SHAMITHA REDDY

Contact: +1 (940)765-3763 | Email: cheedushamithareddy@gmail.com | LinkedIn

DATA SCIENTIST | MACHINE LEARNING ENGINEER | DATA ENGINEER

Data Science graduate with strong hands-on experience in machine learning, multimodal AI, data analytics, and scalable data systems. Proven ability to design end-to-end ML solutions, build privacy-preserving AI systems, and transform complex datasets into actionable insights. Experienced across NLP, deep learning, federated learning, and cloud-based data workflows, with a strong focus on real-world, production-style problem solving.

TECHNICAL SKILLS

- **Programming Languages:** Python, SQL, Java, C, C++ (working knowledge)
- **Machine Learning & AI:** Supervised & Unsupervised Learning, Logistic Regression, Random Forest, XGBoost, SVM, KNN, LSTM, Transformers, Multimodal AI, Federated Learning
- **Natural Language Processing:** TF-IDF, Word2Vec, BERT, RoBERTa, Text Embeddings, Semantic Similarity
- **Data Engineering:** ETL Pipelines, Data Cleaning, Feature Engineering, Data Validation, Apache Spark (PySpark)
- **Databases:** MySQL, PostgreSQL, MongoDB
- **Cloud & DevOps:** Google Cloud Platform (BigQuery, Dataproc), AWS (EC2, S3 – working knowledge), Docker, Git, GitHub
- **Data Visualization**: Tableau, Power BI (basic), Matplotlib, Seaborn
- **Tools:** Jupyter Notebook, Google Colab, VS Code
- **Methodologies:** Agile, SDLC, CRISP-DM

PROJECTS

AGENTIC MULTIMODAL MEDICAL ASSISTANT | Federated Learning–Based Clinical Intelligence Platform | OCT 2025 – PRESENT
Role: Machine Learning Engineer / Graduate Research assistant / Data Annotations

- Designed and implemented a privacy-preserving multimodal medical assistant combining clinical text and medical images for real-time decision support.
- Integrated federated learning to enable distributed model training across decentralized medical datasets while ensuring patient data never left local nodes.
- Built multimodal pipelines leveraging vision-language representations to align medical images with textual clinical context.
- Applied feature fusion and embedding-based retrieval to support accurate and context-aware medical responses.
- Evaluated models using Precision-Recall AUC (PR-AUC) and recall-focused metrics, prioritizing reduction of false negatives in clinical scenarios.

DIGITAL ARCHIVE AI PLATFORM | Metadata Quality & LLM-Based Optimization | UNIVERSITY OF NORTH TEXAS | JAN 2025 – MAY 2025
Role: EDA, API Accessing, Built OCR to LLM Platform / Graduate Student Assistant / Research Data Scientist

- Designed and implemented an LLM-driven metadata quality evaluation system for large-scale digital archives, focusing on accuracy, completeness, and semantic consistency.
- Built OCR-to-LLM pipelines to extract, evaluate, and enrich metadata from unstructured document collections.
- Applied text embeddings and semantic similarity techniques to identify missing, inconsistent, and low-quality metadata fields.
- Integrated GPT-4–powered prompt engineering workflows to automatically improve metadata descriptions and classifications.
- Improved downstream search, retrieval, and archival usability through vector-based similarity scoring and metadata normalization.

SOCIAL MEDIA ANALYTICS PLATFORM | Real-Time Sentiment Intelligence | AUG 2024 – DEC 2024
Role: Machine Learning Engineer / NLP & Deep Learning

- Built a real-time Twitter sentiment analysis platform using LSTM-based deep learning models to classify user sentiment at scale.
- Designed NLP pipelines for text cleaning, tokenization, sequence padding, and feature extraction.
- Trained and evaluated models using precision, recall, F1-score, and confusion matrices, prioritizing robustness over raw accuracy.
- Enabled real-time sentiment trend analysis to support brand monitoring and social intelligence use cases.
- Demonstrated the effectiveness of deep learning models for capturing temporal dependencies in text streams.

ONLINE GAMING PLATFORM | Leaderboard & Database Management System | JAN 2024 – MAY 2024
Role: Database Design & SQL

- Designed and implemented a relational database system to manage player scores and real-time leaderboard rankings.
- Created normalized MySQL schemas, optimized SQL queries, and indexing strategies for high-performance reads.
- Integrated Python-based data ingestion pipelines for leaderboard updates and analytics.

- Improved query efficiency and data integrity through schema optimization.

NETWORK SECURITY PLATFORM | Intrusion Detection & Threat Prediction | DEC 2022 – APR 2023
Role: Machine Learning Engineer

- Built a machine learning-based intrusion detection system (IDS) to classify network traffic as normal or malicious.
- Performed data preprocessing, feature engineering, and supervised model training.
- Evaluated models using classification metrics with emphasis on reducing false negatives in security-critical scenarios.
- Improved query efficiency and data integrity through schema optimization.

EDUCATION

UNIVERSITY OF NORTH TEXAS | MASTERS IN DATA SCIENCE | JAN 2024 – DEC 2025

CERTIFICATIONS

AWS Certified Cloud Developer – Associate | AWS ML Foundation | IBM Data Visualization with Python | Tata GenAI-Powered Data Analytics | NPTEL Data Science for Engineers