

ERNEST ESSEL-KAITOO, *PhD*

LinkedIn: [linkedin.com/in/ernselito/](https://www.linkedin.com/in/ernselito/)
Website: <https://ernselito.github.io/>

Phone: 928-255-2192
Email: ernselito@gmail.com

PROFESSIONAL SUMMARY

PhD Data Scientist & AWS ML Specialist driving end-to-end ownership of customer-obsessed AI systems. Delivered 40% faster insights via SageMaker pipelines, 90%+ accuracy predictive models (PyTorch/XGBoost), and MLOps automation eliminating manual workflows.

SKILLS

ML/MLOps:	PyTorch, TensorFlow, SageMaker, XGBoost, Model Deployment, A/B Testing
AWS:	Lambda, Glue, S3, Redshift, EC2, DynamoDB, SageMaker Pipelines
Languages:	Python (NumPy, Pandas, Scikit-learn), SQL, Java (for scalable systems)

PROFESSIONAL EXPERIENCE

Data Scientist (Research Analyst) | UNCO Aug 2021 – Present

- Deployed patient risk stratification models (Logistic Regression, Gradient Boosting) on SageMaker using EHR data, enabling proactive care interventions for 15K+ high-risk patients.
- Drove cross-functional requirements sessions with researchers via JIRA/Confluence, translating clinical needs into PySpark-compatible technical specs for ML pipelines.
- Engineered serverless data pipeline (Python, Glue, S3) reducing prep time 40%, accelerating model iterations from weekly to daily.

Data Scientist | NAU Aug 2018 – Aug 2020

- Shipped Random Forest model (Python/Scikit-learn) automating end-of-month sales prioritization (AUC: 0.85), capturing 18% incremental revenue through data-driven deal sequencing.
- Architected predictive framework (XGBoost, PyTorch) for novel scintillator luminosity prediction (90% accuracy), reducing experimental costs 25% and accelerating R&D cycles 3X.
- Engineered ensemble regressor (Linear Models/SVM) correlating material composition to optical properties (96% accuracy), enabling ML-driven design of high-efficiency compounds.

Data Scientist | Vodafone Aug 2016 - Aug 2018

- Streamlined the path-to-production for ML models by re-architecting data science projects into object-oriented code; this initiative, driven by feedback from operations, created reusable components that eliminated their need for manual refactoring.
- Reduced customer churn 15% by implementing regularized regression models (Ridge/Lasso) with 81% accuracy.
- Engineered self-service Tableau dashboards and Python-driven reports automating KPI monitoring, eliminating 10+ hours/week manual reporting effort and accelerating data-driven decisions.
- Identified and diagnosed root causes of network failures by developing ARIMA and RNN models on speed test data, reducing incident occurrence by 25% through predictive maintenance.

Data Analyst | KCCR in Tropical Medicine Aug 2015 - Aug 2016

- Created a web application to analyze and predict quantifiers such as stocks, crude price, trade revenue etc. in the energy market from news headlines using NLP, topic modelling and text classification with an 89% accuracy.
- Researched and developed machine learning models to address business problems.

EDUCATION & CERTIFICATION

- | | | |
|---|--|---------------------|
| • PhD in Applied Statistics | University Northern Colorado, Greeley | Aug 2021 - Jun 2025 |
| • Master of Science in Statistics | Northern Arizona University, Flagstaff | Aug 2018 - Aug 2020 |
| • AWS Certified Machine Learning Specialty | | Aug 2025 - Aug 2028 |
| • IBM Data Science Professional Certificate | | Feb 2020 - Present |

RELATED PROJECTS

Healthcare Demand Prediction | AWS SageMaker, Python, Streamlit

- Deployed XGBoost classifier on SageMaker to predict patient readmissions (F1: 0.88), reducing costs by 15% in pilot.
- Architected CI/CD pipeline (CodeCommit/CodePipeline) for model retraining, ensuring <5% data drift.

Admissions Prediction Engine | Python, Scikit-learn, Flask

- Engineered ensemble classifier (Logistic Regression/SVM/Naïve Bayes) optimizing hyperparameters via grid search, achieving 84% cross-val accuracy