Venkata Revanth Jyothula

New York City, New York | jyorevanth@gmail.com | (917)-618-5892 | LinkedIn | Github

EDUCATION

New York University, **Master of Science in Computer Engineering,** ***GPA: 3.7/4, Cum Laude*** **August 2023 - May 2025**
Relevant Coursework*: Machine Learning, Computer Systems Architecture, Data Structures and Algorithms, Advanced Machine Learning, Deep Learning, Programming for Data Science, Computer Vision, Big Data, ML Operating Systems*

Sardar Patel Institute of Technology, **B.Tech. in Electronics and Telecommunication,** ***GPA: 3.4/4*** **August 2019 - July 2023**
Relevant Coursework*: DBMS, Python, Artificial Intelligence, Computer Networks, Computer Architecture, Soft Computing*

EXPERIENCE

Data Engineering Intern**,** 8th Element AI **January 2023 - July 2023**

- Built and deployed a scalable **ETL pipeline** using **Python**, **Apache Spark**, and **Azure Databricks**, reducing data processing time by **40%** and optimizing storage across distributed systems for both **batch and streaming** workloads.
- Curated and transformed **30K+ healthcare records** from diverse sources by designing big data schema models using **dimensional modeling**, **partitioning**, and **Delta Lake**, improving data quality, consistency, and analytical readiness for BI reporting.
- Validated and cleansed structured/unstructured data using **SQL**, enhancing trust in analytics pipelines by reducing data errors by **60%**, and supporting a smooth system upgrade in collaboration with engineering and analytics teams.

Data Engineering Intern**,** Larsen and Toubro Infotech (Now LTIMindtree) **June 2022 - August 2022**

- Led a multi-source data migration project to **Snowflake**, transforming fragmented business data into a unified training-ready schema using **SQL** and **Python**, improving processing efficiency by **20%**.
- Automated preprocessing and quality checks on **50K+ records** using **Pandas**, **SnowSQL**, and **Jupyter**, boosting data accuracy by **30%** and reducing pipeline failures by **50%** during M&A consolidation—collaborated with the Data Principal and cross-functional teams to ensure smooth integration.

PROJECTS

Twitter Content Moderation System, MLOps Engineer, Project **January 2025 - May 2025**

- Designed and implemented an end-to-end MLOps pipeline on Chameleon for content moderation using BERT, achieving **95% classification accuracy**; incorporated modular preprocessing, automated training, and experiment tracking with MLflow.
- Deployed scalable model inference via **Docker**, **Kubernetes**, and **Terraform**, enabling **70% faster rollout** of retrained models through staged CI/CD (canary, staging, production) and reproducible versioning via MinIO.
- Orchestrated distributed training and ETL pipelines using **Ray** and **Argo Workflows**, supporting weekly retraining cycles, cutting manual effort by 40% and boosting pipeline reliability by 50% across cloud compute nodes.

NBA Player and Performance Analysis, Data Engineer, Project **September 2024 - January 2025**

- Engineered an end-to-end data pipeline with Python, Pandas, and **NumPy** to clean, preprocess, and analyze NBA player statistics, improving dataset processing efficiency by **50%**. Conducted extensive EDA to uncover feature correlations and distributions.
- Integrated and harmonized multi-source player data from regular season and playoffs, applying transformation logic and **statistical feature engineering** to enable comparative performance analytics across advanced metrics.

Music Recommendation System Using NLP and Spotify API, Data/ML Engineer, Project **September 2024 - January 2025**

- Deployed a music recommendation system using Spotify and **Genius APIs**, processing data with **Spark** for scalable ingestion.
- Improved recommendation accuracy by 25% using **TF-IDF** vectorisation and cosine similarity, optimized API calls to reduce latency by 35%, and integrated user behavior to boost relevance.

Enhanced Resume/CV Matching Using Doc2Vec and Deep Learning, ML Engineer, Project **January 2024 - May 2024**

- Developed a job-matching system using **Doc2Vec**, processing large text datasets to achieve 90% accuracy in semantic matching between resumes and job descriptions, optimizing recruitment for IT and software roles.
- Enhanced data preprocessing with **NLTK** and **SpaCy** (tokenization, lemmatization), optimizing data handling by 25%, and integrated Gensim to boost matching precision by 45%, ensuring high-quality input for downstream modeling.

TECHNICAL SKILLS

Programming Languages: Python, SQL, R, Lua, C, Linux, MATLAB, Bash
Tools/Technologies: MLflow, Docker, Kubernetes, Terraform, Ansible, FastAPI, Git, Azure, Snowflake, MongoDB, AWS
Machine Learning: TensorFlow, Pytorch, Scikit-learn, Pandas, Spark, NLP, CNNs, RNNs, Transformers, XGBoost, NumPy

PUBLICATIONS IN IEEE

- **Fault Detection and Classification in Microgrid Using Machine Learning Techniques,** Publication **ICCCNT 2023**
- **Tropical Cyclone Intensity Classification Using Convolutional Neural Network,** Publication **I4Tech 2023**
- **Effective Malware Detection Using Gradient Boosting and Convolutional Neural Network,** Publication **IBSSC 2023**