Nikhil Doye

+1-857-214-9074 doye.n@northeastern.edu LinkedIn GitHub

SUMMARY

Data Engineer with 2+ years of experience designing scalable data pipelines and automating performance dashboards. Expert in SQL and Python, with a proven record of optimizing ETL processes and enhancing data accessibility. Demonstrated ability in building reliable, quality-controlled systems and automated alerts to support data-driven decision making. Adept at developing tools for real-time field analysis to improve product reliability.

ACCOMPLISHMENTS

- **World No.1 SQL Developer**: Ranked as the World No.1 SQL Developer (Practice) on HackerRank.
- **Silver Medalist in NeuroHack**: Led a team of 6 and secured a Silver Medal in NeuroHack(AI Hackathon) by developing a BERT-based model for IT service ticket classification, later integrating HDBSCAN for automated clustering.
- **Bronze Medal in a Kaggle Competition**: Earned a Bronze Medal in a Kaggle Competition for predicting student dropout rates using random forests, optimizing model performance through hyperparameter tuning.

EDUCATION

Northeastern University **Sep 2023 - Mar 2025**

Master of Science, Artificial Intelligence

- **GPA:** 3.9/4.0

WORK EXPERIENCE

Fix-It 24/7 Group LLC **Jul 2024 - Dec 2024**

Data Engineer Intern *Denver, CO*

- Optimized SQL-based ETL pipelines to consolidate customer and service request data, reducing query execution time by 40%.
- Designed Power BI dashboards for KPI tracking, streamlining data visualization and automating reporting processes.
- Developed a scheduling optimization algorithm in Python using Google OR-Tools and Pandas to implement constraint-based optimization and dynamic resource allocation.
- Achieved a 35% improvement in technician efficiency and significantly reduced customer response times.
- Implemented data quality validation scripts to enhance reporting accuracy and support data-driven decision making.

Northeastern University **Sep 2024 - Dec 2024**

Graduate Teaching Assistant (MKTG3501 - Marketing Analytics) *Boston, MA*

- Led workshops on SQL-based data transformation and visualization, showcasing real-world marketing data analysis using Power BI.
- Developed and presented an Airflow-driven automated ETL pipeline project, integrating real-time campaign performance data into a centralized data warehouse.
- Provided mentorship to 40+ students, enhancing their proficiency in SQL queries, Python data pipelines, and visualization tools.

LTI-Mindtree **Dec 2021 - Jun 2023**

Data Engineer II *India*

- Designed and productionized scalable ETL pipelines using Apache Airflow and Amazon Redshift, improving data throughput efficiency by 50% for processing high-volume ITSM incident data.
- Built and deployed a Python-based automated problem classification system leveraging NLP and rule-based logic, cutting manual ticket triage time by 40% and generating $200K+ in annual operational savings.
- Refactored and optimized MongoDB aggregation queries and executed workload migration to Snowflake, resulting in a 25% reduction in query execution latency and measurable decrease in infrastructure spend.
- Engineered an unsupervised NLP-driven incident clustering pipeline using BERT embeddings and HDBSCAN, increasing IT incident correlation accuracy by 30% and enabling faster root cause identification.

LTI-Mindtree **Jul 2021 - Dec 2021**

Software Engineer *India*

- Designed and developed RESTful API endpoints for a custom data orchestration platform using Flask and FastAPI, enabling productized integration between internal services and automating pipeline interactions to support scalable, self-serve data workflows.
- Implemented modular ETL automation scripts using Python (Pandas, SQLAlchemy) and SQL, reducing data processing errors by 20%, accelerating pipeline runtimes by 30%, and enhancing system observability with integrated logging and exception handling.
- Wrote detailed technical documentation and established best practices for developing Airflow DAGs, integrating version control (Git) and CI/CD workflows (GitLab CI, Docker) to support scalable, testable, and maintainable data pipelines.

Tarah Technologies **May 2020 - Jul 2020**

Machine Learning Intern *India*

- Created predictive models for image classification using CNN, achieving an F1-score of 0.96 to support real-time decision-making.
- Engineered a segmentation model for biomedical imaging using advanced techniques, attaining an F1-score of 0.89.

TECHNICAL SKILLS

- **Data Engineering & Workflow Automation**: ETL Pipelines, Data Modeling, Airflow, Batch & Real-Time Processing, REST/gRPC APIs, CI/CD
- **Big Data & Cloud Technologies**: AWS Glue, Redshift, Snowflake, Databricks, Apache Spark, Hadoop, Kafka
- **Programming & Tools**: Python, SQL, Scala, PL/SQL, HiveQL, SparkSQL, FastAPI, pandas, Jupyter, matplotlib, NumPy, SciPy, Git
- **Database & Data Distribution**: PostgreSQL, NoSQL (MongoDB, DynamoDB), Real-Time Data Processing
- **Business Intelligence & Reporting**: Power BI, Tableau, Automated Dashboards, Financial Data Analytics