Siddharth Hiraou

Buffalo, NY 7167509361 sidhiraou@gmail.com LinkedIn GitHub Website

Education

University at Buffalo, Buffalo, New York **August 2024 – January 2026**
Masters in Data Science *GPA: 3.6/4.0*
Coursework: Data Intensive Computing, Machine Learning, Statistical Learning, Data Models & Query Language, Database Fundamentals

International Institute of Information Technology, Pune, India **June 2020 – June 2024**
Bachelor of Engineering in Information Technology *GPA: 4.0/4.0*
Coursework: Data Structures & Algorithms, Big Data Analytics, Cloud Computing, Object-Oriented Programming, Artificial Intelligence

Technical Skills

Programming Languages: Python, Java, Javascript, C/C++, R, Java, SQL, Bash Shell Scripting
Database: Oracle SQL, MySQL, MongoDB, PostgreSQL, Firebase, Hadoop
Framework/Libraries:Seaborn, Numpy, Pandas, Matplotlib, Tensorflow, Keras, Pytorch,Bigquery, Talend, Altery,Kubernetes
Cloud: Amazon Web Services (AWS), Google Cloud Platform (GCP), Oracle Cloud, MongoAtlas
Dev Tools: Git, Advanced Excel, Tableau, Power BI, Snowflake,Matlab, AWS, GCP, Azure DevOps, Airflow, Kafka, Apache Spark

Experience

Hanker Analytics **May 2023 – Jul 2023**
Data Engineering Intern *Pleasanton, California*

- Engineered **Data Quality Management (DQM) model** in Python, leading **36.9%** improvement in analysis report accuracy.
- Implemented an **ML model** using Random Forest to predict causes of disease, with an accuracy of **86.1%.**
- Developed **interactive Power BI dashboard** to visualize client application metrics,driving **300%** increased client engagement.
- Designed **statistical visualization charts** to track company's economic progress, enhancing decision-making & engagement.

IsquareIT **Jun 2022 – May 2023**
Research Assistant *Pune, India*

- Created a **real-time poultry monitoring system** for Venky's Chicken farm using **ML algo (RCNN)**, achieving model accuracy of **93.1%.**
- Utilized **Python, OpenCV, and CNN-based pattern recognition** to monitor flock behavior, analyzing over **41,000 images** of chicks for health trend identification and risk mitigation
- Conducted **on-ground research** on chicken herd pattern monitoring at IPMT(Institute of Poultry Management and Technology) and implemented attendance monitoring system for farm workers, enhancing workforce accountability and operational efficiency.
- Co-authored the following **research paper** published in April 2023 in the **International Journal of Technological Exploration and Learning (IJTE)**.

Internlabs **Dec 2022 – May 2023**
Data Analyst Intern *Pune, India*

- Developed a **Real Estate Data Analysis application**, developed **ETL pipelines** integrating data from multiple informal sources including Facebook, Telegram channels and Whatsapp communities on Google Analytics.
- Consolidated data from diverse sources into a unified data frame structure, processing over **90,000** real estate listings in Pune.
- Developed an **automated data pipeline** that removed manual data entry, saving approximately 48 hours of work per week.

Orgpedia **Feb 2023 – Apr 2023**
Software Engineering Intern *Pune, India*

- Corrected code on debugging python for document analysis model of company, Increased accuracy of the model by 22.61 percent.
- Collaborated with the data science team to identify and resolve critical issues in the existing codebase and mentored juniors regarding future work.

Projects

Flock Vision | *AWS Lambda, Roboflow, CNN, Git, DynamoDB, Python, PyTorch, IoT*

- Designed a smart poultry farm management system using AI-driven flock behavior monitoring and disease detection to optimize productivity for Venky's Chicken, a leading poultry producer.
- Leveraged Python, OpenCV, and CNNs integrated via Roboflow for real-time group pattern recognition, enabling early health alerts and trend analysis.
- Deployed IoT-based temperature regulation and implemented a cloud-synced attendance tracking system to improve worker accountability and maintain optimal farm conditions.

AutoETL | *GPT-4 API, LangChain, Airflow, Python, Pandas, Streamlit, Docker*

- Built a natural language-driven ETL pipeline generator that converts user prompts (e.g., "Group by region and average sales") into executable Pandas or PySpark code using GPT-4 via LangChain.
- Integrated with Apache Airflow to auto-generate and schedule DAGs, enabling seamless deployment of repeatable data workflows with zero manual coding.
- Provided an interactive Streamlit interface for users to preview transformations, validate data, and export results to CSV or cloud storage, accelerating ETL prototyping and collaboration between engineers and analysts.