Sergio Ramírez Gallego

Data Scientist

✎ Born in 1988 in Jaén, Spain

📞 (34) 699 72 40 43

✉ sramirez@decsai.ugr.es
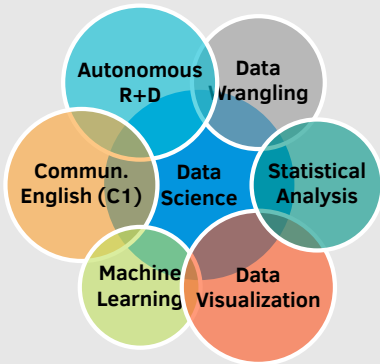
in /in/sramirezg

○ 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 - An optimized version of the mRMR feature selector for Spark and Nvidia-GPU | 57 ★

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

FeuerFreiKiller - A production-ready deep learning project to automatically detect fire in nature from spectogram images
1st and 3rd projects are Spark-packages. Check my GitHub profile for more info.

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 Assistant** University of Granada

Thesis: Distributed Data Reduction Algorithms for Big Data
- Development of scalable data reduction (discretization, feature and instance selection) algorithms for Spark. Achivements: 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. 3 months each.
- Publications: `https://scholar.google.com/citations?user=cEkfaW4AAAAJ`

Experience

March 2019 - Now	**Data Scientist & ML Engineer**	Docomo Digital Spain Inc.

- Default prediction in mobile billing. Focused on developing machine learning models, production deployment, unit testing, scaling and performance tuning. **Tools**: Spark, Scala, Python, MongoDB, Jenkins, LDA, LightGBM, Jupyter.
- Conducted interviews for junior data scientist roles.

Jun 2018 - March 2019	**Data Scientist & ML Engineer consultant**	Stratio Big Data Inc.

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

Sep 2014 - Feb 2018	**Researcher & Graduate Teaching Assistant**	University of Granada
Dec 2016 - May 2017	**Data Scientist**	University of Granada & CETAQUA foundation

- Time-series forecasting applied to water-consumption in Spanish cities. **Tools**: R, ggplot2, forecast, tsExpKit, caret, tidyr, knitr.

Oct 2013 - Mar 2015	**Data Scientist**	University of Granada & LaPalma Cooperative

- Time-series forecasting for production in greenhouses. **Tools**: R, ggplot2, forecast, tsExpKit, caret, tidyr, knitr.

Apr 2013 - Dec 2013	**Data Scientist**	University of Granada

- Time-series forecasting of railway deficiencies.

Mar 2012 - Apr 2013	**Full Stack Developer & Data Scientist**	Univ. of Granada & Indra Systems