Sujana Kabir

Arlington, Texas | 903-867-9137 | kabirsujana0@gmail.com | LinkedIn | PortFolio | Github

Education

University of Texas at Arlington — Arlington, TX
Bachelor of Science in Computer Science — *Aug. 2022 – May 2026*

- GPA: 3.96/4.00
- Relevant Coursework: Project Management, Database Systems, Data Visualization & Storytelling, Machine Learning, Data Analytics, Statistical Analysis, Data Mining

Skills

Languages: SQL, Python, Java, R, HTML5, CSS, .NET
Data Engineering & Cloud: Apache Kafka/Spark, AWS (EC2, S3, RDS), Azure, SnowFlake, ETL/ELT Pipelines
Data Analysis: Minitab, MS-Access, Copilot, Tableau, Power BI
Continuous Improvement & Quality: Lean Six Sigma (DMAIC), Root Cause Analysis (Fishbone, 8D), ISO 9001, Quality Management Systems, Process Control

Experience

Data Engineer Intern — August 2025 - Present
Recaro Aircraft Seating — *Fort Worth, TX*

- Engineered high-throughput ETL data pipelines using Apache Kafka and Spark, orchestrating real-time data streaming from inspection line UIs into data warehouses, increasing KPI by **30%**.
- Increased productivity by saving 10 hours/week of manual reporting by creating a real-time **Power BI** dashboard using DAX and Python to visualize airline seat production lifecycles and resolve manufacturing bottlenecks.
- Reduced time tracking per task by **60%** by deploying a dedicated analytics portal on **AWS EC2** and optimizing PostgreSQL queries to provide teams with fast, real-time data visibility.

Quality Engineer Intern — May 2025 - August 2025
Recaro Aircraft Seating — *Fort Worth, TX*

- Facilitated the adoption of AI technologies through proof-of-concept (POC) initiatives, reducing product inspection cycle times by 50% via Computer Vision (YOLOv8) and Machine Learning models.
- Developed Python-based Agentic AI workflows utilizing **LangChain**, **LangGraph** to automate complex data analysis for weekly First Article Inspection (FAI) reports, reducing time for reports by 50% and improving data accuracy.

Data Strategist - Research Assistant — January 2025 - May 2025
University of Texas at Arlington — *Arlington, TX*

- Contributed to a lab researching integration of LLMs with map navigation systems under REU research grant
- Integrated REST API to adapt routes in a dynamic manner in response to real-time conditions and changes
- Migrated local PostgreSQL to **RDS** to support 70+ users, increasing concurrent user base by over 35%

Projects

Ad CTR Predition Model | *Python, Scikit-Learn, Pandas, Streamlit* — March 2026

- Developed a machine learning pipeline using Random Forest and **Logistic Regression** to predict digital ad click-through rates (CTR) and help optimize real-time bidding strategies.
- Cleaned and engineered a **10,000+** record dataset, using one-hot encoding on user demographics and web behavior to significantly improve the model's accuracy.
- Built an interactive web dashboard using Streamlit to visualize top-level campaign metrics (Total Impressions, Average CTR) and run live bid predictions for AdOps managers.

Insurance Claim Prediction | *Data Mining, R, Power BI, Python* — September 2025 - December 2025

- Developed a predictive regression model using Python and **R** to forecast individual insurance costs, achieving an R-squared score of 0.98 and a Mean Absolute Error of 10
- Engineered an analytical Power BI dashboard using data mining techniques to uncover and visualize critical cost trends across age, region, and health markers for over 1,300 individual records.