Michelle D. Davies

Greater Boston, MA
Portfolio: michelledominiquedavies.info
LinkedIn: https://www.linkedin.com/in/
michelle-dominique-davies/

Senior Data Engineer with experience designing and leading large-scale data pipeline architectures, integrating ETL and ML workflows, and mentoring engineering teams. Skilled in Appian, IBM DataStage, Python, TensorFlow, and cloud-based analytics (Azure, AWS). Passionate about optimizing data infrastructure to accelerate machine learning model delivery and insight generation.

Education

University of California, Berkeley – Master of Information & Data Science, 2025
Cornell University, College of Engineering – Bachelor of Science in Electrical and Computer Engineering, 2022

Experience

IBM: Senior Data Engineer; Cambridge, MA – April 2024 - Present
- Led the end-to-end design, development, and deployment of scalable data pipelines and ETL frameworks using IBM DataStage, enabling seamless ingestion, transformation, and delivery of high-volume structured and unstructured data to machine learning and analytics environments.
- Created and maintained BPM applications using Appian, integrating them with enterprise data systems to streamline workflows and support AI-driven decision processes.
- Mentored junior data engineers, conducted code reviews, and established best practices for data modeling, version control, and pipeline reliability, fostering a culture of technical excellence and continuous learning.
- Architected data ingestion and transformation processes to feed machine learning models, optimizing feature pipelines for training and inference in Python, R, and TensorFlow environments.
- Collaborated cross-functionally with data scientists, business analysts, and DevOps teams to define data requirements, ensure high-quality datasets, and operationalize ML models in production.
- Implemented security and compliance frameworks to protect data integrity, ensuring adherence to corporate governance and data protection standards.
- Monitored, optimized, and automated data workflows, reducing latency, improving throughput, and cutting operational costs through proactive issue detection and performance tuning.
- Provided on-call and weekend production support, rapidly resolving critical incidents to maintain high system uptime and reliable data delivery for machine learning applications.

IBM: Data Engineer & BPM Application Developer; Cambridge, MA – October 2023 - April 2024
- Designed, developed, and deployed business process management applications using Appian.
- Designed, developed, and deployed ETL pipeline applications using IBM DataStage.
- Created intuitive and user-friendly interfaces for BPM applications to facilitate easy navigation and use by non-technical users.
- Developed and optimized data processing pipelines using big data technologies to cleanse, structure, and enrich raw data into usable formats.
- Integrated new technologies and data management tools into the existing infrastructure to improve data processing and analytics capabilities.

IBM: Associate Data Scientist; Cambridge, MA – October 2022 - October 2023
- Wrote programs to cleanse and integrate data in an efficient and reusable manner using leading-edge and open-source tools such as Python, R, and TensorFlow, combined with IBM tools and our AI application suites.

- Worked in an Agile, collaborative environment, partnering with other scientists, engineers, consultants, and database administrators of all backgrounds and disciplines to bring analytical rigor and statistical methods to the challenges of predicting behaviors.
- Communicated with internal and external clients to understand and define business needs and appropriate modeling techniques to provide analytical solutions.
- Assisted in leading workshops to gather insight from users about how to improve applications.
- Evaluated modeling results and communicated the results to public sector audiences.

IBM: Client-Facing Technical Support Developer; Littleton & Lowell, MA – May 2021 - December 2021
- Resolved support requests and debugged issues with the IBM InfoSphere Information Server (IIS) product.

Academic & Technical Projects

RealPage Rent Impact, University of California-Berkeley Capstone
Skills: SQL · Python · Pandas · Data Engineering · Business Intelligence
Analyzed 100M+ rental listings with SQL and Python to uncover pricing anomalies linked to RealPage's rent-optimization algorithm.

Context-Enriched NER for Identifying Emerging Trends in Video Comments, University of California-Berkeley
Skills: NLP · Deep Learning · Transformers (BERT) · Python · PyTorch
Fine-tuned BERT models to extract emerging trends and locations from YouTube comments using NLP and PyTorch.

Applications of 5G Cellular Communication Systems, Cornell University
Skills: MATLab · Communication Engineering
MATLAB simulations compare 5G and LoRaWAN for IoT healthcare, analyzing bit error rate, latency, retransmissions, and reliability, alongside practical and socio-economic impacts.

Machine Learning Systems Engineering: Deploying an API, University of California-Berkeley
Skills: ML Ops · Docker · FastAPI · Model Deployment · Cloud (AWS)
Built and deployed a containerized ML prediction API with FastAPI, Docker, and cloud infrastructure.

Non-Linear Regression for Housing Prices, University of California-Berkeley
Skills: Regression Analysis · scikit-learn · XGBoost · Feature Engineering · Model Evaluation
Compared linear vs. non-linear models (e.g. XGBoost) to boost R^2 on real estate datasets.

Certifications

- IBM Data Engineering Specialty - Experienced
- Appian Certified Senior Developer
- Neo4j Graph Data Science Certification
- Azure AI Fundamentals Certification
- Azure AI Engineer Associate Certification
- AWS Partner Accreditation (Technical)
- Databricks Generative AI Fundamentals Accreditation
- IBM Information Server - Datastage ETL Developer
- IBM Quantum Machine Learning