Kranthi Chaithanya Thota

New York, USA | kranthichaithanya3000@gmail.com | +1 315-603-7581 | linkedin.com/in/kranthi-chaithanya-thota
kc-3000.com | github.com/kc-3

Summary

Highly analytical Data Scientist & Master's Candidate (GPA 4.0) with 3+ years of experience in Machine Learning, Deep Learning, NLP, Computer Vision, and statistical analysis. Proficient in Python, SQL, TensorFlow, PyTorch, Scikit-learn, and cloud platforms (GCP, AWS). Skilled in developing predictive models, extracting insights from large datasets, and driving AI solutions that deliver measurable impacts, including >90% F1-scores, 85%+ model accuracy, and 22% operational improvements. Eager to apply advanced analytical skills to complex data science challenges.

Professional Experience

NLP - Graduate Research/Teaching Assistant, *Clarkson University, Potsdam, NY* Feb 2024 – Present

- Designed and delivered hands-on assignments for graduate-level NLP coursework covering text processing, embeddings, RNNs/LSTMs, Transformers (BERT), Large Language Models (LLMs), Retrieval-Augmented Generation (RAG), and chatbot architectures, enhancing student understanding of SOTA techniques.
- Mentored and guided over 50 graduate students in developing complex NLP projects using **Python**, **Hugging Face**, and **TensorFlow/PyTorch**, resulting in projects achieving **85%+ evaluation accuracy/F1-scores**.
- Developed a complete data pipeline for an energy grant analysis project: scraped **10k+ records** from public websites (**Beautiful Soup**), performed data cleaning/structuring (**Pandas**), automated ingestion into **PostgreSQL**, and applied a fine-tuned **Hugging Face Transformer** for Named Entity Recognition (NER), achieving a **90% F1-score** on identifying key grant entities.

C3G Consultant/Intern, *Clarkson University, Potsdam, NY* May 2024 - Present

- Engineered and deployed a Computer Vision model using **YOLOv8** and **Python/OpenCV** to analyze drone survey imagery of campus parking lots, achieving **95% accuracy** in real-time empty space detection and providing critical data for parking resource optimization.
- Led data collection efforts using drone surveys and metrocount devices; analyzed resulting traffic flow and volume data using **Python** and **SQL**, identifying bottlenecks and delivering insights that directly contributed to initiatives reducing peak-hour campus traffic congestion by **22%**.
- Developed and maintained real-time operational dashboards using **Tableau**, connecting to various data sources to visualize parking availability and traffic trends, enabling campus safety teams to optimize resource allocation and response.

Institutional Research Data Analyst, *Clarkson University, Potsdam, NY* Aug 2024 – Present

- Developed a data pipeline to convert IPEDS datasets from Access to SQL Server, boosting scalability and efficiency. Improved data integrity and reduced manual processing by 60%, streamlining workflows.
- Conducted in-depth statistical analysis of sensitive Harvard COACHE faculty satisfaction survey data using **Python (Pandas, SciPy)** and visualized findings in **PowerBI**, generating actionable insights on faculty retention drivers for university leadership while ensuring strict data governance and confidentiality.
- Designed, developed, and deployed interactive dashboards integrating diverse institutional data sources (student enrollment, degrees awarded, sponsored research, IPEDS) using **PowerBI/Tableau**, enabling data-driven planning and reporting across academic departments.
- Engineered an automated ETL pipeline using **Python** scripts to ingest, clean, validate, and transform federal IPEDS reporting data, loading into a dedicated **SQL Server** , improving data accuracy and reducing manual processing time by **70%**.
- Designed and developed the institutional research website, improving data discovery and reducing average data retrieval time by **30%** for over 500 faculty members through optimized design and data integration.

Associate Data Engineer, *Egen, Hyderabad, India* Jul 2022 – Dec 2023

- Led the design and development of an NLP-driven solution for patent analysis; utilized **Google Document AI** API and engineered complex **Regex** patterns in **Python** to extract key structured fields (inventors, claims, dates) from over 2,000 unstructured patent PDFs with high accuracy.
- Developed custom machine learning models and business logic integrated into the patent analysis pipeline to classify patent types and identify novelty metrics, delivering key insights not available through simple extraction.
- Built and deployed interactive dashboards using **Looker Studio** and **BigQuery** (connected via **Python** scripts) visualizing extracted patent data and model outputs with 25+ KPIs, reducing manual reporting time for the client by **40%**.
- Engineered and automated complex ETL workflows using **Apache Airflow** on **GCP**, orchestrating data ingestion, cleaning, ML model inference (patent analysis), and loading into **BigQuery**, improving overall data processing efficiency by **30%** and ensuring data reliability for analytical systems.
- Implemented automated CI/CD pipelines using **GCP Cloud Build** and **Git** for managing deployments of Cloud Functions,

Dataflow jobs, and BigQuery updates, reducing manual deployment errors by **40%** and enabling rapid iteration for 15+ microservices supporting data applications.

