

WORK EXPERIENCE

Proofpoint

Durham, NC

Machine Learning Manager and Tech Lead (Apr 2022 - Present)

- **Text-to-SQL:** NL querying of data in the Cloud Protection Platform. Leveraging knowledge distillation and low-rank LMs. Slated for production Q1 2024.
- **Form Recognition for DLP:** leading a team of 3 engineers and data scientists. Similarity search algorithm for customer-specific sensitive forms. Slated for production Q4 2023.
- **Machine Assisted Triage for ITM:** led a team of 2 engineers. Instance-based anomaly detection algorithm for high-FP insider threat alerts. In prod Q1 2023, precision > 99%.

Data Scientist (May 2019 - Jul 2021), Senior Data Scientist (Aug 2021 - Mar 2022)

- **Proofpoint Patrol data science tech lead:** responsible for the maintenance, evaluation, and latency of 54 classifiers across 100's of customers, processing 100k's of social media posts daily. Trained my replacements in 2022.
- **Promissory Images Classifier:** designed and delivered an image classification system that detects images that may make a promise of financial gain. Organized an in-house annotation team to acquire labels. Published a Proofpoint blog post and a patent.
- **Physical Locations Classifier:** delivered an NER model to detect mentions of a location with an address. Significant increase in accuracy and latency by replacing BiLSTM-CRF with DistilBERT. Delivered the required 90% precision.
- **NLP Classifier Framework:** ideated and implemented a spaCy-centric framework to serve all classifiers (rules-based and ML). Patented one classifier.

Automated Insights

Durham, NC

Senior Data Scientist (Dec 2018 - Apr 2019), Data Scientist (Aug 2017 - Nov 2018)

- **Augmented NLG:** Designed and prototyped a natural language generation system as an add-on feature for BI dashboards. Published a patent.
- **Contextual Synonym Suggester:** productionized the custom word vectors (trained using gensim and Common Crawl) with AWS services and Flask and Docker.

TECHNICAL SKILLS

Languages: Python, Bash, Git, SQL, Terraform, Javascript, Go

Technologies: PyTorch, TensorFlow, scikit-learn, XGBoost, NumPy, Pandas, spaCy, Docker, OpenCV, MLflow, DVC, Elasticsearch, Postgres, Cassandra, LangChain, AWS stack

Techniques: Linear & Tree-based Models, CNNs, RNNs, GNNs, Transformers, Approximate NN, Clustering, Knowledge Distillation, Semi-Supervised (PU, Weak, Metric)

ACADEMIC ENGAGEMENT

Duke MIDS Capstone Advisor: advised different groups of 3-5 Masters students since 2020. Logo Detection (2021), ML Churn Reduction (2022), Low-latency Transformers (2023).

Washington State University Collaborator: managing the joint-collaboration with Proofpoint on conformal predictions for binary classifiers.

Summer Intern Mentor: managed one undergrad/grad student during summers of 2021-2023.

EDUCATION

M.S. in Engineering, Duke University 2017

Abdicated PhD Student

Dr. Nimmi Ramanujam (BME): Point-of-Care Optical Cancer Imaging

Dr. Larry Carin (ECE): Deep Learning for Threat Detection in TSA Bag Scans

B.Sc. in Biomedical Engineering, NC State University 2014

Bioinstrumentation Concentration. Minor in Piano Performance.