Influence of Parsimony and Work-related Psychological Constructs in Predicting Turnover Intention when Using Machine Learning VS Linear Regression

A DISSERTATION

Submitted to the Faculty of

Montclair State University in partial fulfillment

of the requirements

for the degree of Doctor of Philosophy

by

Diego Figueiras

Montclair State University

Montclair, NJ

May 2024

Dissertation Chair: Dr. Michael Bixter

MONTCLAIR STATE UNIVERSITY

THE GRADUATE SCHOOL

DISSERTATION APPROVAL

We hereby approve the Dissertation

Influence of Parsimony and Work-related Psychological Constructs in Predicting Turnover Intention when Using Machine Learning VS Linear Regression

of

Diego Figueiras

Candidate for the Degree:

Doctor of Philosophy

Graduate Programs:
Counseling

Certified by:

Dr. Scott Herness
Vice Provost for Research and
Dean of the Graduate School

Date

Dissertation Committee:

Dr. Michael Bixter
Dissertation Chair

Dr. Kevin Askew

Dr. John T. Kulas

Abstract

This dissertation explores the ongoing debate between traditional statistical regression models and machine learning (ML) algorithms in predictive modeling, focusing on the impact of sample size and the number of variables. Study 1 investigates the relationship between sample size and predictive accuracy, proposing hypotheses regarding the advantages of ML over regression as sample size increases. Additionally, the study examines the influence of the number of variables on predictive accuracy, emphasizing the trade-off between ML and regression models. Using data from the Federal Employee Viewpoint Survey, the research aims to contribute insights into the conditions favoring each modeling approach. Study 2 shifts the focus to incremental validity, exploring whether work-related psychological constructs enhance ML models' predictive accuracy in turnover intention compared to biodata alone. The proposed hypotheses suggest that incorporating psychological constructs will improve predictive accuracy, addressing the "garbage in garbage out" concern prevalent in ML applications. The methods involve diverse datasets, including responses from federal employees and an online survey through Amazon's MTurk, with machine learning algorithms such as Gradient Boosting Trees, Random Forest, Neural Networks, Support Vector Machines, and logistic regression. The dissertation seeks to advance understanding in the field, offering practical insights for researchers and practitioners navigating the dynamic landscape of predictive modeling.