



Announces the Ph.D. Dissertation Defense of

## Brian Heredia

for the degree of Doctor of Philosophy (Ph.D.)

# “Machine Learning Algorithms for the Analysis of Social Media and Detection of Malicious User Generated Content”

July 5, 2018, 9 a.m.  
Engineering East (EE), Room 405  
777 Glades Road  
Boca Raton, FL

#### DEPARTMENT

Computer and Electrical Engineering and Computer Science

#### CHAIR OF THE CANDIDATE'S PH.D. COMMITTEE

Taghi M. Khoshgoftaar, Ph.D.

#### PH.D. SUPERVISORY COMMITTEE

Mehrdad Nojournian, Ph.D.

Dingding Wang, Ph.D.

Xingquan Zhu, Ph.D.

#### ABSTRACT OF DISSERTATION

##### **Machine Learning for the Analysis of Social Media and Detection of Malicious User Generated Content**

One of the defining characteristics of the modern Internet is its massive connectedness, with information and human connection simply a few clicks away. Social media and online retailers have revolutionized how we communicate and purchase goods or services. User generated content on the web, through social media, plays a large role in modern society; Twitter has been in the forefront of political discourse, with politicians choosing it as their platform for disseminating information, while websites like Amazon and Yelp allow users to share their opinions on products via online reviews. The information available through these platforms can provide insight into a host of relevant topics through the process of machine learning. Specifically, this process involves text mining for sentiment analysis, which is an application domain of machine learning involving the extraction of emotion from text. Unfortunately, there are still those with malicious intent and with the changes to how we communicate and conduct business, comes changes to their malicious practices. Social bots and fake reviews plague the web, providing incorrect information and swaying the opinion of unaware readers. The detection of these false users or posts from reading the text is difficult, if not impossible, for humans. Fortunately, text mining provides us with methods for the detection of harmful user generated content. This dissertation expands the current research in sentiment analysis, online review detection and election prediction. We determine the ability for cross-domain sentiment analysis using tweets and reviews. Novel techniques combining ensemble and feature selection methods are proposed for the domain of online spam review detection. We investigate the ability for the Twitter

platform to predict the United States 2016 presidential election. In addition, we determine how social bots influence this prediction.

#### BIOGRAPHICAL SKETCH

Born in Newark, New Jersey, U.S.A.

B.S., Florida Atlantic University, Boca Raton, Florida, 2012

M.S., Florida Atlantic University, Boca Raton, Florida, 2014

Ph.D., Florida Atlantic University, Boca Raton, Florida, 2018

#### CONCERNING PERIOD OF PREPARATION & QUALIFYING EXAMINATION

**Time in Preparation:** 2015 - 2018

**Qualifying Examination Passed:** Fall 2015

#### Selected Published Papers:

Heredia, Brian, Joseph Prusa, and Taghi M. Khoshgoftaar. "Location-Based Twitter Sentiment Analysis for Predicting the U.S. 2016 Presidential Election." In the 30th Florida Artificial Intelligence Research Society (FLAIRS), International Conference. pp. 265-270. 2018.

Heredia, Brian, Joseph Prusa, and Taghi M. Khoshgoftaar. "Exploring the Effectiveness of Twitter at Polling the United States 2016 Presidential Election." In Collaboration and Internet Computing (CIC), 2017 IEEE 3rd International Conference on. pp. 283-290. 2017.

Heredia, Brian, Taghi M. Khoshgoftaar, Joseph Prusa, and Michael Crawford. "Improving the detection of untrustworthy online reviews using ensemble learners combined with feature selection." Journal of Social Network Analysis and Mining, vol. 7(1):37, 2017.

Heredia, Brian, Taghi M. Khoshgoftaar, Joseph Prusa, and Michael Crawford. 2017. "Integrating multiple data sources to enhance sentiment prediction." In Collaboration and Internet Computing (CIC), 2016 IEEE 2nd International Conference on. pp. 285-291. 2016.