SURYA TEJA JAKKA

sj888@nau.edu | (928) 863-9423 | https://www.linkedin.com/in/sjakka14/

OBJECTIVE

Applied Data Scientist and Embedded ML Engineer with hands-on experience building **end-to-end data and machine learning pipelines** on real-world sensor datasets in environmental monitoring and IoT systems. Strong track record of transforming raw, high-frequency time-series data into **validated, calibrated, and model-ready datasets** using Python, SQL, and lightweight ML frameworks including TensorFlow Lite and scikit-learn.

Designed and deployed **anomaly detection models** trained on **500,000+ labeled sensor data points**, achieving **92% precision, 94% recall, and 78% false-positive** reduction in real-time edge environments. Adept at collaborating with cross-functional research teams, defining data requirements, and implementing **robust data quality, calibration, and validation protocols** supporting **50,000+ daily sensor readings** and **200+ weekly precision measurements**. Seeking a full-time Data Scientist role focused on time-series ML, anomaly detection, and data engineering for sensor- and event-driven applications.

CORE COMPETENCIES

Programming & Data

- Python (pandas, NumPy, scikit-learn, TensorFlow Lite, Matplotlib, Seaborn, Plotly)
- SQL (SQL Server, PostgreSQL, MySQL): schema design, complex queries, performance optimization
- MATLAB for numerical computing, modeling, and visualization

Machine Learning & Statistics

- Supervised learning and anomaly detection on time-series data
- Feature engineering, normalization, windowing, and threshold optimization
- Model evaluation and tuning (precision/recall tradeoffs, false-positive reduction)
- Edge ML deployment on resource-constrained devices (ESP32 / microcontrollers)

Data Engineering & Pipelines

- Web scraping and ETL pipelines using Python (Selenium, BeautifulSoup) processing **50,000+ records/day**
- Automated reporting and analytics workflows using Python and SQL
- High-frequency time-series data logging and management (50,000+ sensor records/day)

Visualization & Reporting

- Exploratory data analysis, trend visualization, and anomaly dashboards
- Calibration reporting and sensor performance analysis

Domain & Tools

- IoT and sensor data (environmental monitoring: dendrometers, tensiometers, soil moisture sensors)
- Precision metrology, calibration workflows, and measurement system analysis
- Git/GitHub, Linux/Windows environments, working knowledge of embedded C/C++ for edge data pipelines

WORK EXPERIENCE

Northern Arizona University, Flagstaff, AZ, USA
Research Associate – Datalogger Development (ECOSAIL Lab) *(August 2024 – Present)*

- Designed and implemented the **data architecture** for a multi-node environmental monitoring platform, ingesting **50,000+ time-series measurements per day** at 1-minute resolution from **5+ heterogeneous sensor types**.
- Built **data preprocessing and quality-control pipelines** incorporating range validation, missing-data handling, and consistency checks to ensure reliable downstream analysis and modeling.
- Developed a **scalable data logging and storage pipeline** handling **500MB+ weekly sensor data**, enabling efficient ingestion, retrieval, and longitudinal analysis.
- Collaborated closely with ecologists and climate scientists to translate research questions into **data requirements, sampling strategies, and analytical metrics**, ensuring data directly supported hypothesis-driven modeling.
- Supported time-series analysis and anomaly detection workflows used to assess environmental conditions and sensor behavior over extended deployments.

Northern Arizona University, Flagstaff, AZ, USA
Research Associate (MRTL Lab) *(August 2024 – Present)*

- Processed and analyzed **200+ precision measurement records weekly** from coordinate measuring machines (CMM) and related instruments.
- Designed **metrology-based data validation frameworks** used to calibrate and validate environmental sensors deployed in field monitoring systems.
- Improved measurement accuracy by **15%** through systematic error analysis, calibration experiments, and statistical validation.
- Connected sample-level metrology data to **sensor performance thresholds**, directly informing data quality standards for downstream ML models.

TATA Consultancy Services, Hyderabad, India
Assistant System Engineer – Trainee *(July 2021 – August 2022)*

- Automated **15+ manual reporting workflows** using SQL Server, reducing report generation time from **4 hours to 15 minutes (94% improvement)**.
- Designed normalized relational schemas supporting **10,000+ borrower records** and **50,000+ transactional entries** for analytics and reporting use cases.
- Migrated **25+ Excel-based workflows** into SQL-driven reporting pipelines, improving data consistency and accessibility for **50+ business users**.
- Improved analytical query performance by **60%** through indexing strategies, join optimization, and query refactoring.

ACADEMIC PROJECTS

Embedded Machine Learning for Real-Time Sensor Anomaly Detection | 2024

- Built an end-to-end **time-series anomaly detection pipeline** using **500,000+ labeled sensor data points** spanning diverse environmental conditions.
- Trained and evaluated lightweight TensorFlow Lite models (50–200KB), achieving **92% precision and 94% recall** on held-out datasets.
- Reduced false positives by **78%** through adaptive thresholding informed by historical context and temporal patterns.

- Deployed models on ESP32 microcontrollers with **<50ms inference latency**, enabling real-time, on-device anomaly flagging without cloud dependency.
- Documented feature engineering, modeling assumptions, and evaluation methodology to support interpretation by non-ML collaborators.

Intelligent Web Scraping and Automation Pipeline for Job Search Optimization | 2024

- Designed a distributed Python-based scraping system using Selenium and BeautifulSoup to collect **50,000+ job postings daily** from **20+ sources**.
- Implemented data cleaning, normalization, and structured storage to support downstream analysis and ranking.
- Developed a rules- and similarity-based matching algorithm across **50+ criteria**, achieving **~85% relevance accuracy** in internal evaluation.
- Built monitoring metrics and dashboards tracking responses and outcomes, reducing time-to-hire by **60%**.

Smart Home IoT Control System (Data & Control Layer) | 2020

- Collected and managed real-time device state data for **5+ smart devices** via a MATLAB-based interface connected to Arduino microcontrollers.
- Designed a monitoring and control layer achieving **~50ms response time** and **99.5% state consistency** across restarts.
- Demonstrated reliable logging, state tracking, and device coordination in a multi-device environment.

EDUCATION

Northern Arizona University, Flagstaff, AZ, USA — May 2024
Master of Science, Computer Science — GPA: 3.55/4.0
GITAM Deemed to be University, Hyderabad, Telangana, India — June 2021
Bachelor of Technology, Electrical Electronics and Communication Engineering — GPA:8.76/10.0