Oran Looney

Data Scientist

olooney@gmail.com
linkedin.com/in/oran-looney
www.oranlooney.com
Madison, WI

Summary

I am a pragmatic, hands-on data scientist seeking technically challenging projects—projects that can't be solved with off-the-shelf methods but instead demand that we roll up our sleeves, break problems down to their fundamentals, and devise a novel approach. I excel at projects that need extensive optimization, involve unusual data or requirements, or call for lesser-known models or statistical techniques.

I hold master's degrees in math and physics, providing a strong foundation for advanced data science work. I have extensive experience in statistics and machine learning, focused on model development with healthcare data. I spent five years as a full-stack developer, giving me excellent programming and database skills. This unique combination has given me a proven track record of solving hard problems.

Experience

Company	Job Title	Time Period
Frazier Healthcare Partners	Director, Data Science and AI	2022-Present
Healthgrades	Senior Data Scientist	2014-2021
Healthgrades	Senior Software Architect	2013-2014
CPM Marketing Group Inc.	Director of Software Development	2012-2013
CPM Marketing Group Inc.	Software Engineer	2007-2012
Epic Systems	Interface Analyst	2005-2007

Education

Degree Earned	Institution	Graduated
M.S. Physics	Indiana University	2004
M.A. Mathematics	Indiana University	2004
B.S. in Physics & Mathematics	University of Oregon	2001

Skills

Core Skills: Machine Learning • Statistics • Software Development • Data Visualization

Tools: Python • NumPy • Pandas • matplotlib • Keras • Hyperopt • scikit-learn • SciPy • statsmodels • SQL • PostgreSQL • Linux • Docker • Databricks • Spark • AWS • Azure • R • Tidyverse • ggplot2

Modeling Expertise: Neural Networks • Generalized Linear Models • Regularized Regression • Hypothesis Testing • ANOVA • Random Forest • Boosted Trees • GAM • ARIMA • k-means • PCA • GMM • MARS • LLMs (e.g. GPT, Llama, BERT) • Vector Embeddings

Soft Skills: Mentoring Data Scientists • Statistical Methodology • Machine Learning Best Practices • Experiment Design • Software Architecture • Scientific Communication

Recent Projects

• Developed a classification model to identify 'at-risk' students given clinical notes and demographics. Evaluated random forest and neural net models before settling on regularized logistic regression due to limited sample size. Advised on deploying to Azure and handed off retraining and model monitoring scripts so the client could be fully self-sufficient going forward.

• Wrote several small chatbot POCs to demonstrate cutting edge techniques such as RAG, asynchronous token streaming, voice (Twilio and retell.ai,) and tool use (allowing the LLM to run SQL.) These POCs helped internal business stakeholders understand the potential (and limitations) of current-gen LLMs.

• Ran a statistical analysis of the impact of operational quality metrics on third-party outcome measures for treating children with autism with ABA therapy. We used a logistic regression model to control for confounding variables such as age and case complexity and were able to demonstrate statistically significant effects for explanatory variables including quality of treatment and clinician experience.

• Advised on a model to predict medication adherence from clinical interviews conducted by pharmacists. The focus was on using interpretable ML techniques to derive insight into which questions had the greatest effect on medication adherence for different classes of drugs.

• Developed a neural net model to estimate the probability of a physician responding to a survey based on facts about the physician, timing and length of the survey, and the incentive offered.

• Advised on a project to use an LLM to identify and flag sections of free-text medical notes relevant to certain conditions such as COPD or Diabetes. The project was designed to save time when reviewing medical records for correctness.

• Worked on a large suite of neural net models used to estimate a patient's future healthcare utilization and trained on more than ten million medical records. Performed hyperparameter optimization, evaluating new data features such as biometrics (height/weight) and census data, and combining hundreds of similar models into a single multi-label model, vastly improving performance.

• Developed an LDA topic model to automatically group providers into sub-specialties using only historical claims data. This could be used to cross-check and augment other specialty data because the sub-specialties discovered by the model were more granular and less reliant on self-reporting.

• Developed a time series model to predict website traffic by geography and category. The model was used to estimate feasible contract sizes and set delivery guarantees for promoted search campaigns.

Selected Writings

Large Language Models: GPT-4o Image Embeddings, LLM Benchmarks

Machine Learning Fundamentals: PCA From Scratch, Backpropagation

Numerical Programming: Playfair Ciphers, Fast Fibonacci

Recreational Mathematics: Angle Addition, Kaprekar