Simone Robutti

Personal Information

Place and Date of Birth:	Italy | 2 December 1988
Address:	Keibelstr. 37, 10178, Berlin
Phone:	(Ask Privately)
email:	simone.robutti@gmail.com

Keywords

Scala
Apache Flink
Apache Spark
Python
Machine Learning
ML Engineering
Big Data/Fast-Data

Education

2013-2015 — Master's Degree in Computer Science, **Università degli Studi di Milano**, Milan

Research Thesis on the design, development and evaluation of a Support Vector approach to the disambiguation of fuzzy and shadowed datasets in a semi-supervised context.

Advisor: Prof. Dario Malchiodi

2009-2012 — Bachelor Degree in Computer Science, **Università degli Studi di Milano**, Milan

Thesis on the development of a highly-modular, resource agnostic tagging system for a CMS platform oriented to e-Democracy and e-participation.

Advisor: Prof. Fiorella DeCindio

Work Experience

from Nov 2016 — Data Engineer in MotionLogic, Berlin

Development and testing of the company's core product: a system to ingest and process cellular network signals to extract complex informations through ad-hoc models and traditional aggregations. My role as a Data Engineer covered many different tasks over time: productization of existing algorithms developed by the Data Science Team, development of new models and heuristics on my own, development and maintenance of the pipeline, development of a data quality monitoring system, production and improvement of geospatial support data, performance optimization at scale.

from Feb 2016 — Machine Learning Engineer in Databiz - Radicalbit, Milan

As the only Machine Learning Engineer, my work involved designing new solutions for our customers to make use of Flink, core component of our distribution. Also I've been involved in the open-source development of FlinkML and FlinkJPMML. I had the possibility to implement many different algorithms from scratch and experiment with the many possibilities offered by the framework to carry out ML tasks at scale. Other relevant components of the distribution were: Cassandra, Kafka, Zeppelin, Alluxio.

from Apr 2015 — Machine Learning Engineer in Databiz, Milan

Design and implementation of a near-real-time scoring streaming system. The end-goal was to provide the user with a flexible system based on Apache Spark to design ML pipelines for data preparation and batch learning on Knime and deploy the models created for scoring in streaming. The project involved a model repository with a front-end component, a data validation and reconciliation process. Other tools employed in the project were: HBase, JPMML, Hive, Impala.

Skills

Programming Languages	Python	Advanced
	Scala	Intermediate
	Java	Intermediate
	Javascript	Basic
Frameworks and Tools	Apache Spark	Intermediate-Advanced
	Apache Flink	Intermediate
	Pandas	Intermediate
	Play! Framework	Intermediate
	Keras	Basic
	Akka	Basic
	breeze	Basic
Machine Learning	SVM	Intermediate
	Deep Learning	Basic
	Genetic Algorithms	Basic
	ALS	Basic
	PMML	Intermediate
Database	PostgreSQL	Intermediate
	MongoDB	Intermediate
	Cassandra/Scylla	Basic
	MariaDB/MySQL	Basic
	InfluxDB	Basic
Building systems	Maven	
	SBT	
	setuptools	
Non-technical skills	Online Communication	
	Web-Oriented Writing	
	Social Media Management	
	Community Management	

Interests and Activities

Programming, History and Philoshopy of Science, Mathematics, Open Source

History, Communication of Science, Geopolitics, Philosophy of the Mind

Internet Culture, Martial Arts, Videogaming and Videoludic Arts

Languages

English:	Fluent
Italian:	Native Speaker