Siham Elmali

siham.elmali@gmail.com | SF Bay Area, CA | (443) 488-4923 | siham-elmali | shadowfax42

Summary

Data Scientist and Engineer with 8+ years of experience specializing in end-to-end analytics, data governance, and silicon validation within the semiconductor industry. Expert in designing and deploying scalable data pipelines, predictive models, and interactive dashboards that drive measurable business impact. Proven track record in improving data quality, reducing product defects, and enabling real-time decision-making.

Technical Skills

Programming & Development: Python, R, SQL, Bash, JavaScript, TypeScript, HTML/CSS, Flask, FastAPI

Databases: PostgreSQL, MySQL, MongoDB, Firebase

Data Engineering & Visualization: PySpark, Pandas, Seaborn, Matplotlib, Plotly (Dash), PowerBI, Tableau

Machine Learning: scikit-learn, NumPy, XGBoost, SciPy, PyTorch

LLM & AI Tools: OpenAI, Claude, Gemini, Ollama, DeepSeek

MLOps & Infrastructure: Azure, Databricks, Docker, Kubernetes, Celery

DevOps & Tools: Git/GitHub, GitHub Actions, CI/CD, Linux/Unix, Azure DevOps, JIRA, Scrum, Agile

Spoken Languages: English (Fluent), Arabic (Fluent), Moroccan Darija (Native), French (Conversational)

Professional Experience

Intel Corporation Santa Clara, CA

Senior Data Scientist — Analytics & Insights Management Team *Apr 2023 – Present*

- Currently developing automation tools on Linux using Perl and Python to replace manual Excel-based processes, ensuring engineers have up-to-date, reliable data for analysis and decision-making.
- Modularized and modernized a GUI-based risk analysis tool using object-oriented programming, improving usability, interface design, and maintainability, while implementing a new Monte Carlo simulation for predictive risk modeling.
- Co-developed a VS Code extension integrating an internal LLM using TypeScript and the VS Code API, enhancing spec document navigation for 14,000+ engineers.
- Designed and deployed a full-stack, Dockerized Flask app on Cloud Foundry to monitor silicon anomalies, providing real-time visibility into product health and accelerating data-driven decision-making across teams.

Data Scientist — Data Governance Team *Nov 2021 – Apr 2023*

- Created advanced predictive models and dashboards using Python, Power BI, and SQL on Azure, empowering the Data Governance team to proactively address schedule risks and enhance data quality and compliance.
- Implemented data pipelines in SQL and Python with embedded governance (quality checks, naming standards, access controls), boosting compliance from 30% to 70%, with 100% projected by year-end.
- Applied machine learning methods using SparkML and R on multi-source data with Spark distributed processing to identify early indicators of silicon issues, contributing to a 25% reduction in bad die rates. This work was featured in an internal white paper and served as my MS Analytics capstone project.

Product Development Engineer — Cache Memory Team *Jul 2017 – Nov 2021*

- Built and deployed forecasting models using Python and XGBoost, orchestrated with Kubernetes and Docker, and integrated with SQL databases and Tableau dashboards. Achieved forecast accuracy improvements projected to deliver $15M–$100M in business impact per 1% accuracy gain.
- Led a team of 4 engineers to validate MBIST (Memory Built-In Self-Test) for 4th Gen Intel Xeon Scalable and Xeon Max processors, achieving 100% test coverage and reducing defect rates by 90% through early detection of critical issues.
- Streamlined regression testing for SystemVerilog code using Python and Bash, reducing manual effort and improving validation speed.
- Created visual analytics tools with Python and Power BI for post-silicon data analysis, improving debug workflows and shortening issue resolution timelines.

Johns Hopkins University Applied Physics Laboratory Laurel, MD

Engineering Intern *Jun 2016 – Jan 2017*

- Supported a classified radar research initiative for U.S. defense applications, leveraging prior active security clearance to ensure compliance and project success.
- Enhanced MATLAB-based radar simulation tools with signal processing libraries, improving analysis accuracy and runtime performance.
- Led SME interviews and use-case modeling to inform the design of a large-scale program management platform. Contributed to a case study that led to the adoption and further development of the platform post-internship.

Research And Academia

Georgia Institute of Technology Atlanta, GA

Instructional Associate/Staff (remote — part-time) *May 2025 – Present*

- Serve as Teaching Assistant for Georgia Tech's OMS Analytics course *Simulation for Scientists and Engineers*, supporting 650+ graduate students through online office hours, Piazza discussions, and project feedback, leveraging both academic and industry experience to reinforce statistics, probability, simulation, and analytical modeling concepts.

Johns Hopkins University Baltimore, MD

Research Assistant *May 2015 – Dec 2015*

- Analyzed lung recordings using MATLAB for feature extraction and data cleaning; applied machine learning to explore an audio-based pneumonia detection system.
- Designed PCB using Eagle for a digital stethoscope to implement the signal processing algorithm.

Utah State University Logan, UT

Research Assistant, Engineering Education Research *May 2014 – Aug 2014*

- Cleaned and analyzed survey data from 435 high school students across four states using SPSS and Excel; applied statistical modeling to evaluate MESA program impact on self-efficacy, perception of engineering, and STEM interest.
- Performed correlation, regression, and t-test analyses to assess relationships between program participation and student outcomes.
- Contributed to an unpublished NSF-supported research paper on STEM engagement and educational outcomes.

Selected Projects

Semiconductor "Burn-In" Process Optimization (GT Master's Project) Fall 2022

- Led an end-to-end master's capstone project for Intel's Data Warehousing Team, focusing on predictive modeling of the "burn-in" phase of high-volume semiconductor manufacturing.
- Applied ML techniques to detect subtle process variations and parameter shifts, enabling earlier corrective action by identifying latent defects post-fabrication.
- Analysis and models contributed to a 25% reduction in bad die rates, improving product reliability and manufacturing efficiency; work was featured in an internal white paper.

Energy Hub

- Led development of a full-stack energy management platform using Python (Dash, Plotly), JavaScript/HTML/CSS, and Docker, deployed on Azure Web Apps; focused on data visualization and cloud computing to monitor electricity usage across non-residential buildings.
- Designed interactive dashboards and 168-hour forecasting models to support energy operations teams; applied time-series forecasting techniques and data visualization best practices across multiple facilities.

Education

Georgia Institute of Technology Atlanta, GA

Master of Science in Analytics *Aug 2020 – May 2023*

Johns Hopkins University Baltimore, MD

Master of Engineering in Electrical and Computer Engineering *Jan 2016 – May 2017*

Bachelor of Science in Electrical Engineering *Aug 2014 – May 2017*

Certifications

Lean Six Sigma Green Belt, Intel Corporation (Issued Sep 2025)

Lean Six Sigma Yellow Belt, Georgia Institute of Technology (Issued Dec 2021)