- Automated the provisioning and management of **GCP** infrastructure (Compute Engine, GCS, BigQuery datasets, Pub/Sub) for data processing and analytics using **Terraform (IaC)**, deploying **50+ resources** consistently across environments and reducing setup time by **70%**.
- Developed monitoring dashboards using **GCP Cloud Monitoring** and **Looker** to track system/pipeline performance (latency, throughput, error rates), enabling proactive issue detection and resolution.

Consultant Data Engineer, *SpringML, Hyderabad, India* Jan 2022 – June 2022

- Developed and automated ETL pipelines using **Google AppsScript** and **Python** to integrate data from Google Workspace applications into **BigQuery** for reporting and analysis.
- Developed internal budget management web tools using **Flask** and **Jinja** templates to streamline data management tasks for project teams.

Technical Skills and Certifications

Languages: Python (Pandas, NumPy, Scikit-learn, sciPy, TensorFlow, PyTorch, Flask, Requests, duckdb, Beautiful Soup), SQL (Expert: Joins, CTEs, Window Functions, Query Optimization), R (ggplot2, dplyr, tidyr, shiny, tidyverse), JavaScript, C++
Machine Learning & Deep Learning: Predictive Modeling (Regression, Classification, Time Series), NLP (Text Preprocessing, Embeddings, Topic Modeling, Sentiment Analysis, Transformers, BERT, GPT, RAG, NER), Computer Vision (Image Classification, Object Detection - YOLO, Image Processing, OpenCV, EfficientNet), Reinforcement Learning (Concepts, Libraries), Unsupervised Learning (Clustering - K-Means, Dimensionality Reduction - PCA), Feature Engineering, Model Evaluation & Selection, Hyperparameter Tuning
Statistics & Experimentation: Statistical Modeling, Hypothesis Testing (ANOVA, Chi-Square, T-tests), A/B Testing Design & Analysis, Regression Analysis, Probability Distributions, Data Segmentation, EDA (Exploratory Data Analysis)
Data Visualization: Tableau, PowerBI, Looker, Data Studio, Matplotlib, Seaborn, Plotly, D3.js, Data Storytelling
Cloud Platforms & Data Tools: GCP (BigQuery, Vertex AI, Dataflow, Cloud Run, Functions, GCS, Pub/Sub, Composer, Document AI), AWS (SageMaker, RedShift, Athena, S3, EMR, Lambda), Azure ML, Apache Airflow, Apache Spark (PySpark), Hadoop, Docker, Kubernetes, Terraform (IaC)
Databases: PostgreSQL, MySQL, SQL Server, BigQuery, RedShift, SQLite, MongoDB, Vector Databases
Certifications: Google Cloud Professional Data Engineer, Associate Cloud Engineer, Hashicorp Certified Terraform Associate

Education

Clarkson University Jan 2024 - Aug 2025
Master of Science in Applied Data Science (CGPA: 4.00/4.00) New York, USA
Kakatiya Institute of Technology and Science Jul 2018 - May 2022
Bachelor of Technology in Electronics and Communication (CGPA: 3.45/4.00) Warangal, India

Projects

Healthcare RAG Question Answering System (*Hugging Face Transformers, TF-IDF, Vector DBs, Python*)

- Architected & fine-tuned a Retrieval-Augmented Generation (RAG) system integrating **TF-IDF/Vector DB** retrieval with fine-tuned **Transformer models (BERT/T5)** to provide accurate, context-aware answers based on a corpus of healthcare documents, achieving high relevance scores in evaluations.

Named Entity Recognition (NER) Model Fine-tuning (*BERT, Hugging Face, TensorFlow/PyTorch, Python*)

- Developed a high-performance NER solution by fine-tuning **BERT** on a domain-specific dataset using **Hugging Face/TF/PyTorch**, achieving a **>92% F1-score** and demonstrating expertise in adapting large language models for specific information extraction tasks.

BRFSS Health Risk Prediction & Analysis (*TensorFlow, Python, Scikit-learn, Statistical Tests*)

- Engineered and validated multiple ML models (**TensorFlow, Scikit-learn**) on **400,000+** BRFSS health records, achieving **80% accuracy** in predicting chronic disease risk; conducted rigorous feature importance analysis using statistical tests (ANOVA, Chi-Sq) and model explainability techniques.

DeepFake Detection (Advanced Computer Vision) (*EfficientNet, TensorFlow, PyTorch, Python*)

- Developed and trained a deep learning model using **EfficientNet** architecture for DeepFake image detection, achieving **85% accuracy**; improved performance by **15%** through implementing advanced image preprocessing techniques like Adaptive Error Level Analysis (ELA).

Reinforcement Learning for Autonomous Control (*Python, RL Libraries, Simulation Environment*)

- Applied Proximal Policy Optimization (PPO) / Deep Q-Network (DQN) algorithms using **Python** and **RL libraries** within a simulation environment to successfully train an agent for stable autonomous quadcopter control, designing effective reward functions to guide learning.

AI Therapeutic Chatbot for Seniors (*Python, Flask, Ollama/Gemma, NLP Techniques*)

- Engineered a full-stack AI therapeutic chatbot prototype using local LLMs (**Ollama/Gemma**), implementing empathetic conversational flows, context management, and secure user handling (**Flask** backend, voice I/O).