Zhixing (Jason) He

Cell phone: +1(703) 625-4756
E-mail: hezhixing957@gmail.com
LinkedIn: linkedin.com/in/Zhixing-he
GitHub: github.com/hhzzxx957

SUMMARY

Ph.D. in Physics. 3+ years of experience in data analysis from cross-disciplinary research and project work. Proven skills in developing efficient programs. Proficient in Python and R. Experienced in machine learning and statistics. Excellent in communicating with a variety of audiences and explaining technical details. Highly motivated and self-disciplined.

EDUCATION

Virginia Tech	Ph.D. in Physics	*Aug 2013 - Mar 2020*
Zhejiang University	B.S. in Physics	*Aug 2009 - May 2013*

SKILLS

Programming: Python (Pandas, PySpark, numpy, scikit-learn, Matplotlib), R, MATLAB, SAS, SQL, Bash, AWS
Machine learning: Regression, Classification, Clustering, NLP, Deep neural network
Stats and Math: Hypothesis Testing, Linear Algebra, Bayesian Inference, Stochastic Process, Multivariate Calculus
Soft skills: Problem-solving, self-motivated, excellent communication, team-working, time management, creative

PROJECTS AND RESEARCH EXPERIENCE

Nano-structure detection by quantitative optical anisotropy imaging — *Oct 2017 - Dec 2019*

- **Designed** a novel optical **system** with fast acquisition and high precision by combining the classical optical detection with fast Fourier transform **(**FFT)-based digital **signal processing**
- **Automated the workflow** of cleaning and transferring data and programs between different platforms
- **Developed efficient programs** to extract nanostructure key features with various **statistical tools** (autocorrelation analysis, non-linear regression, hypothesis testing)
- Improved the speed of scattering **Monte Carlo simulation** by 8 times via a **parallel computing** solution on clusters
- **Collaborated** with different teams in several departments

Toxic Comment Classification and Analysis — *Jan 2018 - Jun 2019*

- Identified and classified **half million** toxic Wikipedia comments by Bidirectional **LSTM neural network**
- Developed a **text cleaning** program through Python NLTK, and implemented **word embedding** through GloVe to learn the features of text corpus
- Achieved **top 5% on Kaggle** through a 2-stage **stacking** strategy on out-of-fold predictions
- Built up a toxic word **testing website** backed by **SQL** database with pretrained model through Python Flask
- Optimized toxicity detection across **two million** conversations by using the state-of-art **BERT** algorithm

Data Mining and visualization on Virginia Car Accidents — *Jun 2018 - Dec 2018*

- Parallelized the preprocessing of nearly **1 million** Virginia car accidents through PySpark on AWS EMR
- **Performed visualized analysis** on residents' activities through K-means clustering of crashes
- Applied a **LightGBM classification** to predict crash severity based on road, light and weather conditions
- Identified factors for fatal accidents via **feature importance analysis**