# 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.
- Designed and trained Deep Siamese architecture with PyTorch to extract aesthetic features from listing pictures.
- Performed efficient similarity search with FAISS library to find nearest neighbors for recommendations.
- Trained Faster R-CNN and Mask R-CNN for object detection and instance segmentation for Home & Living pictures.

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

- Trained Deep Learning models (BERT, RNN-LSTM) for NER and document classification on legal domain, processed 10K documents monthly with 96% accuracy, saving client 250 hours of data labeling per month.
- Pre-trained state-of-the-art transformer language models (BERT, RoBERTa, Electra) on large Spanish corpus from scratch on multi-GPU AWS machine and appied transfer learning on clients' Spanish NLP projects.
- Deployed BART and MobileBERT models for news summarization for client in media industry, achieving 5.5x faster speed than original research paper while retaining 95% ROUGE score.
- Built large-scale sentiment analysis API for news articles and tweets with Twitter API and Flask.

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

- Collaborated with Aurubis Buffalo as part of final practicum project at University of Rochester.
- Developed data preprocessing and regression pipeline to predict yield percentage of coil production to improve scheduling efficiency and lowering inventory cost.
- Built web app to deploy regression pipeline to generate yield predictions for future production.

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

**Kaggle: Global Wheat Detection Competition with Detectron2**

- Preprocessed images and bounding boxes on 3,000 outdoor images of wheat plants worldwide.
- Trained Faster R-CNN model to estimate number and size of wheat heads, achieving high accuracy with 52 mAP.

**Humana-Mays Healthcare Analytics Competition: Opioid Abuse Prediction – 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.

**Kaggle: Advanced Regression Techniques in House Price Prediction – Top 0.6% on leaderboard**

- Performed comprehensive data analysis, data cleaning and feature engineering on Ames, Iowa housing dataset.
- Ensembled Ridge, Lasso, XGBoost, and LightGBM models to predict house prices.

## 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 (Hive, Spark), R, SQL, Cypher
**NLP:** Sentiment Analysis, NER, Language Generation, Summarization
**Others:** AWS EC2, GCP, Tableau, SAS, SPSS, Microsoft Office, Adobe Suite