

Vinyas Naidu Karri

+1 (857) 506-4664 — karri.vi@northeastern.edu — <https://www.linkedin.com/in/vinyasnaidukarri/> — Boston, MA - 02120

EDUCATION

Northeastern University, Boston, USA

Dec 2025

Master of Science in Data Science and Analytics

GPA: 4.0/4.0

Course work: Foundations for Data Analytics Engineering, Data Mining Engineering, Data Management for Analytics, Computation and Visualization for Analytics, Statistical Learning for Engineering, Neural Networks and Deep Learning, Machine Learning Operations, Generative AI (Audit Course)

WORK EXPERIENCE

Data Scientist, ZS Associates (Full-Time), Pune, India

July 2022 - Dec 2023

Data Scientist, Consulting — Advanced Excel SQL, Pyspark, PowerBI, Machine Learning, Automating reports

- Led data science projects for Fortune 500 clients, aligning project goals to business objectives through data-driven insights and client engagement.
- Created interactive dashboards with PowerBI and Tableau, delivering key insights for cross-functional teams to improve strategic decision making.
- Developed a machine learning model for a new product launch, achieving a 2% market share increase through targeted segmentation of high-risk healthcare providers. (Predictive Modelling)
- Utilized logistic regression and A/B testing to identify key drivers, improving therapy adherence by 93%.

ZS Associates (Internship), Pune, India

Jan 2022 - June 2022

Junior Data Scientist, Consulting — Advanced Excel SQL, R, Machine Learning

- Developed a marketing mix model for US pharmaceutical products using Bayesian methods and generalized linear models, improving forecasting accuracy by 15%.
- Conducted data cleaning, preprocessing, and exploratory data analysis also Utilized R and Python to build predictive models and optimize promotional strategies for healthcare clients.

PROJECTS

Human Activity Recognition (Video Data): TensorFlow, CNN, LSTM, OpenCV

- Developed human activity recognition system using ConvLSTM and LRCN, achieving 87% test accuracy.
- Processed 5,000+ video frames using OpenCV for frame extraction, resizing to 64x64, and normalization.
- Optimized a model with 1.2M parameters using dropout (0.4), max pooling, early stopping to prevent overfitting.

End-to-End Text Summarization System Using Hugging Face and Fast API:

- Engineered and deployed a text summarization tool by fine-tuning Hugging Face's Pegasus model, utilizing Google Colab for development and FastAPI for real-time API deployment.
- Implemented a robust data processing pipeline and conducted rigorous model training and evaluation, consistently monitoring ROUGE metrics to assess and improve summarization performance in production environments.

End-to-End RAG LLM App Using LlamaIndex and OpenAI

- Built a LLM application using 2 APIs LlamaIndex and OpenAI, focusing on retrieval-augmented generation to enhance querying capabilities and manage multiple PDFs.
- Integrated OpenAI NLP for optimizing indexing techniques for improved retrieval accuracy and high quality outputs.

End-to-End Machine Learning & ETL Pipeline Development: Python, AWS (EC2, S3, RDS), Docker, MLflow, GitHub Actions

- Engineered and deployed an end-to-end ML and ETL pipeline using Python, MongoDB Atlas, and Docker for robust data ingestion, validation, and transformation (Data pipelines).
- Optimized ML models using hyperparameter tuning and managed the models with MLflow integrated with DagsHub.
- Automated CI/CD pipelines using GitHub Actions, streamlining model training, evaluation, and deployment on AWS EC2. Leveraged GitHub for code version control and collaborative development.

TECHNICAL SKILL

Analytics Data Science: Python, R, A/B Testing, Tableau, Power BI, Advanced Excel (Pivots, VBA, Functions), SQL, NoSQL, Pyspark, Hive, Hadoop, Redshift, Snowflake, Agentic AI, Langchain, AutoGen, Semantic Kernel

Programming Machine Learning: Pandas, NumPy, Matplotlib, Seaborn, Machine Learning Algorithms, TensorFlow, PyTorch, CNN, RNN, LSTM, Transformers, LLMs, Natural Language Processing

Data Engineering Cloud Technologies: AWS, Docker, Apache Airflow, Kafka, Astronomer, PostgreSQL, Amazon Redshift, DVC, ETL, CI/CD, Git, Flask API, Fast API, Kubernetes