BIOGRAPHICAL SKETCH

Born in Port-au-Prince, HAITI
B.S. December 1990, Florida Atlantic University
M.C.S. December 1993, Florida Atlantic University
M.B.A. August 1996, University of Miami

QUALIFYING EXAMINATION & PUBLICATIONS

Time in Preparation: 2008-2011

Qualifying Examination Passed: Spring 2009

Selected Publications:

W. Altidor, T. Khoshgoftaar, and A. Napolitano. A noise-based stability evaluation of threshold-based feature selection techniques. In *Proceedings of the IEEE International Conference on Information Reuse and Integration*, IRI '11, Las Vegas, Nevada, USA, 2011. In Press.

W. Altidor, T. M. Khoshgoftaar, J. Van Hulse, and A. Napolitano. Ensemble feature ranking methods for data intensive computing applications. In B. Furht and A. Escalante, editors, *Data Intensive Computing*. Springer, 2011. In Press.

W. Altidor, T. Khoshgoftaar, and J. Van Hulse. Robustness of filter-based feature ranking: A case study. In *Proceedings of the Twenty-Fourth Florida Artificial Intelligence* Research *Society Conference*, FLAIRS-24, pages 453–458, 2011.

W. Altidor, T. Khoshgoftaar, and K. Gao. Wrapper-based feature ranking techniques for determining relevance of software engineering metrics. International *Journal of Reliability, Quality, and Safety Engineering (IJRQSE)*, 17(5):425–464, October 2010.

W. Altidor, T. Khoshgoftaar, and A. Napolitano. Wrapper-based feature ranking for software engineering metrics. In *Proceedings of the Eighth IEEE International Conference on Machine Learning and Applications*, ICMLA '09, pages 241–246, Miami, FL, USA, 2009.

W. Altidor, T. M. Khoshgoftaar, and J. Van Hulse. An empirical study on wrapper-based feature ranking. In Proceedings of the 21st IEEE International Conference on Tools with Artificial *Intelligence*, ICTAI '09, pages 75–82, Newark (New York Metropolitan Area), New Jersey, USA, 2009.

FAU

COLLEGE OF ENGINEERING & COMPUTER SCIENCE
Florida Atlantic University

THE FLORIDA ATLANTIC UNIVERSITY

COLLEGE OF ENGINEERING & COMPUTER SCIENCE

announces the

Ph.D. Dissertation Defense

of

WILKER ALTIDOR

for the degree of

DOCTOR OF PHILOSOPHY (PH.D.)

JUNE 17, 2011 / 10:00 AM

In

EE405 (CEECS Conference Room)

777 Glades Road

Boca Raton, FL

DEPARTMENT: Computer and Electrical Engineering and Computer Science

DISSERTATION TITLE: “Stability Analysis of Feature Selection Approaches with Low Quality Data”

CHAIR OF THE CANDIDATE’S PH.D. COMMITTEE:
Dr. Taghi M. Khoshgoftaar

PH.D. SUPERVISORY COMMITTEE:
Dr. Martin K. Solomon
Dr. Hanqi Zhuang
Dr. Bassem Alhalabi

STABILITY ANALYSIS OF FEATURE SELECTION TECHNIQUES WITH LOW QUALITY DATA

One of the greatest challenges to data mining is erroneous or noisy data. Several studies have noted the weak performance of classification models trained from low quality data. This dissertation shows that low quality data can also impact the effectiveness of feature selection, and considers the effect of class noise on various feature ranking techniques. It presents a novel approach to feature ranking based on ensemble learning and assesses these ensemble feature selection techniques in terms of their robustness to class noise. It presents a noise-based stability analysis that measures the degree of agreement between a feature ranking technique’s output on a clean dataset versus its outputs on the same dataset but corrupted with different combinations of noise level and noise distribution. It then considers classification performances from models built with a subset of the original features obtained after applying feature ranking techniques on noisy data. It proposes the focused ensemble feature ranking as a noise-tolerant approach to feature selection and compares focused ensembles with general ensembles in terms of the ability of the selected features to withstand the impact of class noise when used to build classification models. Finally, it explores three approaches for addressing the combined problem of high dimensionality and class imbalance. Collectively, this research shows the importance of considering class noise when performing feature selection.