Claudio **Scalzo**

Software Engineering and Data Science student

☐ (+39) 346 24 55 116 • ✉ claudio.scalzo@outlook.com • ⌂ github.com/claudioscalzo • in 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*

• with Master Thesis on: *"Leveraging Machine Learning driven Data Matching techniques for an Azure Data Lake"*

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

• 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 1st position in the final ranking. Secured great comprehension of metaheuristics. Improved algorithmic and team-working skills.

Virtual Assistant for answering music related questions

GITHUB.COM/D2KLAB/MUSIC-CHATBOT • CHATBOT.DOREMUS.ORG

• 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 Messenger*, *Slack* and *Google Assistant* clients, using the respective APIs.
• Did two pull requests (accepted and merged) to the `botkit-middleware-dialogflow` author, for concurrency and language support.

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

GITHUB.COM/LOMLUCA/AML

• 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 stacked 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

Languages	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)
...	Internship Skills...
Machine Learning	Deep Learning (NN, CNN, RNN) • Probabilistic Machine Learning (Bayesian Classification, Regression, Mixture Models)
OS & Tools	GNU/Linux (+ Bash, AWK, Sed) • Git • Jupyter Notebook • Dialogflow (+BotKit) • MATLAB • LaTeX