

## **ABSTRACT**

**DISSERTATION:** Fixed Effects or Mixed Effects?: Predictive Classification in the Presence of Multilevel Data

**STUDENT:** Anthony A. Mangino

**DEGREE:** Doctor of Philosophy

**COLLEGE:** Teachers College

**DATE:** July 2021

**PAGES:** 183

Some of the data in the social sciences features a nesting structure in which cases (level-1 units) are nested within higher-level clusters (level-2 units). This structure violates a fundamental assumption of many statistical models, namely the independence of cases, and thus necessitates the use of multilevel modeling techniques. Little research has yet been done assessing the efficacy of fixed and mixed effects models for supervised classification, where the outcome groups are known. The present study sought to compare fixed and mixed effects models for the purposes of predictive classification in the presence of multilevel data with small sample sizes. The first part of the study utilizes a Monte Carlo simulation methodology to systematically manipulate conditions within multilevel data across several different classifiers, including fixed and mixed effects logistic regression and random forests. Following the simulation study, an applied examination of the prediction of student retention in the public use Program in International Student Achievement (PISA) dataset will be considered to further bolster findings from the simulation study. Collectively, the results of both the simulation study and PISA data examinations will be used to provide recommendations to researchers for use when implementing classifiers for the purpose of prediction. Results of this study indicate that despite the use of fixed effects models, their predictions were nearly equivalent to mixed effects models across both the simulation and PISA examinations regardless of sample size. Taken holistically, these results

suggest that researchers should be more cognizant of the type of predictors and the data structure being used, as these factors carried more weight than did the model type in accuracy metrics.