Kewen Peng

Cell: 336-251-9877 |**Email:** kpeng@ncsu.edu | **GitHub:** kpeng2019 | **ResearchGate**

Education

North Carolina State University — Raleigh, NC
Ph.D. in Computer Science | Advisor: Dr. Tim Menzies | Lab: RAISE Lab — Aug. 2019 – May 2024

Wake Forest University — Winston-Salem, NC
B.S. (Honor thesis) in Computer Science | B.A. (Honor thesis) in Mathematics — Aug. 2015 – May 2019

Skills

Language: Python, Java, C/C++, SQL, JavaScript
Data analysis tools: Pandas, SciPy, Scikit-learn
DevOps tools: Docker, Jenkins, Ansible, Travis-CI
Research Interests: Data analytics, ML Optimization, Supervised/Unsupervised Modeling
Coursework: DevOps | Data Structure | Algorithm | Data Mining | Advanced AI | Automated SE

Work Experience

Data Scientist Intern — May 2022 – August 2022
Indeed Inc. — Remote

- Participated in multiple analysis-driven experimental studies to diagnose latent features contributing to positive feedback from job seekers.
- Designed, configured, and deployed machine learning pipelines to predict positive outcomes for job seekers in 11 new international markets; Obtained significant improvement (over **12%**) in predicting positive outcomes.
- Maintained the team-owned data index in Java; Proposed a new weighting strategy for hierarchical classification.

Research Assistant — Jan 2020 – Present
North Carolina State University, RAISE Lab — Raleigh, NC

- Designed a code refactoring recommendation system to reduce the probability of defect-prone code patterns. The new model achieved up to 100% improvement compared to SOTA methods (applied on Random Forest, MLP, and SVM learners via Scikit-learn).
- Experimented with active learning and curriculum learning tactics in NER and emotion identification tasks. Proposed a router-based pipeline that reduced training and inference costs for deep learning models (e.g., BERT, LSTM) to 10%.

Teaching Assistant — Aug 2019 – Dec 2019
North Carolina State University, Computer Science Department — Raleigh, NC

- Coordinated with the professor & other TAs as a team to structure the (SE, Programming Language) courses, design tests, and facilitate labs.

Selected Research Experience

Taming Deep Learning — Jan. 2021 – Present
RAISE Lab, research project, funded by LAS — Raleigh, NC

- Explored the viability of using model-agnostic methods to reduce the training time of deep learning models (via Keras) on NetFlow attack detection data.
- Achieved significant performance improvement and reduction in training time by 50% compared to benchmarks.

Machine Learning Fairness — Oct. 2019 – Present
RAISE Lab, research project, funded by LAS and NSF — Raleigh, NC

- Explored fairer results in machine learning software via different approaches of bias mitigation. Achieved significant reduction in bias by 67% without compromising model performance.
- Explored reliable and robust explanation generation tools for fairer SE. Designed the recommendation system that enhances explanation generation with actionable analysis.

Lung Cancer Survival Prediction using TCGA Clinical Data — Aug. 2018 – May. 2019
Undergraduate honor thesis — Winston-Salem, NC

- Designed a multi-learner pre-processor for missing data imputation when using machine learning models to predict lung cancer survival.
- The hybrid framework combining imputed data and the encoded imputation information achieved significantly better performance by 20% on average in all selected machine learning models.

Publications

- Luigi Ferraro, Ellen Kirkman, W Frank Moore, **Kewen Peng**, *On the Noether Bound for Noncommutative Rings*, **PAMS journal** (Accepted).
- Joymallya Chakraborty, **Kewen Peng**, Tim Menzies, *Making fair ML software using trustworthy explanation*, **ESEC/FSE 2020** (Accepted).
- **Kewen Peng**, Tim Menzies, *Defect Reduction Planning (using TimeLIME)*, **TSE journal** (Accepted).
- **Kewen Peng**, Christian Kaltenecker, Norbert Siegmund, Sven Apel, Tim Menzies, *VEER: Disagreement-Free Multi-objective Configuration*, **EMSE journal** (Submitted).
- **Kewen Peng**, Joymallya Chakraborty, Tim Menzies, *FairMask: Better Fairness via Model-based Rebalancing of Protected Attributes*, **TSE journal** (Accepted).
- Time Menzies, **Kewen Peng**, Andre Lustosa, *Fairer Software Made Easier (using "Keys")*, **ASE 2021 RAISE Workshop** (Accepted).

Honors and Service

- Member in Upsilon Pi Epsilon Honor Society, 2017
- Honorable Mention in ICPC Mid-Atlantic Regional, 2018
- Honorable Mention in COMAP MCM Contest, 2018
- Undergraduate Summer Research Fellowship, Wake Forest University, 2018
- Teaching Assistant, North Carolina State University, 2019
- Research Assistant, North Carolina State University, 2020-Present
- Journal Reviewer, Empirical Software Engineering (EMSE), 2021
- Keynote Speaker, International Conference on Software Engineering (ICSE), 2022
- Web Chair, International Conference on Software Engineering (ICSE) Fairware Workshop, 2023