

ABSTRACT

Author: Art Marciano
Title: Explainable AI and Data Optimization
Thesis Advisor: Imad Antonios
Department: Department of Computer Science
Year: 2022

In machine learning it is believed that more data help models perform more accurately. Larger data, however, might potentially present issues with too many patterns or the model not learning what the user might want it to learn. Since the internal logic is concealed, it is impossible to see what the models might be learning. Explainable AI (XAI) helps to address these problems and allows for management and oversight in the event of negative or undesirable results, such as biased decision making. XAI is typically used for describing an AI model, its expected impact and potential biases, in other words, it helps humans to understand the reason for the model's decisions. XAI examines the dataset and informs the user of what the model is learning from the dataset. While XAI's primary goal is to shed light on a machine learning model, it can be adapted to produce importance measures of individual data items relative to a model's performance. This study explores the joint use of the XAI method SHAP (short for SHapley Additive exPlanation) and the dimensionality reduction technique RFE as a method for reducing data size. The research demonstrates that in some cases this approach can maintain, or sometimes improve, model performance using a small fraction of the dataset.