

Yue Zhao

CONTACT INFORMATION Hamburg Hall 2005 zhaoy@cmu.edu
4800 Forbes Ave <https://github.com/yzhao062>
Pittsburgh, Pennsylvania <https://www.linkedin.com/in/yzhao062>
United States, 15213 <https://www.andrew.cmu.edu/user/yuezhao2>

RESEARCH INTERESTS Data mining and knowledge discovery *algorithms, systems, applications*, and *real-world implications*. After I joined the Heinz College, **Information systems**, especially how machine learning may impact on our daily life, arises as my new interest.

Specifically, I focus on proposing ensemble learning, outlier detection, and clustering **algorithms**. Additionally, I design and implement scalable machine learning **systems and libraries** for accessibility and efficiency. Last but not least, I am excited about apply learning algorithms to real-world problems (such as healthcare and finance), i.e., building **applications** and understanding their **implications**.

- Ensemble Learning
- Outlier & Anomaly Detection
- Outlier Ensembles
- Scalable Machine Learning Systems
- Clustering
- Active Learning
- Semi-supervised Learning
- Machine Learning Applications

EDUCATION **Carnegie Mellon University** Aug. 2019 - May. 2024 (Expected)
H. John Heinz III College
Ph.D. in Information Systems & Management (Primary)
Joint Ph.D. in Machine Learning & Public Policy (Expected)
• **Research Advisors:** Prof. Leman Akoglu and Prof. Amelia Haviland

University of Toronto Sep. 2015 - Dec. 2016
Department of Computer Science GPA: 3.83/4.00
Master of Science in Computer Science (Applied Computing)
• **Applied Research:** *Human Resource Management Analytics with Machine Learning*¹
• **Research Advisor:** Prof. Anthony Bonner

University of Cincinnati Sep. 2010 - May. 2015
College of Engineering and Applied Science GPA: *Magna Cum Laude* (3.85/4.00)
Bachelor of Science in Computer Engineering Student Marshall
• **Minor:** *Computer Science and Mathematics*
• **Senior Thesis Advisor:** Prof. Paul Talaga

Shanxi Experimental Secondary School Sep. 2007 - Jul. 2010
Experimental class (Honor Class) GPA: 3.84/4.00
• **Concentration:** *Science*

PUBLICATIONS See my [Google Scholar](#), [ORCID](#), and [ResearchGate](#).

Peer-reviewed Journal Papers

1. [Yue Zhao](#), Zain Nasrullah, Zheng Li
PyOD: A Python Toolbox for Scalable Outlier Detection
Journal of Machine Learning Research (JMLR), 2019.

¹This work is partly supported by Mitacs-Accelerate Research and Development Funding (IT07884).

Peer-reviewed Conference & Workshop Papers

1. Zain Nasrullah, Yue Zhao
Music Artist Classification with Convolutional Recurrent Neural Networks
IEEE International Joint Conference on Neural Networks (IJCNN), 2019, Hungary.
Accepted, to appear.
2. Yue Zhao, Zain Nasrullah, Maciej K. Hryniewicki, Zheng Li
LSCP: Locally Selective Combination in Parallel Outlier Ensembles
SIAM International Conference on Data Mining (SDM), 2019, Calgary, Canada.
Acceptance rate 22.7% (90/397).
3. Yue Zhao, Maciej K. Hryniewicki
DCSO: Dynamic Combination of Detector Scores for Outlier Ensembles
ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD Workshop on Outlier Detection De-constructed), 2018, London, UK.
4. Yue Zhao, Maciej K. Hryniewicki
XGBOD: Improving Supervised Outlier Detection with Unsupervised Representation Learning
IEEE International Joint Conference on Neural Networks (IJCNN), 2018, Rio, Brazil.
5. Yue Zhao, Maciej K. Hryniewicki, Francesca Cheng, Boyang Fu, Xiaoyu Zhu
Employee Turnover Prediction with Machine Learning: A Reliable Approach
Intelligent System Conference (Intellisys), 2018, London, UK.
Acceptance rate 34% (194/568).
6. Yue Zhao*, Zhongtian Qiu*, Yiqing Yang*, Weiwei Li*, Mingming Fan
An Empirical Study of Touch-based Authentication Methods on Smartwatches
ACM International Symposium on Wearable Computers (ISWC), 2017, Maui, USA.
Acceptance rate 25.6% (23/90). (*equal contribution)

Preprints & Working Papers

1. *Combining Machine Learning Models and Scores using combo library*
AAAI Conference on Artificial Intelligence (AAAI), Demo track
Submitted, under review.
2. Colin Wan, Zheng Li, Alicia Guo, Yue Zhao
SynC: A Unified Framework for Generating Synthetic Population with Gaussian Copula
Submitted, under review.
3. *HD-Cluster: Synthesized Clustering and Outlier Detection on High-dimensional Data*
In preparation, working paper.

