

# Zhihui Shao

702 Verano Rd, Irvine, CA, 92617

E-mail: [zshao006@ucr.edu](mailto:zshao006@ucr.edu) Phone No. (951)231-8045

## EDUCATION

---

- Ph.D. in Computer Science, University of California, Riverside, California 09/2017-06/2021(*expected*)
- M.Sc. in Computer Science, University of California, Riverside, California 09/2017-03/2019

## TECHNICAL SKILLS

---

- **Programming Languages:** Python, C/C++, JAVA, SQL;
- **Machine Learning Framework:** Scikit-learn, TensorFlow, Keras, Pytorch;
- **Data Analysis Tools:** Pandas, Seaborn, Spark, NumPy, Matplotlib;
- **Database Management:** MySQL, SQLite, MongoDB;
- **IDE Tools:** VS code (C/C++), IntelliJ (java), Eclipse (java), Jupyter Notebook(Python), PyCharm (Python);

## SELECTED PROJECT

---

**Design of search engine for Wikipedia (*demo web:* <http://3.87.174.210/>)** January - April 2018

- **Craw web pages:** multiple threads to increase speed; noise removal to optimize the database storage;
- **Index web pages:** Hadoop MapReduce for higher throughput of index process;
- **Score algorithm:** BM25, Proximity and Page-Rank; **User Interface design:** MVC framework on Node.js;
- **Database:** SQLITE to store web pages, MongoDB to store index;

## EXPERIENCE

---

### Amazon Applied Scientist Internship (DEX-ML)

**Project 1: Customer embedding on delivery preference** June - September 2020

- Customer behavior analysis with SQL, Spark, Pandas and Seaborn;
- Embedding architectures include XGBoost, CNN, RNN (LSTM), and Auto-encoder;
- Evaluation metrics include accuracy, F1 score, ROC, and AUPR;

**Project 2: Amazon Day customer behavior and explore machine learning opportunity** June - September 2019

- Data collection and data analysis with SQL and PySpark;
- Customer acquisition model is implemented on Random Forest and XGBoost;

### Research Assistant, Non-linear Computing Lab, UC Riverside

*Sep 2017 - present*

**Project 1: Efficient server sprinting using deep reinforcement learning (DRL)**

- Target: burst cloud servers dynamically with deep reinforcement learning;
- Accepted in IEEE Cloud's20 (<https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9284233>);

**Project 2: Increasing trustworthiness of deep neural networks via accuracy monitoring**

- Target: monitor DNN performance on operational domain;
- Accepted in AI-Safety (IJCAI), ([http://ceur-ws.org/Vol-2640/paper\\_3.pdf](http://ceur-ws.org/Vol-2640/paper_3.pdf));

**Project 3: Calibrating deep neural network classifiers on out-of-distribution datasets**

- Target: calibrate DNN models without re-training and improve the prediction trustiness of deployed model;
- Report: (<https://arxiv.org/abs/2007.01472>);

**Project 4: Your noise, my signal: exploiting switching noise for stealthy data exfiltration from desktop computers**

- Accepted in SIGMETRICS'20 (<https://dl.acm.org/doi/10.1145/3379473>);

### Teaching Assistant, *Algorithm and Data Structure*, UC Riverside

*Fall 2018 - present*

- Teaching CS010, CS012, CS014 (e.g. C++ basic, Stack, Trees, Heaps, Graphs, and binary search, DFS, BFS);

## SELECTED GRADUATE COURSES

---

- Design and Analysis of Algorithms
- Data Mining Techniques
- Machine Learning and Data Mining
- Probabilistic Models for Artificial Intelligence
- Information Retrieval and Web Search
- Databases and SQL for Data Science
- Neural Networks and Deep Learning
- Improving Deep Neural Networks: Hyperparameter tuning, Regularization and Optimization
- Structuring Machine Learning Projects
- **Convolutional Neural Networks**
- **Sequence Models**

**Online course:**