XIAOTIAN ZHAN

Phone: (+1) 734-926-7082 **Address:** 327 West Tremont Ave Apt 265, Charlotte, NC 28203
Email: billtsan1994@gmail.com **LinkedIn:** https://www.linkedin.com/in/xiaotian-zhan/

EDUCATION

UNIVERSITY OF MICHIGAN Ann Arbor, USA
MS in Applied Statistics (August 2017 – April 2019) ***GPA:*** 3.8/4.0
SUN YAT-SEN UNIVERSITY Guangzhou, China
BS in Statistics (September 2013 – June 2017) ***GPA:*** 3.7/4.0

SKILLS

Programming Skills: Python(anaconda, numpy, pandas, NLTK), R, SQL, Matlab, SAS, Java, Linux.
Data Science Skills: Web Crawler, AWS, Scikit-Learn, TensorFlow, Keras, NumPy, Pandas, ggplot2

WORK EXPERIENCE

Wells Fargo Charlotte, USA
Quantitative Analytics Specialist (July 2019 – Now)

- Built an XGBoost model to compare with traditional scorecard model on credit data. Performed parameter tuning and obtained 2% lift on AUC score ($0.903 \rightarrow 0.921$) and 4.64% lift on KS score ($64.7 \rightarrow 67.7$).
- Leading an ML curriculum project, developing team tutorial documents, building Python interactive demos and making presentations to help team members learn python machine learning contents.

ProQuest LLC Ann Arbor, USA
Data Science (Natural Language Processing) Intern (June 2018 – August 2018)

- Cleaned and processed 10 GB user logs data with AWS in Linux. Analyzed the data and made visualizations with R and Tableau satisfying the needs of the product team. Helped find drawbacks of a current product.
- Developed an automatic medical literature processing system in Python to help literature reviewers improve their efficiency. It used web crawlers to fill missing information, used NLP package spaCy and google map API to analyze authors' geography information and highlighted medical terms automatically.
- Developed a hybrid Bi-LSTM neural network NLP model in Python with Keras. Trained it in AWS with GPU acceleration. The model could label different sections of abstracts in medical papers, such as Background, Methods and so on. The final accuracy was 85% and the model was submitted to ProQuest github repo.

PROJECTS

ProQuest LLC – Student Multidisciplinary Design Team Ann Arbor, USA
Improve Drug Safety Using Machine Learning Techniques (January 2018 – December 2018)

- Wrote SQL queries to retrieve and filter medical paper abstracts data provided by ProQuest. And wrote Python web crawling scripts to collect information about drugs.
- Used Naive Bayes and SVM in scikit-learn to classify articles in terms of different seriousness levels, then implemented Text CNN in Python with TensorFlow to improve the accuracy, which was 80% finally.

UNIVERSITY OF MICHIGAN Ann Arbor, USA
Recommendation System For Movies in Apache Spark (November 2018 - December 2018)

- Built a data ETL pipeline with Spark for big data of Movielens which has 27 million movie rating records.
- Implemented Alternative Least Square recommendation algorithm in Spark and used K-fold crossvalidation to tune hyperparameters with a smaller dataset. The best model reached RMSE of 0.86 on test set.

Emotion Identification on Twitter Data (October 2018 - December 2018)

- Built a Twitter query API in Python to help search for tweets and processed the text data with NLTK.
- Used wordcloud and LDAvis in Python to visualize the distribution of keywords from different emotions.
- Built LSTM and CNN models in Python with Keras to classify tweets into 6 different emotion: love, anger, surprise, joy, sadness, fear. The best accuracy was 57.6% and beat the baseline of 51.86% described in the paper we referred to.