Claudio **Scalzo**

Data Science and Software Engineering student

(+39) 346 24 55 116 • claudio.scalzo@outlook.com • github.com/claudioscalzo • linkedin.com/in/claudioscalzo

Born in Italy in 1994. Passionate about data science and software development. Strong analytical thinking and team-working skills. Successful academic results due to the passion for these topics and the high precision in on-time fulfillment of projects, tasks and assignments.

Education

Master's Degree in Data Science and Engineering — *Sophia Antipolis, France*

Eurecom / Telecom ParisTech — *Sep. 2017 - Mar. 2019*

- Double Degree programme between Telecom ParisTech and Politecnico di Torino

Master's Degree in Software Engineering — *Turin, Italy*

Politecnico di Torino — *Sep. 2016 - Mar. 2019*

Bachelor's Degree in Software Engineering — *Turin, Italy*

Politecnico di Torino — *Sep. 2013 - Jul. 2016*

- with the highest grade, 110/110

Experience

SAP France — *Paris, France*

Data Scientist / Software Engineer intern — *Jul. 2018 - Dec. 2018*

- Provisioned...
- Built...
- Implemented...

Projects

Team leader for an optimization project for TIM / SWARM Joint Open Lab

github.com/claudioscalzo/coiote-oma-project

- Solving of a *VRP (Vehicle Routing Problem)* optimization problem proposed by *TIM* and the *SWARM Joint Open Lab*, for an IOT project named *CoIoTe*. Multistart tabu-search approach, written in Java (with the *OpenTS* Java library).
- Achieved first position in the final ranking. Secured great comprehension of metaheuristics and, overall, improved logical/algorithmic competences and team-working skills.

Virtual Assistant for answering music related questions

github.com/d2klab/music-chatbot

- Developing of a virtual assistant (in the chatbot and vocal assistant forms), capable of answering music related questions and providing detailed graphical results. Informations extracted from the *DOREMUS* knowledge base, queried using the *SPARQL* language.
- Built using *Node.js* code with the *BotKit* framework. Used and trained Google's *Dialogflow* as NLP. Interfaced with the *Facebook Messanger* and *Slack* clients, using the respective APIs.
- Did a pull request (accepted and merged) to the `botkit-middleware-dialogflow` author, introducing language support in the middleware.

House prices *Kaggle* challenge: predicting sales prices with advanced regression techniques

github.com/claudioscalzo/aml/tree/master/challenge

- Solution of one of the most famous Kaggle challenges, developed during the *AML (Algorithmic Machine Learning)* course at *EURECOM*.
- Top grade on the course ranking, thanks to smart preprocessing techniques (like *PCA* and *DBSCAN* for outlier removal), and thanks to advanced stacked regressors models.
- Written in *Python*, using *Pandas DataFrames* for the data structures and *scikit-learn* for the modeling phase.

Challenge on the `Fashion-MNIST` and `CIFAR-10` datasets: Naive Bayes Classifier and Bayesian Linear Regression

github.com/claudioscalzo/asi-challenge

- Solution of the *ASI (Advanced Statistical Inference)* course. Implemented (from scratch) the Naive Bayes Classifier and the Bayesian Linear Regression. Used in the classification tasks of the Fashion-MNIST and CIFAR-10 images datasets.
- Written in *Python* using *NumPy* and *Pandas*. Achieved extremely satisfying results in terms of accuracy and computational efficiency.

Skills

Programming	Python (+ NumPy, Pandas, scikit-learn, Keras, TensorFlow, PySpark, Spark MLlib) • C • Java • SQL
Big Data	Apache Spark • MapReduce & HDFS • NoSQL Architectures • Data processing (cleansing, analysis, mining)
Machine Learning	Deep Learning (NN, CNN, RNN) • Probabilistic Machine Learning (Bayesian Classification, Regression, Mixture Models)
OS & Tools	GNU/Linux (+ Bash, AWK, Sed) • Git • Jupyter Notebooks • Dialogflow (+BotKit) • MATLAB • LaTeX