Mounika Boggavarapu

Irving, TX/ mounikaboggavarapu123@gmail.com / 813-568-7138 / GitHub / LinkedIn

SUMMARY

Results-driven Data Scientist with expertise in machine learning, risk analytics, and portfolio insights across large-scale datasets. Skilled in statistical modeling and deep learning, with hands-on experience in deploying data solutions on AWS and Azure. Certified in AWS Data Engineering and Cloud Foundations.

SKILLS

Programming Languages: Python, SQL, R, Java, Scala, Julia, C++, C, TypeScript, Go, Rust, Bash, Shell, MATLAB, SAS

Machine Learning & Deep Learning: TensorFlow, PyTorch, SparkML, LightGBM, XGBoost, Scikit-learn, Transformer Models, LLMs, Generative AI, A/B Testing, Model Evaluation, Feature Engineering, Model Fine-Tuning

Domains & Applications: Natural Language Processing (NLP), Computer Vision, Audio & Video Analysis, Time Series Forecasting, Data Science

ML Workflow & Tools: ML Pipelines, Data Preprocessing, Model Training, Hyperparameter Tuning, Evaluation, Model Deployment, SHAP, MLflow, Jupyter, GitHub Actions

Big Data & Cloud: Apache Spark, Hadoop, Airflow, AWS (S3, SageMaker, Lambda), Azure (ML Studio), Databricks

Data Visualization & Business Intelligence: Tableau, Looker, Power BI, Excel, Salesforce

Collaboration & Documentation: Agile, Cross-functional Collaboration, Version Control, Communication, Problem Solving

EDUCATION

Master of Science in Computer Science: The University of South Florida (USF) *Aug 2023 - May 2025*

CERTIFICATIONS

IBM Data Science Professional Certificate

AWS Certified Data Engineer – Associate

AWS Certified Cloud Practitioner

WORK EXPERIENCE

The University of South Florida (USF) Tampa, Florida

Data Science Research Assistant - Advisor: Dr. Yicheng Tu *Aug 2024 - Dec 2024*

Tech Stack: Python, PyTorch, CUDA, cuDNN, TensorFlow, Transformers, NumPy, Pandas, Matplotlib, Seaborn, Linux

- Developed a genomic classifier using DNA-BERT and PyTorch on 1.5M+ sequences, achieving 85% ROC-AUC and 22% higher precision via embedding optimization and prompt tuning for biological text.
- Accelerated training on 10M+ samples by 40% using CUDA/cuDNN parallelization, and improved generalization by 18% through statistical diagnostics and attention-based error analysis.

Tata Consultancy Services (TCS) Hyderabad, India

Data Science Engineer *Aug 2022 - Jul 2023*

- Developed fraud detection models using logistic regression and random forests across 15M+ transactions, increasing detection precision by 31% and reducing false positives by 20% with SHAP-based interpretability.
- Constructed AWS-native ETL pipelines (S3, Lambda) using LSM trees and Python to process high-volume logs, reducing query latency by 25% and improving system uptime to 99.99%.
- Segmented 10M+ users using K-means and hierarchical clustering to enhance marketing efforts, driving a 12% increase in targeted campaign conversion rates through tailored strategies.
- Devised live dashboards in Dash and SQL to monitor 10+ KPIs, cutting reporting delays by 100% and enabling real-time visibility into business performance metrics.

Mee Buddy Nuzvid, India

Data Science Engineer *May 2020 - Jul 2022*

- Spearheaded the development of time series and regression models to forecast delivery times for over 10,000+ weekly orders, improving SLA adherence by 17% and reducing the average delivery delay by 1.8 hours across 15 districts.
- Designed and deployed survival analysis models to estimate churn risk and customer lifetime value across over 25,000+ user profiles, reducing churn by 12% and increasing average revenue per user by 9% within a 3-month intervention period.
- Automated over 20 recurring analytics workflows using SQL, Python, and GitHub Actions, cutting the manual reporting workload by 50% and accelerating stakeholder insight delivery from 3 days to same-day availability.
- Engineered and optimized asynchronous backend pipelines using Celery and Redis to support over 50,000+ weekly task executions, ensuring stable system performance during 3x traffic surges and maintaining 99.98% job success rate.

PROJECTS

Cloud-Native Portfolio Analytics Platform

- Built a cloud-native portfolio analytics system using Java, Spring Boot, ReactJS, and Azure, handling 10K+ daily trades. Reduced trade processing time by 40% using asynchronous APIs and improved system uptime to 99.95% via container orchestration with Docker and Kubernetes.

High-Performance Data Analytics with CUDA

- Implemented a GPU-accelerated dynamic parallel processing system using CUDA C++ to analyze 10M+ records in genomics and imaging datasets. Reduced kernel execution time by 55% through shared memory optimization and warp-level parallelism. Benchmarked performance against CPU implementations and achieved a 6.2× speedup for data-intensive analytical workflows.