Khanh (Chris) Tran

khanh.tran@simon.rochester.edu | (716) 513-4637 | linkedin.com/in/chriskhanhtran | chriskhanhtran.github.io

EXPERIENCE

Etsy, Inc. **Brooklyn, NY**
Computer Vision Intern July 2020 – Present

- Applied cutting-edge Computer Vision research to improve personalization and image-based recommendations.
- Developed DL architecture to extract aesthetic features from images uploaded by users to find style-matching items.
- Trained object detection and image segmentation models on Etsy's datasets with Detectron2.

Skim AI Technologies, Inc. **New York, NY**
Machine Learning Research Intern (NLP) Oct. 2019 – Present

- Trained BERT model for document classification and NER on legal documents, processed and performed inference on 10K documents monthly with 96% accuracy, saving client 250 hours of data labeling per month.
- Deployed BART and T5 models for news summarization for client in media industry.
- Built large-scale sentiment analysis API for English and Spanish articles with Flask.

Aurubis Buffalo, Inc. **Rochester, NY**
Data Science Capstone Project Feb. 2020 – May 2020

- Developed data preprocessing and regression pipeline to predict yield percentage of coil production with high accuracy, improving scheduling efficiency and lowering inventory cost.
- Built web app to deploy regression pipeline to generate yield predictions for future production.

Tax Technologies, Inc. **Buffalo, NY**
Tax Intern Mar. 2019 – July 2019

- Performed application testing and collaborated with software engineers to build enhancement update for Tax Series.
- Provided technical supports to Fortune 500 clients and assisted on implementation engagements for new clients.
- Conducted in-depth research on tax regulations and e-file requirements in 32 states and 4 foreign countries.

PROJECTS & COMPETITIONS (more details at https://chriskhanhtran.github.io/)

Extractive Summarization with BERT

- Implemented paper Text Summarization with Pretrained Encoders (Liu & Lapata, 2019).
- Trained MobileBERT for extractive summarization and build web app to scrape and summarize news articles.

Social Media Analytics for Airline Industry: Fine-tuning BERT for Sentiment Analysis

- Vectorized tweets with fastText vectors and trained CNN model for sentiment classification using PyTorch.
- Fine-tuned BERT model to detect negative tweets, achieving 10% accuracy improvement over TF-IDF baseline.

Humana-Mays Healthcare Analytics Competition – Top 50 out of 460 teams

- Preprocessed 7M EHR records of 20K patients, performed feature engineering from past diagnoses and medical claims.
- Built LightGBM model to predict patients with long-term opioid therapy, achieving 0.88 AUC score.

EDUCATION

UNIVERSITY OF ROCHESTER **Rochester, NY**
Master of Science in Business Analytics (STEM); GPA: 3.96/4.00 Dec. 2020 (Available for full-time from May 2020)

- Coursework: Core Statistics, R Programming, Predictive Analytics with Python (Machine Learning), Causal Analytics with R (A/B Test), Social Media Analytics (NLP), Database Management (SQL, Cypher), Big Data (Hive, Spark)

NIAGARA UNIVERSITY **Niagara University, NY**
Bachelor of Business Administration in Accounting; GPA: 3.99/4.00 2019

- Dean's List (all attended semesters); Top 5 graduated student
- Coursework: Business Analytics, Linear Models, Management Information Systems, Econometrics

SKILLS

Programming: Python (NumPy, Pandas, Scikit-learn, PyTorch, TensorFlow), Big Data (Spark, Hive), R, SQL, Cypher
NLP: Sentiment Analysis, NER, Language Generation, Summarization
Others: AWS EC2, GCP, Tableau, SAS, SPSS, Microsoft Office, Adobe Suite