



Curriculum vitae

PERSONAL INFORMATION

Simone Di Gregorio



WORK EXPERIENCE

August 2022 – September 2022

Data Science Intern

KNIME GmbH, Konstanz, Germany

KNIME GmbH is the German branch of KNIME, the data science software company developing the homonymous open source analytics platform, which enables data manipulation/handling and machine learning applications with directed graphs/workflows built through a graphical interface.

Main responsibilities were:

- Development of a KNIME component for Word2Vec and of a KNIME workflow showing a possible use case for the resulting embeddings.
- Development of a KNIME node (Python based) implementing a similar pipeline, but far more focused on performance and low-level Tensorflow.
- Writing blog articles describing theoretical and practical aspects of the above implementations.
- Collaboration with the internal Python development team in testing and perfecting software extensions for user development.

September 2023 - March 2024

Teaching Assistant/Tutor

Sapienza University of Rome

Statistical Methods for Data Science & Laboratory I

Statistical Methods for Data Science and Laboratory is one of the main courses, spanning two semesters, in the Data Science Master's Degree, dealing with an introduction to Probability Theory before delving into Frequentist and Bayesian Inference. I won a grant (Bando n.5 2023 - I3S faculty) for tutoring in the first part of the course (9 CFUs out of 12).

This was a part time job where the main activities were:

- Holding lectures/office hours once a week to explain specific topics/revise already covered ones and tackle exercises; topics spanned from basic probability theory (non-analytic) to frequentist inference with practical examples and simulations using the R language.
- Grading homeworks (there are three in the first part of the course): each group of students was required to submit a markdown report with R code, exploring a specific (possibly advanced) topic or procedure related to what they covered in class (e.g. confidence sequences, correlation graphs with appropriate testing procedures...).
- Helping out with exams.

EDUCATION AND TRAINING

2019–2022

Bachelor's Degree in Management and Computer Science

LUISS Guido Carli University

Mathematical Analysis, Multivariate Calculus, Linear Algebra, Graph Theory, Statistics, Statistical/Machine Learning, Micro/Macroeconomics, Python/R programming, Theory of Algorithms and Relational Database Design/SQL.

Thesis (statistical learning topics):

Movie box office analysis and inference with time series models and tree-based methods;

Supervisor: Dr. Francesco lafrate

Grade: 110/110

Date of graduation: 02/11/2022

2022–2024 Master's Degree in Data Science

Sapienza Università di Roma

Frequentist and Bayesian Inference, Probability Theory, Stochastic Processes, Convex Optimization for machine learning applications, Signal Processing, Statistical Learning, Theory of Algorithms, Deep Learning, Computer Vision, Telecommunications Networks.

Exams:

- **18/07/2024** - Digital Epidemiology and Precision Medicine - 30/30
- **10/07/2024** - Signal Processing for Big Data - 30/30 cum laude
- **03/06/2024** - Data Management for Data Science - 30/30
- **12/04/2024** - Neural Networks for Data Science Applications - 30 cum laude
- **31/01/2024** - Advanced Machine Learning - 30/30 cum laude
- **18/01/2024** - Stochastic Processes [one of the main courses from the *Statistical Methods and Applications* Master's Degree] - 30/30 cum laude
- **14/09/2023** - Statistical Learning - 30/30 cum laude
- **21/07/2023** - Economics of Network Industries - 27/30
- **13/07/2023** - Networking for Big Data and Laboratory - 30/30 cum laude
- **10/07/2023** - Optimization Methods for Data Science - 30
- **28/06/2023** - Statistical Methods in Data Science and Laboratory - 30/30 cum laude
- **23/01/2023** - Algorithmic Methods of Data Mining and Laboratory - 30/30 cum laude
- **19/01/2023** - Fundamentals of Data Science and Laboratory - 30/30 cum laude

Thesis (mathematical statistics topics):

Estimation of the coefficients of a vector stochastic differential equation: from classical parametric methods to neural network estimators; **Supervisors:** Prof. Alessandro De Gregorio, Dr. Francesco Iafrate and Prof. Pierpaolo Brutti

Honours Programme: Collaboration with Prof. Piccialli and the French Institute for Research in Computer Science and Automation on data augmentation and machine learning frameworks for a follow-up publication to *Data augmentation driven by optimization for membrane separation process synthesis* [2023].

Grade: 110/110 with honors

Date of graduation: 28/10/2024

2024–Currently PhD in Data Science

Sapienza Università di Roma

PhD in Data Science under the internal supervision of Professor Stefano Leonardi.

PERSONAL SKILLS

Language Italian (mother tongue) and English.

Language Certification English C2 CEFR certificated by C1 Advanced Cambridge Exam in June 2018/Score: 201/210

- Computer skills
- **Python and R programming for data science** tasks and projects, specifically for what concerns data manipulation, data visualization and training and performance evaluation of all traditional machine learning models and techniques. For what concerns Python, strong confidence in using scikit-learn and the whole data science stack (NumPy, Pandas, SciPy, etc.), relatively in-depth knowledge of Tensorflow (and of course of Keras) and general knowledge of vanilla PyTorch. For what concerns R, fluency is both with native syntax and tidyverse-based syntax.
 - **KNIME proficiency** across all the main levels of certifications (L1, L2, L3), both for local use cases and production, with additional knowledge concerning the software due to employment at KNIME GmbH in Konstanz, Germany for two months.
 - General knowledge of SQL and relational database management systems.

- Additional expertise, experience and/or skills
- Quite in depth knowledge of **advanced and rigorous probability theory and real analysis**, specifically for stochastic processes. This is due to self-studying carried out during the Master's Degree, in parallel to exams and courses. The Stochastic Processes exam from the *Statistical Methods and Applications* Master's Degree is to be framed in this context. This kind of knowledge gave the foundations for my thesis in mathematical statistics topics and is related to the topics I am planning to focus my PhD research on.
 - Years of experience in entertainment journalism, having worked for important Italian web magazines, handling reviews, previews and videos/streams related to important releases. Care for **writing style, content exposition and content layout** stems from this professional experience.