

BHANU TEJA KURAKULA

Chicago, IL | bhanutejakurakula@gmail.com | [Linkedin](#) | [Github](#) | +1-(217)926-2758



WORK EXPERIENCE

Data Analyst - Polygon Health Analytics

May 2025 - Present

- Developed a user-friendly **FinOps dashboard** to monitor token utilization, cost per API call, and usage trends, with a drill-down capability by model, user and service. Also, created the **data model** supporting this dashboard.
- Integrated LLM-based classification and auto-tagging features into an existing Java based vaccine-newsletter using OpenAI and Gemini APIs and rolled it out after rigorous testing to improve tag quality.
- Implemented near-duplicate detection with OpenAI **embeddings**, performed cosine similarity to reduce redundancy, conducted **code-reviews**, validated with backfill runs before a staged rollout to production.

Data Engineer - AGCO Corporation

June 2024 - May 2025

- Developed **data models**, designed a **Tableau** dashboard & built a **SQL** based **ELT** pipeline supporting **40** stakeholders to track product metrics and other operational KPIs, enabling better decision making.
- Engineered **10 ETL** data pipelines on **AWS**, standardized large volumes of sales data across EU with schema enforcement & rigorous data quality checks, used CloudFormation to automate deployment..
- Led migration of **20** legacy **alteryx** workflows to **Tableau Prep**, cutting licensing costs by **20000USD** annually, documented, trained and supported the end users from the business team in adopting the migrated workflows.
- Implemented a batch data pipeline on **Databricks** using **Delta Live Tables (DLT)** to process and transform an internal dataset, and created interactive dashboards using **Databricks SQL** for data visualization and reporting.

Data Engineer - Cognizant

September 2022 - July 2023

Client : CapitalOne

- Ensured **Snowflake** warehouse reliability by monitoring **200 key metrics** and **50+ jobs**, actively troubleshooting job failures and ensured timely report generation ahead of business use, meeting SLAs.
- Developed a POC using **MLaaS** with an aim to offer predictive analytics to the business users, performed EDA on a complex internal dataset, performed A/B testing for deploying the best model.
- Identified recurring bottlenecks, collaborated with stakeholders to **optimize SQL queries**, resulting in a **20 percent** reduction in query execution time, reducing delay by **20 minutes** on a year over year basis.

Software Engineer Intern - PhysioAI

September 2021 - Feb 2022

- Developed a full-stack ReactJS + Django yoga assistant using MediaPipe and pose-specific MLP classifiers, enabling accurate detection of **30** yoga poses at **30 FPS** with less than **40 ms** end-to-end latency.
- Productized a **ML model** with one binary model per pose, improving classification accuracy and reducing overlap between similar postures, collected **18K+** predictions per 10-minute session.
- Designed and implemented a **star schema-based dimensional model** to track **40 session-level KPIs** (pose hold duration, average confidence, recovery count), enabling detailed progress analytics across users.

Data Science Intern - Neuron.ai

September 2020 - January 2021

- Spearheaded the life cycle of **3** product categories of a warehouse by scraping over **10K+** images per category.
- Ensembled, trained and deployed the best performing CNN models to production, enhancing ops efficiency.

Web Analytics Intern - Eezyyweb

May 2020 - June 2020

- Developed 40+ websites for small to midsize businesses, optimizing them for top search engine rankings.
- Implemented strategies to improve key digital marketing metrics, including CTR, Impressions, and CPM.
- Integrated metric tracking plugins and developed a **centralized Google Analytics dashboard** to monitor performance, enabling data-driven decision-making and providing ongoing support to clients.

EDUCATION

University of Illinois, Urbana-Champaign (UIUC)

August 2023 - May 2025

Master of Science - Information Management

Relevant Coursework : Data Statistics, Data Modeling & Database Design, Data Warehouse & Business Intelligence

Amrita University

July 2018 - June 2022

Bachelor of Technology - Electronics and Communication Engineering

SKILLS

Programming Languages	: Python, SQL, Pyspark, JavaScript
Platforms & Frameworks	: AWS, Snowflake, Databricks, Kafka, Airflow, Flask, Tensorflow
Tools	: Tableau, PostgreSQL, Docker, Github, Alteryx
Miscellaneous	: Data Modeling, OOPs, Data Structures, Algorithms
Certifications	: Databricks Certified Data Engineer AWS Solutions Architect Associate AWS Machine Learning Specialty

PROJECTS

[Kafka-Based Real-Time Data Processing Pipeline](#) | Confluent Kafka, docker

- Developed a Kafka-based real-time data processing pipeline that efficiently consumes, processes, and publishes messages while ensuring robust error handling and summary statistics.
- Implemented modular components including a consumer, producer, transformer, and summary printer, utilizing multiple Kafka topics for effective data flow management.
- Designed the pipeline to anonymize sensitive data through hashing, ensuring compliance with privacy standards while providing real-time insights into processing metrics.

[Coupon Accepting MLOps Pipeline](#) | MLflow, EC2, S3, RDS, Terraform, Docker, Prefect, Github Actions

- Developed an end-to-end machine learning pipeline for vehicle coupon acceptance prediction using XGBoost, implementing robust data preprocessing, training, and deployment.
- Utilized MLflow for experiment tracking, Prefect for workflow orchestration, Docker for containerization, and AWS services for cloud infrastructure, ensuring scalability, reproducibility, and efficient collaboration across teams.
- Designed and implemented automated CI/CD pipelines with GitHub Actions, incorporating continuous monitoring, alerting, and model retraining mechanisms to mitigate data drift and sustain long-term model performance.

PUBLICATION

[Decentralized Web App for covid-19 using AI, IEEE-Xplore](#)

- Developed a Web3-based medical data storage solution integrating blockchain with deep learning to create a secure and tamper-proof environment for patient records.
- Built a full-stack application using Dash-Plotly, TensorFlow, Flask, ReactJS, Heroku, Ganache, Google Sheets, and Solidity to manage data and visualize results.
- Trained deep learning models to detect COVID-19 from chest X-rays and stored patient details and test results on an Ethereum blockchain via Ganache for decentralized and transparent access.

ACHIEVEMENTS

[Honeywell Hackathon Winner - 2021](#) : Secured **1st position out of 180 teams** for developing a POC on smart traffic system.
[Rising Star Award - 2022, Cognizant](#) : Awarded for being the **Best College Hire** based on leadership ratings.