Tian Yun

1121 Crowne Oaks Circle | Winston-Salem, NC 27106 | tttyuntian@hotmail.com | +1-530-551-9835 (US)

EDUCATION

Brown University Providence, RI
Master of Science in Computer Science *September 2020 – May 2022*
- **Relevant Coursework**: Deep Learning, Special Topics in Computational Linguistics

Wake Forest University Winston-Salem, NC
Bachelor of Science in Mathematical Statistics *GPA: 3.92/4.00* *August 2016 – May 2020*
Bachelor of Science in Computer Science *GPA: 3.96/4.00*
- **Relevant Coursework**: Database Management System, Parallel Computing, Machine Learning, Statistical Learning, Time Series Analysis, Categorical Data & Multilevel Models, Multivariate Statistics
- **Honor & Award:** Dean's list all semesters; 4th place in ICPC Mid-Atlantic USA Regional Contest 2019 at UNC-Chapel Hill University; Honorable Mention in ASA DataFest 2019 at Duke University; Honorable Mention in COMAP's Mathematical Contest in Modeling 2019

PROFESSIONAL EXPERIENCE

Tencent Shenzhen, China
Data Analyst Intern *May – August 2019*
- Aggregated relevant data from HDFS with SQL (more than 10 billion rows per day) to monitor apps' performances
- Developed back-end service with Python Flask and Git Webhook to help colleagues to automatically organize and upload scripts to task scheduling system
- Evaluated potential factors leading to the loss of daily active users of QQ Kandian
- Trained GBDT with PySpark and Sklearn libraries to predict if daily active users will leave in the following week

China Minsheng Bank Beijing, China
Data Analyst Intern *May – July 2018*
- Aggregated and processed data for 8 million customers and obtained 500 thousand records with all information of the customers (over 200 features) using SQL on DBM2
- Fitted models (12.2% recall rate in GBDT) with Sklearn library to find the target customers (48,000/8 million)

RESEARCH EXPERIENCE

Detecting Latent Topics and Trends in Pediatric Clinical Trial Research using Dynamic Topic Modeling Winston-Salem, NC
Wake Forest University | Advisor: Dr. Natalia Khuri *June 2020 – September 2020*
- Extracted and preprocessed pediatric research papers from S2ORC dataset using NLTK library
- Evaluated research trends of pediatric clinical trial research with DTM in Gensim library
- Yun, Tian and Khuri, Natalia. *Detecting Latent Topics and Trends in Pediatric Clinical Trial Research using Dynamic Topic Modeling*. Drafted.

Mining Drug Labels for Pediatric Information Winston-Salem, NC
Wake Forest University | Advisor: Dr. Natalia Khuri *May 2020 – June 2020*
- Extracted and preprocessed drug label dataset using NLTK library
- Fitted SVM, decision tree, and KNN models with Sklearn library to predict whether a drug label contains pediatric information and analyzed performances of the models
- Yun, Tian and Garg, Deepti and Khuri, Natalia. *Mining Drug Labels for Pediatric Information*. Submitted.

Natural Language Processing and Topic Modeling, Statistics Senior Thesis Winston-Salem, NC
Wake Forest University | Advisor: Dr. Nicole Dalzell *August 2019 – May 2020*
- Conducted sentiment analysis on literature works and Twitter comments dataset using ideas of bag-of-words, n-grams, word hashing, and TF-IDF
- Derived latent Dirichlet allocation (LDA) with Gibbs sampling process
- Implemented LDA with Gibbs sampling process to classify documents with latent topics on R

Object Classification and Segmentation, Computer Science Honor Program Winston-Salem, NC
Wake Forest University | Advisor: Dr. Natalia Khuri *August 2019 – May 2020*
- Preprocessed DSTL satellite images with data augmentation and min-max normalization
- Constructed U-Net architecture to segment and detect roads in satellite maps with Keras library
- Set up experiments to explore the relationship between hyperparameters and performance of U-Net

COMPUTER AND LANGUAGE SKILLS

Programming Language: Python, Java, R, SQL, C++, C, Matlab
Framework & Tool: (Python) scikit-learn, numpy, pandas, genism, nltk, keras; (C++) lpthread