Rohit Kulkarni

📞 +1 (732)-496-6422 rohit.kulkarni@rutgers.edu in rohitak8 ⌂ rohitkulkarni08

SKILLS

Big Data & Cloud:	Microsoft Azure (Data Factory, DevOps, Data Lake, SQL), Google Cloud Platform (VertexAI, BigQuery), Postman, Docker, Databricks, Apache Spark
Deep Learning & NLP:	Scikit-Learn, TensorFlow, Keras, PyTorch, Sentiment Analysis, PineCone, Transformers (GPT, BERT), OpenAI, LLM Guardrails, RAG, Prompt Engineering, Finetuning
Programming & Database:	Python, PySpark, SQL, R, PostgreSQL, MongoDB, MS SQL Server
Data Science:	Statistical Modeling, Bayesian Analysis, Data Mining, Feature Engineering, Time Series Forecasting, A/B Testing, Ensemble Methods, Gradient Boosting
Visualization:	PowerBI, Tableau, Matplotlib, Seaborn
Certifications:	Microsoft Certified Azure Data Engineer Associate, Databricks Lakehouse Fundamentals

Professional Experience

Data Science Intern — *Nov 2024 - Jan 2025*
MediCodio Inc. — *Pleasanton, CA (Remote)*

- Improved medical coding accuracy by **37%** using **GPT-4o** and **Pinecone**, implementing domain-specific prompt engineering and multi-level similarity scoring.
- Boosted profitability by **70%** by developing an **OCR**-based **text extraction** pipeline, eliminating third-party dataset costs through automated entity recognition and rule-based validation.
- Optimized query performance and reduced storage costs by **24%** by designing a scalable, multi-layered **Azure Data Lake** architecture integrated with **Databricks** for efficient data pipelines.
- Reduced data redundancy and improved runtime by **30%** developing a **Python-PostgreSQL** ingestion framework with batch scheduling and parallel processing, streamlining high-volume EMR data pipelines.

Machine Learning Engineer — *May 2024 - Oct 2024*
AI Assistant Co. — *San Jose, CA (Remote)*

- Developed data models for an **LLM**-powered **ATS** using **GPT-4o**, **Claude**, and **Llama3** for structured resume parsing and categorical scoring, increasing prescreening accuracy by **38%** and enhancing candidate-job matching.
- Boosted extraction accuracy by **80%** for resume-job parsing by constructing **OCR**-workflows using **PyPDF** and **PyTorch**.
- Built data pipelines with **AWS S3**, **Flask** and **PostgreSQL**, improving data integrity and reducing processing time by **25%**. Tested **REST APIs** with **Postman** to ensure reliability and scalability.
- Enhanced candidate-job matching accuracy by **40%** through **A/B testing** and advanced feature engineering.

Analytics Consultant — *Jul 2021 - Jul 2023*
Fractal Analytics — *Bengaluru, India*

- Delivered **$50M+** in cost savings for **Unilever**'s retail division by developing a **Garrett Ranking** and **KNN**-based product delisting recommendation engine, optimizing SKU Simplification and enabling data-driven decision-making.
- Led technical initiatives for the North American track, **restructuring** US market analytics workflows, improving new product performance tracking and business decision making by refining data sources and KPIs.
- Consolidated **2TB+** of multi-source data to create a uniform **system of record**, delivering complex KPIs and critical product life-cycle insights, boosting user confidence.
- Cut **ETL** runtime by **85%** by migrating legacy **Python** workflows to **PySpark** and **SparkSQL** with **Delta Tables**, improving modularity and processing efficiency.
- **Automated** ingestion, execution, validation, and publishing workflows with **Azure Data Factory** and **Databricks**, eliminating **98% of manual tasks** and significantly improving efficiency.

Personal Projects

Advanced American Options Pricing Chatbot: *(ongoing)* Developed and deployed a neural network-powered American options pricing platform combining stock data with trader-intuition features, improving pricing accuracy over binomial models. Built a synthetic data engine simulating market regimes and early exercise behavior to enhance robustness. Deployed on GCP using Vertex AI and BigQuery, and delivered an NLP-based tool for scenario analysis and comparative insights.

HR Lookup System for Narrative-Driven Employee Insights: Built a natural language query system for HR background checks, integrating relational databases with LLMs to deliver narrative-driven employee insights. Designed an automated schema adaptation mechanism, enabling efficient query execution while enforcing PII protection standards.

Education

Masters of Science - Statistics, Data Science — 2023 - 2025
Rutgers, The State University of New Jersey — *New Brunswick, NJ*

Bachelors of Engineering - Electronics and Telecommunication Engineering — 2017 - 2021
R.V. College of Engineering — *Bengaluru, India*