition in cooperation with the business team. Interviewed junior roles.
- **Stack:** Spark, Scala, Python, Cloudera, HDFS, MongoDB, Jenkins, LDA, XGBoost, LightGBM, Kibana.

Jun 2018 - **Data Scientist & Big Data Developer**

Stratio Big Data Inc.

Mar 2019

- Projects: Default prediction in mobile carrier billing with Spark and XGBoost. Customer churn prediction of millions of customers with PySpark (Databricks). Ads optimization in social media with reinforcement learning (Thompson sampling, OpenAI gym).
- **Stack:** Spark, Scala, Python, MongoDB, Jenkins, XGBoost, Kibana.

Dec 2011 - **Data Scientist and Research Consultant**

University of Granada

Mar 2018

- Projects:
  - **Indra Systems:** Workload prediction in cloud systems with Google-based data. 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:** Time-series forecasting of greenhouse climate conditions. Automatic code generation of analytic plots with ggplot2.
  - **CETAQUA foundation:** Time-series forecasting applied to water consumption in Spanish cities. Data analysis and modeling.
- **Stack:** R, ggplot2, forecast, tsExpKit, caret, tidyr, knitr.

# Sergio Ramírez Gallego

## Machine Learning Engineer



Born in 1988



(34) 699 72 40 43



srg00017@gmail.com



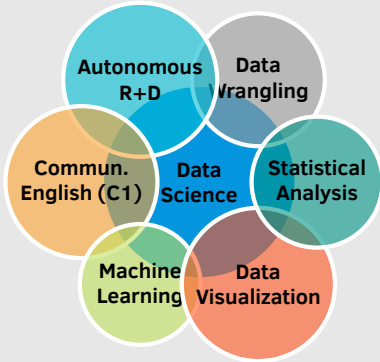
/in/sramirezg



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**

### Courses:

- Coursera Deep Learning Specialization (by deeplearning.ai): CNN, Sequence Models, Structuring ML projects, NNs, Improving Deep NNs.
- Natural Language Processing (by National Research University Higher School of Economics). Final project: Telegram chatbot (chitchat + answering from StackOverflow). Deployed in AWS EC2.
- How to Win a Data Science Competition: Learn from Top Kagglers (by National Research University Higher School of Economics).
- Open Data Science free-course (by ods.ai).

### Challenges and competitions:

- Omdena challenge: Improving the Aftermath Management of an Earthquake with AI.
- Kaggle competition: Understanding Clouds from Satellite Images, among others.

2013 - 2018 **Ph. D., Artificial Intelligence 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 Big Data with Apache Spark. Fast feature selection on 30M cols (kdd2010), and discretization on 60M rows.
- **Tools:** Spark, Scala, Java, MOA, Flink,  $\LaTeX$ , Git, Maven.
- **Internships:** Wroclaw University of Science and Technology, Poland & Mälardalen University, Västerås, Sweden. Duration: 3 months each.
- **Publications:** <https://scholar.google.com/citations?user=cEkfaW4AAAAJ>
- More than 1K cites in several papers published in top-tier international scientific journals.
- **Books:** Big Data Preprocessing.