

# Celdrick Kuta

Data Scientist

(309) 989-0419 Loganville, GA [kutaceldrick880@gmail.com](mailto:kutaceldrick880@gmail.com)

Website: [kuta-ndze.github.io](https://kuta-ndze.github.io) GitHub: [github.com/kuta-ndze](https://github.com/kuta-ndze) LinkedIn: [linkedin.com/in/kutaceldrick](https://linkedin.com/in/kutaceldrick)

## TECHNICAL EXPERIENCE

---

<b>Data Scientist</b>	3rd Oct 2022 – Present
<i>Mayo Clinic</i>	<i>Rochester, MN</i>
<i>(Current scope)</i>	<i>5th Jun 2024 – Present</i>

- Built **Inference Engine**, an AI-driven, context-aware, ontology-driven reasoning engine that generates patient-centric clinical views from structured EHR data with grounded semantic context.
- Established a pipeline to create and maintain a **patient centric knowledge graph** with a **grounded semantic layer** and **ontology linking** to support interoperability across patient-centric applications and role-based views in EPIC. Achieved a 30% reduction in time for physicians to access relevant patient data.
- Built fast prototyping workflows to spin up agents using **Google ADK** to detect and contextualize potential **adverse drug events**, enriching the semantic layer for patient medications.
- Delivered clinical decision support using **RAG** over care process models: built embedding index and retrieval, improving retrieval accuracy by 20% and improving matching quality by 25%.
- Developed ingestion pipelines using HL7 **FHIR R4** and LPR data in **BigQuery** to populate patient views in near real time, improving retrieval efficiency by 40%.

<i>(Earlier scope)</i>	<i>3rd Oct 2022 – 5th Jun 2024</i>
------------------------	------------------------------------

- Built “Find The Right Expert,” an NLP search solution recommending expert physicians based on user queries.
- Developed and deployed GCP microservices for entity negation and ICD/CPT search with full test coverage and API documentation, reducing response time from 2s to  $\leq 500$ ms.
- Implemented hybrid search in Elasticsearch using BioSent2Vec and ELSER, reducing latency by 60%.
- Trained Learning-to-Rank models (XGBoost, LTR) on 390MB+ datasets; improved precision, recall, and NDCG by 5%, 4%, and 3%.
- Built a Google BQML framework to speed model development, reducing modeling time by 60%.

<b>Data Scientist</b>	7th Jun 2021 – 26 Sept 2022
<i>Elevated Technologies</i>	<i>Houston, TX</i>

- Built an insurance model to predict claim severity after unforeseen events using supervised ML in Python.
- Reduced claim processing time by 20% and improved model accuracy by 10%.
- Used AWS SageMaker for training and AWS Glue for ETL pipelines.

<b>Data Scientist</b>	3rd Feb 2020 – 28th May 2021
<i>Optimum</i>	<i>Houston, TX</i>

- Built deep learning models for real-time defect detection and localization in steel manufacturing imagery.
- Reduced sales cycle from 2 months to 1 month and increased sales team performance by 40%.
- Deployed data and inference workflows using AWS (S3, Lambda).

## Data Scientist

7th Nov 2016 – 29th Jun 2018

*Panus Software*

*Buea, Cameroon*

- Built customer segmentation using k-means and DBSCAN with feature engineering and dimensionality reduction (PCA, t-SNE).
- Delivered Tableau and Power BI dashboards; expanded sales to nearby markets, increasing revenue by 250%.

## EDUCATION

---

<b>MSc. in Applied Statistics</b>	Illinois State University	May 2021
<b>MSc. in Computer Science</b>	University of Yaoundé	Jul 2016
<b>BSc. in Mathematics   Computer Science</b>	University of Buea	Jul 2010

## SKILLS

---

**Tools and Languages:** Python, R, SQL (PostgreSQL, MS SQL Server, MySQL), TopBraid, GraphDB, Git, Linux/Bash, Docker, Kubernetes, BigQuery, Databricks, Snowflake, Elasticsearch, MATLAB, JavaScript

**ML and GenAI:** NLP, RAG, embeddings, Learning-to-Rank, deep learning (CNN), prompt design, agent prototyping (Google ADK), knowledge graphs, ontology linking

**Collaboration:** Clinical stakeholder collaboration, requirements translation, technical documentation, solution design, rapid prototyping, debugging and incident triage, mentoring, presentations