

# ARYAN MISHRA

+1 (646) 994-9732 | [am5696@columbia.edu](mailto:am5696@columbia.edu) | <https://www.linkedin.com/in/mishraaryan/> | <https://aryanmsr.github.io/>

## LANGUAGE AND IT SKILLS

---

- Programming: Python (Advanced), SQL (Advanced), R (Intermediate), JavaScript (Intermediate), C++ (Basic).
- Tools: AWS (SageMaker, Bedrock, Glue, Athena), Adobe (Analytics, CDP, Marketo), Git, Docker, Jenkins (CI/CD).
- Libraries & Frameworks: NumPy, Pandas, Matplotlib, Scikit-learn, TensorFlow, PyTorch, LangChain, LangGraph, FastAPI, React.
- Spoken Languages: English, Hindi, French, Italian.

## WORK EXPERIENCE

---

### Prudential Financial

Newark, NJ

*Lead Data Scientist – Retirement Strategies*

Feb 2023 – Present

- Creating AI agents using graph-based techniques for a sales co-pilot tool and designing scalable RAG pipelines with hybrid search and re-ranking techniques, combining dense and sparse retrieval to optimize knowledge management.
- Developed and deployed Markov-based models to mine 20+ TB of clickstream data, delivering real-time actionable insights.
- Led end-to-end deployment on AWS SageMaker using Docker, ECR, and SageMaker Endpoints for real-time inference.

*Data Science Intern*

May 2022 - Aug 2022

- Created a “Next Best Action” model optimizing 2M+ lead interactions using LSTM neural networks to maximize conversation rates.

### Unilever

Trumbull, CT

*Data Science Intern - R&D*

Sep 2022 - Dec 2022

- Automated knowledge retrieval of genes associated with dermatological conditions (e.g., atopic dermatitis, xerosis) from dense biomedical literature using a fine-tuned BERT language model, significantly enhancing research efficiency by cutting average research time from months to days.
- Built and deployed a fuzzy-matching model using Python that accelerated identification of 40K+ product variants from 200K+ e-commerce reviews.
- Streamlined code and productionized preprocessing, training, and prediction endpoints through FastAPI.
- Contributed to deployment pipeline integration and load testing APIs - effectively debugging 100% of issues.

### Focus Project Europe

Madrid, ES

*Product Data Science Intern*

Jun 2021 - Sep 2021

- Spearheaded design and implementation of a prototype grade prediction feature using a XGBoost model for an Ed-Tech start-up, achieving 95% accuracy and scaling it to a user base of 5,000 students.

## RESEARCH AND PROJECTS

---

### Ardian

New York, NY

*Data Scientist - Capstone Project*

Jan 2022 - May 2022

- Forecasted one-month electricity price spikes using XGBoost model, potentially saving business \$8M on an annualized basis.
- Created a Power BI dashboard to visualize backend model results and empower stakeholders to make business-critical decisions.

### Bocconi University

Milan, IT

*Research Assistant - Computer Vision (Bocconi-Gucci Lab)*

Jun 2020 - Aug 2020

- Analyzed and web-scraped over 50K images of various fashion brands to design an in-house labeled dataset for fashion research.
- Built and tested the architectures of various image classification (CNN) models using TensorFlow, including VGG-16 and ResNet.

## EDUCATION

---

### Columbia Engineering & Columbia Business School: M.S. Business Analytics, GPA: 3.7/4.0

New York, NY | Feb 2023

Courses: Machine Learning, Deep Learning, Marketing Analytics, Optimization Models and Methods, Probability Statistics & Simulation.

### Bocconi University: B.S. Economics & Computer Science, GPA: 3.8/4.0 (105/110)

Milan, IT | Jul 2021

Courses: Advanced Statistics, Game Theory, Mechanism Design, Data Structures & Algorithms, Computer Programming, Econometrics.

## ADDITIONAL INFORMATION

---

- Co-Founder of Relevant Projects - an EdTech startup that enabled disadvantaged students to engage with companies on high-impact projects, fostering creativity and bridging industry-academia gaps.
- Runner-Up Vodafone Italy "TOBi" Hackathon 2019 – enhanced Vodafone's chatbot using unsupervised machine learning algorithms and Natural Language Processing. Subsequently endorsed by Vodafone on a national level.