Bruno Stefoni

+1 646 203 7124 | brunostefoni12@gmail.com | New York | LinkedIn | GitHub | Scholar

Professional Summary

Five years of experience in Data Science. Building data pipelines, delivering predictive models, designing metrics and hypothesis tests to present insights and participate in data-driven business decisions.

EDUCATION

NEW YORK UNIVERSITY — New York City, NY
Master of Science in Data Science (GPA: 3.8/4.0) — Sep 2021 – May 2023

- Coursework: **Deep Learning, Machine Learning, Big Data**, Optimization & Lineal Algebra, Probability & Statistics.
- **TA of Data Science** Bootcamp Class at NYU Business school during two semesters.
- **Student Representative** for the entire Master's Program, meeting faculty and communicating student issues.

UNIVERSIDAD DE CHILE — Santiago, Chile
Master in Engineering Sciences (Final grade 6.6/7.0) — Jan 2016 – Aug 2018

- **TA of five different courses**, e.g., Stochastic Processes, Network Optimization, Networks and Circulation.
- Graduated with highest distinction.

UNIVERSIDAD DE CHILE
Engineer Degree (Final grade 6.6/7.0) — Mar 2010 – Dec 2015

- **Graduated in the top 10%** of my class cohort. Granted **Outstanding Student award** in 2010.
- Key Modules: **Computer Science I and II, Probability, Statistics, Optimization, Operations Research, Data Mining.**

PROJECTS AND PUBLICATIONS

- Meera Health: fine-tuning Lage Language Model **Llama 2** from clinical trial protocols for a client facing Chatbot.
- **S&P Global** Capstone Project: News classification by fine tuning **Large Language Models (BERT) from unlabeled data** using Python and Hugging Face. Presented final results (98% accuracy) and a live demo to stakeholders.
- **Published two papers** based on my Master in Engineering thesis related to computer simulation: **GScholar link.**
- Deep Learning Competition: next frame video prediction using a **Transformer and CNN (Resnet)** in a JEPA network.
- Text topic analysis over time of a large academic thesis corpus using **LDA** and GloVe word embeddings.
- Twitter sentiment and classification analysis using fastText and f**ine-tuning BERT** to assess best content strategy.
- Won Hackathon competition. Led an interdisciplinary group of 6. Designed and developed an app for transit users.

EXPERIENCE

❖ **Sortile** — Manhattan, New York
Contractor | Sortile's purpose is to provide sorting infrastructure using ML to identify fibers. — Jun 2023 – Aug 2023
Data Scientist

- Built a 92% accurate XGBoost classifier to identify black/white fibers using light wavelength sensor data.
- Developed and benchmarked XGBoost, LGBM and Deep Learning models. Deployed XGB model on AWS Lambda.

❖ **Google** — Mountain View, California
Internship | Platforms & Ecosystems - Play Store — Jun 2022 – Aug 2022
Data Scientist

- Developed clear understanding of App Developers, conducted a detailed market research and graph analysis by building a network dataset from diverse data sources using PLX, **Google BigQuery, Python and R data pipeline.**
- **Detected up to 15% of user churn** by continuous survival and time series analysis of Software adoption rate data.
- Presented insights to several internal stakeholders and management. Proposed a new methodology for analysts.

❖ **STP Santiago** — Santiago, Chile
FTE | STP is a transit company in Chile's capital operating 980 buses, 250 of which are electric. — May 2018 – May 2022
Data Scientist

- **Built and implemented a company-wide chat bot** in a Linux private server to support management and on-site workers to obtain real-time updates on fleet status and manage electric charging stations efficiently.
- Drove STP's revenue and established strong KPIs by identifying routes with a difference of up to 25% between planned and real bus speeds, **automating A/B tests** in data and generating custom alerts.
- Prevented $500K investment on tech less accurate than traditional methods by **launching a statistical study**.
- Build **machine learning model for battery and fuel consumption** to predict future costs for contract negotiations.
- **Detected data outliers** due to faulty batteries in 20% of STP's electric fleet **via a statistical analysis using Python.**
- Implemented daily monitoring of **six different KPIs** via automation of data ingestion and processing, these KPIs determined subsidy amounts according to the transit concession's contract.
- Increased 6% the main contract KPI by scraping twitter data to alert service disruptions using **text processing**.

❖ **Telefonica R&D and Universidad de Chile** Santiago, Chile

FTE | Research project with Telefonica, one of the largest telecom companies in Latin America. Nov 2017 – May 2018

Data Analyst

- Designed a new public transit route service for forecasted utilization by 60K people weekly by utilizing Machine Learning and **large cell phone datasets**.
- Created framework of extraction and **visualization tools from spatial data** by leveraging Python and R from 9M+ spatial data points and street network data.

SKILLS

- Advanced Python for Data Analysis, Machine Learning and Deep Learning:
  - Pandas, NumPy | PyTorch, PyTorch-Lightining | XGBoost, LGBM, scikit-learn | Plotly, Seaborn, Matplotlib
  - NLTK, SpaCy, fastText, GloVe | Transformers, Hugging Face | LangChain, openAI API
- SQL | R | MATLAB | QGis | C++ | Java | Bash | Git
- Linux | WandB | Docker | Snowflake, Dask, pySpark
- AWS (S3, EC2, Lambda) | Microsoft Office Suite (Excel, PowerBI)
- Native Spanish speaker; fluent in English; intermediate Italian.