Felipe Operti

Turin, Italy
operti.felipe@gmail.com • github.com/opertifelipe
linkedin.com/in/felipeoperti

Passionate Data Scientist, an expert in statistical analysis, machine learning (supervised algorithms such as linear, logistic regressions, support vector machines, random forest, gradient boosting among others, and unsupervised algorithms such as k-means, DBSCAN, hierarchical clustering, etc.), and deep learning, especially in Natural Language Processing (NLP) (Named Entity Recognition, Relation Extraction, transformers, language models, classifications, etc.). Deep knowledge of churn models, pricing models, and time series analysis. High level of Python, Scala, Spark, SQL, and Bash. Main libraries: Pandas, Sci-Kit Learn, Numpy, PyTorch, Spacy, HuggingFace, Plotly, Matplotlib, Seaborn, FastAPI, etc. Experience leading end-to-end projects with teams composed of data engineers, data scientists, and machine learning engineers. Furthermore, knowledge and experience using big data frameworks such as Spark, Pyspark, Hive, Impala, Cloudera, Databricks, etc. Dual-citizenship: Italian and Brazilian. Three published papers.

EMPLOYMENT

Data Science Specialist — Turin, Italy
Reale Mutua — From Jul 2022

As a Data Scientist Specialist at Reale Mutua, my primary focus is on leveraging cutting-edge technologies such as Natural Language Processing (NLP) and Deep Learning models to extract insights from complex data sets. I work extensively on Azure cloud infrastructure, utilizing its powerful capabilities to design, develop, and deploy advanced data science solutions that drive business outcomes.
Technologies: Azure, AzureML, Azure Cognitive Services, Python, Docker, Pytorch, HuggingFace, Spacy, etc.

Data Scientist | Machine Learning Engineer — Turin, Italy
Inda - Intervieweb S.r.l. - Zucchetti Group — Mar 2021 - Jul 2022

Data scientist at INDA (Intervieweb S.r.l. solution, a Zucchetti company) working principally with Natural Language Processing (NLP) and Deep Learning. Responsible for the development, deployment, and maintenance of machine learning solutions.
Technologies: AWS EKS, AWS ECS, AWS Lambda, AWS Sagemaker, Python, Docker, Kubernetes, FastAPI, ElasticSearch, MongoDB, etc.

Data Scientist Specialist — São Paulo, São Paulo, Brazil
Samsung SDS — Nov 2020 - Mar 2021

Data Scientist Specialist in CRM area at Samsung SDS. Responsible for the development and deployment of propensity models, churn models, sentimental analysis, and RFV models.
Technologies: Data Lake, HDFS, Brightics, Adobe Campaign, Python, Scala, Docker, Spark, and PySpark.

Data Scientist Specialist — São Paulo, São Paulo, Brazil
ViaVarejo — Jul 2020 - Oct 2020

Responsible for the development of machine learning models for price optimization and retail assortment.
Technologies: Data Lake, Cloud, DataBricks, Azure, Python, Scala, Jenkins, Docker, Spark, and PySpark.

Data Scientist Consultant — São Paulo, São Paulo, Brazil
Santander Bank — Aug 2019 - Jul 2020

Data Scientist in Santander DataLab. Data Lab is an innovative area whose aim is to develop large cross-areas end-to-

end projects. Data Lab works as a consultancy society within the bank. In the lab, the interaction between product owners, data engineers, data scientists, and machine learning engineers is strongly encouraged. My responsibilities there are: from one side to develop machine learning models (as data scientist consultant). Due to the heterogeneity of the projects, also the applied machine learning and deep learning models are heterogeneous. On the other side, I am also responsible for checking the entire pipeline (from the ETL, the development of the models, and finally the deployment).
Technologies: Data Lake, Hadoop, Cloudera, Python, Scala, SQL, Hive, Impala, Spark, PySpark, and Oracle.

Senior Data Scientist — São Paulo, São Paulo, Brazil
Santander Bank — Dec 2018 - Aug 2019

Data scientist at Santander Bank in Customer Relationship Management (CRM). My position there was of senior data scientist and my responsibility was the development of machine learning models for different purposes (propensity and churn models for different products among others). Such models were typically machine learning supervised algorithms such as Random Forest, Gradient Boosting, Logistic Regression, and Artificial Neural Networks as well as unsupervised algorithms such as hierarchical, k-means, and DBSCAN. Furthermore, we worked with NLP algorithms for sentimental analysis of reclamations data.

Technologies: Data Lake, Hadoop, Cloudera, python, R, Scala, SQL, Java, ShellScript, Hive, Impala, Spark, PySpark, and Oracle.

EDUCATION

PhD in Physics, Complex Systems — Fortaleza, Brazil
Federal University of Ceará — Feb 2015 - Dec 2018
Thesis: Computational analysis for socio-economic sciences.

Visiting PhD in Physics, Complex Systems — New York City, United States
City College of New York — Apr 2018 – Jun 2018

Visiting PhD in Physics, Complex Systems — Rome, Italy
University La Sapienza — Apr 2017 – Sep 2017

MSc in Physics, Complex Systems — Fortaleza, Brazil
Federal University of Ceará — Jul 2014 – Jan 2015
Thesis: Interpolation strategy based on Dynamic Time Warping.

BSc in Physics — Turin, Italy
University of Turin — Sep 2009 – Oct 2012
Thesis: Monte Carlo Simulation of the radiation distribution emitted by a CT scan in the field of the radioprotection.

SKILLS

Machine Learning and Deep Learning: Scikit-Learn, Pytorch, Keras, Spacy, HuggingFace, Gensim, NLTK, BERT, OpenAI, Azure Cognitive Services, etc.

MLOps: FastAPI, Kubernetes, Docker, GitLab CI/CD, Typer, Pydentic, Poetry, etc.

Cloud: Azure, AzureML, Azure DevOps, AWS, AWS ECS, AWS EKS, AWS ElasticBeanstalk, AWS Sagemaker, AWS RDS, AWS Aurora, etc.

Data Analysis: Python (Jupyter Notebook, Pandas, NumPy, StatsModels, etc), Scala, SQL, Spark, PySpark, Hadoop, Hive, Impala, Excel, etc.

Data Visualization: Matplotlib, Seaborn, Plotly, Streamlit, Grafana, Gephi, QGis, etc.

Databases: MySQL, PostgreSQL, HDFS, ElasticSearch, and MongoDB.

Other computer skills: Java, C++, and Bash.

Operating systems: Linux (Arch Linux, Debian, and Ubuntu), MacOS, and Windows.

Languages: Italian (native), Portuguese (fluent), and English (fluent).

PUBLICATIONS

- Felipe G. Operti et al. Dynamics of racial segregation and gentrification in New York City. Frontiers in Physics. 2022.
- Felipe G. Operti et al. Dynamics in the Fitness-Income plane: Brazilian states vs World countries. PlosOne. 2018.
- Felipe G.Operti et al. The light pollution as a surrogate for urban population of the US cities. Physica A. 2018.