CRISTIANA PRINCIPATO, Ph.D.
Senior Data Scientist

cristiana.principato@gmail.com | 773 709-2118 | linkedin.com/in/cristiana-principato

Senior Data Scientist with a Ph.D. in Experimental Physics and 5+ years of experience in data analytics and science. Experienced in developing and deploying predictive algorithms, machine learning models, advanced statistical analysis, and data visualization. Proficient in Python, SQL, with a strong foundation in programming and database environments. Adept at working cross-functionally in fast-paced, agile environments.

Skills

Machine Learning
- NLP
- Anomaly Detection
- Clustering
- Time Series
- Regression
- Classification
- Bayesian Modeling

Databases
- MongoDB
- PostgreSQL
- Azure
- AWS S3

Data Science
- Data Modeling
- Predictive Analytic
- Advance Statistics
- Hypothesis Testing (t-test, KS)
- A/B Testing
- Data Quality
- ETL

Frameworks
- ChatGPT API
- AWS, SageMaker

Languages & Libraries
- Python
- PyTorch
- SQL
- Numpy
- Scipy
- PyMC

Artificial Intelligence
- Large Language Models
- Retrieval Augmented Generation (RAG)
- Automatic Speech Recognition

Experience

COMPASS PATHWAYS, San Francisco — Aug 2022 - Present

Senior Data Scientist

Biotechnology company that leverages AI and predictive modeling to enhance mental health interventions

- Developed a patient health and well-being snapshot using **biometrics data from wearable devices** such as the Apple Watch, Oura Ring, Fitbit, Garmin, etc.
  - Created an **Anomaly detection algorithm** using a **Hiearchical Bayesian Model**
  - Researched **time series data** collection processes on iOs and Android
  - Analyzed data quality to determine reliability
  - Designed an **LLM-generated summary** that combines data from multiple sources into a single easy-to-use metric for healthcare providers
- Led the development of an education-focused **AI agent** to prepare patients for large-scale clinical trials by:
  - **Leveraging the CHATGPT API as the LLM framework** for the agent
  - Developing an **Advanced RAG framework** to customize responses to clinical needs
  - Performing prompt engineering to improve model performance on a clinical domain
  - Benchmarking the response accuracy from state-of-the-art **embedding models**
- **Developed an 85% accurate predictive model** for depression treatment outcomes, harnessing NLP techniques integrating LLM and ASR models to extract patient sentiment from recorded therapy sessions.

    - Prototyped algorithm for transcribing psychotherapy sessions in clinical trials with an outstanding **Word Error Rate of 3%** and a **Diarization Error Rate of 5%** by:
        - Implementing state-of-the-art translation, transcription, and diarization algorithms
        - Leveraging automatic speech recognition (ASR) technologies
    - Developed an AWS-based transcription pipeline to integrate partner data and improve model confidence, creating
        - Procedures for data ingestion and cleaning
        - Quality and performance benchmarks.
- Innovated a cutting-edge data-splitting algorithm for small datasets that **improves predictive model accuracy in stratified datasets by 15%**.

MakerSights, San Francisco Sep 2021 - Nov 2022
Senior Data Scientist & Consultant
SAAS company that uses customer data to improve product decisions & improve profitability for fashion brands

- Generated over **$1M in Annual Recurring Revenue (ARR)** by leading the development of a groundbreaking price optimization project:
    - Architecting new data collection, quality, and analysis methodologies
    - Championed the use of a state-of-the-art consumer preference choice-based model
    - Leveraged **MongoDB** and **PostgreSQL** to extract and transform data

Smiths Digital Forge, San Francisco Sep 2020 - Sep 2021
Artificial Intelligence Scientist
Center of excellence for AI technology and software R&D for the oil & gas industry

- Achieved **90%** precision in autonomous remote operational monitoring of oil pumps by devising a novel time series classification algorithm using a **Kalman Filter, clustering techniques, and a probabilistic approach**:
    - Created a brand-new comprehensive process for data processing, including receiving data from 10 distinct sensors, data resampling and standardization, feature computation, and efficient storage in MongoDB.
    - Collaborated with an external engineering team to devise effective solutions to operating challenges.
- Pioneered the development of an innovative application for anomaly detection in mechanical systems by leveraging acoustic emission sensor data, marking a first in the industry.

MakerSights, San Francisco Sep 2019 - May 2020
Data Scientist

- Innovated and conducted rigorous **statistical hypothesis testing** (**ANOVA, t-tests, A/B testing, Kolmogorov–Smirnov test**s) to refine the model selection, assess survey significance, and deliver premium insights on product evaluations, enhancing strategic decision-making.

Education

Insight Data Science **2019**
Fellowship, Insight Health Data Science

University of Virginia **2019**
PhD, High Energy Physics

University of Roma "Tor Vergata" **2014**
MS, Nuclear, and Sub Nuclear Physics