Zekun Wang

zekun.wang.us@gmail.com | +1 (615) 972-0128 | waittim.github.io | in/zekun-wang

EDUCATION

Vanderbilt University — **Nashville, TN**
*Master of Science in **Data Science**, GPA: 3.96/4* — *Aug. 2019 - May. 2021*
Core courses: Data Science Algorithms, Data management system, Modeling & Machine Learning, Big Data & Scaling

Central South University — **Changsha, China**
*Bachelor of Science in **Statistics*** — *Sep. 2014 - Jun. 2018*

EXPERIENCE

DiDi (Uber China, UEBA SaaS) — **Mountain View, CA**
Data Science Intern — *Apr. 2020 - Aug. 2020*

- Established 1st version of Login-Risk detection for the company. Provided real-time/monthly user risk score to controller.
  - Monitored 15K user's behavior through building real-time models with Kernel Density Estimation to identify baseline.
  - Performed feature engineering and built automatic pipelines for key metrics. Identified monthly risky users with anomaly detection model.
- Built an automatic program to informed sensor testers on software updates by checking common software hash daily. Reduced Mean Time To Detection by 90%.
- Constructed features based on agent log data and accelerated the data pipeline for verifying the cause of high CPU load.

Latin American Public Opinion Project (LAPOP), Vanderbilt University — **Nashville, TN**
Research Assistant - Data Science — *Feb. 2020 - Apr. 2020*

- Built intuitive and interactive tool to strengthen understanding of Latin America's public opinions for the general public.
- Created a dashboard with back-end ETL and front-end data visualization for unskilled users by Shiny R, allowing users to select time, area, and topic of interest to learn key statics, time-series trends and cross-analysis.
- Created effective ETL pipelines for 230K survey data from 38 countries, connecting with iterative Tableau dashboard.

Hunan Fenghuang Taifeng Cultural Tourism Investment Ltd. — **Fenghuang, China**
Intern, Business Analysis Intern Group Leader — *Jan. 2018 - Mar. 2018*

- Led a team of 15 on the pioneering analysis for a $300M tourism development project. The strategy was adopted and received the highest employee evaluation.
- Predicted tourism market development with ARIMA. Built User Portrait by analyzing tourist demand & behavior.
- Increased profit by projected $5M through implementing strategies such as featured brands and experiential consumption.

PROJECTS

Hurricane Catastrophic Loss Prediction - The General — *Mar. 2020 - Apr. 2020*

- Predicted the potential loss of hurricane with $R^2 = 0.8$ based on the policy&claim data for advanced decision making.
- Established a composite pipeline to address unbalanced data including cleaning, feature engineering and modeling.
- Identified high-risk vehicles and predicted loss using Neural Networks and a Random Forest Model.

Trending Video Analysis and Prediction — *Oct. 2019 - Nov. 2019*

- Provided guidelines for creating trending videos with sentiment analysis for 41K trending videos on YouTube.
- Built an NLP model with Naïve Bayes to predict how trendy a video would be with Python(nltk, sklearn).

Natural Background Leaf Family-Classification with Deep Learning — *Mar. 2018 - Jun. 2018*

- Earned 1st Prize (Undergraduate Excellent Graduation Thesis) through developing algorithm to classify leaf families.
- Improved Automatic Marker-controlled Watershed Segmentation Algorithm by Matlab and increased accuracy by 50%.
- Fine-tuned deep Convolutional Neural Networks on Caffe to identify the leaf family with 94% accuracy.

SKILLS

Statistics: Linear, Logistic, Ridge, Lasso, Probability
Machine learning: Random Forest, Isolation Forest, k-means, PCA, Neural Networks (CNN)
Programming & Others: Python (Pandas, sklearn, Keras, PySpark, BeautifulSoup), R(ggplot2, Shiny, H2o.ai), Git, Matlab, SPSS, Tableau, RDBMS(MySQL)