Kranthi Chaithanya Thota

kranthichaithanya3000@gmail.com | +1 315-603-7581 | linkedin.com/in/kranthi-chaithanya-thota | kc-3.github.io/portfolio

Summary

Passionate Data Scientist & Master's Candidate (GPA 4.0) with 3+ years experience applying Machine Learning, Deep Learning, NLP, Computer Vision, and statistical analysis to solve complex problems and drive data-informed decisions. Proficient in Python, SOL, TensorFlow, PyTorch, Scikit-learn, Hugging Face, and cloud platforms (GCP, AWS). Proven ability to build predictive models, extract insights from diverse datasets, and develop innovative AI solutions, delivering impacts such as >90% F1-scores, 85%+ model accuracy, and 22% operational improvements. Eager to contribute advanced analytical and modeling skills to challenging data science initiatives.

Education

Clarkson University, New York, Master's in Applied Data Science, GPA 4.0

Professional Experience

NLP - Graduate Research/Teaching Assistant, Clarkson University, Potsdam, NY

- Elevated 50+ students' NLP model building skills (Transformers, RAG, Hugging Face) achieving 85%+ project accuracy.
- Developed data pipeline scraping 10k+ records (Beautiful Soup), applying NER (Hugging Face, 90% F1) to structure data. C3G Data Specialist/Intern, Clarkson University, Potsdam, NY May 2024 - Present
- Deployed a **YOLOv8**-based Computer Vision model for **95% accurate** parking space detection, optimizing resource allocation.
- Analyzed campus traffic patterns using drone/metrocount data, delivering insights that reduced peak-hour congestion by 22%. Aug 2024 – Present
- Institutional Research Data Analyst. Clarkson University. Potsdam. NY
- Analyzed Harvard COACHE faculty satisfaction survey data using PowerBI, Python and Pandas, ensuring data governance standards and maintaining data integrity to generate actionable insights for university leadership.
- Developed the institutional research website and interactive dashboards, adhering to data governance policies and enhancing data accuracy, reducing data retrieval time by **30%** for **500+** faculty members.
- Associate Data Engineer, Egen, Hyderabad, India
- Led patent data extraction/analysis using Google Document AI and text processing for 2,000+ records, developing models/logic and delivering insights via Looker Studio.
- Automated ETL workflows (Apache Airflow) delivering reliable data for analytical systems, boosting efficiency by 30%.
- Developed dashboards (Looker, Python, BigQuery) analyzing system/pipeline performance, cutting manual reporting by 40%.
- Automated infrastructure (Terraform IaC) for data processing and analytics environments on GCP.

Projects

Healthcare RAG Question Answering System (Hugging Face (T5/BERT), TF-IDF, Vector DBs)

- Architected & fine-tuned a Retrieval-Augmented Generation (RAG) pipeline using Transformers and Vector DBs for document-grounded healthcare QA, ensuring accurate, context-aware responses.
- Named Entity Recognition (NER) Model Fine-tuning (BERT, Hugging Face, TensorFlow/PyTorch)
- Developed high-performance NER solution by fine-tuning BERT on domain-specific data, achieving >92% F1-score and demonstrating expertise in leveraging large language models.
- BRFSS Health Risk Prediction (TensorFlow, Python, Statistical Tests)
- Engineered and validated a TensorFlow model on 400,000+ health records, achieving 80% accuracy in predicting health risks using rigorous statistical feature validation (ANOVA, Chi-Sq).

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

• Developed EfficientNet model for DeepFake detection (85% accuracy); enhanced performance 15% via advanced preprocessing (Adaptive ELA), demonstrating CV expertise.

Twitter Sentiment Analysis Dashboard (Python, NLP Libraries, Looker/Tableau)

- Built automated sentiment analysis pipeline (Python, NLP) for Twitter data, visualized insights in Looker/Tableau, reducing manual reporting by 40%.
- E-commerce Customer Segmentation (Python, Scikit-learn, SQL, Visualization)
- Performed customer segmentation using K-Means (Scikit-learn, Python) on e-commerce data, providing actionable insights for targeted marketing campaigns.

Technical Skills & Certifications

Languages: Python, SQL, R, JavaScript, C++

Machine/Deep Learning: TensorFlow, PyTorch, Scikit-learn, Keras, Hugging Face Transformers, spaCy, LLMs, RAG, BERT, YOLO, EfficientNet, RL Libraries, Predictive Modeling, Feature Engineering, Model Evaluation

Statistics & Analysis: Statistical Modeling, Hypothesis Testing (ANOVA, Chi-Sq, T-tests), A/B Testing, Regression Analysis, Data Segmentation, EDA

Data Viz & Reporting: Tableau, PowerBI, Looker, Data Studio, Matplotlib, Seaborn, D3. js, Report Development, Dashboarding Cloud & Data Tools: GCP (BigQuery, Dataflow, Vertex AI, GCS, Pub/Sub), AWS (RedShift, Athena, S3, SageMaker), Azure, Apache Airflow, Spark, Hadoop, Docker, Kubernetes, Terraform, Pandas, NumPy

Databases: PostgreSQL, MySQL, SQL Server, BigQuery, RedShift, SQLite

Certifications: Google Cloud Professional Data Engineer, Associate Cloud Engineer, Hashicorp Certified Terraform Associate

Jan 2024 - Aug 2025

Aug 2024 – Present

Jul 2022 – Dec 2023