

Karim El-Sharkawy

248-275-3220 | Troy, MI, US | karimelsharkawy2002@gmail.com | [linkedin.com/in/karimme/](https://www.linkedin.com/in/karimme/) | github.com/karim-sharkawy

EDUCATION

Purdue University

Bachelor of Science in Mathematics

Data Science Certificate; Atmospheric Science Minor

West Lafayette, IN

08/2020 - 12/2024

EXPERIENCE

Sprout

Machine Learning Engineer (Volunteer)

Remote

04/2025 - Present

- Developed NLP pipelines to analyze 270GB text data to recognize climate anxiety for 2 mental health organizations.
- Extracted themes and emotions from 21M social media posts using transformers to inform educators and counselors.
- Cut processing time from 2 weeks to 15 hours while ensuring interpretable outputs by automating Azure ML workflows.
- Led a 5-person cross-functional Agile team to deliver production NLP solutions aligned with stakeholder needs.
- Designed emotionally-aware presentations and reports to surface interpretable insights to non-technical stakeholders.

Data Scientist (Volunteer)

01/2025 - 04/2025

- Developed an LDA model on 50k posts as a baseline for improving mental health intervention planning.
- Improved coherence from 0.344 to 0.432 out of 1 by fine-tuning model parameters using a grid search.
- Delivered sentiment insights on +75 themes that guided stakeholder outreach and informed decisions.
- Developed accessible data-collection protocols that connected technical methods with community needs.

Purdue University

Linear Programming Research Assistant

West Lafayette, IN

06/2023 - 12/2024

- Co-developed an optimization-based classifier using linear programming and SVMs in Python, tested on 100K points.
- Implemented and benchmarked scikit-learn classifiers to study separability and decision boundaries in high dimensions.
- Applied convex optimization methods to improve interpretability in structured classification tasks.
- Co-authored 2 forthcoming papers and presented results to 40+ faculty and students at a Purdue mathematics seminar.

Planetary Climate Dynamics Research Assistant

05/2022 - 12/2022

- Analyzed 90GB reanalysis datasets (EMARS, MACDA) in Python researching dust storm track variability.
- Developed data processing workflows and visualizations to support atmospheric dynamics research and reporting.
- Presented findings to the research team and contributed an abstract submission to the AGU conference.

PROJECTS

Customer Feedback Classification System

09/2025 - Present

- Developing scalable training and inference pipelines for text classification using PySpark and HuggingFace transformers.
- Deploying batch and real-time prediction workflows on Databricks with experiment tracking via MLflow.
- Designing feature stores and structured labeling pipelines to support continuous model improvement.
- Containerizing model services with Docker and integrating CI/CD for automated retraining and deployment.

Ensemble Model for Time Series Climate Forecasting

06/2024 - 10/2024

- Built an ensemble forecasting system combining LSTM and Random Forest models for hourly climate prediction.
- Implemented end-to-end training workflow with Bayesian hyperparameter optimization and experiment tracking.
- Achieved 92% error reduction over baseline through feature engineering and model selection.
- Evaluated model robustness and limitations to guide deployment-ready probabilistic forecasting improvements.

SKILLS

Languages

Python (Pandas, PySpark), SQL

Machine Learning

Transformers (HuggingFace), TensorFlow/Keras, Scikit-learn, BERTopic

MLOps / Deployment

MLflow, Docker, FastAPI, CI/CD

Data + ML Infrastructure

Databricks, Azure ML, PostgreSQL, MinIO