Caroline Smith

Data Science Engineer

—

Contact

91 Bristol Rd, Apt 2,
Somerville, MA 02144

304.982.0928
carolinesmith527
@gmail.com

—

Technology

Python, Databricks,
Scikit-learn,
Pytorch, Pyspark, AWS,
Azure, Tensorflow,
Pandas, Geopandas,
Numpy, LLMs, APIs,
ArcGIS, GIS, R,
Sentence-transformers,
SQL, SAP S4 HANA,
Oracle, Stata, Tableau,
PowerBI

—

Skills

Machine Learning,
Data Science,
Data Engineering,
Generative AI, API
Development,
Data Modeling,
Applied Mathematics,
Mathematical
Programming &
Modeling, Entity
Resolution, Graphical
Data Modeling,
Application Design and
Build, Data Analysis,
Spatial Analysis, Web
Application
Development

Employment History

Accenture / Data Science and Engineering Specialist
OCT 2019 - PRESENT, BOSTON

- Worked on API and Python development for a generative AI "Platform in a Box" in Azure, which enabled generative AI governance, monitoring, responsible AI checks, load balancing, proxy connections, and standardization for a large biopharma client.
- Designed and developed machine learning analysis applications in AWS to automate master data management systems (using LLMs and Graphical Machine Learning) — improving teams accuracy, scalability, and productivity.
- Developed Python and SQL data quality monitoring and automations to resolve backlogs and ensure continued data quality at various commercial clients.
- Deployed & owned calculation views for the master data management team which merged data across systems and applied data quality rules to transform data in SAP.
- Deployed and owned Python environments on Databricks/S3 for analyses, script execution and automation.
- Developed Python analyses to ingest materials data from legacy databases and perform web scraping and data enrichment for drug information from external reference sites and APIs.
- Performed ad-hoc statistical analyses on research questions regarding client data.
- Developed text processing algorithms and toolkits to perform entity resolution and clustering on client and materials data across multiple systems using Python and SQL.
- Oversaw the creation and implementation of novel data solutions from prototype stage and onward.
- Performed QA testing, analysis, and automation for SAP implementation for large public utility and electric provider – analyzing live data ("Net Metering") from renewable energy infrastructure. Led creation of technical documents for client education.

Otis College of Art & Design Office of Institutional Research / Data Analyst
OCT 2018 - JUNE 2019, LOS ANGELES

- Database creation, data analyses/scripting, & research project data support
- Created ad-hoc SQL data solutions in support of institutional research and effectiveness, strategic planning, and assessment activities institution wide.

University of Kentucky Center for Health Sciences Research Data, Analytical, and Statistical Core / Data Analyst Intern
SEP 2017 - JUNE 2018, LEXINGTON KY

- Developed and executed database mining, analysis, querying, algorithms, visualization, and research project support using SQL, SPSS, and Tableau for healthcare applications.

University of Kentucky Thesis Project / Machine Learning
AUG 2017 - MAY 2018, LEXINGTON KY

- Analyzed large academic datasets to build a machine learning integer programming model to calculate optimal busing systems for students while accounting for student location, cost constraints and other criteria.

Education

University of Kentucky / Mathematical Economics B.S. / 2014 - 2018, LEXINGTON KY
Minor in Dance

Udemy / Zero to Mastery Complete A.I. & Machine Learning, Data Science Bootcamp / 2024