

FABIO PETRONI, PhD

Researcher in Computer Science



📍 London, UK

✉️ fabio.petroni.1986@gmail.com

🌐 <http://www.fabiopetroni.com>

I'm an AI & Deep Learning Enthusiast. My broad areas of interest include Machine Learning, Natural Language Processing, Recommender Systems and Big Data. Currently, I am developing new deep learning and embedding models for Information Extraction in the corporate R&D department at Thomson Reuters. Before that, I have worked as Big Data Engineer at KPMG. I hold a Ph.D. in Engineering in Computer Science from Sapienza University of Rome. I also have a Master's in computer engineering summa cum laude and a Bachelor's degree from Sapienza University in Rome. I have been a Visiting Researcher at the University of Mannheim, and I have spent some time at Rijksuniversiteit Groningen during an Erasmus exchange. I'm an HDP Certified Apache Spark Developer. In my spare time, I'm currently looking at blockchain technologies.

PROFESSIONAL EXPERIENCE

Research Scientist

Thomson Reuters

📅 January 2017 – Present

📍 London



THOMSON
REUTERS

Machine Learning, Big Data, Natural Language Processing, Deep Neural Networks, Distributed Computation, Blockchain

TECHNOLOGIES:

- TensorFlow, scikit-learn, python
- Spark, Kafka, NLP4j, OpenNLP, scala
- Elasticsearch, AWS, GPU computation
- Smart Contracts

Big Data Engineer

KPMG

📅 April 2016 – December 2016

📍 London



Machine Learning tasks on Big Data frameworks (Hadoop and Spark)

- Extensive use of Hadoop and Spark (MLlib, GraphX, SQL, Streaming)
- Build a proprietary ChatBot and the relative knowledge base (graph)
- Natural Language Processing - entity extraction, sentiment analysis
- Dimensionality reduction techniques (PCA) for customer segmentation
- Refactor R/ Python scripts into parallel Spark jobs - 1000X faster
- Scale up graph analytics algorithms on large graphs (million of nodes)
- Store and index data on Elasticsearch, Hive and Accumulo
- Geospatial analysis on news stories with Spark + GeoMesa + Accumulo

TECHNOLOGIES:

- **Programming Languages:** Scala (60%), Java (30%), Python (10%)
- **Big Data Frameworks:** Apache Spark, Hadoop, Pig
- **Machine Learning tools:** TensorFlow, MLlib, H2O, Deeplearning4j
- **Databases:** Hive, Accumulo, MongoDB, Elasticsearch
- **Other:** GeoMesa, GeoServer, Apache Kafka, AWS

CERTIFICATIONS



HORTONWORKS
HDP CERTIFIED
APACHE SPARK
DEVELOPER

CORE SKILLS

Machine Learning	Artificial Intelligence	
Natural Language Processing	Big Data	
Recommender Systems	Data Science	
Algorithms	Architectures	Data Mining
Social Media Analysis	Graph Analytics	
Distributed Systems	Computer Science	

CODING SKILLS

Programming Languages: Scala, Java, Python, C, C++, Bash, etc
Big Data Frameworks: Apache Spark, Hadoop, Spark, GraphLab
Machine Learning tools: TensorFlow, MLlib, H2O, Deeplearning4j, Mahout, scikit-learn, nltk, libFM, NLP4j, etc
Databases: Hive, Accumulo, MongoDB, Elasticsearch, PostgreSQL, mysql, etc
Other: Knowledge of algorithms and data structures, cost analysis, multithreading, unit tests, stream processing, AWS.

OPEN SOURCE PROJECTS:



**libFM:
Factorization
Machine Library**

<https://github.com/fabiopetroni>

HONORS AND AWARDS



RESEARCH GRANT
Natural Language Processing 2015



FINALIST
Telecom Big Data Challenge 2014



OUTSTANDING GRADUATE STUDENT AWARD
Sapienza University of Rome 2012



PATH OF EXCELLENCE
Sapienza University of Rome 2011

SELECTED PUBLICATIONS

M. D. Vicario, A. Bessi, F. Zollo, F. Petroni, A. Scala, G. Caldarelli, H. Stanley, W. Quattrociocchi:

The spreading of misinformation online.

In: Proceedings of the National Academy of Sciences (PNAS), 2016.

F. Petroni, L. Querzoni, K. Daudjee, S. Kamali and G. Iacoboni:

HDRF: Stream-Based Partitioning for Power-Law Graphs.

In: Proceedings of the 24th ACM International Conference on Conference on Information and Knowledge Management (CIKM), 2015.

F. Petroni, L. Del Corro and R. Gemulla:

CORE: Context-Aware Open Relation Extraction with Factorization Machines.

In: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2015.

F. Petroni and L. Querzoni:

GASGD: Stochastic Gradient Descent for Distributed Asynchronous Matrix Completion via Graph Partitioning.

In: Proceedings of the 8th ACM Conference on Recommender Systems (RecSys), 2014.

For a complete list of publications see <http://www.fabiopetroni.com/publications>

EDUCATION

Ph.D. in Engineering of Computer Science

SAPIENZA UNIVERSITY OF ROME



November, 2012 - February, 2016

Rome, Italy

- **Topics:** Big Data, Machine Learning, Data Mining
- **Thesis:** "Mining at scale with latent factor models for matrix completion"

MAIN RESULTS:

- Proposed a graph partitioning algorithm that provides the smallest vertex replication with optimal load balances (currently the best performing graph partitioning algorithm)
- Improved the scalability of matrix factorization algorithms for recommender systems (up to 100x faster when running on a cluster)
- Proposed a knowledge-base completion model for nlp that incorporates generic contextual information to improve prediction performance
- Analyzed the spreading of misinformation in online social networks

EDUCATION

MSc in Engineering of Computer Science

SAPIENZA UNIVERSITY OF ROME

October, 2010 - October, 2012

Rome, Italy

- **Main Topic:** Architectures and Distributed Systems
- **Final Mark:** 110/110 summa cum laude (first-class honours)

MAIN SUBJECTS:

Distributed Systems, Database, Operating Systems, Big Data, Computer Networks, Software Engineering, Artificial Intelligence, Computer and Networks Security, Machine Learning, Capacity Planning and others.

BSc in Engineering of Computer Science

SAPIENZA UNIVERSITY OF ROME

October, 2005 - July, 2010

Rome, Italy

MAIN SUBJECTS:

Mathematics, Statistics and Probability, Computer Science, Operations Research and others.

ACADEMIC EXPERIENCE

Research Scientist

UNIVERSITY OF MANNHEIM

December, 2014 - May, 2015

Mannheim, Germany

Studied Machine Learning techniques based on factorization machines for relation extraction (knowledge-base completion), a natural language processing task that aims at predicting new real facts of the world.

Erasmus Program

RIJKSUNIVERSITEIT GRONINGEN

February, 2008 - September, 2008

Groningen, The Netherlands

LANGUAGES

Italian ●●●●●

English ●●●●●

REFERENCES

References available upon request.