Tian Yun

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

EDUCATION

Wake Forest University Winston-Salem, NC

Bachelor of Science in Mathematical Statistics *August 2016 – May 2020*

Bachelor of Science in Computer Science

- **Major GPA: (Mathematical Statistics)** 3.92/4.00 (**Computer Science)** 3.96/4.00
- **Computer Science Relevant Coursework**: Database Management System, Data Structures and Algorithms, Parallel Computing, Machine Learning
- **Mathematical Statistics Relevant Coursework:** Statistical Learning, Time Series Analysis, Computational & Non-parametric Statistics, Categorical Data & Multilevel Models, Multivariate Statistics
- **Honor:** Dean's list all semesters; Member of Pi Mu Epsilon Honor Society; Member of Upsilon Pi Epsilon Honor Society; Member of Phi Beta Kappa Honor Society
- **Award:** 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*

- Extracted relevant data from HDFS (over 10 billion rows of data per day); processed data to monitor the performances of QQ Kandian and Kandian APP using Hive-SQL or Spark-SQL
- Used Flask, Python and Git Webhook to develop back-end service to help colleagues to upload the scripts to the task scheduling system by only one command "git push"
- Evaluated the causal relationships between the loss of daily active users and the potential factors; Predicted if the daily active users will leave in the following week by fitting models, like random forest, GBDT and etc., by using PySpark and Python (GBDT has AUC score of 0.9173)

China Minsheng Bank Beijing, China

Data Analyst Intern *May – July 2018*

- Aggregated and evaluated data for 8 million customers and obtained 500 thousand records with all the information of the customers (i.e. over 200 features) using SQL on DBM2
- Fitted models (i.e. 12.2% recall rate in GBDT) in Spyder to find the target customers (i.e. 48,000/8 million); helped the bank to target these customers and to sell the financial products

RESEARCH EXPERIENCE

Natural Language Processing and Sentiment Analysis, Statistics Senior Thesis Winston-Salem, NC

Wake Forest University *Auguest 2019 – Present*

- Conducted sentiment analysis on literature works and Twitter comments dataset using ideas of bag-of-words, n-grams, word hashing, TF-IDF, etc.
- Studied and derived latent Dirichlet allocation (LDA) with Gibbs sampling process and used R to finish coding my own version of LDA with Gibbs sampling process to classify documents with latent topics

Object Classification and Segmentation, Computer Science Honor Program Winston-Salem, NC

Wake Forest University *September 2019 – Present*

- Studied methods for image data preprocessing and data augmentation
- Used Keras library to construct U-Net architecture to segment and detect the map-related objects in satellite maps
- Set up experiments to explore the relationship between the hyperparameters and the performance of U-Net

COMPUTER AND LANGUAGE SKILLS

Programming Language: Python, Java, R, SQL, C++, C, Matlab

Software: Tableau, DBM2, SQLiteStudio, Microsoft Excel, Word, PowerPoint

Python Library: scikit-learn, numpy, pandas, keras

Language: Fluent in Mandarin, English, and Cantonese