

# RAHUL SHARMA

College Park, Maryland | +1-240-610-7281 | [rahul11s@umd.edu](mailto:rahul11s@umd.edu) | [@LinkedIn](#) | [@Git](#) | [@Portfolio](#)

## Summary

Data Scientist with 4+ years of experience applying statistical modeling & ML/DL to large-scale problems. Proven track record across NLP, CV, recommender systems, and time-series modeling driving impact in healthcare, risk, and operational analytics.

## Education

### University of Maryland - College Park, US

May 2026

*Master of Science in Data Science*

- Relevant Courses: Probability & Statistics, Machine Learning, Data Science, Algorithms, Data Modeling, Big-Data Systems, Computer Vision, Time-Series, Cloud Computing

### Indian Institute of Technology (BHU) Varanasi, India

July 2020

*Bachelor of Technology in Engineering*

- Relevant Courses: Eng. Mathematics, Computer Programming, Numerical Methods, Applied Mathematics.

## Work Experiences

### Bilbo.ai, Baltimore, US

Sept 2025-Present

*AI Engineer Intern*

- Designed and deployed a **retrieval-augmented NLP system** that ingested medical PDF manuals to generate step-by-step MARCH-PAWS checklists, producing stage-specific Q&A grounded in source files with citations.

### Jet2 Travel Technology, Pune, India

June 2022- Aug 2024

*Data Scientist*

- Processed **1M+ call transcripts**; adapted pretrained LLMs via supervised fine-tuning and applied few-shot prompting at inference, achieving **88% Jaccard recall** and enabling real-time decision-making in contact centre operations.
- Implemented **unsupervised topic modeling** and anomaly detection pipelines; identified disaster patterns at **92% specificity** and translated findings into prescriptive playbooks for operational leaders.
- Deployed **LSTM/GRU** models with walk-forward validation in Dataiku, scheduling automated jobs and alerts; ensured data integrity with Pydantic and sustained **<2% model drift** through continuous monitoring, integrating **Tableau** for insights.

### Fibe (Formerly EarlySalary), Pune, India

Sept 2021-June 2022

*Data Scientist*

- Led feature-engineering & scorecard modeling on 1000+ bureau vars; lifted **ROC-AUC to 0.94** and translated model outputs into executive dashboards and decision frameworks, driving credit-policy changes.
- Automated Python scripts with Knime, added **AI-governance checks** (data integrity, security), shrinking turnaround **15x**.

### Advanced Technology Consulting Service, Jaipur, India

Nov 2020-Sept 2021

*Associate Data Scientist*

- Streamlined PySpark to MSSQL ETL pipeline, integrating data quality validation prior to analytics.
- Trained Mask R-CNN model using transfer learning in TensorFlow; delivered to Azure blob storage for model inference, achieving **85% mAP**.

### Axio (Formerly CapitalFloat), Bangalore, India

May 2020-Nov 2020

*Research Intern*

- Built address dataset, applied **clustering & statistical evaluation**, achieving **95% F1**, and informed fraud-risk strategy.
- Designed OCF metrics and developed **Power BI storytelling dashboards**, translating analytics to non-tech executives.

## Projects

- MIMIC-CXR Radiology Report Generation:** Fine-tuned **LLaVA-NeXT (Mistral-7B)** with **LoRA** to generate structured JSON radiology reports from chest X-rays. Designed curriculum learning and conducted **A/B evaluation** comparing image-only vs image+EHR conditioning to assess report quality and clinical reasoning.
- Multimodal Fashion Recommendation System, Fall'24, UMCP:** Fine-tuned CLIP vision and text encoders with BLIP-generated captions and segmentation masks to improve text-image alignment. Integrated adaptive query reweighting and FAISS retrieval, achieving **88% Mean Opinion Score** against human relevance judgments.

## Publications

- Co-author, IEEE ICSCDS 2025.** "Deep Learning for Reliable Routing in WANETs." *Peer-reviewed paper.* [Link](#)

## Skills

- Programming:** Python (Adv.), SQL (Adv.), PySpark (Int.), C/C++ (Int.), MATLAB (Int.)
- ML & Modeling:** Scikit-learn, XGBoost, Anomaly Detection, Feature Engineering, Time-Series, Recommendation Systems
- Statistics & Experimentation:** A/B Testing, Hypothesis Testing, Causal Analysis, Bayesian Methods, Statsmodels
- DL & Generative AI:** PyTorch, TensorFlow, Transformers, LLMs, RAG, Fine-Tuning, Quantization
- Cloud & Analytics:** AWS, GCP, Azure, Snowflake, Databricks, Docker, Tableau, Power BI, Dataiku, KNIME