Naitik Shah

313-487-5218 | naitik@wayne.edu | https://www.linkedin.com/in/naitik-shah-49baba1a1/ | https://github.com/naitik2314

Skills

Python, SQL, Alteryx, GitHub, Databricks, PyTorch, Pandas, AWS (Glue, Lamda), Azure, Power BI, Algorithms, ETL, API, XGBoost, Numpy, Sklearn, Statistical & Regression Analysis, Data Validation, TensorFlow, Predictive Modeling, Matplotlib, NLP, Spark, Hive, Hadoop, CI/CD, PySpark, Feature Engineering, Docker, A/B Testing, PyTest, Behave

Work Experience

DTE Energy Jan 2024 - Present
Data Scientist Co-op Detroit, USA

- Built predictive models using **LightGBM** and **TensorFlow**, with **regression analysis** and hyperparameter tuning (**Optuna**), achieving 75% accuracy for early customer risk identification, supporting strategic decision-making. Applied **data validation** using automated testing frameworks to ensure robust model accuracy.
- Analyzed historical and external data using **SQL**, **Python**, and **Spark** to create a seasonality-based predictive model for call volume forecasts, reducing allocation errors by 20% and optimizing call center resource allocation. Developed **ETL pipelines** using **AWS Glue** to enable dynamic data processing.
- Conducted hypothesis testing and regression analysis of complex datasets in **Python** and **SQL** to provide data-driven insights, increasing decision-making efficiency by 10% for business units. Integrated **AWS Lambda** functions to automate data retrieval and storage
- Reduced financial losses by 15% through ETL pipelines built with **Spark** and **Hive**, enabling accurate identification of high-risk customers for targeted intervention.
- Analyzed call volume fluctuations and high average handle times using **Databricks SQL**, **PySpark**, and **Hadoop**. Visualized findings in **Power BI** and **Matplotlib**, identifying patterns leading to actionable recommendations.
- Resolved data inconsistencies that hindered analysis accuracy by building automated **Alteryx** workflows for data wrangling and validation, maintaining 99.9% accuracy. Improved data integrity, ensuring reliable insights for reporting and decision-making.
- Conducted **A/B testing** on pilot initiatives to evaluate their effect on average handle time, providing insights that supported strategies to reduce call volume by 25% and operational costs by 10%.
- Automated schema migration processes using **Python** and **Git**, reducing the need for manual input and enhancing unstructured data extraction with **NLP** techniques. Cut migration time by 30%, improving data quality and efficiency.

LW Informatics Jan 2020 - Dec 2021
Software Developer Intern Jaipur, India

- Enhanced data processing efficiency by 15% through **SQL** optimization and **data wrangling**, facilitating faster, more reliable analysis for product development. Developed automation scripts with **Numpy**, **Pandas**, and **APIs** for real-time analysis resulting in a further 20% increase in processing efficiency.
- Improved application performance by 25% by optimizing **algorithms** and **data structures**, leading to faster response times and enhanced overall reliability.
- Engineered data-driven applications with **Python**, **APIs** and **Flutter**, enabling secure data exchange integrating and data analytics features to enhance user experience and support real-time data processing and visualization capabilities.

Education

Wayne State University Detroit, USA
Master of Science, Computer Science, GPA: 4.0/4.0 Jan 2023 - Dec 2024

- Courses: Design and Analysis of Algorithms, DBMS, Intelligent System Algorithms, Cybersecurity, Data Networks

Mumbai University Mumbai, India
Bachelor of Engineering, Computer Engineering, GPA: 3.43/4.0 Aug 2018 - Aug 2022

Certifications

Data Science Professional Certificate: IBM, 10/09/2024

AWS Cloud Fundamentals: AWS, 04/24/2021

Projects

Energy Anomaly Oct 2024

- Designed an **LSTM**-based Anomaly detection model with, Feature Engineering, and Matplotlib for energy theft and faulty meters achieving 85% accuracy

RxPredict Aug 2024

- Engineered a patient adherence prediction model using **XGBoost**, achieving 78% accuracy. Ability to support healthcare teams in identifying patients at high risk of non-adherence, enhancing intervention strategies.