

# Sergio Ramírez Gallego

## Data Scientist



Born in 1988 in Jaén, Spain



(34) 699 72 40 43



sramirez@decsai.ugr.es



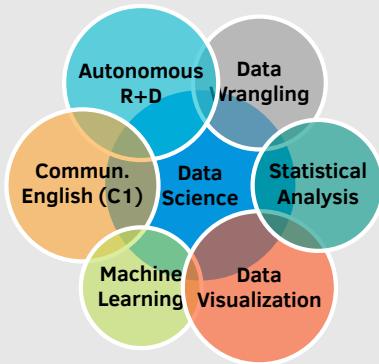
/in/sramirezg



sramirez

## Skills

### Overview



## Programming

Novice —————> Daily

Scala • Spark •  $\text{\LaTeX}$

Java • Git • Flink • R

C • C++ • Python • Maven

## Software

**Spark-IT-FS** - A distributed feature selection framework for Apache Spark | 82 ★

**fast-mRMR** - An optimized version of the mRMR feature selector for Spark and Nvidia-GPU | 34 ★

**Spark-MDLP** - A Spark implementation of the MDLP discretizer (presented in Spark Summit '17) | 29 ★

**MOAReduction** - A library for streaming data reduction in MOA | 2 ★

1<sup>st</sup> and 3<sup>rd</sup> projects are Spark-packages. Check my GitHub profile for more info.

## Education

- 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 Assistant** University of Granada  
**Thesis:** Distributed Data Reduction Algorithms for Big Data
- Proposed several scalable data reduction (discretization, feature and instance selection) techniques for Apache Spark. Aimed at decreasing the complexity of large-scale static and dynamic DBs.
  - Tools:** Spark, Scala, Java, MOA, Flink,  $\text{\LaTeX}$ , Git, Maven, Eclipse.
  - Internships:** PWR, Wrocław, Poland & MDH, Västerås, Sweden. 3 months each.
  - Summary of publications: <https://scholar.google.com/citations?user=cEkfaW4AAAAJ>

## Experience

- Sep 2014 - Feb 2018 **Researcher & Graduate Teaching Assistant** University of Granada
- FPU Spanish Research and Teaching Fellowship @ Department of Computer Science and Artificial Intelligence.
- Dec 2016 - May 2017 **Data Scientist** University of Granada & CETAQUA foundation
- Water-consumption forecasting in Spanish eastern cities. Data analysis, visualization and time-series forecasting. **Tools:** R, ggplot2, forecast, tsExpKit, knitr.
- Oct 2013 - Mar 2015 **Data Scientist** University of Granada & LaPalma Cooperative
- Crop forecasting and detection of plagues in greenhouses using weather conditions. Data analysis, visualization and time-series forecasting. **Tools:** R, ggplot2, forecast, tsExpKit, knitr.
- Apr 2013 - Dec 2013 **Data Scientist** University of Granada
- Time-series forecasting of railway deficiencies. Multi-objective optimization for railway infrastructure maintenance. In cooperation with VIAS Inc., ADIF, and others. **Tools:** R, Java, PMML, C++.
- Mar 2012 - Apr 2013 **Full Stack Developer & Data Scientist** Univ. of Granada & Indra Systems
- Development of a Eclipse plugin to code/deploy Java apps, and manage non-relational DBs in APLaCA, a cloud platform. Time-series forecasting of platform load. **Tools:** Java, Eclipse, R, HTML.
- Dec 2011 - Mar 2012 **Data Scientist** University of Jaen & OroLiveSur Inc.
- Extraction and analysis of user-patterns in an e-commerce site. Rule learning from user-patterns. **Tools:** Java, Weka, KEEL.
- Sep 2013 - Feb 2018 **Linux Server Administrator** SCI2S research group, UGR
- Tools:** CentOS, LDAP, SGE, Spark, Hadoop, Ganglia, KickStart, PXE.