COMMUNITY ACTIVITIES & MEMBERSHIP

Reviewer

- Knowledge and Information Systems (KAIS)
- IEEE Computational Intelligence Magazine (CIM)
- The Journal of Open Source Software (JOSS)

Membership: *ACM, IEEE, SIAM, SIAG/SDM*

PROFESSIONAL EXPERIENCE

Consulting & Deals, PwC Canada

- | | |
|------------------------------------|---------------------|
| Senior Consultant (Data Scientist) | Aug. 2017–Jun. 2019 |
| Consultant (Data Scientist) | Feb. 2017–Jul. 2017 |
| Research Associate (Intern) | May. 2016–Jan. 2017 |
- Designed fraud analytic solutions for major Canadian banks and insurance firms.
 - Led various applied data mining projects, e.g., client segmentation and churn analysis.
 - Developed multiple pricing optimization models with statistical methods.

Siemens PLM Software USA

Software Engineer (Intern & Contract) Mar. 2012–Dec. 2014

- Managed a Java project to transition the LabManager system to vCloud Director.
- Refactored outdated automation code and added new modules and JUnit test cases.
- Led a C++ Code Coverage project on Teamcenter platform to strengthen its stability.

TEACHING EXPERIENCE **Teaching Assistant**, Embedded Systems (Prof. Philip Anderson, Toronto) F 2015
Teaching Assistant, Intro to Programming (Prof. George Purdy, Cincinnati) F 2014

FUNDS AND AWARDS Mitacs-Accelerate Research and Development Funding \$30,000 2016–2017
University Global Award and Scholarship \$32,000 2010–2015
Mantei/Mae Award & Scholar (<https://manteimaeawards.com/>) \$40,000 2012–2015
Engineer of the Month (University of Cincinnati) Jun. 2014

OPEN-SOURCE WORKS **A Python Toolbox for ML Model Combination** (combo) Jul. 2019–Present

- Implemented various key combination algorithms, averaging, majority vote, and AOM.
- Developed advanced combination frameworks, e.g., Dynamic Classifier Selection.
- Covered a wide range of applications (classification, clustering, and outlier detection.)
- Provided Unified APIs, detailed documentation, and interactive examples for all implemented algorithms.
- Achievements & Highlights:

1. **3,000 PyPI downloads** and **150 GitHub stars** and **20 forks** since July 2019.
2. A corresponding demo paper is under submission at AAAI 2020.

Python Outlier Detection Toolbox (PyOD) Oct. 2017–Present

- Implemented various key outlier detection algorithms, such as ABOD and AutoEncoder.
- Developed advanced outlier ensembling frameworks, such as Feature Bagging and LSCP.
- Provided Unified APIs, detailed documentation, and interactive examples for all implemented detection algorithms.
- Achievements & Highlights:

1. **100,000 PyPI downloads** and **2,300 GitHub stars** and **400 forks** since 2018.
2. Emerged as **the most popular anomaly detection toolbox** and ranked **top 10 data mining toolkit** on GitHub.
3. Featured by **KDnuggets**, **Analytics Vidhya**, and **Computer Vision News**.

anomaly-detection-resources (GitHub Repository) Jun. 2018–Present

- Compiled a repository with outlier related books, papers, courses, datasets, and tools.
- Summarized and shared the latest outlier detection advancement for practitioners.
- Achievements & Highlights: **1,400 GitHub Stars** and **400 forks** since June 2018.

RELEVANT SKILLS **Technical:** Python (expert), C++ (advanced), Database (proficient)
Languages: English (fluent), Mandarin (native)

GRADUATE
COURSEWORK

Courses at Carnegie Mellon University

- Advanced Intro to Machine Learning
- Probability and Statistics

- Ph.D. Seminar I
- Ph.D. Microeconomics

Courses at University of Toronto

- Machine Learning and Data Mining
- Big Data Analytics in Healthcare
- Communications for Computer Scientists

- Human Computer Interaction
- Systems Thinking for Global Problems
- Technical Entrepreneurship