

Moez BEN HAJHMIDA

1715 Spruce Street, Berkeley, CA 94709 | C: (510) 520-7215 | moez.benhajhmida@gmail.com

Summary

Data scientist with thorough hands-on experience in web scraping, data cleaning, statistics, machine learning techniques, and predictions. Experienced with high performance computing and large scale data analytics and mining in distributed and multicore systems. Supportive and enthusiastic team player dedicated to new creative ways of data collection and knowledge extraction. Willing to learn new technologies.

Skill Highlights

- Fluent in C/C++, Java
- Fluent in Python, Shell scripting
- Mastering Multi-threading
- Proficient with distributed systems
- Analytical thinking, planning
- Data analytics and mining
- Ant, Maven, Gradle
- Experienced Hadoop and Spark
- MySQL, MongoDB
- Well organized

Projects

- Prediction of mobile applications breakout** 2015-2016
 - Identified 1768 mobile applications by Web scraping using Scrapy and Python
 - 31 K posts and ~2 M comments extracted from Facebook API and stored in MongoDB
 - User sentiment derived from emoticons and bag of words
 - Features engineered and a dataset created
 - 50% of breakouts explained by users interactions and sentiments
- High performance grid computing in genome-wide association studies** 2009-2012
 - Framework that supporting distributed execution of the MDR method
 - Help biologists to automate multiple tests of gene-gene interactions
 - 2 pair of gene combinations distributed among all computation nodes
 - Efficiency (82%) and scalability (50K attributes) on 400 2-core computing nodes
- Weka Extension for distributed data mining** 2008-2010
 - Web Services Resource Framework technology
 - Distributed execution of the meta-learning methods
 - 2.84 speedup on 4 computing nodes on real datasets
 - Similar predictions performance compared to the C4.5

Experience

- Assistant Professor** Sep 2005 to Aug 2016
National Engineering School, ISSAT, and FST, Tunisia
 - Big Data, Virtualization, Parallel programming
 - Operating systems, Networking, Java programming
- Java/J2EE developer** Jan 2004 to Jul 2004
Marekha, Tunisia
 - Web based software for Network Element Information Management
 - Wrote java code based on Struts framework
 - Led the development team to plan, design, develop, and test robust solutions
- System administrator** Sep 2002 to Dec 2003
Faculty of Sciences of Tunis, Tunisia
 - IBM SP2 parallel machine system administration and network administration
 - Wrote and debugged parallel programs in MPI and OpenMP

Education

- Ph.D.**, Computational and Statistical Learning 2012
- Master of science**, Computational and Statistical Learning 2005
- Bachelor of Engineering**, Software Engineering 2002
Faculty of Sciences of Tunis, Tunisia

Publications

Journals

- Ben Haj Hmida, M. and Slimani, Y.: High performance grid computing for detecting gene-gene interactions in genome-wide association studies. *International Journal of Grid and Distributed Computing*, Volume 5, Number 3, pp. 33-44, March 2012.
- Ben Haj Hmida, M. and Slimani, Y.: Improving Classification Accuracy Using Code Migration. *International Journal of Advanced Science and Technology*, Volume 38, pp. 25-36, January 2012.
- Ben Haj Hmida, M. and Slimani, Y.: Grid-Enabled Framework for Large-Scale Analysis of Gene-Gene-Interactions. *Communications in Computer and Information Science*, June 2011, Volume 162, Part 3, pp. 348-357. Springer Berlin Heidelberg.
- Ben Haj Hmida, M. and Slimani, Y.: Meta-learning in grid-based data mining systems. *International Journal of Communication Networks and Distributed Systems*. Volume 5, Number 3, pp. 214-228, August 2010.

Refereed conference proceedings

- Ben Haj Hmida, M. and Slimani, Y., A scalable framework for large-scale analysis of gene-gene interactions, in *IEEE International Conference On Cluster Computing*, Madrid, 2014.
- Ben Haj Hmida, M. and Slimani, Y.: WSRF services for learning classifiers from data grid. *IEEE/ACS International Conference on Computer Systems and Applications*, Rabat, Morocco, May 2009, pp.27-32.

Book chapter

- Ben Haj Hmida, M. and Congiusta, A.: *Parallel, Distributed, and Grid-Based Data Mining: Algorithms, Systems, and Applications*. Handbook of Research on Computational Grid Technologies for Life Sciences, Biomedicine, and Healthcare, IGI Global, USA, pp. 90-119, May 2009.

Scientific activities

Software

- Grid Computing MDR: Grid-enabled version of the MDR software. The objective of this project is to make available an open-source of a gridified version of the Multifactor Dimensionality Reduction (MDR) software. <http://sourceforge.net/projects/gridmdr>.

Tutorials

- Tutorial on Parallel and Distributed Data Mining at 10eme Conférence Internationale Francophone sur l'Extraction et la Gestion des Connaissances (EGC) 2010.

Conference organization

- International Conference on Knowledge Discovery and Data Analysis (KDDA) 2015. Organizing committee member.
- International Workshop on Advanced Information Systems for Enterprises (IWAISE) 2014. Organizing committee chair.
- MPSoC Winter School on "Design, programming and Applications of Multi-Processor Systems on Chip" 2014. Funded by the German Academic Exchange Service (DAAD). Organizing committee member.
- International Conference on the Extraction and Management of Knowledge - Maghreb (EGC-M) 2012. Organizing committee member.

Program committee member

- International Conference on Knowledge Discovery and Data Analysis (KDDA) 2015.
- International Conference on the Extraction and Management of Knowledge-Maghreb (EGC-M) 2012.