NIKOLAI (COLE) MOROKHOVICH

Quantitative Ecologist and Data Scientist | nmorokhovich@gmail.com | GitHub | LinkedIn

SUMMARY

Data scientist with experience building Machine Learning and AI pipelines for high-dimensional data, developing statistical models for constrained systems, and processing big scientific datasets for signal extraction. Skilled in Python, ML fundamentals, and automated workflows for scientific applications including raw sensor data processing, feature extraction, and anomaly detection. I work well with different types of people: coding with engineers, explaining results to executives, or partnering with scientists who don't write code.

EXPERIENCE

PUNT LAB, UNIVERSITY OF WASHINGTON 2023-Present

Graduate Research Assistant

Fisheries Risk Assessment for Insurance Products

- Built ETL pipeline processing raw sensor data with automated anomaly detection flagging marine heatwaves
- Developed PML and Loss Exceedance Curves for catastrophic marine heatwave scenarios using Monte Carlo simulation
- Compiled fisheries disaster dataset linking catch records, temperature data, and economic impacts from multiple stakeholders
- Ran Monte Carlo simulations generating 10K+ scenarios comparing economic loss in custom climate scenarios
- Documented data gathering, modeling, and visualization methodology and processes for reproducibility and validation
- Presented findings at Ocean Modeling Forum workshop in Tasmania to researchers, industry partners, and stakeholders

Spatiotemporal Modeling

- Built 3 spatiotemporal forecasting models (EOF-SVD, Bayesian GLLVM-SPDE, spectral clustering), improved predictions 15%
- Coded custom Bayesian C++/TMB likelihood functions for zero-inflated data with spatial correlation on 12K+ observations
- Applied statistical methods (decomposition) to separate signals from noise in messy, real-world environmental datasets
- Designed validation framework testing model performance across 349 locations using multiple evaluation metrics

Additional Projects

- Developed machine learning pipelines in Python for clustering, dimensionality reduction, anomaly detection, and forecasting.
- Processed raw scientific sensor data, including quality control, feature extraction, and anomaly detection.
- Combined data from 3 databases with inconsistent schemas into unified datasets, standardizing 20+ parameters across 16 devices
- Built clustering pipeline (10K samples×600 features) using DPGMM and HDBSCAN to identify distinct patterns in data
- Engineered dimensionality reduction pipeline (PCA, UMAP) reducing data 99% while preserving pattern structure for classification
- Implemented distributed and parallelized machine learning model training across CPU cores for large scientific datasets

EVOLVED BY NATURE 2020-2023

Senior Associate Sales Manager (Promoted 3x from Sales Account Associate)

- Built analytics system tracking user behavior and purchasing data, growing revenue 1,100% ($360K to $4M) in 8 months
- Pitched to C-suite executives at multinational brands, translating technical concepts into business recommendations
- Ran 10+ A/B tests across pricing and user flows, analyzing results to increase conversion 30%
- Created real-time Looker dashboards monitoring 12+ KPIs and alerting teams to anomalies in user behavior or system performance
- Analyzed purchase patterns identifying high-value user segments, leading to 400% increase in subscription adoption
- Built demand forecasting models using time series analysis, reducing out-of-stock time 80% and guiding ad spend allocation
- Optimized $2.4M paid acquisition spend increasing ROI from 0.8 to 1.6 by identifying true user engagement vs. low-quality traffic.
- Owned end-to-end e-commerce operations for 14-product portfolio, reporting directly to CEO on analytics, strategy, and growth
- Worked with engineering, marketing, and operations teams using data to guide product and marketing decisions

SKILLS

Programming & Data Engineering	Statistical & Analytical	Visualization & Communication
• Python (scikit-learn, PyTorch)\	• Machine Learning	• Power BI/Looker/RShiny/Streamlit
• SQL	• Statistical Modeling & Inference	• Data Storytelling
• R (tidyverse, Stan, ggplot2)	• Time Series Analysis	• Stakeholder Communication
• Java, Git, C++/TMB	• Monte Carlo Simulations	• Technical Documentation
• ETL Pipeline Development	• Risk Modeling & Simulation	• Dashboard Development
• Database Architecture	• Bayesian & Hierarchical Models	• Cross-Functional Collaboration

EDUCATION

UNIVERSITY OF WASHINGTON 2023-Present

MS: Quantitative Ecology and Resource Management | Statistics and Data Science Focus

Thesis: *Quantifying Uncertainty in Dynamic Systems: Statistical Approaches for Forecasting and Fisheries Risk Assessment.*

Relevant Coursework:

- Machine Learning for Big Data (Neural networks, distributed computing, large-scale model training, LLMs, gradient descent)
- Data Structures and Algorithms
- Database Systems
- Currently exploring LLM applications/training and modern deep learning frameworks through self-study and online courses

PRINCETON UNIVERSITY 2016-2020

BA: Ecology and Evolutionary Biology, Magna Cum Laude

Varsity Men's Soccer Team, Academic All-Ivy