

(315) 949-8690 | [dederaj@gmail.com](mailto:dederaj@gmail.com) | [LinkedIn](#) | [GitHub](#)

Data Scientist with a PhD in Physics and a proven track record in machine learning, and data-driven decision-making. Experienced in developing statistical models, analyzing large datasets, and delivering actionable insights.

## Skills

**PROGRAMMING LANGUAGES:** Python | C | SQL | LATEX | HTML

**VERSION CONTROL:** GIT

**MACHINE LEARNING LIBRARIES:** Scikit-Learn | PySpark | Statsmodels | CatBoost | XGBoost

**DATA PROCESSING & VISUALIZATION :** PostgreSQL | Numpy | Pandas | GeoPandas | Matplotlib | Seaborn | Plotly | Flask | Streamlit

**INDUSTRY KNOWLEDGE:** Supervised Learning | Unsupervised Learning | Quantitative Research | Statistical Analysis | Deep Learning | Natural Language Processing | Large Language Modeling

## Education

**Ph.D. in Physics** | Syracuse University – May 2023

**Data Scientist Certification** – The Data Incubator – March 2023

**Python for Data Science AI & Development** | IBM

**M.S. in Physics** | Syracuse University – June 2017

## Experience

**Research Scientist** | Syracuse University | Syracuse, NY | Sept 2018 – December 2022

- Successfully conducted and completed research on four projects on geometric studies on thin films, with one of the projects featured in American Physics Society Magazine.
- Reduced simulation runtime by 40%+ by applying new approach in mesh models optimizing energy of floating thin films using gradient descent .
- Managed over 400 GB of experimental data and over 600 GB simulation data by building a data ingestion pipeline with data transformation steps to make the data ready for data processing.

## Projects

### 1. [Anomaly Detection](#)

- Developed solutions for monitoring client's business metrics in real time for instant detection of incidents that may impact their revenue.
- Leveraged ensemble models and conducted comprehensive performance analysis utilizing precision, recall, and F1 score metrics. Identified an optimal model with a 15% increase in true positive identification, leading to enhanced anomaly detection capabilities.

### 2. [Trails To Health](#)

- Designed and implemented "Trails to Health," a recommendation system for trails across 230 New York State Parks.
- Built unsupervised learning model to cluster trails by difficulty, accounting for factors like elevation and distance. and employed cosine similarity for user recommendations.
- Integrated USGS Rest API, Flask, and geolocation services for streamlined user access and personalized trail recommendations. ([Trails To Health](#)).

## Volunteer Work:

**DataKind September 2023 Event**

- Collaborated with a diverse team of over 200 volunteers to support the IMF Gender-Based Project, aimed at promoting gender equality for sustainable economic progress across 5 countries.
- Spearheaded the development of a document translation pipeline, facilitating the conversion of non-English documents to English for seamless communication and data analysis.
- Leveraged OpenAI technologies through Azure cloud computing to create an interactive PDF insights chatbot providing users with a more accessible and interactive approach to exploring data and conducting in-depth analysis.