Uday Sai Bela

571-530-9100 | udaysaibela@gmail.com | LinkedIn

Education

George Mason University — Fairfax, VA
Master of Science in Data Analytics Engineering | GPA: 3.89 — *Jan 2024 – Dec 2025*

Jawaharlal Nehru Technological University Hyderabad — Hyderabad, India
Bachelor of Technology in Chemical Engineering — *Aug 2019 – May 2023*

Professional Experience

Data Analyst — Jun 2022 – Dec 2023
Claro Software Solutions Pvt Ltd — *Hyderabad, India*

- Conducted in-depth fraud detection analysis on over 250,000 transaction records, identifying behavioral patterns and implementing data-driven algorithms that improved detection accuracy by 18% .
- Collaborated with a cross-functional team to enhance existing fraud detection models, reducing false positives by 22% and improving precision in identifying legitimate transactions.
- Presented insights and recommendations to senior management, leading to the adoption of new fraud detection workflows and faster anomaly response times by 15%
- Utilized Python, SQL, and machine learning techniques to analyze large datasets and extract actionable insights for identifying suspicious transactions.
- Developed dashboards and visual reports to monitor fraud trends and model performance, enabling data-driven decision-making across teams.

Technical Skills

Programming: Python, SQL, MATLAB, R, Bash
Big Data & Distributed Systems: Apache Spark, PySpark, Spark SQL, Hive, Hadoop, Databricks
ML & LLM Data: Scikit-learn, PyTorch, LLM evaluation, Prompt Engineering
Visualization: Tableau, Power BI, Matplotlib, Seaborn, Plotly
Cloud: AWS (S3, EC2, Glue, RDS), GCP (BigQuery), Azure Data Studio
ETL & Pipelines: Data Wrangling, PySpark ETL, API Integration, Data Curation, Privacy-preserving ETL
Databases: MySQL, PostgreSQL
Tools: Git, VS Code, Jupyter, Docker (basic), Agile/Scrum
Other Technical Skills: REST APIs, JSON, Regex, Web Scraping (BeautifulSoup, Selenium), Data Pipeline Development, Statistical Modeling

Projects

AI-driven FDA Form 483 Compliance Analysis System — November 2025

- **Tech Stack: Python, PyPDF2, Regex, Large Language Models (LLMs), JSON, Pandas**
- Integrated structured metadata and unstructured PDF inspection reports to form the foundation of an AI-driven compliance system, enabling organized inspection tracking and regulatory intelligence extraction.
- Developed a preprocessing pipeline using PyPDF2 and regex-based segmentation to extract, clean, and standardize narrative observations from long-form FDA Form 483 PDFs.
- Cross-validated extracted observations with key FDA metadata (e.g., firm name, FEI number, media ID) to ensure high accuracy and alignment with official records.
- Implemented LLM-based classification to determine overall inspection status (OAI, VAI, NAI), violation severity (Critical, Significant, Standard), compliance program mapping, and recommended follow-up actions.
- Structured all outputs in strict JSON format, providing consistent, machine-readable data for dashboards, analytics, and chatbot integrations.
- Enabled automated regulatory intelligence, transforming raw inspection narratives into actionable insights for risk assessment, compliance monitoring, and decision-making support.

Market Basket Analysis on Distributed Platform (Spark + Databricks) — April 2025

- **Tech Stack: PySpark, MLlib, Databricks, SQL, Pandas**
- Processed large-scale transactional datasets on Databricks Spark clusters, enabling fast and scalable analytics.
- Designed and implemented end-to-end distributed ETL workflows for data ingestion, cleaning, and transformation.
- Generated optimized association rules using FP-Growth, improving recommendation quality and model efficiency.
- Developed parameterized notebooks to automate the pipeline, ensuring repeatability and scalability across experiments.
- Delivered actionable insights on purchasing patterns, supporting data-driven marketing and product placement strategies.

Bridge Condition Deterioration Prediction (Structured Data Pipeline) — October 2024

- **Tech Stack: Python, Pandas, Scikit-learn, Matplotlib**
- Engineered a dataset of 59,500 rows and 17 features through automated cleaning and feature transformation.
- Developed preprocessing pipelines for missing values, encoding, normalization, and scaling for ML modeling.
- Implemented a Random Forest model with hyperparameter optimization via GridSearchCV for accurate predictions.
- Produced visual and tabular reports to communicate predictive insights for bridge maintenance planning.
- Enhanced model interpretability and decision-making by analyzing feature importance and performance metrics.