Siraj Patnam

St. Louis, MO | 314-393-9371 | rahamansirajahmedpatnam@gmail.com | LinkedIn | GitHub | Portfolio

PROFESSIONAL SUMMARY

Results-driven Data Scientist and ML Engineer with expertise in scalable ML pipelines, AWS cloud deployment, and data workflow optimization. Experienced in translating complex data into actionable insights through NLP, computer vision, and advanced analytics. Skilled in delivering production-ready solutions that reduce costs and improve business performance across real-world applications. Looking for opportunities to learn, grow and contribute to the world.

TECHNICAL SKILLS

Programming Languages: Python, SQL, R, Java, C++, JavaScript
Machine Learning & Deep Learning: TensorFlow, PyTorch, Scikit-Learn, Keras
Cloud Platforms: AWS SageMaker, AWS ML Services, Google Cloud Platform (GCP)
Data Engineering & Processing: Pandas, NumPy, Apache Spark, Hadoop, MongoDB
Data Visualization: Tableau, Power BI, Matplotlib, Seaborn
Tools & DevOps: Git, Docker, Kubernetes, MLflow, CI/CD
Specialized Areas: Natural Language Processing (NLP), Computer Vision, ETL Pipelines
Frameworks & Platforms: Streamlit, FastAPI, Large Language Model (LLM) Frameworks

WORK EXPERIENCE

Data Analyst Intern — **Jun 2024 – Aug 2024**
Washington University, University Advancement, St. Louis, MO

- Developed an end-to-end ETL pipeline using **Python**, **Pandas**, and **ETL Pipelines** to automate data extraction from Salesforce CRM, eliminating manual delays and accelerating availability.
- Designed executive-level dashboards using **Power BI** and **Tableau** to visualize real-time advancement metrics, reducing data retrieval time by 30% and enhancing decision-making for leadership meetings.
- Devised donor retention and giving score algorithms using **Scikit-Learn** in **Python**, enabling predictive donor segmentation and improving assignment accuracy by 15% and 25% respectively.
- Automated the full donor assignment lifecycle with **Python** scripting and **CI/CD** deployment, cutting manual workload and reducing turnaround time from three months to two weeks.
- Conducted deep-dive analysis of frozen fund allocations using **SQL** queries and **data visualization** dashboards, surfacing actionable insights that increased mobilizable grant funding by over 20%.

Data Science Intern — **Apr 2023 – Jul 2023**
Data Exposys Labs, Hyderabad, India

- Built a diabetes prediction model using **Random Forest**, **Decision Trees**, and **Scikit-Learn** in **Python**, achieving 91% accuracy and validating the model's robustness across cross-validation sets.
- Engineered robust data pipelines using **SQL**, **Pandas**, and custom **ETL Pipelines** to clean, transform, and enrich raw health records into structured datasets optimized for machine learning workflows.
- Created detailed exploratory visualizations and model performance insights using **Matplotlib** and **Seaborn**, compiling reports that improved stakeholder understanding and decision confidence.
- Organized and maintained structured datasets in **MongoDB** and **Hadoop**-based repositories, ensuring high data availability, scalability, and smooth collaboration across distributed teams.

PROJECTS

MLOps Network Security Pipeline | End-to-End ML Implementation — **Apr 2025**

- Evolved a phishing URL detection pipeline using **Random Forest**, **Gradient Boosting**, and **Scikit-Learn**, achieving 97.8% F1 score and optimizing end-to-end model performance.
- Deployed the solution with **FastAPI**, **Docker**, and **GitHub Actions** on **AWS EC2**, integrating **MLflow** tracking and real-time data ingestion from **MongoDB Atlas**.

Twitter Sentiment Analysis with BERT | Production NLP Pipeline — **Mar 2025**

- Fine-tuned a **BERT**-based transformer model using **Hugging Face Transformers** to classify tweets into six emotion categories, achieving 94% accuracy and a 0.92 F1-score on real-world Twitter sentiment datasets.
- Constructed and deployed a scalable **NLP inference pipeline** using **FastAPI**, **Docker Compose**, and **Nginx**, enabling reliable sub-100ms real-time predictions with horizontal scaling support.

Semantic Segmentation with Amazon SageMaker | IIIT-Oxford Pets Dataset — **Feb 2025**

- Made a semantic segmentation model using **Amazon SageMaker**'s **FCN** with **ResNet-50**, tuning hyperparameters to reach 92.6% pixel accuracy and 0.605 mIOU on the IIIT-Oxford Pets dataset.
- Deployed the trained model as a real-time inference service using **SageMaker Endpoint**, enabling scalable and low-latency image segmentation predictions for downstream **computer vision** applications.

ML Model Comparison Sandbox | Interactive Tool — **Mar 2025**

- Assembled an interactive model comparison app using **Streamlit** with support for **SVM**, **Random Forest**, and Logistic Regression, displaying real-time performance metrics.
- Enabled custom hyperparameter tuning and live feature selection using **Scikit-Learn**, with visualizations built in **Matplotlib** and **Seaborn** to aid interpretability.

EDUCATION

Saint Louis University, College of Engineering **Expected May 2025**
Master of Science in Information Systems | GPA: 3.8/4.0

Muffakham Jah College Of Engineering And Technology **Jul 2023**
Bachelor of Technology in Computer Science, Hyderabad, India

CERTIFICATIONS & ACHIEVEMENTS

Google Advanced Data Analytics Specialization *– Google* **May 2025**
Credential ID: DW8P1VAQFR2P
AWS Certified Machine Learning - Specialty *– Amazon Web Services* **Apr 2025 – Apr 2028**
Accelerating End-to-End Data Science Workflows *– NVIDIA* **Apr 2025**
Credential ID: TREGJIuITZOpt565qKLtBg
AWS Certified AI Practitioner *– Amazon Web Services* **Mar 2025**
Best Use Case of AI Award *– e-Hacks Hackathon* **Apr 2024**
3rd Place – NASA Space Settlement Contest **2020**
Winner – Know For Sure Quiz Competition *– Britannica* **2022**