PRAHLAD SIWAKOTI

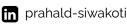
Data Engineer

Data Scientist



Harrisburg, PA





SUMMARY -

As a Physicist turned Data Scientist/Engineer, I am passionate about combining rigorous scientific methodology with technical expertise to transform complex data into actionable insights. In my research career, I applied advanced analytics to complex problems and published findings in academic journals and conferences. I have built statistical and machine learning models using Python, R, and SQL, while also designing robust data pipelines with tools like dbt, Airflow, and Snowflake. I have leveraged cloud functions (Azure, Google Cloud) to automate data ingestion, transformation, and event-driven processing. With experience in both predictive modeling and data architecture, I'm passionate about creating end-to-end data solutions that drive informed decision-making. I am looking forward to integrating my educational background, research experience, and technical skills to solve challenging data problems.

SKILLS

Languages:

- Python
- R
- · SQL
- SAS

Visualization:

- Power BI
- ggplot
- Matplotlib
- R-Shiny

Machine Learning:

- · Scikit-learn
- Tensorflow
- · Pytorch

Database:

- · Database Design
- PostgreSQL
- MongoDB

Engineering:

- · Docker, SFTP
- · Data Ingestion, Migration
- · SQL Scripting, dbt
- · Airflow, FastAPI
- Cloud Functions (Azure, GCP)
- · Cloud Storage (Azure Blob, MINIO)
- Snowflake
- · Databricks / Dremio

Additional:

- Debugging
- · Time-Series Analysis
- Deep Learning
- Version Control (Git/GitHub)

EDUCATION

1/2024-5/2025

Masters in Data Science

University of Texas at Austin

Courses taken: Machine Learning, Probability and Statistical Inference, Data visualization, Algorithms, Advanced Predictive Models, Deep Learning, Natural Language Processing, ReInforced Learning, Data Science for Health.

8/2015 - 11/2021

PhD in Physics

Louisiana State University, Baton Rouge, LA

Dissertation:

Effects of Structure, Crystallographic Orientation, and Dimensionality on Emergent Properties of Transition Metal Oxide Thin Films

DATA SCIENCE EXPERIENCE —

05/2025 - Present **Data Engineer Apprenticeship**

Nashville Software School

- · Automated extraction, transformation, and loading (ETL) of structured and unstructured data from REST APIs and local storage sources
- · Designed and implemented data pipelines using Airflow, dbt, and Snowflake, following the medallion architecture for scalable and reliable data processing
- · Developed workflow orchestration and ETL processes for efficient data movement and transformation with Airflow.
- · Built a FastAPI-based frontend with integrated Swagger documentation for interactive API exploration and testing

9/2023 - 7/2024

Data Scientist Apprenticeship

Nashville Software School

- · Wrangled data and performed exploratory data analysis using Python's pandas library and R's tidyverse packages
- · Created data visualizations using matplotlib, seaborn, and ggplot2 which helped understand complex dataset and problems.
- · Built and evaluated statistical and machine learning models using the scikit-learn and statsmodels libraries
- · Developed and evaluated machine learning models for classification and clustering tasks, with hands-on experience interpreting confusion matrices, ROC curves, and precision-recall metrics.
- · Applied natural language processing using the nltk and spaCy libraries to enhance text analysis capabilities, improving data insights
- · Performed network analysis on graph data using Neo4j to identify key relationships and pat-
- Built and deployed interactive data visualizations using the R Shiny library
- Managed source code version control with Git/GitHub, ensuring code integrity and facilitating team collaboration

PROFESSIONAL EXPERIENCE -

11/2021 - 12/2023

PostDoctoral Researcher

University of Tennessee at Knoxville

- · Developed and maintained a data analysis pipeline for large-scale synchrotron data using Python and R
- · Wrote python scripts to simulate observed data and to perform statistical analysis
- · Collaborated with researchers from various disciplines to analyze, interpret data and deduce conclusions
- · Provided mentorship and training to graduate students with research, instrumentation, and troubleshooting

01/2018 - 11/2021 Graduate Research Assistant

Louisiana State University, Baton Rouge

· Explored non-trivial physics of transition metal oxide perovskite thin films with respect to their symmetry and growth orientation and studied various two-dimensional defects.

SELECTED PROJECTS -

Marvan Research Data Pipeline

LINK

Built an automated data pipeline for the Marvan research project using Airflow, dbt, and Snowflake to streamline data ingestion, transformation, and analysis. Developed a FastAPI data access layer with Swagger docs for easy querying and interaction with research datasets. Skills: Data Pipeline Automation (Airflow, dbt), Snowflake, Medallion Architecture, FastAPI (Swagger Docs)

NPPES - project

LINK

A comprehensive data processing pipeline for analyzing healthcare providers across US counties using the National Plan and Provider Enumeration System (NPPES) data.

Skills: Big Data Ingestion, ETL, Azure Functions, Polars, PostgreSQL, Pipeline Automation with HTTP triggers and stored procedures, Complex SQL joins to create views for stakeholders.

Air Quality: Machine learning models applied to air quality data

LINK

Constructed a predictive model for air-quality monitoring from data obtained from inexpensive air-sensors by PurpleAir and various meteorological data. I have utilized various tree-based spatio-temporal models as well as neural networks to predict the air quality.

Skills: Time-Series Analysis, Spatial regression, Kriging interpolation, Machine Learning, Deep Learning (Pytorch), Data Visualization

Other Projects

Portfolio

PEER REVIEWED PUBLICATIONS	PEER	REVIEWED	PUBLIC	CATIONS
----------------------------	-------------	-----------------	---------------	---------

Google Scholar:

Prahlad Siwakoti