Khanh (Chris) Tran

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

EDUCATION

UNIVERSITY OF ROCHESTER — **Rochester, NY**
Master of Science in Business Analytics (STEM); GPA: 3.97/4.00 — Dec. 2020

- Coursework: Core Statistics, R Programming, Predictive Analytics with Python (Machine Learning), Causal Analytics with R (A/B Testing), Social Media Analytics (NLP), Data Management, Big Data, Pricing Analytics

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

EXPERIENCE

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

- Trained and fine-tuned Deep Learning models (Transformers, CNN, RNN-LSTM) for multiple NLP tasks including Document Classification, Sentiment Analysis, Information Extraction, Summarization and Language Modeling.
- Trained Spanish BERT model from scratch on 150 GB of Open Super-large Crawled Corpus (OSCAR) on AWS EC2 instances and fine-tuned pretrained model for downstream tasks in clients' projects.
- Vectorized documents with pretrained word vectors such as Word2vec, fastText and ELMo to feed classification algorithms.
- Read latest NLP research papers, reported findings to CTO and wrote codes to reproduce research results.

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

- Provided technical supports to Fortune 500 clients utilizing Tax Series – an all-inclusive SaaS global data collection, tax compliance and provision software, and assisted on implementation engagements for new clients.
- Conducted essential application diagnostics on client financial data, including periodically generating technical reports, maintaining data integrity and monitoring client databases.
- Conducted in-depth research on tax regulations and e-file requirements in 32 states and four foreign countries.
- Performed application testing and collaborated with software engineers to build enhancement update for Tax Series.

Business Analytics Competition & Conference @ Manhattan College — **New York City, NY**
Data Analytics Team Leader — Feb. 2018 – May 2018

- Led a team of four students to analyze NYC and Boston government spending and contract data, winning 2nd best research poster out of 18 participating colleges.
- Cleansed (missing data, outlier detection, duplications), integrated (merge, join, subset) large datasets of 6 million records, and performed exploratory data analysis and visualization using Python and Tableau.
- Built statistical models to predict government spending, crime rate and education quality.

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

Detect Negative Airline Tweets: CNN and fine-tuned BERT for Sentiment Analysis

- Vectorized tweet data with pretrained fastText word embedings and trained CNN model using PyTorch.
- Fine-tuned BERT model to detect negative tweets, achieving 10% accuracy improvement over TF-IDF.

Credit Risk Prediction Web App (https://credit-risk.herokuapp.com/)

- Preprocessed dataset of 10,000 credit applications and built machine learning models to predict credit default risk.
- Built interactive user inteface using Streamlit and deployed web app on GitHub and Heroku server.

Kaggle Competition: 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.

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

- Preprocessed 7 million medical records of 20,000 patients, identified and labeled patients with long-term opioid therapy and performed feature engineering from past diagnoses, medical claims and prescriptions.
- Built LightGBM model to predict patients with long-term opioid therapy, achieving 0.88 AUC score.

Predict Breast Cancer with PCA, RF and SVM using Python

- Performed comprehensive exploratory data analysis and PCA on the Breast Cancer Wisconsin dataset.
- Trained Random Forest and SVM models to detect breast cancer, achieving 97% accuracy rate.

SKILLS

Programming: Python (NumPy, Pandas , Scikit-learn, TensorFlow, PyTorch), R, SQL, MATLAB
Visualization & Statistical Software: Tableau, Python (Matplotlib, Seaborn), SAS, SPSS, Adobe Suite