Siriluk Abelson

Data Scientist | Research Scientist

Hillsboro, OR 97124, siriabelson@gmail.com, 612.991.0982, LinkedIn | Portfolio | GitHub

Experienced Data Scientist and Research Scientist with 5+ years of expertise in machine learning, predictive modeling, and computational simulations, including advanced applications in computational chemistry. Proven track record of optimizing business performance through data-driven insights, automation, and cloud-based deployment. Industry experience spans market research, financial, healthcare, semiconductors, chemicals, and automotive, delivering impactful solutions across R&D and manufacturing. Skilled in Python, SQL, and deploying scalable ML pipelines, with a deep understanding of chemical data, materials design, and cross-functional collaboration in a fast-paced environment.

Skills

Core Expertise: Machine Learning, Deep Learning, Statistical Analysis, Data Preprocessing, Feature Engineering, Data Visualization, Cloud-Based ML Deployment, Team Leadership
Tool & Technology: Python, SQL, Bash, Linux, AWS, Git/GitHub/GitLab, Power BI, Power Query, MySQL, HPC, VS Code
Libraries & Frameworks: Scikit-Learn, TensorFlow, PyTorch, SciPy, Pandas, NumPy, Seaborn, Matplotlib, MLflow, Docker, LangChain, Hugging Face, Streamlit
Model & Techniques: Linear/Logistic Regression, Decision Trees, Random Forest, SVM, PCA, K-Means, Gradient Boosting, XGBoost, CNN, NLP, LLM, Agents, Generative AI, Clustering

Professional Experience

Motley Data, Remote
Data Scientist October 2024 – Present

- Built LLM-driven synthetic datasets for market share forecasting, eliminating $50K–100K in external data spend; integrated insights into a client-facing chat platform.
- Designed and implemented a confidence scoring pipeline for user queries using data quality metrics and ML models, helping clients assess response reliability.
- Prototyped an LLM-based agent to interpret financial data from the SEC and Yahoo Finance APIs, enabling real-time analysis of earnings reports.
- Partnered with software engineers to design an end-to-end data science pipeline (ingestion, preprocessing, model training, deployment), ensuring a scalable, reliable analytics workflow.

University of Minnesota, Department of Chemistry, Minneapolis, MN, Remote
Research Scientist/Specialist (Part-time) February 2024 – Present

- Developed a benchmark database for spin-splitting energy calculations of transition metal complexes, setting a new accuracy standard using diverse DFT functionals and basis sets.
- Collaborated with researchers to downselect from large datasets using a genetic algorithm, reducing error by 8% and delivering an efficient benchmark for testing and training electronic structure models.

Citrine Informatics, Remote
Interim Senior Data Scientist (Contract) January 2025 – May 2025

- Led technical execution on 10+ ML projects across 6 customers in materials and chemicals, accelerating product development through tailored AI models.
- Provided direct technical support to customers using Citrine's AI platform, guiding them through statistical analyses, ML model development, defining search spaces, and candidate selection strategies for materials innovation.
- Collaborated cross-functionally with Product, Engineering, and Customer Success teams to onboard new customers, resolve technical issues promptly, and develop solutions tailored to customer workflows and business objectives.
- Mentored two Materials Informatics Engineers, training them in ML, Python analytics, and technical design for independent project execution.

DUPONT, DuPont Silicon Valley Technology Center, Sunnyvale, CA
Senior Data Scientist January 2023 – January 2024

- Led 4+ data analysis and machine learning projects across various business units, impacting over $20M in revenue through end-to-end model development and data-driven decision-making.
- Developed models using regression, tree-based methods, and optimization algorithms in Python, improving electronic product performance by 15% and cutting customer research time by 50%.

- Applied Bayesian Optimization and EDA (SQL, Python) to optimize healthcare materials R&D, reducing research time by 80% and removing dependency on external vendors within 1 month.
- Created a Python-based ML tool for chemists to accelerate formulation design, streamlining experimental workflows..
- Applied BERT-based NLP to complex chemical datasets, extracting actionable insights that supported the formulation development team.
- Partnered with cross-functional teams to analyze adsorption patterns on metal surfaces, identifying corrosion-inhibiting molecules that improved semiconductor reliability.

UNIVERSITY OF MINNESOTA, Department of Chemistry, Minneapolis, MN
Graduate Research Assistant September 2017 – December 2022

- Conducted computational chemistry research, developing advanced models with machine learning and data analysis to enhance computational chemical property prediction, tool development, and scientific discovery.
- Optimized computational model using active learning on a dataset of 20+ databases, achieving second-ranked accuracy in material property predictions among leading quantum methods.
- Led collaborative research analyzing chemical databases and evaluating optimal model chemistry for catalytic mechanism calculations, producing the field's most cited paper in the domain for 2020–2021.
- Enhanced the Python program FREQ for vibrational frequency calculations, upgrading its compatibility with the latest Gaussian Software, significantly streamlining computational chemistry workflows. Widely adopted by scientists and cited in numerous research papers.
- Authored 8+ peer-reviewed scientific publications, demonstrating expertise in scientific writing and research collaboration.

NATIONAL NANOTECHNOLOGY CENTER (NANOTEC), Bangkok, Thailand
Research Intern, National Simulation Laboratory March 2015 – March 2016

- Conducted data analysis using Material Studio Project and VASP to model dye adsorption on TiO2 surfaces, performing DFT calculations for properties and energy alignment to improve solar cell efficiency. Presented at the conference, winning the best poster award.

PROJECTS

- **Hospital Staffing:** Analyzed staffing data using EDA and time series forecasting to uncover trends, seasonality, and stable efficiency per patient day, while predicting declining productivity from fluctuating admissions.
- **AWS SageMaker Deployment:** Deployed XGBoost models for visa approval and hotel reservation predictions, managing end-to-end pipelines with SageMaker and AWS CodePipeline.
- **Malaria Detection:** Implemented CNN-based models for malaria detection from blood cell images, achieving 98.6% accuracy with image augmentation techniques.
- **Predictive Health Insurance Model**: Built and deployed XGBoost models via Streamlit, predicting health insurance premiums with >97% accuracy and <10% error between actual and predicted values.
- **Credit Risk Model for Loan Approval:** Deployed a credit risk model with linear regression and Optuna, achieving >95% accuracy and high AUC and Gini scores, streamlining loan approval processes using Streamlit.
- **Sales & Finance Analytics:** Built advanced data models in SQL, Power BI, and Excel to uncover insights that drove data-informed business performance improvements.

EDUCATION

- **PhD** in Computational Chemistry**,** University of Minnesota, Minneapolis, MN, 2017 – 2022
- **MS** in Chemistry**,** Mahidol University, Bangkok, Thailand, 2014 – 2017
- **BS** in Chemistry, Mahidol University, Bangkok, Thailand, 2011 – 2014

CERTIFICATION

- **Certificate of Achievement: Fundamentals of Agent,** 2025
- **5-Day Gen AI Intensive**, Kaggle, 2024
- **Data Science on Cloud with AWS**, Great Learning, 2023
- **Applied Data Science**: **Leveraging AI for Effective Decision-Making,** MIT Professional Education, 2022
- **Python Data Structure**, University of Michigan (Coursera), 2022

SELECTED PRESENTATION AND PUBLICATION

- "A Day in the Life of a Data Scientist", SEMICON WEST 2023, Moscone Center, San Francisco, CA
- M11pz: A Nonlocal Meta Functional with Zero Hartree-Fock Exchange and with Broad Accuracy for Chemical Energies and Structures*, J. Chem.Theory Comput*., **2023**, *19*, 9102