

SUMMARY

I am an aspiring Data Scientist with a background in experimental physics. In my career as a researcher, I have experience combining rigorous scientific methodology with advanced analytics to solve complex problems and presented my findings through publications in academic journals and research conferences. As a data scientist, I have built statistical and machine learning models to derive insights from complex datasets using Python, R, SQL, and other open-source tools. I have leveraged cloud functions (Azure, Google Cloud Functions) to automate and scale data pipeline tasks such as data ingestion, transformation, and event-driven processing. I am looking forward to integrating my educational experience, research background, and programming tools to solve complex problems.

SKILLS

<b>Languages:</b> <ul style="list-style-type: none"><li>• Python</li><li>• R</li><li>• SQL</li><li>• SAS</li></ul> <b>Visualization:</b> <ul style="list-style-type: none"><li>• Power BI</li><li>• ggplot</li><li>• Matplotlib</li><li>• R-Shiny</li></ul> <b>Machine Learning:</b> <ul style="list-style-type: none"><li>• Scikit-learn</li><li>• Tensorflow</li><li>• Pytorch</li></ul>	<b>Database:</b> <ul style="list-style-type: none"><li>• Database Design</li><li>• PostgreSQL</li><li>• MongoDB</li></ul> <b>Engineering:</b> <ul style="list-style-type: none"><li>• API Data Retrieval</li><li>• Docker</li><li>• Data Ingestion</li><li>• SQL Scripting</li><li>• Data Migration</li><li>• SFTP</li><li>• Cloud Functions</li></ul>	<b>Additional:</b> <ul style="list-style-type: none"><li>• Debugging</li><li>• AWS</li><li>• Time-Series Analysis</li><li>• Deep Learning</li><li>• Version Control (Git/GitHub)</li></ul>
--	--	--

EDUCATION

1/2024–5/2025	<b>Masters in Data Science</b> 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.	<b>University of Texas at Austin</b>
9/2023 – 7/2024	<b>Data Science Apprenticeship</b> Intensive part-time boot-camp focusing on data science fundamentals and problem solving	<b>Nashville Software School, Nashville, TN</b>
8/2015 – 11/2021	<b>PhD in Physics</b> Dissertation: Effects of Structure, Crystallographic Orientation, and Dimensionality on Emergent Properties of Transition Metal Oxide Thin Films	<b>Louisiana State University, Baton Rouge, LA</b>
8/2011 – 3/2014	<b>Masters of Science in Physics</b> Dissertation: <i>First-Principles Study of Neutral <math>\langle(N_2)_n\rangle</math> and Singly Cationic <math>\langle(N_2)_m^+\rangle</math> Molecular Nitrogen Clusters; (<math>n = 1, 2</math> and <math>m = 1, 2, 3, 4, 5</math> and <math>6</math>)</i>	<b>Tribhuvan University, Kathmandu, Nepal</b>

DATA SCIENCE EXPERIENCE

05/2025 – Present	<b>Data Engineering Bootcamp</b> <ul style="list-style-type: none"><li>• Designed and implemented data ingestion pipelines to collect data from REST APIs and local storage sources</li><li>• Automated extraction, transformation, and loading (ETL) of structured and unstructured data</li><li>• Working with cloud-based data storage and processing tools</li><li>• Implementing ETL processes and workflow orchestration</li><li>• Hands-on experience with tools such as Airflow, dbt, and Docker</li><li>• Collaborating on real-world data engineering projects in a team setting</li></ul>	<b>Nashville Software School</b>
-------------------	--	----------------------------------

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
- Performed geospatial analysis using geopandas and folium
- Gathered data through APIs and webscraping
- Retrieved and analyzed data using PostgreSQL and sqlalchemy
- 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
- Performed network analysis on graph data using Neo4j
- Built and deployed interactive data visualizations using the R Shiny library
- Source code version control with Git/GitHub
- Project management/tracking with GitHub project boards and issue tracking
- Interacted with AWS using the CLI and ssh

## 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

---

- **Time-Series Forecasting : A python Implementation** [LINK](#)  
Exploration of various time-series forecasting methods using Python. Various statistical and machine learning models were implemented to predict the future values of a time-series data and compared with the actual values. The models include ARIMA, XgBoost, and LSTM.  
**Skills:** Time-Series Analysis, Data Wrangling, Data Cleaning, Data Visualization, LSTM, XgBoost, ARIMA, Pytorch
- **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, Data Visualization
- **Wildland fires and their effects on visitation data in US National Parks** [LINK](#)  
Created an interactive R Shiny app of various National Parks in the US featuring wildfire events in the past to visualize the effect of these events in the park visitation statistics.  
**Skills:** ARIMA forecasting, R Shiny, Data Wrangling, Data Cleaning, webscraping
- **Other Projects** [Portfolio](#)

## PEER REVIEWED PUBLICATIONS

---

Google Scholar :

[PrahlaD Siwakoti](#)