

EXPERIENCE

Proofpoint

Durham, NC

Email cybersecurity leader for the Fortune 100. Tech Lead for five years in central machine learning organization.

Machine Learning Manager

April 2022 - Present

- **LLM Initiatives Team:** Leading the research and POCs on low-latency techniques for LLMs (quantization, pruning, finetuning, hardware), agentic workflows for security operations, RAG architectures for business operations.
- **ML Platform Team:** Designing a new environment for model development to replace SageMaker-centric platform. Leveraging Pachyderm, Prefect, MLflow, all running on kubernetes. Managing 30 different projects across the company.
- **DLP ML Team:** Currently releasing Custom Form classification feature for 1000's of DLP accounts based on single instance of a customer-uploaded sensitive form. Leveraging deep metric learning running in AWS Lambda and ECS.

Senior Data Scientist & Data Scientist

May 2019 - April 2022

- **Triage Alert Classifier:** automatic categorization of security alerts for analysts. Leveraging XGBoost and approximate NN. Saving average of 3 hours per account with accuracy of 99%.
- **Promissory Images Classifier:** designed and delivered an image classification system for financially promissory image detection. Leveraging object detection, classical ML, and approximate NN. Key differentiator for all finserv customers. Precision/recall maintaining over 24+ months with retrainings and annotation. Created, tuned autoscaling deployment.
- **NLP Classifier System:** implemented and maintained and scaled a spaCy-centric framework to serve 54 NLP classifiers over 100Ks social media posts daily across 1000's of customers. Addressed customers FP/FNs weekly.
- **Physical Locations Classifier:** constructed custom dataset of 1000's of spans for an NER model to detect mentions of a location with an address. Selected DistilBERT over BiLSTM-CRF. Enabled for hundreds of Patrol Security accounts.

Automated Insights

Durham, NC

Natural language generation startup before NLP had its Imagenet moment. Lead Data Scientist on two projects.

Senior Data Scientist & Data Scientist

July 2017 - May 2019

- **Augmented NLG:** Designed and prototyped a natural language generation system using Node.js, React as an add-on feature for BI dashboards. Published a patent.
- **Contextual Synonym Suggester:** Productionized custom word vectors trained using gensim and Common Crawl for synonym suggestion as a Flask microservice. Core ML feature of Wordsmith product.

Duke University

Durham, NC

PhD Program in Machine Learning Lab led by Larry Carin. Mastered out to take first job at Automated Insights.

PhD Research Assistant

Aug 2014 - July 2017

- **Weapons Detection in TSA Bag Scans:** Research on VAE and applied object detection (i.e. YOLO, Faster-RCNN) to X-ray images for high-precision detection of guns, knives. Published open-source VAE library in Tensorflow.
- **Optical Detection of Cancer:** Research on multiple instance learning and applied tree-based models for classifying optical scans of breast cancer tissue. Published open-source MIL library in Tensorflow.

TECHNICAL SKILLS

- **Languages:** Python, Bash, Git, SQL, Terraform, Javascript, Go
- **Techniques:** Linear & Tree-based Models, CNNs, RNNs, GNNs, Autoencoders, Transformers, Approximate NN, Clustering, Knowledge Distillation, Semi-Supervised Classification (PU, Weak, Metric)
- **Technologies:** PyTorch, TensorFlow, scikit-learn, XGBoost, NumPy, Pandas, spaCy, Docker, OpenCV, Pachyderm, DVC, MLflow, Prefect, k8s, Elasticsearch, Postgres, MySQL, Cassandra, LangChain, VertexAI, SageMaker, AWS/GCP Stack

EDUCATION

- **Duke University**

Master of Science in Electrical Engineering

Durham, NC

2014 - 2017

- **North Carolina State University**

Bachelor of Science in Biomedical Engineering

Raleigh, NC

2010 - 2014