Announces the Ph.D. Dissertation Defense of

Justin M. Johnson

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

“Addressing Highly Imbalanced Big Data Challenges for Medicare Fraud Detection”

November 7, 2022, 10:30 a.m.
Dissertation Defense

Zoom
Meeting ID: 938 9466 9496
Passcode: dissert101

DEPARTMENT:
Electrical Engineering and Computer Science

ADVISOR:
Taghi M. Khoshgoftaar, Ph.D.

PH.D. SUPERVISORY COMMITTEE:
Taghi M. Khoshgoftaar, Ph.D., Chair
Xingquan Zhu, Ph.D.
DingDing Wang, Ph.D.
Mehrdad Nojoumian, Ph.D.

ABSTRACT OF DISSERTATION
Addressing Highly Imbalanced Big Data Challenges for Medicare Fraud Detection

Access to affordable healthcare is a nationwide concern that impacts most of the United States population. Medicare is a federal government healthcare program that aims to provide affordable health insurance to the elderly population and individuals with select disabilities. Unfortunately, there is a significant amount of fraud, waste, and abuse within the Medicare system that inevitably raises premiums and costs taxpayers billions of dollars each year. Dedicated task forces investigate the most severe fraudulent cases, but with millions of healthcare providers and more than 60 million active Medicare beneficiaries, manual fraud detection efforts are not able to make widespread, meaningful impact. Through the proliferation of electronic health records and continuous breakthroughs in data mining and machine learning, there is a great opportunity to develop and leverage advanced machine learning systems for automating healthcare fraud detection.

This dissertation identifies key challenges associated with predictive modeling for large-scale Medicare fraud detection and presents innovative solutions to address these challenges in order to provide state-of-the-art results on multiple real-world Medicare fraud data sets. Our methodology for curating nine distinct Medicare fraud classification data sets is presented with comprehensive details describing data accumulation, data preprocessing, data aggregation techniques, data enrichment strategies, and improved fraud labeling. Data-level and algorithm-level methods for treating severe class imbalance, including a flexible output thresholding method and a cost-sensitive framework, are evaluated using deep neural network and ensemble learners. Novel encoding techniques and representation learning methods for high-dimensional categorical features are proposed to create expressive representations of provider attributes and billing procedure codes. Finally, meticulous performance evaluation and statistical analysis is used to highlight common pitfalls related to evaluating predictive models in the context of highly imbalanced big data and to make recommendations for future works in classifying highly imbalanced big data. As such, this dissertation covers the entire machine learning life cycle and provides meaningful contributions in areas related to data preparation, modeling techniques, and performance evaluation.
BIOGRAPHICAL SKETCH

Born in New Jersey, USA
B.S., Florida Atlantic University, Boca Raton, Florida 2016
M.S., Florida Atlantic University, Boca Raton, Florida 2019
Ph.D., Florida Atlantic University, Boca Raton, Florida 2022

CONCERNING PERIOD OF PREPARATION
& QUALIFYING EXAMINATION

Time in Preparation: 2019 - 2022
Qualifying Examination Passed: Fall 2019

Published Papers:


