Khanh (Chris) Tran

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

EDUCATION

SIMON BUSINESS SCHOOL, University of Rochester — Rochester, NY
Master of Science in Business Analytics (STEM) — May 2020

- Coursework: Core Statistics, Programming for Analytics, Causal and Predictive Analytics using R, Advanced Predictive Analytics with Python, Text Analytics, Data Management, Big Data, Pricing Analytics

NIAGARA UNIVERSITY — Niagara University, NY
Bachelor of Business Administration in Accounting — 2019

GPA: 3.99/4.00; Dean's List (all attended semesters); Top 5 graduated student

- Coursework: Business Analytics, Linear Models, Management Information Systems, Econometrics
- Awards: Beta Gamma Sigma Honor Society, Everett Ockerman Award for Academic Excellence

EXPERIENCE

SKIM AI TECHNOLOGIES — New York City, NY
NLP and Deep Learning Intern — Oct. 2019 – Present

- Implemented state-of-the-art word embedding techniques including Word2vec, BERT, and massive pre-trained language models to transform natural language data into useful features to feed classification algorithms.
- Utilized PyTorch to build and fine-tune deep CNN and RNN models for NLP tasks including Named Entity Recognition, Sentiment Analysis, Fact Checking, Language Generation and Multi-document Summarization.
- Reviewed and reported latest NLP research papers to CTO and applied latest research on enterprise-level data.

TAX TECHNOLOGIES, INC. — Buffalo, NY
Tax Intern — Mar. 2019 – July 2019

- 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, logged technical reports 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 discover insights from NYC and Boston government's spending datasets and won runner-up prize for best research poster out of 18 participating colleges.
- Cleansed (missing data, outlier detection, duplications) and integrated (merge, join, subset) large data sets (6 million records) of governmental spending, contracts and KPI metrics.
- Utilized Python and Tableau to perform exploratory data analysis to understand governmental spending distribution.
- Built statistical models to determine socioeconomic factors affecting government spending and predict government KPI metrics, crime rate and education quality.

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

Kaggle Competition: Predict House Prices – Advanced Regression Techniques – Top 0.6% on leaderboard

- Performed comprehensive EDA, data cleaning and feature engineering on Ames, Iowa housing data set.
- Built and tuned 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 EDA on the Breast Cancer Wisconsin data set.
- Trained Random Forest and Support Vector Machine models to detect breast cancer, achieving 97% accuracy rate.

Detect Spam Messages – Natural Language Processing with Python

- Tokenized and vectorized text messages using TF-IDF and bag-of-words model.
- Utilized Naïve-Bayes algorithm to classify text messages into normal and spam, achieving 95% accuracy rate.

SKILLS

Programming: Python (NumPy, Pandas , Scikit-learn, TensorFlow, PyTorch), R, SQL, MATLAB
Visualization and Statistical Software: Tableau, Python (Matplotlib, Seaborn), SAS, SPSS, Adobe Suite
Machine Learning: Supervised Learning (Regressions, Random Forest, SVM), Unsupervised Learning (Clustering, PCA)