Raahul Narayana Reddy Kummitha

knraahul@umd.edu | linkedin.com/in/raahul-narayana-reddy-k-7904a21aa | github.com/Knraahul | +1-240-791-7348 | College Park, USA 20740

EDUCATION

University of Maryland, College Park *College Park, MD, USA* Aug 2024 – May 2026
Master of Science (M.S.) in Data Science and Analytics

- **Relevant Coursework:** Data Science, Machine Learning, Advanced Machine Learning, NLP, Cloud Computing

New Horizon College of Engineering *Bengaluru, Karnataka, India* Aug 2017 – May 2021
Bachelor of Engineering (B.E.) in Information Science and Engineering (GPA: 3.55/4.0)

- **Relevant Coursework:** Data Structures, DBMS, Data Mining, Big Data, Python

EXPERIENCE

Senior System Engineer, *Infosys Limited — India* Nov 2021 – Jul 2024

- Built and optimized Talend-based ETL pipelines by translating business requirements into scalable designs and robust transformation logic; improved large-batch processing using chunk-based loads, cutting runtime by **∼50%** (from 4–6 hrs to 2–3 hrs).
- Performed Python-driven EDA and data quality checks to segregate null/irrelevant records, remove non-required data early, and ensure only validated datasets reached downstream processing and client delivery.
- Implemented error handling and reconciliation workflows by flagging failing records and routing them to a separate reject stream/log for reprocessing, improving pipeline stability and enabling smoother client deliveries.
- Built and enhanced Data Mart mappings across heterogeneous sources (including DB2 XML) and standardized reusable ETL framework components to improve reporting consistency and reduce repeated development effort.

Intern, *Defence Bio-Engineering & Electro Medical Laboratory (DEBEL), DRDO — India* May 2021 – Jul 2021

- Built an end-to-end skin cancer classification pipeline using Python, TensorFlow/Keras, and OpenCV, including dataset cleaning, image–label mapping, preprocessing (resize/normalize), label encoding, and train/validation/test splits.
- Trained and evaluated a CNN model with a structured validation workflow, using validation accuracy, confusion matrix analysis, and misclassification review to assess performance and improve model reliability.
- Conducted EDA (class distribution and demographic patterns such as age/gender/location) to understand dataset characteristics and support clear, data-driven reporting of results.
- Designed the solution to support low-resource screening, enabling image-based decision support without continuous access to medical experts and improving accessibility for rural communities.

Volunteer, *Elconics Automation Pvt. Ltd. — India* Jan 2024 – Jan 2024

- Supported multiple hands-on projects during a 3-day IoT and Robotics workshop at NHCE College by helping participants understand tools/software, troubleshoot implementation issues, and guide basic model development to ensure successful project completion.

PROJECTS

Student Performance Prediction System with AI Guidance - FastAPI, SHAP, Local vs Cloud Deployment

- Built the local FastAPI inference API with clean request/response schemas and integrated SHAP for per-prediction explainability.
- Connected predictions + SHAP outputs to an AI guidance agent to generate actionable student-focused recommendations.
- Contributed to documentation, local vs cloud performance/reliability comparison, and end-to-end architecture presentation in slides.

Speech-to-Text for Accessibility in Education - Streamlit, Whisper Large V3 (CPU), Transformers, BART, Docker

- Built a real-time accessibility app that captures live audio, transcribes speech continuously, and generates concise lecture summaries for hearing-impaired students.
- Implemented non-blocking audio capture (sounddevice callback + thread-safe queue) and CPU-only ASR using Whisper Large V3 via Hugging Face Transformers with incremental transcript assembly.
- Added BART Large CNN summarization plus a Streamlit UI (Start/Stop, live transcript, summary, Sessions page) with session storage (transcript.txt/summary.txt) and Docker deployment (port 8501, audio passthrough).

SKILLS

- **Programming Languages & Frameworks:** Python, SQL, C, C++, HTML5, FastAPI (REST APIs), Jupyter/Anaconda
- **Data Engineering & ETL:** Talend Open Studio (ETL jobs, reusable ETL framework patterns), Azure Data Factory, PySpark, Data Mart mappings, DB2 XML source integration, MySQL
- **Machine Learning & Modeling:** scikit-learn; Decision Tree, Random Forest, Gradient Boosting, SVM, Bayesian Ridge Regression, Polynomial Regression, model evaluation (Accuracy, Precision/Recall, F1, R^2, MSE/RMSE)
- **NLP & Statistical Testing:** TF-IDF, Multinomial Naive Bayes, DistilBERT, LSTM; hypothesis testing, Chi-Square tests, distribution analysis (box plots/bar charts)
- **Computer Vision:** CNN-based image classification, TensorFlow/Keras, OpenCV, validation/testing, confusion matrix, misclassification analysis
- **Explainability & Model Serving:** SHAP explainability, FastAPI Swagger docs (/docs), health monitoring (/health), AI recommendations via Hugging Face Inference API
- **Cloud & Containerization:** AWS (S3, SageMaker, Lambda, API Gateway, ECR, CloudWatch, Secrets Manager, KMS, IAM), Docker, Docker Compose
- **Data Analysis & Visualization / BI:** Pandas, NumPy, MS Excel, Statistics, Data Cleaning/Validation, Matplotlib, Tableau, Power BI