Caitlin Kuhlman

cakuhlman@wpi.edu

PhD candidate in Computer Science studying Data Mining and Machine Learning,
with a focus on using open data to solve social and economic problems.

Education

PhD Computer Science, expected 2019
Worcester Polytechnic Institute *(GPA 3.9)*

MS Computer Science, 2017
Worcester Polytechnic Institute (*GPA 3.9*)
Thesis: Pivot-based Data Partitioning for Distributed k Nearest Neighbor Mining

Major Certificate in Computer Science, 2013
University of Massachusetts Boston (*Cumulative GPA 3.6, Major GPA 3.8, Dean's List 2011, 2012*)

BFA Fine Arts 3D, 2007
Massachusetts College of Art and Design

Research Experience

Research Assistant in the Computer Science Department, *Worcester Polytechnic Institute August 2014 – present*
Advisor: Elke Rundensteiner

- Member of the Data Science Research Group working to develop scalable data mining and machine learning techniques using distributed computing technologies including **Hadoop and Spark**.
- **Lead researcher and developer** on the Massachusetts Technology, Talent, and Economic Reporting System (MATTERS), an online analytics dashboard measuring the economic competitiveness of U.S. states. Work closely with domain experts from the Massachusetts High Tech Council on the identification and integration of high fidelity datasets, the design of intuitive visual analytics and user interfaces, and development of novel data analysis tools.
- **Supervise teams** of undergraduate students for their Major Qualifying Projects. Past Projects include:
  o System to automatically extract and integrate heterogeneous data from different online sources.
  o Suite of tools for data management and curation, currently in use by a student team at Brandeis University.
  o Public-facing API, providing access to a collection of over 30 datasets.

Data Science for Social Good Fellow *IBM Research Yorktown Heights, NY May – August 2016*
- One of six researchers selected to work with IBM and partner organizations on data analytics solutions for social good. Focus of project: to better quantify innovation in countries around the world.
- Worked closely with a small team of data scientists to analyze over 1400 publicly available economic, demographic, and environmental datasets.
- Developed models to predict and measure the level of innovation in a country, and identify contributing factors that facilitate innovation, resulting in a publication in the IBM Journal of Research and Development.

Technical Intern *MITRE Corporation Bedford, MA June – August 2014 and June – August 2015*
- Member of a small team of researchers investigating new techniques for automatic feature learning leveraging the Hadoop distributed computing framework.
- Worked with a lead data scientist to perform analytics on TB scale data for cybersecurity applications.
- Developed and implemented a MapReduce algorithm in Java for feature extraction using Hadoop, and provided proof-of-concept for a supervised learning-based system for intrusion detection.

Research Assistant *Knowledge Discovery Lab UMass Boston May – August 2013 and January – May 2014*
- Contributed to a collaborative project with George Mason University and the US Department of Energy for the automatic analysis of geospatial imagery, leading to a journal publication.
- Implemented a system for detecting buildings in satellite images, developed a web client for image analysis, and released an open source command line tool for image preprocessing.

Caitlin Kuhlman

cakuhlman@wpi.edu

Skills

Programming Languages: Java, Python, C, R, Bash, SQL, Javascript
Technologies & Platforms: Hadoop, Spark, Jupyter, Git
Operating Systems: Linux/Unix, Windows
Software & Libraries: Weka, Scikit-learn, Pandas, Eclipse IDE, Adobe Photoshop, Illlustrator
Design: Fine arts training and working knowledge of visual design principles and best practices.

Honors and Awards

ORISE Fellowship *2014 – 2016*
Appointment to the Student Research Participation Program at the U.S. Army Natick Soldier Research, Development and Engineering Center, administered by Oak Ridge Institute for Science and Education and DOE.

Graduate Research Innovation Exchange *Worcester Polytechnic Institute 2017*
2nd Place PhD level Data Science and Cybersecurity Poster Competition

Invited Student Speaker *Boston University 2014*
New England Undergraduate Computing Symposium

Travel Awards

Amazon Graduate Research Symposium *2017*	Women in Machine Learning *2015*
WPI Graduate Student Travel Award *2017*	Broadening Participation at UbiComp *2015*
CRA-W Grad Cohort Workshop *2017*	CRA-W Grad Cohort Workshop *2015*
Broadening Participation in Data Mining *2016*	UMass Boston Undergraduate Research *2014*

Leadership

Tutorials Chair Broadening Participation in Data Mining Workshop (BPDM) *August 2017*
- Organized two tutorial sessions at the BPDM Workshop co-located with the ACM SIGKDD Conference.
- Collaborated on content for tutorial on Algorithmic Fairness, provided as an interactive Jupyter notebook on Github.

Publications

Caitlin Kuhlman, Yizhou Yan, Lei Cao, and Elke Rundensteiner. Pivot-based Distributed K-Nearest Neighbor Mining. *European Conference on Machine Learning, Principles and Practice of Knowledge Discovery (ECML-PKDD) Research Track, Springer LNCS, 2017* (pdf)

Caitlin Kuhlman and Elke Rundensteiner. Towards an Interactive Learning-to-Rank System for Economic Competitiveness Understanding. *KDD Workshop on Interactive Data Exploration and Analytics (IDEA) 2017* (pdf)

Caitlin Kuhlman, Karthikenyan Natesan Ramamurthy, Prassana Sattigeri, Aurelie C. Lozano, Lei Cao, Chandra Reddy, Aleksandra Mojsilovic, Kush R. Varshney. How to Foster Innovation: a Data-Driven Approach to Measuring Economic Competitiveness. *IBM Journal of Research and Development. In press*

Yizhou Yan, Lei Cao, **Caitlin Kuhlman**, and Elke Rundensteiner. Distributed Local Outlier Detection in Big Data. *SIGKDD Conference on Knowledge Discovery and Data Mining (KDD) 2017* (pdf)

Lei Cao, Yizhou Yan, **Caitlin Kuhlman**, Qingyang Wang, Elke Rundensteiner and Mohamed Eltabakh. Multi-tactic Distance-based Outlier Detection. *IEEE International Conference on Data Engineering (ICDE) 2017* (pdf)

Rodica Neamtu, **Caitlin Kuhlman**, Ramoza Ahsan, and Elke Rundensteiner. The impact of Big Data on making evidence-based decisions. *Book chapter in Frontiers in Data Science. CRC Press Reference 2017*

Joseph Paul Cohen, Wei Ding, **Caitlin Kuhlman**, Aijun Chen, and Liping Di. Rapid Building Detection using Machine Learning. *Applied Intelligence 45, no. 2: 443-457 2016* (pdf)