Chris **Addy**

Machine Learning Engineer

484-464-1249 | chris.william.addy@gmail.com | christopheraddy | chrisaddy

Skills

Languages	Python, R, SQL, Javascript, Clojure
Statistics/ML Domains	Bayesian Statistics, LSTMs, encoder-decoder networks, time-series forecasting
NLP	document classification, text extraction, named-entity recognition, masked language modeling
Technologies/Frameworks	Docker, asyncio, aiohttp, boto3/aioboto, GNU Make, CircleCI, FastAPI, dagster, datadog, sentry
ML Technologies/Frameworks	NumPy, Pandas, Plotly, Scikit-Learn, TensorFlow, PyTorch Lightning, Pyro, MLFLow, spaCy
AWS Services	Lamba, S3, DynamoDB, ECR/ECS, SageMaker, CloudWatch, Kinesis

Experience

Pepsico E-Commerce — *Philadelphia, PA*

Machine Learning Engineer — *July 2021- Present*

- Working with stakeholders to gather a priori expectations about data and model outputs to build business expectations into code. Using tools like **Great Expectations** for data expectation testing and **Streamlit** and **Plotly** for interactive visulazations to give business team greater insights into their data.
- Hyperparameter tuning and automating model selection for marketing mix ROI models for Pepsico brands including Pepsi, Mountain Dew, Brisk.
- Optimizing batch jobs to reduce training time and reduce costs.
- Introducing MLOps tools like **MetaFlow** and **MLFlow** to increase reproducibility, system visibility, and increase collaboration among data scientists.

Penn Interactive — *Philadelphia, PA*

Lead Data Scientist - Barstool Sportsbook — *June 2020 - April 2021*

- Building infrastructure for continuous delivery of machine learning algorithms so data science team can focus on model building. Introducing **SageMaker** and **MLFlow** into workflow decreased model ideation and iteration from weeks to days. Introducing dagster as a trusted data pipeline framework between data science and data engineering teams decreased painful overlap and led to both teams being able to coordinate with lighter coupling.
- Production-hardening of RESTful HTTP ML services, including moving to asynchronous frameworks **aioboto**, **aiohttp**, **FastAPI**.
- Building Bayesian models for multivariate testing and pipelines for automated inference so that product can proactively test ideas across multiple measurements.
- Introducing stochastic processes to testing, especially at load; uncovered dozens of critical infrastructure changes missed by deterministic tests and led to zero downtime launch weekend that saw nearly **75k** user registrations and over **$8M USD** in bets.
- Starting as a team of one, built the data science team from the ground up and now leading a group of 4 other machine learning engineers and data scientists.
- Distilling statistical learning models for stakeholders at the highest levels of Penn Interactive as well as our parent company, Penn National Gaming, so teams across the organization feel empowered by data and analysis, not overwhelmed.

Penn Interactive — *Philadelphia, PA*

Senior Data Scientist — *November 2019 - June 2020*

- Implemented recommender system to surface implied user preferences in order to personalize in-app experience and target cohorts for marketing team.
- Masked-Language Modeling with BERT and similar algorithms for transfer learning from **English-to-"Stoolese"** to better understand how our users speak. This language is used as the basis for a number of tasks, including sentiment analysis, sarcasm detection, conversation summarization and collusion detection.
- Introduced Bayesian approach to AB testing so that stakeholders can make statistically-valid decisions without having to rely on p-values and null-hypothesis statistical testing/frequentist style analysis.
- Created real-time datadog dashboards for use across the organization, so everyone could have basic insights into our users.

Heavywater, AI Division of Black Knight *Philadelphia, PA*

Data Scientist *November 2018 - October 2019*

- Expanded the original document-classification algorithm that was capable of classifying 20 document types to classify over **300 document types**.
- Repurposed that original algorithm to serve as a pre-processor for a larger composite model capable of classifying high-volume document types at over **96% accuracy**, find probabilistic boundaries between pages to group and format documents for reviewers. Composite model, which contained multiple machine learning algorithms and pre-processors across multiple HTTP microservices, averaged **7 seconds** to process a **1200 page mortgage**. Led to increase in reviewers' document processing rate from an average of **3-4** documents per hour to **10-20** per hour.
- Introduced named-entity recognition and text extraction to retrieve identity, income, and expense information from tax and banking documents, further decreasing operator time per mortgage.
- Built data pipelines, mostly serverless in design, to collect and store training data in a central environment so that continual improvements to the model were replicable and team members could focus on algorithm design.

Algorhythm *Philadelphia, PA*

Research Manager/Statistical Consultant *December 2016 - November 2018*

- As direct report to CEO, led all statistical modeling of core products. Data primarily consisted of youth self-reported psychometric evaluations.
- Managed a small data science team, including a data engineer and quantitative psychologists, to bring contemporary modeling and data analysis techniques to these and other social science problems.
- Developed unsupervised learning models, notably **confirmatory factor analysis**, **principal component analysis**, and **k-nearest neighbors**, to score and categorize young people in youth development programs on key social and emotional learning metrics.
- Built supervised learning models, including **support vector machines**, **random forests**, **shallow neural nets**, to predict youth growth and give recommendations to those working with youth to help boost their key social and emotional learning metrics.

Projects

Neural Painting GAN • PyTorch Lightning

Generative Adversiarial Network development to produce artwork similar to artwork from the Late Muromachi Period to the Early Edo Period of Japanese portaiture. The data set is interesting in that the styles of art differ drastically.

rrr R • regression • PCA

R library for reduced-rank regression with applications to principle component analysis, linear discriminant analysis and canonical variate analysis. While I don't maintain the library as much as I would like, it has been downloaded over 16,000 times and is used by researchers in the US, Japan, and Australia in a number of fields including oncology, nutrition, and geriatrics.

Education

Temple University *Philadelphia, PA*

Master of Science, Statistics *2015 - 2017*

Notable coursework in statistical machine learning, survival analysis, structural equation modeling, multivariate time-series modeling

Lehigh University *Bethlehem, PA*

Bachelor of Arts, Economics *2006-2010*

Notable coursework in time-series forecasting, causal modeling.