Zekun Wang

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

EDUCATION

Vanderbilt University **Nashville, TN**
*Master of Science in **Data Science**, GPA: 3.85/4* *Aug. 2019 - May. 2021*

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

EXPERIENCE

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

- Established 1st version of Login-Risk detection for the company. Provided real-time/monthly user risk score for early warning of data leakage, account sharing, external hackers, etc.
  - Monitored 15K user's behavior through building real-time models with Kernel Density Estimation to identify baselines.
  - Performed feature engineering pipelines for key metrics. Identified monthly risky users with anomaly detection.
- Automated software update notifications for sensor testers' convenience. Reduced Mean-Time-to-Detection by 90%.
- Constructed pipeline to extract sensor usage data from service log to investigate CPU overheating issue on user device. Identified causes and provided recommendations to optimize detection rules, reduced user complaint reports by 10%.

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

- Built intuitive and interactive tools to strengthen understanding of Latin America's public opinions for the general public.
- Participated in building an automated ETL to process 230k of survey data in different formats from 38 countries.
- Established dashboards for unskilled users by Shiny and Tableau, enabling them to perform basic analysis by clicking.

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

- Led a team of 14 on the pioneering analysis for a $300M tourism project. Strategy was adopted with the highest 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

WearMask(web-based efficient AI recognition of mask: facemask-detection.com) *Nov. 2020 - Dec. 2020*

- Modified and trained the Yolo-Fastest model with 9K collected Mask/NoMask images based on PyTorch.
- Converted the model to NCNN through ONNX and compiled it by WebAssembly to run in the JavaScript environment.
- Established a website and supported all common devices, has gained more than 2000 users within 1 month.

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 by compound models including Neural Networks and Random Forest.

Think and Get Jobs - H1B Visa Analysis *Oct. 2019 - Dec. 2019*

- Completed cleaning, text process feature engineering with R based on 3 million H1B Visa application data.
- Analyzed and visualized the distribution and trend of different data roles in various industries, companies, and regions.

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 & ML: Linear, Logistic, Ridge, Lasso, Random Forest, Isolation Forest, k-means, PCA, CNN
Programming & Others: Python (Pandas, sklearn, PySpark, BeautifulSoup), R(ggplot2, Shiny, H2o.ai), Git, Matlab, Tableau, RDBMS(MySQL)