

# Bruno Stefoni

+1 646 203 7124 | brunostefoni12@gmail.com | New York | [LinkedIn](#) | [GitHub](#) | [Scholar](#)

## Professional Summary

Experience in Deep Learning, NLP, Data Science, and storytelling. Creating data pipelines, designing metrics, delivering predictive models, and performing hypothesis tests to present insights and participate in data-driven business decisions.

## EDUCATION

### NEW YORK UNIVERSITY

New York City

Master of Science in Data Science (GPA: 3.8/4.0)

Sep 2021 – May 8<sup>th</sup>, 2023

- Relevant coursework: **Deep Learning, Machine Learning, Big Data**, Optimization and Lineal Algebra.
- **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.

Engineering Degree (Final grade 7.0/7.0)

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

- Capstone Project with **S&P Global**: News classification by fine tuning **Large Language Models from unlabeled data**.
- Deep Learning Competition: next frame video prediction using a **Transformer and CNN (Resnet)** in a JEPA network.
- **Text analysis** of large academic thesis corpus using **LDA** and GloVe word embeddings.
- Twitter sentiment and classification analysis using fastText, LGBM, and **fine tuning BERT**.
- **Published two papers** based on my Master's in Engineering thesis: [Google scholar link](#).
- Won Hackathon competition. Led an interdisciplinary group of 6. Designed and developed an app for transit users.

## EXPERIENCE

### ❖ Google

Mountain View, California

Internship in Platforms & Ecosystems

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 results to several internal stakeholders and management. Proposed a new methodology for analysts.

### ❖ STP Santiago

Santiago, Chile

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 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**.
- Refined core business cost estimates by **building a machine learning model for battery and fuel consumption**.
- **Detected data outliers** due to faulty batteries in 20% of STP's electric buses **via quantitative analysis** in a **SQL and Python data pipeline**, alerting the fleet provider contractor.
- Spearheaded automation of data ingestion, processing, and calculations of **six different KPIs** that played a key part in the revenue stream from concession's contract.
- Increased 6% the main contract KPI by scraping twitter data to alert service disruptions using **text processing**.
- Optimized drivers' work shift **scheduling using a machine learning** model with employee attendance data.

### ❖ Telefonica R&D and Universidad de Chile

Santiago, Chile

Research project with Telefonica, one of the largest telecom companies in Latin America.

Nov 2017 – May 2018

#### Data Scientist

- 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 and Deep Learning:
  - Pandas, NumPy | PyTorch, PyTorch-Lightning | 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 | Excel | Snowflake, Dask, pySpark, Hadoop | BeautifulSoup, Selenium, requests
- Native Spanish speaker; fluent in English; intermediate Italian.