Vikas Reddy
Data & Analytics Engineer | ETL Pipelines, Data Warehousing
(443)-992-2554 | wh22215@umbc.edu | linkedin.com/in/vvikasreddy | github.com/vvikasreddy | Portfolio

EDUCATION

University of Maryland Baltimore County (UMBC)   Baltimore, MD
Master's in Data Science; GPA: 4.0   *Aug. 2023 – May 2025*

Institute of Aeronautical Engineering (IARE)   Hyderabad, India
Bachelor's of Technology in Computer Science and Engineering; GPA: 3.41   *May 2019 – May 2023*

EXPERIENCE

Brain-Machine Interface Lab at UMBC   Baltimore, USA
ML Researcher, Data & Analytics Pipeline Development   *Jan. 2024 – Present*
- Developed an integrated data pipeline combining ConvLSTM and Transformer models to process multimodal EEG and facial data; improved workflow scalability and accuracy (97%), featured at **BSN Conference**.
- Created EmoFormer, a SegFormer-based analytics layer for emotion recognition; enabled consistent, benchmarked output supporting unified data views across research projects.
- Partnered with industry collaborator (Brainwave Science) to develop a Kafka-powered real-time data pipeline for EEG data; increased live classification accuracy to 81% while ensuring continuous data quality monitoring
- Led data sourcing and transformation for 25+ EEG participants; developed a repeatable framework for data validation, cleaning, and operational readiness.

Epam Systems   Hyderabad, India
Intern, Data Engineering and Automation   *Jan. 2023 – June 2023*
- Built cloud-native ETL workflows using AWS (Lambda, S3, Athena) and Hadoop to support high-volume operational dashboards (50GB processed); used in release readiness reviews by QA teams.
- Automated backend data quality checks and reporting pipelines in Python/Java; integrated Agile workflows to iterate directly with cross-functional business stakeholders on key metrics.
- Enhanced CI/CD operations using Kubernetes and Jenkins, reducing deployment cycles by 35% and improving time-to-insight for internal tools.
- Contributed to scalable, test-driven development (SOLID principles); cut production issues by 30% while aligning technical solutions with client delivery timelines.

IndicWiki Project   Hyderabad, India
Intern, Client-Facing Data Analyst   *Mar. 2022 – June. 2022*
- Built and maintained an automated pipeline to ingest, clean, and normalize 10,000+ multilingual Wikipedia entries; boosted content accessibility for 50M+ users.
- Optimized ETL operations using Python and SQL, cutting processing time by 30% while sustaining 98% data accuracy for dashboard-ready outputs.
- Collaborated with translators to align data processing with business goals; delivered insight reports and visualizations via Tableau and Matplotlib.
- Created reusable documentation for data pipeline processes, cutting new analyst onboarding time by 40% and promoting consistent data handling.

PROJECTS

Stock Price Prediction | *Python, TensorFlow, LSTM, NumPy, Matplotlib*   GitHub
- Built and deployed an LSTM model for stock price prediction; achieved 84% of predictions within $\pm10\%$ error, demonstrating robust forecasting.
- Improved model performance via data preprocessing, tuning, and cross-validation for higher prediction reliability.

Lexically Constrained Machine Translation | *Python, PyTorch, Hugging Face*   GitHub
- Built and fine-tuned large language models (LLMs) with custom beam search to guide word choice, enabling more accurate and user-controlled translations.
- Improved translation performance on WMT (TR-EN) data, yielding BLEU gains over 300+ tests.

TECHNICAL SKILLS

Languages: Python, SQL, Java, R, C++, JavaScript, JSON
Frameworks & Platforms: TensorFlow, PyTorch, Keras, PySpark, Flask, Hugging Face Transformers, LangChain
Developer Tools: Docker, Kubernetes, Jenkins, Git, AWS, VS Code, JIRA
Data Tools: Tableau, Power BI, Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn, OpenCV
Other Skills: ETL Pipelines, Data Modeling, CI/CD, NLP & LLMs, Agile (Scrum/Kanban), SOLID Principles