Applications of Explainable Artificial Intelligence in Medicine

Benjamin Bowers Christopher Sheppard Jae Won Choi Syxri Mazoch Rohan Dave

Introduction

Explainable Artificial Intelligence, or XAI, provides a method of solving a frequently encountered issue in the field of machine learning. As AI becomes increasingly more complex, humans have difficulty comprehending and retracing how algorithms produced by machine learning models come to be. The whole calculation process turns into what is commonly referred to as a "black box" because it is impossible to interpret. These black box models are created directly from the data, and even the engineers or data scientists who create the algorithm cannot understand or explain what exactly is happening inside them or how the AI algorithm even arrived at a specific result. XAI poses a solution to this obstacle which allows human users to take a closer look into this black box and trust the output generated by it. Explainable AI can describe a machine learning model and its potential biases. It helps characterize model accuracy, fairness, transparency, and outcomes. Thus, it is beginning to see increased development in a variety of different fields. So what are they?

Applications

Well, the use cases are limitless, but just to name a few, explainability in AI is becoming especially important in areas such as the automotive industry. Due to the potential safety concerns that autonomous vehicles hold, the explainability of AI algorithms, especially in dealing with safety-critical decisions, would increase the trustworthiness of a self-driving car and reduce the number of crashes resulting from it. In the military field, discussions to use XAI for things like autonomous turrets are arising. A turret unexplainably shooting at objects and potentially causing lethal harm brings up many ethical concerns that explainable AI has the ability to address. Lastly, the field that our research project attempts to introduce XAI into is healthcare and medicine. Diagnosis through machine learning algorithms has become a reality. However, the potential unknown biases in data used to create diagnosis algorithms and the specific reasoning used to reach conclusions are all hidden behind a black box. These factors are particularly important in medicine since incorrect judgment could lead to significant physical harm. Thus, we are applying XAI techniques to datasets related to medicine in an attempt to expand development in the field of explainable artificial intelligence.

References

Gopal G., Wang H., Shakerin F..(2022). Fold-RM: A scalable and inductive learning algorithm for Multi-Category Classification of Mixed Data. Retrieved from Github: https://github.com/hwd404/FOLD-RM

Teboul, Alex. "Diabetes Health Indicators Dataset." Kaggle, 8 Nov. 2021, https://www.kaggle.com/datasets/alexteboul/diabetes-health-indicators-dataset.

Fold-RM and Prolog

The main XAI technique we are applying in our project is the FOLD-RM algorithm. This algorithm is written in Python and works by generating a set of rules based on inputted tabular data, like a spreadsheet. These rules can then be viewed and analyzed by humans in order to notice patterns, correlations, and even biases that are implicitly present in the data. A justification module is present and can be turned on in order to see more clearly how the rules for the dataset are being generated by the algorithm. These rules that are generated can be used to make predictions for new data as well. This is because FOLD-RM outputs its rules in the Prolog language, which is a programming language designed to generate predictions based off of a set of provided rules. Therefore, in regards to healthcare, one could use the FOLD-RM algorithm in tandem with Prolog in order to figure out whether a patient has a particular disease.

Our Dataset

The Fold-RM algorithm works especially well on tabular data as a classification task. We chose our tabular dataset from the Behavioral Risk Factor Surveillance System survey that is collected annually by the CDC, which is publically available through Kaggle. Our dataset is composed of 21 feature variables, namely High Blood Pressure (HighBP), HIgh colesteral (HighChol), Do they participate in Physical activity (PhysActivity), Do they have any Healthcare (AnyHealthcare), and what is their age (Age). The dataset's values consist of 250,000 people, i.e., 4.5 thousand have prediabetes, 35 thousand have diabetes, and 215 thousand have no diabetes.

Why diabetes? Using the Fold-RM algorithm on this diabetes dataset is particularly useful because diabetes is an exceedingly common chronic disease in the United States and is very often undiagnosed. Since the study, it has been found that 88 million adults in the U.S., more than 1 in 3 people, were living with prediabetes. 84% of those people did not know they had it. Of the 31 Million people living with Diabetes in the U.S., about half did not know they had it. Analyzing this dataset allows us to point out which key features are the primary indicators of diabetes.

For our research we propose the following goals: We aim to analyze the prolog rules produced by the model, explore XAI possibilities, and examine the extents of the algorithm.

Brief Analysis

To decide whether he/she are having *Diabetes*, we need 21 numbers such as BMI, Age, and Income. Suppose a man named John visited to test Diabetes and provided 21 data as follows

HioghB	HighCol	CholCheck
1.0	0.0	1.0
BMI	Smoker	Stroke
26	0.0	0.0
HeartDiseaseOrAttack	PhysActivity	Fruits
0.0	1.0	0.0
veggies	HvyAl coholConsump	AnyHealthCare
1.0	0.0	1.0
NoDocbccost	GenHlth	MentHlth
0.0	3.0	5.0
physHlth	diffwalk	sex
30	0.0	1.0
Age	Edu	Income
4.0	6.0	8.0

He got the result that he doesn't have a diabetes. According to our model, there are 3 cases for no Diabetes, and in this case, it is second case. Value for highbp is 1 so it is not the first case. For second case, we need to look at 5 conditions. Value for highbp is 1 so it is true, and for other 4 conditions, value for highchol is 0 so ab3, ab9, ab11, and ab16 is all False. Which means second case is true so this person doesn't have a diabetes.

Conclusion

Dataset	Accuracy	Size
Pre-Diabetes and Diabetes	84.67%	253k
Acute Binary Dataset	72.44%	70K
Health Indicators Binary Dataset	86.26%	253K

What we can see from our results is that, while these models cannot yet compete with statistical machine learning in terms of accuracy, they can provide valuable insights into how to shape our data to reduce biases and remove biased data. For instance, in one of our diabetes datasets, there was a small correlation found between having healthcare and being diagnosed with prediabetes. It is likely that the cause of this is that someone with healthcare is more likely to go see a doctor and have symptoms diagnosed than someone without healthcare, not that they are more susceptible to diabetes. Because we have the insight that the link between healthcare and diabetes is weak, and that it is more likely due to how the data was measured than the meaning of the data, we could remove that datapoint before training a more accurate statistical model. Overall, while the research in the field is rapidly accelerating, with technology like fold-RM vastly increasing accuracy over prior XAI models, it is currently best used in conjunction with more conventional learning methods to gain both the accountability and clear insight that Explainable AI can produce and the strong accuracy that statistical learning methods can produce.