Dan Salo

dancsalo [at] gmail.com • (919) 698-2742
Github • LinkedIn • Website

WORK EXPERIENCE

Proofpoint   Durham, NC

Machine Learning Manager and Tech Lead (Apr 2022 - Present)
- **ML Platform 2.0**: leading a team of 4. Migrating from SageMaker to K8s ecosystem. Managing data access for and supporting 8 models in production.
- **KD for LLMs**: leading a team of 3. Distilling LLMs to low-latency surfacing of semantic labels relevant to phishing for threat analysts. Processing 1M emails daily.
- **Form Recognition for DLP**: led a team of 3. Visual similarity search algorithm for customer-specific sensitive forms. Processing 10Ks images daily. Precision > 95%.
- **Machine Assisted Triage for ITM**: led a team of 3. Instance-based multi-class algorithm for insider threat alerts. Processing 1Ks alerts daily. Precision > 99%.

Data Scientist (May 2019 - Jul 2021), Senior Data Scientist (Aug 2021 - Mar 2022)
- **Promissory Images Classifier**: designed and delivered an image classification system for financially promissory images. Leveraging object detection and classical ML. Published a Proofpoint blog post and a patent.
- **Physical Locations Classifier**: delivered an NER model to detect mentions of a location with an address by fine-tuning DistilBERT instead of BiLSTM-CRF. Precision > 95%.
- **NLP Classifier Framework**: implemented and maintained a spaCy-centric framework to serve 54 NLP classifiers over 100Ks social media posts daily. 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 custom word vectors trained using gensim and Common Crawl for synonym suggestion. Processing 100s of requests daily.

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.