Nitheeshwar Kalaivani Vetriselvan

+1(716)-939-6902 • nitheeshwarkvn2091@gmail.com • www.linkedin.com/in/nitheeshwar-k-v • www.github.com/Nitheeshwar

PROFESSIONAL SUMMARY

Master's student in Computer Science with experience building scalable data pipelines and cloud-native ETL workflows. Proficient in Python, SQL, and AWS, with strong skills in data modeling and performance optimization. Proven contributor to open-source and cross-functional data projects. Seeking Data Engineering roles to drive efficient, production-ready data solutions.

EDUCATION

MS in Computer Science and Engineering ***Aug 2023- Dec 2024***
University at Buffalo, Buffalo, New York CGPA: 3.50

- ***Focused Course work:*** *Database system, Operating System, Data Intensive Computing, Information retrieval*

B. Tech in Electronics and Computer Engineering ***Aug 2019-May 2023***
Vellore Institute of Technology, Chennai, TN, India CGPA: 8.31

- ***Focused Course Work:*** *Machine learning using python, Cloud and distributed computing*

CERTIFICATION

- **AWS Certified Cloud Practitioner**

TECHNICAL SKILLS

- **Programming & Data Handling:** Python, SQL, PySpark, C, C++, JavaScript
- **Frameworks & Tools:** Airflow, FastAPI, Streamlit, TensorFlow, Git, Linux
- **Cloud & Big Data:** AWS, GCP, Snowflake, Hadoop, Spark, PostgreSQL, DynamoDB, Kubernetes

EXPERIENCE

Machine Learning Engineer *Greenstand, Remote, United states* *Jan 2025 - Present*

- Developed and optimized **ETL pipelines** using **Python** and **Airflow** for cleaning and annotating tree image datasets.
- Automating species identification with **AWS**, **Apache Airflow DAGs** and **Kubernetes** for efficient dataset tagging.
- Training and deploying image segmentation models on **AWS Sage Maker**, leveraging **S3 Buckets** for dataset storage.
- Supporting **CI/CD pipelines** and **contributing to open-source** tree-tracking software for global reforestation efforts.
- Built lightweight pipelines to monitor model drift and maintain **real-time accuracy** in species classification.

Research Intern *Vellore Institute of Technology, Chennai, TN, India* *Jan 2022 - Jan 2023*

- Designed a **mobile fundoscopy device** integrated with **Jetson Nano** and a camera module for **retinal anomaly detection**.
- Developed a two-stage ensemble **CNN** model using **TensorFlow**, achieving **98% accuracy** through design optimizations.
- Lead **Agile-driven product development**, including requirements analysis, hardware prototyping, and model optimization.
- Published a **design patent** through collaboration with the University of Liverpool, UK, and Sankara Nethralaya, India.
- Aligned ML insights with stakeholder goals and **cross-communicated** outcomes across technical and non-technical teams.

ACADEMIC PROJECTS

Job Resume Matcher & Skill Gap Analyzer
Python • FastAPI • SQLite • FAISS • Streamlit • Airflow • PostgreSQL

- Built an end-to-end **ETL+LLM pipeline** to extract resume-job skills, calculate match scores, and visualize gaps.
- Leveraged **Sentence Transformers** and **FAISS** to compute semantic similarity; orchestrated via **Airflow DAGs**.
- Deployed a frontend with real-time feedback, using **Streamlit**, connected to a resume database (**PostgreSQL**) for persistence.

News ETL Data Pipeline with Airflow
Python • Pandas • News API • SQLite • Apache Airflow

- Built an **ETL pipeline** to extract, clean, and load news data into SQLite using **Pandas**.
- Automated data ingestion and transformation with **Apache Airflow DAGs** for regular scheduled runs.
- Designed **structured schema** and applied cleaning to ensure accurate, query-ready data storage**.**

Building a Mini SQL Database (Taco-DB)
C++ • SQL • Docker

- Developed a SQL-based mini database in C++ with **B-tree indexing**, containerized in **Docker for scalability**.
- Optimized **SQL JOIN algorithms** and utilized **POSIX I/O** for high-performance data storage and retrieval.
- Enhanced query execution performance through advanced manual optimization of **SQL query execution plans**.

Personalized Anime Recommendation System
FastAPI • AWS EC2/S3 • Content Filtering • Pandas

- Developed a **content-based recommendation engine** using genre matching to improve recommendation accuracy.
- Implemented real-time recommendations with **FastAPI backend**, optimizing queries for parallel execution.
- Integrated **AWS EC2** and **S3** for scalable, low-latency data retrieval and efficient model hosting.