Venkata Krishna Anirudh Nuti

Phone: 857-919-1884 | nuti.krish4@gmail.com | linkedin.com/in/nvkanirudh/ | Portfolio Website

SKILLS

- **Programming Languages**: SQL (*Advanced*), SparkSQL, Python (*Proficient*), R
- **Data Engineering:** ETL/ELT, Data Pipeline Development, Data Modeling, Data Warehousing, Apache Airflow, Apache Kafka
- **Cloud & Big Data:** AWS (S3, EC2, ECR, CodePipeline), Azure Databricks, Snowflake, Apache Spark
- **Data Visualization:** Tableau, Power BI
- **Databases:** MySQL, PostgreSQL, SQL Server, NoSQL (MongoDB), Vector Stores
- **ML Frameworks:** TensorFlow, PyTorch, Scikit-learn, Keras, MLFlow

EXPERIENCE

Data Analyst, *ATC - Boston, MA* *Feb 2023 - Jan 2024*

- Designed and implemented ETL processes, transitioning from Excel to **RDBMS**, reducing data access time by **40%** and improving reporting accuracy.
- Developed a cloud-based data warehouse solution using **Snowflake**, enabling the extraction of valuable insights and improving strategic planning.
- Optimized SQL Server queries, resulting in **30%** faster performance and enhanced decision-making capabilities.
- Conducted in-depth data analysis using **SnowSQL**, producing reports that directly informed strategic decisions.
- Improved data management processes, enhancing data integrity and reducing errors by **25%**.

Machine Learning Technical Engineer, *BU Spark - Boston, MA* *Sept 2022 - Dec 2022*

- Integrated CRAFT object detector and Tesseract OCR using **TensorFlow**, boosting text recognition efficiency by **30%** and reducing manual data entry time.
- Implemented name extraction system for Herbarium specimens, cutting identification time by **40%** and increasing processing throughput by **50%**.
- Mentored 20 students of Data Science with Python course, leading 4 projects with a 90% completion rate and enhancing student understanding of data science projects.

Data Scientist, *Intain Technologies - Chennai, India* *Feb 2019 - Aug 2020*

- Developed and deployed a Mask-RCNN model for document field extraction with **95%** precision, reducing data processing time by **70%** and enhancing data accuracy.
- Automated field extraction with Docker and Flask, cutting manual data entry by **80%**, saving **200+** work hours monthly, and improving workflow efficiency.
- Enhanced field extraction accuracy to **97%** with Mask-RCNN, surpassing RegEx by **7%**, reducing errors by **20%**, and increasing data reliability.
- Managed version control with **GitHub** and communicated technical insights to stakeholders, leading to better project alignment and informed decision-making.

PROJECTS

Toll Data Streaming: [Website]

- Optimized Highway Toll data processing using Apache Airflow for ETL, reducing **PostgreSQL** data loading times by **40%.**
- Implemented Kafka for real-time data streaming, improving **SQL Server** ingestion speed by **30%** and reducing latency for real-time updates by **20%.**
- Enhanced shell scripts to ensure seamless data loading, resulting in a more efficient toll data management system with improved processing time and responsiveness.

HR Analytics Dashboard: [Website][Tableau]

- Conducted HR analytics using **MySQL** and **Tableau** on a real-world dataset of employee records to surface insights on attrition trends and job satisfaction.
- Developed and executed **SQL** queries independently to analyze **key metrics** including attrition rates, performance ratings, and demographic factors.
- Created a Tableau dashboard to visualize patterns across departments, age groups, and education fields.

News Sentiment Analysis using Spark: [Website]

- Conducted sentiment analysis on news headlines with **PySpark** and **Azure Databricks**, achieving **85%** accuracy with SVM.
- Implemented topic modeling using LDA, identifying **15** latent topics in the corpus, and enhancing content analysis capabilities.

Medical Charges Prediction: [Website]

- Developed a predictive model using RandomForest, Gradient Boosting, and CatBoost with scikit-learn framework, achieving a **15%** accuracy improvement.
- Deployed model on **AWS EC2** via Docker, reducing inference time by **25%**, and leveraged AWS ECR via CodePipeline, streamlining the deployment process.

EDUCATION

Master of Science in Applied Data Analytics, **Boston University** - *Boston, MA* *Jan 2023*

Bachelor of Technology in Computer Science, **GITAM University** - *Visakhapatnam, India* *April 2019*