BIAO **YIN**

Worcester, MA 01604 | 774-502-7496 | byin@wpi.edu

Summary

- Led: 4 US Army Research Lab AI Material Science projects with 1 iOS app and 1 web platform (deployed), 1 AI software with its patent in Mental Health (issued), 1 statistical report in Education, 2 big data reports in Finance
- Contributed: Over 10 NSF projects in human behavior mining
- Studying: Deep learning in self-supervision, transfer learning, and GAN
- Published: Top Papers in NLP, CV, Big Data, and Educational Data Mining

Education and Training

Ph.D.: **Data Science** **Expected in 05/2023**

Worcester Polytechnic Institute Worcester, MA, U.S.A

- GPA: 3.9/4.0, Research Mentor and Graduate Qualifier Project Mentor
- Advisor: Professor Elke A. Rundensteiner *https://daisy.wpi.edu*
- US Army Research Lab Grant Recipent ***https://arl.wpi.edu/#members***
- NSF CEDAR Trainee *https://wp.wpi.edu/cedar/trainees-2/*

Master of Science: **Statistics** **05/2016**

George Washington University Washington D.C., U.S.A

- GPA: 3.7/4.0, Teaching Assistant: Business & Economics Statistics I, II

Bachelor of Science: **Statistics** **06/2014**

Anhui University of Finance And Economics Bengbu City, China

- GPA: 3.6/4.0, Major GPA: 3.9/4.0, Ranked in Top 1 % of Major
- Graduation with Honor Award, Best Bachelor Thesis, Excellent Student Cadre, Champion of College Debate

Skills

- **Languages:** Python, R, Javascript, HTML, Latex
- **Frameworks:** PyTorch, TensorFlow, Keras, D3, OpenCV, Pandas, Numpy, Sklearn, ReactJS
- **ML Ops:** Docker, Turing, Weights and Biases, Tableau, SAS, Hyper-Parameter Tuning
- **Databases:** Mysql, PgAdmin, GrantaDB
- **Big Data:** Hive, Xshell, Spark, CUDA
- **Project Management:** GitLab, Jira, Confluence
- **Machine Learning:** Causal Forests, t-SNE, Self Supervision, Domain Adaptation, Style Transfer

Experience

Research Mentor (Founding Research Assistant) **01/2019 to Current**

Worcester Polytechnic Institute **Worcester, MA, USA**

- Led and mentored over 50 students and researchers, majoring in data science, computer science, material science, statistics etc., handling 4 applied machine learning research projects on Adhesives, Corrosion, Aviation & Missile Technology, and Cold Spray sponsored by ***US Army Research Lab (ARL) and PPG*** over 3 years
- Designed and developed a ***production-ready iOS APP*** and a Context-Aware Shared Agile Platform (ARL-CAAP) aimed at smart material discovery with Human-in-the-loop based ***active learning*** from data collection, management, advanced visual analytics, machine learning to design of experiments
- Focused on deep learning research in material assessment and generation across indoor and outdoor experiments on small data; published top papers and mentored graduate students sponsored by ARL and NSF CEDAR program on more fellowship fundings

Core Technology Researcher (Intern) **03/2018 to 08/2018**
AI Research, Core R&D, USTC-iFLYTEK CO. **Hefei, China**

- Developed 100+ high-efficient features to detect Alzheimer's Disease via patients' Handwriting, Speech, NLP, and Facial Expression with medical doctors
- Engined the clinical diagnoses via the subtracted features in Machine Learning models such as CART, RF, SVM, CNN with ***90%+*** accuracy; ***built a software embedding the models and issued patents with developers***
- Collaborated with research teams from MIT CISCIL and USTC on multi-modal human cognition projects

Big Data Engineer (Intern) **02/2018 to 03/2018**
Consumer Business Group, USTC-iFLYTEK CO **Hefei, China**

- Created 10 effective features to assess consumers' credits from ***Mobile App behaviors in 200 million daily users' devices*** using Hive in Xshell on a distributed database
- Completed a feasibility study in consumers' Online Financial Risk Control after cleaning DMP data from Consumers' Contexts, App behaviors, and Typed Texts using Genism Word2Vec, LDA, Scikit-learn and Seaborn

Research Assistant **06/2016 to 12/2017**
Worcester Polytechnic Institute **Worcester, MA, USA**

- Proved significant business impact of a web platform to tutor K-12 students via statistical learning on ***200+ Randomized Control Trials*** such as: human involvement, video or text hints, handwriting or typing, etc.
- Published Heterogeneous Treatment Effects for ***personalized learning*** on the platform using causal trees and causal forest; successfully narrowed down confidence interval with massive human-related features using Hierarchical Linear Modeling, Bayesian Hierarchical Regression (Stan), and Logistic Regressions
- Built 24 essential features based on 3 kinds of student performances on 8 types of assignments from 200+ SQL tables; well solved data cleaning issues and improved the usability of the lab database

Publications and Patents

- [Issued patent 2022: CN109409329A] Feiyang X., **Yin, B.**, et al. Automatic Discriminant Analysis of Graphic Problems Based on Tracked Points, USTC-IFLYTEK CO., China
- [BMVC 2021] **Yin, B.**, et al. Corrosion image data set for automating scientific assessment of materials. In British Machine Vision Conference 2021.
- [IEEE BigData 2020] **Yin B.**, et al. Corrosion Assessment: Data Mining for Quantifying Associations between Indoor Accelerated and Outdoor Natural Tests, 2nd International Workshop on Big Data Tools, Methods, and Use Cases for Innovative Scientific Discovery in 2020 IEEE International Conference on Big Data
- [ACL 2020] Sen, C., Hartvigsen, T., **Yin, B.**, et al. Human Attention Maps for Text Classification: Do Humans and Neural Networks Focus on the Same Words?. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics
- [ICEDM 2017] **Yin, B.**, et al. Causal Forest vs. Naive Causal Forest in Detecting Personalization: An Empirical Study in ASSISTments. In the 10th International Conference on Educational Data Mining, Wuhan, China
- [ACM L@S 2017] **Yin, B.**, et al. Observing Personalizations in Learning: Identifying Heterogeneous Treatment Effects Using Causal Trees. In the Fourth, ACM Conference on Learning @ Scale, MIT, Cambridge, MA. ACM
- [ICEDM 2017] Adjei, S. A., Botelho, A. F., Patikorn, T., **Yin, B.**, & Beck, J. E. Using "Snapshots" of Student Performance to Model Wheel Spinning. In the Tenth International Conference on Educational Data Mining, Wuhan, Hubei, China.
- [Chinese Journal 2017] Xin S., **Yin, B.** Statistics Major in Big Data Era, Statistics and Strategy [J]
- [MIT CodeCon 2016] Patikorn, T., Selent, D., Heffernan, N. T., **Yin, B.**, Botelho, A. ASSISTments Dataset for a Data Mining Competition to Improve Personalized Learning. In CodeCon 2016, MIT, Cambridge, MA