

DANIEL STEVEN DOXAZO SITOMPUL

San Francisco, CA · dsdsitompul@gmail.com · [linkedin.com/in/dsdsitompul/](https://www.linkedin.com/in/dsdsitompul/) · github.com/Danielstevens/

Data scientist with 5+ years of experience applying ML and data analysis to real-world problem.

EDUCATION

M.Eng, Development Engineering, University of California Berkeley **2023 – 2024**

- **Concentration:** Data Analytics / AI | Certificate: [Applied Data Science](#) and [Technology Entrepreneurship](#)
- **Capstone Project:** “Generator Climate Impact Assessment using ML Techniques in Health facilities in DR Congo”
- **Publication:** [2025 IEEE GHTC](#) | [2026 Stanford Global Health Research Convening](#)

B.Sc, Electrical Engineering, Institute of Technology Bandung (ITB) **2015 – 2019**

- **Final Project:** Runway Edge Light Photometry System Using Drone-Mounted Instrument for Airports in Indonesia.
- **Publication:** IEEE ([ISESD 2019](#) | [ISITIA 2019](#) | [ICEEI 2019](#)) & [ISICE 2018](#) | **Activities:** ITB Robotics Team (ABU Robocon)

EXPERIENCE

Associate Researcher,

International Council on Clean Transportation, San Francisco, CA **Jun 2024 - Now**

- Performed exploratory data analytics on 3 TB of global aviation trajectory data to quantify CO₂, NO_x, and PM_{2.5} emissions across commercial, business, cargo, and general aviation aircraft using Pandas, Numpy, Matplotlib, Seaborn, and Plotly.
- Designed scalable Python + SQL data pipelines (~70M records) integrating aircraft, engine, and flight metadata for reproducible emissions modelling by utilizing parallel processing (Joblib) resulting up to 3x faster processing.
- Applied statistical analysis, modelling, and aggregation techniques to identify emissions and pollution, operational patterns, and variability of aircraft/flights to produce insight (Cartopy, Folium). ([Private jet](#) | [Vizcon 2050](#))
- Produced datasets and web visualizations for global airport local climate reports using Streamlit. ([AIRLIFT](#))

Data Scientist, **Oct 2023 - Jan 2026**

University of California, Berkeley, Berkeley, United States

- **African Great Lakes project, Lead Data Scientist (+ 1.5 yr)**
 - Led a group of 4 people to do EDA, feature engineering, and develop ML model to detect generator usage which estimate emissions and pollution using electrical signal data in healthcare facilities using regression, random forest, and XG-Boost.
 - Analysed model generalization for 4 sites by using Leave-One-Subject-Out Cross-Validation.
 - Evaluated model accuracy achieving >95% performance metrics using power and fuel efficiency ratio.
- **EpiNu: Data Science, Governance, and App UI/UX Designer (9 months)**
 - Built K-Means clustering models to segment regions to assess nutrition adequacy for pregnant woman.
 - Designed a data governance framework and user-facing analytics concepts for public-health decision.

Chief Technology Officer, **Jan 2024 - June 2024**

Alkelink (Startup), Berkeley, United States

- Designed data schemas and system architecture for Alkelink healthcare platform to integrate pharmacy inventory
- Defined API requirements and data pipeline to support ML-analytics infrastructure using python
- Led early product analytics thinking, balancing technical feasibility with user and business needs

Project Manager, Network Deployment and Solution Partnership **Jan 2020 – Jul 2023**

Telkom Indonesia, Jakarta, Indonesia

- Managed infrastructure deployment projects while tracking performance metrics and resource utilization
- Built operational dashboards to monitor network capacity and improve asset utilization by >20%
- Worked with engineering and operations teams using SQL database (IBM NETEZZA) to plan telecom expansion

SKILLS

Programming: Python (6+ yrs), SQL (2+ yrs), C (1 yr), R (6 months)

Data Science: Pandas, NumPy, Matplotlib, Seaborn, Plotly, Openpyxl, Scipy, Joblib, Airportsdata

Machine Learning: Scikit-learn, XGBoost, Statsmodels, HMMlearn, Ruptures

Visualization: Data visualization (Folium, PyQt5, Cartopy, Kaleido), prototyping (Figma), dashboard (Streamlit)

ACHIEVEMENTS

- **Mercy Corps Disaster Hack Winner (2024):** Developed [HACIT](#) which calculates aid delivery costs by integrating APIs data on transportation, fuel, exchange rates, and emissions, supported by database sourced through web scraping.
- **Indonesia Robotics Championship (2017-2019):** 1st Winner & Favorite Team (2019 National) · 1st Winner & Best Design (2017 Regionals) · 4th Place & Favorite Team (2018 Regionals)