


## Summary

I am currently a data scientist at S&P Global. I analyze data to drive optimization and improvement of product development and business strategies. In the company, I conduct R&D projects to resolve challenging problems in business. I am enthusiastic about state of the art in computer algorithms, particularly machine learning methods. I hold a master's degree in applied mathematics and a doctorate in computer science. See my LinkedIn  for more information.

## Specialities

Machine Learning, Time Series Analysis, Natural Language Processing, Optimization, Numerical Computing

## Skills & Expertise

Programming	Python, C/C++, Scala
Libraries	Scikit-Learn, Tensorflow, Apache Spark, SuiteSparse, OpenGL, QT
Operating Systems	Linux, macOS, Windows
Others	Amazon AWS, Git, Docker, Terraform, Jira, Trello

## Experience

- 2018 - Present **Data Scientist Director**, *S&P Global*, United Kingdom  
As a data scientist, I design and implement data-driven solutions employing cutting edge machine learning techniques to resolve real problems from business lines. The daily work is project-based, and the topics have a wide range, for instance, time series analysis, natural language processing, etc. Besides the hands-on research and development, I conduct project management, typically leading 3-4 people. I am also responsible for crafting the data science blueprint for the team, e.g., tools and best practices.
- 2016 - 2018 **Data Scientist**, *MediaGamma Ltd*, United Kingdom  
I analyse terabytes of data and develop applications for programmatic advertising business by using machine learning and distributed computing. The role is responsible to design and build automatic solutions that address large-scale machine learning problems in digital advertising including click-through rate, customer segmentation, budget allocation, etc.
- 2013 - 2016 **Research associate**, *University College London*, United Kingdom  
I am a contributor to the IQmulus project funded by European Commission. The project aims to create a high-volume fusion and analysis platform for large geospatial datasets. I am developing applications to process large LiDAR datasets using the distributed computing engine Apache Spark.

## Education

- 2010 - 2013 **Ph.D.**, Computer Science, Institut National Polytechnique de Lorraine / INRIA
- 2008 - 2010 **M.S.**, Applied Mathematics, Zhejiang University  
**Excellent Graduation Thesis in Zhejiang University**
- 2004 - 2008 **B.S.**, Mathematics, Zhejiang University  
**Excellent Graduation Thesis in Zhejiang University**

## Languages

English	Professional working proficiency
Mandarin Chinese	Native