

Nathan Frank

EXPERIENCED MACHINE LEARNING LEADER

✉ nathanryanfrank@gmail.com | 🏠 nrfrank.github.io | 📷 nrfrank | 🌐 nrfrank | 📧 nathanfrank



About

Skilled machine learning leader with experience in all aspects of the machine learning and software development life cycle. Former Astrophysicist turned data scientist and machine learning engineer with a proven history of delivering results into production across a wide variety of domains while leading projects with international, cross-functional teams.

Experience

Strong Analytics

DIRECTOR OF MACHINE LEARNING ENGINEERING

Chicago, IL

Apr. 2021 - Present

- Lead and mentor an international team of engineers supporting **delivery across over a dozen Strong projects** overseeing efforts in MLOps and DevOps, model optimization and deployment, ETL pipelines, streaming microservices, and full stack web development.
- Automated model training and evaluation pipeline (AWS EventBridge, Step Function, and Batch), **decreasing deployment lifecycle** from monthly to weekly and improving responsiveness of the client's automated content monitoring system.
- Served as project manager and technical lead for client building out a grocery **product identification and recommendation service**.
 - Implemented request volume monitoring and horizontally auto-scaling infrastructure, enabling **>200 QPS** system throughput.
 - Optimized and deployed feature extraction (DELG) and detection (YOLOv5) models to **Triton** services, **reducing inference time ~ 50%**.
 - Oversaw development of web platform for dataset annotation and management, model analysis, and endpoint generation, enabling the rapid creation of dozens of custom, domain-specific customer endpoints.
- Collaborated to architect and build streaming user classification service. Trained and optimized model using sci-kit learn and ONNX and deployed to Triton service. Integrated model service within API processing Kafka events, utilizing Redis for caching.

SENIOR DATA SCIENTIST

Aug. 2020 - Apr. 2021

- Deployed suite of forecasting models for health system records to **Databricks** and built jobs for dataset generation (**Spark**), model training and inference (**scikit-learn** and **PyTorch**), and API deployment (**MLflow**).
- Led analysis of student activity and course content data for a professional education platform, utilizing text embedding, graph network analysis, and sequential forecasting methods to enable automated course creation and content delivery.

Stats Perform (formerly STATS, LLC)

Chicago, IL

AI SCIENTIST [TECH LEAD, ML PLATFORM]

Jan. 2020 - Aug. 2020

- Architected and led development of an enterprise **machine learning platform**, integrating tools for model experimentation (MLflow), training (Docker and Batch), feature engineering (Athena and Tecton), and deployment (EKS), leading to a reduction in average model deployment time **from months to days**.
- Productized **GNN**-based model to predict minutes played of NBA players with **RMSE lower than human domain experts**.

AI SCIENTIST [TECH LEAD, STATS VQ]

Jan. 2019 - Jan. 2020

- Led international team of 13 data scientists and engineers in building a **predictive player props API** for sports books, generating **\$500k of new business**.
- Provided technical oversight of the VQ product: **cloud data lake** (AWS S3, Glue & Athena), Java Spark **ETL pipeline**, TensorFlow based **model development framework**, cloud infrastructure (AWS), **human-in-the-loop interface** and both internal and client-facing APIs.

DATA SCIENTIST

Jan. 2017 - Jan. 2019

- **STATS Edge**: Collaboratively refactored legacy data science pipeline from Python jobs to **Apache Spark ETL** pipeline in a paired programming (XP) setting, reducing runtime and scaling volume 10x. **Optimized** and deployed existing models to Flask API, **reducing inference time by an order of magnitude**.
- **NBA Live Win Probability**: Achieved **SOTA accuracy (88%)** predicting end of game outcomes for NBA games given in-game context through the use of embedding techniques. Processed and **released a curated dataset** containing 352 features of over 8.7M NBA plays.

Dept. of Physics & Astronomy, UNC - Chapel Hill

Chapel Hill, NC

RESEARCH & TEACHING ASSISTANT

Aug. 2011 - Jan. 2017

- Developed an **analytic model** of afterflow emission from structured, off-axis Gamma-ray Bursts (GRBs) with results comparable to 3D MHD simulations.
- Builder of **PROMPT-SSO** and contributor to the **Skynet Robotic Telescope Network**: a collection of fully automated telescopes, control software, image and data processing pipeline, and web interface.
- Head instructor for **Introductory Astronomy Lab**. Organizer and host of **Morehead Observatory Guest Night**, a weekly public astronomy presentation.

Education

The University of North Carolina at Chapel Hill

PHD CANDIDATE (ABD), PHYSICS & ASTRONOMY

Chapel Hill, NC

Jun. 2014 - Jan. 2017

- Advanced to candidacy by completing all coursework as well as qualifying and preliminary exams.
- Dissertation: *Structured Energy Profiles in Gamma-ray Burst Afterglows: Analytic Calculation of Images and Light Curves for Direct Modeling*

MS, PHYSICS & ASTRONOMY

Aug. 2011 - May 2014

- Thesis: *Application of a Toy Model of Jet Structure to the Afterglow of GRB 080710*

University of California, Santa Cruz

BS, PHYSICS (ASTROPHYSICS), HONORS

Santa Cruz, CA

Sep. 2004 - Jun. 2009

- Thesis: *Searching for Millisecond Pulsars in Gamma-ray Data Using the Fermi LAT*

Honors & Awards

- 2018 **Most Disruptive Entry (2nd Place Overall)**, Vista Consulting Group AI/ML Hackathon
- 2017 **Most Creative Entry (4th Place Overall)**, Automated Insights/Amazon Alexa Hackathon
- 2015 **Graduate Research Fellowship**, NC NASA Space Grant
- 2014 **Graduate Research Fellowship**, NC NASA Space Grant

Chicago, IL

Durham, NC

Chapel Hill, NC

Chapel Hill, NC

Presentations

- Apr 2022 **Basketball Analytics Summit**, "How to Begin Your Journey: Hands on Basketball Analytics in Python"
- Apr 2021 **Basketball Analytics Summit**, "Analytics in the Wild: Advice and Insights from Real World Applications"
- May 2020 **CSBA - Analytics Learning Academy**, "Artificial Intelligence in Sports"
- Apr. 2019 **Basketball Analytics Summit**, "Artificial Intelligence in Sports"
- Nov. 2018 **Promotable Data and Sports**, "Artificial Intelligence in Sports"
- Sep. 2017 **Painless/WISE Sports & Tech Event**, Panel discussion on Sports and Technology.

webinar

webinar

webinar

Durham, NC

Chicago, IL

Chicago, IL