Ethan K. Long

Boston, MA, USA | (617) 449-8371 | ethan.long@uconn.edu

Summary

Data scientist and quantitative analyst with experience building end-to-end models on large, high-stakes datasets across biomedical research, sports analytics, and financial disclosure analysis. Strong foundations in statistical learning, feature engineering, and model validation, with demonstrated judgment translating quantitative results into defensible stakeholder insights.

Education

University of Connecticut **Expected May 2026**

Bachelor of Arts, Statistics

- **Coursework:** Intro to Data Science | Statistical Programming | Deep Learning | Mathematical Statistics | Multivariable Calculus | Differential Equations

Academic Research Experience

Data Analyst Intern | UConn Men's Ice Hockey Team (NCAA Division 1) Aug. 2025 – Present

- Co-designed a custom event-logging framework for in-game breakout sequences, generating a labeled dataset of 1,000+ observations across the last two seasons.
- Analyzed breakout effectiveness under varying opponent pressure, game location, and time context, identifying patterns that contradicted prevailing assumptions.
- Delivered decision support insights that directly informed game planning, including matchup-specific breakout strategies and avoidance of low-success patterns under high defensive pressure.

Undergraduate Researcher | Statistics Department, University of Connecticut Apr. 2025 – Aug. 2025

- Designed end-to-end binary classification modeling on FDA-provided genomic datasets (80,000+ records), predicting phosphorylation events from codon-level and gene-level features under severe class imbalance.
- Engineered 115+ features from raw biological sequences, including selective one-hot encodings and derived codon statistics, while eliminating leakage and multicollinearity through sequential feature selection and correlation analysis.
- Built and validated ensemble models (SVM, RF, XGBoost) using weighted voting, improving performance from 63% –> 87% F1 and 82% –> 95% ROC-AUC.
- Reduced end-to-end runtime by ~35% through vectorization of gene sequences and aggressive feature pruning without degrading predictive performance.
- Maintained and shared, version-controlled repository reviewed by academic and FDA collaborators; presented findings biweekly to FDA research principals and weekly to faculty advisors.

Vice President | UConn Men's Club Ice Hockey Team Sept. 2024 – Present

- Oversaw operations and a $55,000 annual budget for a 30+ member organization, streamlining logistics and expense tracking.

Academic Projects

ML-Based Evaluation of Clutch Performance in MLB Players

- Quantified "clutch" as performance degradation between regular and postseason play, modeling Total Bases vs. At-Bats to assess predictability loss under high-leverage constraints.
- Applied K-Means clustering to segment players into three postseason performance archetypes, validating cluster stability via elbow method and perturbation testing.
- Presented findings at the Connecticut Sports Analytics Symposium (2025), translating technical findings into actionable player evaluation insights.

Analysis of Narrative Bias in M&A Comparables Using 10-K Disclosures

- Constructed time-aligned corpora of acquiring-firm 10-K and proxy filings, enforcing strict event-window controls to prevent forward-looking contamination.
- Extracted and evaluated comparable company lists and associated valuation multiples, applying paired statistical tests to identify systematic differences between narrative selected and size/sector matched baselines.
- Demonstrated that narrative framing in disclosures can materially influence perceived valuation fairness, with implications for investors and deal evaluation.

Boston 311 Service Request Analysis

- Engineered a predictive framework to analyze municipal responsiveness by integrating historical Boston 311 service data with American Community Survey (ACS) socioeconomic indicators.
- Developed linear regression and tree-based machine learning models (Random Forest, XGBoost) to evaluate structural service delays.
- Leveraged SHAP values for feature importance, identifying department assignment as driving over 35% of model variance, and conducted spatial autocorrelation to visualize service disparities across neighborhoods.

Technical Skills

- Languages & Libraries: Python (pandas, numpy, scikit-learn), SQL (MySQL, PostgreSQL), R, SAS
- Quantitative Methods: Statistical testing, regression, clustering, time series, data cleaning, EDA
- Tools: Git, Docker, Jupyter, Tableau, Web scraping (BeautifulSoup, Selenium)