

# ROHINI SONDOLE

[LinkedIn](#) | [GitHub](#) | Email: rsondole21@gmail.com | Phone: (512) 815-0990 | Austin, TX

## SUMMARY

Machine Learning Engineer with 6+ years of experience building and deploying large-scale production ML systems across **AWS and Azure**. Strong background in scalable **data pipelines, MLOps, and cloud-native infrastructure** supporting high-throughput, low-latency AI applications.

## SKILLS

- **Programming Languages:** Python (Pandas, NumPy), SQL, Scala, Golang, Shell Scripting, REST APIs, Flask, Git, Jira, Agile/Scrum
- **ML & Deep Learning:** PyTorch, TensorFlow, Scikit-learn, XGBoost, LightGBM, NLP, Computer Vision, Time-Series Forecasting
- **GenAI & LLMs:** LangChain, RAG, Prompt Engineering, Agentic Workflows, Vector Search, Embedding-Based Retrieval, Fine-Tuning
- **MLOps & Infrastructure:** A/B Testing, CI/CD, MLflow, Apache Airflow
- **Cloud & Infrastructure:** AWS (SageMaker, Glue, Lambda, S3, ECS), Azure (Databricks, Data Factory), Kubernetes, Docker, Terraform
- **Big Data & Databases:** Spark, PySpark, Databricks, Kafka, Snowflake, PostgreSQL, MongoDB, Cassandra, Delta Lake, Elasticsearch, Data Lake
- **Computer Science Fundamentals:** Data Structures, Algorithms, Distributed Systems, System Design
- **Architecture:** Distributed Systems, Microservices, RESTful APIs, GraphQL, Serverless

## PROFESSIONAL EXPERIENCE

### CISCO - MACHINE LEARNING ENGINEER INTERN - Austin, TX

Jan 2026 – Present

- Deployed **predictive autoscaling models** with 50+ features, **improving forecasting accuracy** by ~25%
- Optimized **TensorFlow deep learning models** for real-time inference, reducing scaling latency by 30%
- Architected **fault-tolerant data platform** analyzing 2B+ weekly events across 120+ Kubernetes clusters
- Designed **data models and logging infrastructure** delivering \$6M/year in projected **cost savings**
- Participated in **architecture and code reviews**, documenting solutions with technical specs and presentations

### CYBAGE SOFTWARE - DATA AND MACHINE LEARNING ENGINEER - Pune, India

Sep 2022 – Feb 2025

- Built and optimized ML infrastructure with **Airflow, PySpark, and Azure ML**, processing 500GB+ daily data
- Partnered with **data scientists** to transition models from research to production using Python and Scala
- Owned Azure ML platform for 20+ downstream apps, reducing **model deployment** time by 15%
- Implemented MLOps with automated model validation, **CI/CD pipelines**, and **unit testing**, **boosting reliability** by 20%
- Led **4-engineer team** migrating 10TB+ to Snowflake with Delta Lake, 25% ahead of schedule
- Mentored 3 engineers via pair programming and **code reviews**, **reducing code defects** by 40%

### TATA CONSULTANCY SERVICES - DATA ENGINEER - Pune, India

Sep 2018 – Sep 2022

- Collaborated with **data scientists** to build 20+ validated datasets, enabling **3 ML models to production**
- Built high-performance pipelines processing 1TB+ daily in Python and Java using **AWS Glue and Lambda**
- Led **cloud migration** from IBM DataStage to Azure, cutting operational costs by 40%
- Developed **automated test suites** with mocking frameworks, ensuring **pipeline reliability** across environments
- Built **PySpark Transformations** in Palantir Foundry for **ML training and operational analytics**

## PROJECTS

### Deepfake AI Image Detection ([Article Link](#))

- Trained and evaluated 3 deep learning architectures (**ResNet-50, ConvNeXt-Tiny, MobileNetV3**) on 140,000+ image dataset, achieving 87% validation accuracy and 0.92 AUC with explainable AI (Grad-CAM) for model interpretability

### AI Data Insights Agent ([Project Link](#))

- Engineered full-stack **GenAI** application with **LangChain, OpenAI GPT, and RAG pipeline** enabling natural language to SQL generation, accelerating insights delivery from 15+ minutes to under 10 seconds for non-technical stakeholders

### Connect 4 ML Model Serving & Experimentation ([Website link](#)) (username= dan and password= optimization1234)

- Deployed RESTful **ML serving API** using Flask handling 500+ concurrent sessions with sub-second latency, benchmarking **CNN, Transformer, and MCTS** models with extensive hyperparameter tuning via GridSearchCV, improving precision by ~18%

### YouTube Channel Intelligence Platform: Snowflake Cortex AI

- Built scalable **NLP** analytics platform on Snowflake processing 15K+ channels with Medallion Architecture, **Cortex AI** for sentiment analysis, semantic search, and **Streamlit** dashboards for real-time insights

## CERTIFICATIONS

Cloud: Microsoft Certified: Azure Data Engineer Associate, Snowflake: Data Warehousing Workshop, Databricks

AI: Generative AI Leader Certification, AWS Certified AI Practitioner, IBM: Python for Data Science, AI & Development

## EDUCATION

### Master of Science in Business Analytics

The University of Texas at Austin – McCombs School of Business

July 2025 – May 2026

### Bachelor of Technology in Computer Technology

Yeshwantrao Chavhan College of Engineering, Nagpur, India

July 2014 – May 2018