

# Thomas Hartvigsen

<https://wpi.edu/people/doctoral-student/twhartvigsen>  
twhartvigsen@wpi.edu | 585.245.2767

## EDUCATION

### WORCESTER POLYTECHNIC INSTITUTE

#### PH.D. IN DATA SCIENCE

Expected May 2021 | Worcester, MA

Cum. GPA: 3.83 / 4.0

Advisor: Dr. Elke Rundensteiner

### SUNY GENESEO

#### BA IN APPLIED MATHEMATICS

Minor in BioMathematics

May 2016 | Geneseo, NY

Cum. GPA: 3.07 / 4.0

## LINKS

LinkedIn:// [in/thartvigsen](#)

Github:// [thartvigsen](#)

## COURSEWORK

### GRADUATE

Knowledge Discovery and Data mining

Introduction to Data Science

Statistical Learning

Database Management Systems

Big Data Management\*

Business Intelligence\*

Machine Learning\*\*

\*to be completed Dec 2017

\*\*to be completed May 2018

### UNDERGRADUATE

Modeling Biological Systems

Combinatorics

Differential Equations

Probability and Statistics

Linear Algebra

Calculus I-III

## SKILLS

### PROGRAMMING

Python - TensorFlow for Deep Learning,

Scikit-Learn for Machine Learning • R -

Statistical Learning, Graph mining, ggplot2

• Shell •  $\text{\LaTeX}$  • SQL - PostgreSQL for research, SQLPLUS for DBMS Course

## AWARDS

2016 GAANN Research Fellowship  
U.S. Department of Education

## EXPERIENCE

### WPI | GAANN RESEARCH FELLOW

August 2016 – present | Worcester, MA

- Working with Elke Rundensteiner on Machine Learning/Deep Learning with Sequential Data.

### UNIVERSITY OF ARIZONA | RESEARCH EXPERIENCE FOR UNDERGRADUATES INTERN

June 2015 – Aug 2015 | Tucson, AZ

- Worked in the School of Natural Sciences and the Environment with Dr. Shirley Papuga.
- Designed Machine Learning algorithms to segment sequences of images of Creosote Bushes to record changes through time in regards to drought.
- Presented findings at the Undergraduate Research Opportunities Consortium (UROC) in August 2015.

## RESEARCH

### DATA SCIENCE RESEARCH GROUP Aug 2016 – Pres. | Worcester, MA

Currently working with mainly RNNs/LSTMs in TensorFlow on time series classification, attempting to classify data well in advance of index events. One application of this work will be predicting infections far in advance of actual diagnoses. 2016-2017, worked with Dr. Elke Rundensteiner and Cansu Sen to create CREST, a machine learning tool for the early detection of infections in hospitals. Early warnings of infection were generated using variants of SVMs and Random Forests accounting for imbalanced datasets. Used the MIMIC-III publicly-available EHR database to validate the efficacy of the system. PostgreSQL was used for data management and extraction and Python was used for all transformations and machine learning implementation. Paper accepted at ECML/PKDD in June, 2017.

### BIOMATHEMATICS INNOVATION GROUP Jan 2013 – May 2016 | Geneseo, NY

Focused on Network Analysis related to the spread of infections using R under the advisement of Dr. Chris Leary. Proposed vaccination strategies on many theoretical graph structures built on top of differential equation models. Mentored younger students and brought a total of 6 projects to regional undergraduate research conferences.

## PUBLICATIONS

- Sen, C., Hartvigsen, T., Claypool, K., Rundensteiner, E. (2017, September). CREST - Risk Prediction for Clostridium Difficile Infection Using Multimodal Data Mining. ECML/PKDD 2017.
- Hartvigsen, T., Sen, C., Brownell, S., Teeple, E., Rundensteiner, E. Early Prediction of MRSA Infections using Electronic Health Records. In submission to HealthInf 2018.
- Hartvigsen, T., Sen, C., Teeple, E., Rundensteiner, E. Risk Stratification and Diagnostic Performance of a Machine Learning Algorithm for Clostridium Difficile Detection Using Electronic Health Records Data. Preparing for submission to the New England Journal of Medicine in October 2017.