

Summary

A passionate result driven practitioner with hands-on skills on data mining/machine learning algorithm development and research. Strong eagerness to work with talented and energetic people on promising projects.

Experience

Industrial Experience

- 2020-Now **Tech Lead Manager, Roblox**, San Mateo, CA, USA.
 - Leading the engineering of Roblox homepage, including frontend, backend, and machine learning.
- 2019-2020 **Tech Lead Manager, Machine Learning Platform, Uber Technologies Inc.**, San Francisco, CA, USA.
 - Support the remote GPU inference service used in model evaluation for the use cases of perception, prediction, motion planning, labeling, and more for the autonomous vehicles.
 - Support the ML CI/CD, automating the verification of the ML logic changes regarding dataset generation, model training, model serving, and various model evaluation stages.
- 2014-2019 **Research Scientist, Facebook Inc.**, Menlo Park, CA, USA.
 - **Feed and Story Integrity**
 - Tead lead of News Feed and Story from the ranking anti-manipulation and integrity perspective.
 - Led the design and implementation of the ML-based fake engagement detection system from scratch, reducing hundreds of millions of in-feed impression for contents benefit from fake engagement.
 - Led the design and implementation of ML-based content moderation system for Facebook Story, including ML-based labeling, measurement, detection, and enforcement on both violating contents (bully, drug and firearm selling, etc.) and gray area contents (clickbait, engagement-bait, etc.).
 - **Core Growth**
 - ML tech lead of the core growth org, worked on ML problems across multiple product lines, including Friend recommendation (PYMK), Identity, Acquisition, Login, and Account recovery.
 - Owner of the holistic auction system that is responsible for inserting promotion contents to the News Feed, along with the organic contents and ads. The content types include but not limited to: recommended friends, recommended groups, recommended events, recommended jobs, recommended places, etc.
 - Intern of the News Feed Ranking team
 - Designed and implemented the pipeline as well as the related analytic tools for personalized News Feed ranking at Internet scale, increased the AUC of the prediction by 0.x%.
- 2011-2013 **Intern, Service Management team, IBM T.J. Watson Research Center**, New York, NY, USA.
 - Summer internships in 2011, 2012, and 2013, focusing on ML R&D on the IaaS IBM Cloud. Also closely collaborated with IBM research while at school. Filed 3 patents during these years.

2009 **Intern, *Data Intelligence and Tools team, Microsoft Research***, Beijing, China.

- Worked on distributed data mining and analysis algorithm R&D. Worked closely with MS Office teams to help them better understand the user behaviors on the products.

Open Source Projects

Open Source **Apache Software Foundation.**

- Community ◦ **PMC Member** of Apache Hama: A BSP framework for distributed computing.
 - Working on the distributed machine learning package, including the design and implementation of distributed multilayer perceptron, linear regression, logistic regression, and auto-encoder. [\[GitHub Link\]](#)
- **Committer** of Apache Mahout: A machine learning library for Hadoop.
 - Implemented the multilayer neural network model trained with stochastic gradient descent.

Publication

- Program committee member and reviewer of the top academic conferences/journals in the area of data science, such as: *ACM Conference on Knowledge Discovery and Data Mining (SIGKDD)*, *International Conference on Data Mining (ICDM)*, *ACM Conference on Recommender Systems (RecSys)*, *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, *ACM Transactions on Knowledge Discovery from Data (TKDD)*, *Knowledge and Information System (KAIS)*, *Information Sciences*, *IEEE Transactions on Cybernetics*, *Data and Knowledge Engineering (DKE)*, etc.
- Published 20+ papers at the top journals/conferences in the area of data science with 700+ citations, such as: *ACM Conference on Knowledge Discovery and Data Mining (SIGKDD)*, *International Conference on Data Mining (ICDM)*, *ACM Conference on Recommender Systems (RecSys)*, *ACM Conference on Information Retrieval (SIGIR)*, *ACM Conference on Information and Knowledge Management (CIKM)*, *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, *ACM Transactions on Intelligent Systems and Technology (TIST)*, etc. (Details at: <https://goo.gl/4dRHrb>)
- Published 3 patents during the internships at IBM,

Education

2015 **Ph.D. in Computer Science**, *Florida International University*, Miami, FL, USA.

Research Area: Temporal Data Mining, Distributed Data Mining

2010 **B.S and M.S. in Computer Science**, *Sichuan University*, Chengdu, Sichuan, China.

Research Area: Data Mining