

In the recent years, machine learning has been used in high-stake decision making as medical diagnosis, criminal justice system and loan decisions just to name a few. Therefore, having certain transparency on how these models work is crucial in the event where the model is biased or provide inaccurate predictions.

Black box models have been scrutinized though being highly accurate, it lacks transparency and the comprehensiveness compared to other models like linear regression and decision tree. Despite so, many researchers sacrifice the interpretable machine learning and chose to train the more accurate black box model.

However, it is often that black box models are not compatible with certain situations where even minor information are needed to be combined with risk assessment. For example, the COMPAS recidivism model, where it is accused of being racially biased in the context of it being used widely in the US justice system for bail and parole decisions.

Rashomon effect applies to where a variety of approximately-equally accurate descriptions are used to explain a phenomenon. For a given problem, all hypothetically possible models are referred to as the Rashomon set.

The Rashomon ratio is unique to each dataset, where a large ratio depicts multiple highly accurate models that can be used to solve the particular problem, indicating the simplicity of the problem. Using the Rashomon ratio, the Rashomon curve and elbow is introduced where the ratio is plotted against observed error of a particular model class. The turning point in the curve represents the ideal model class where it is neither too complex nor is it too erroneous.

If the Rashomon set contains a large enough set of models, a simpler function and one that is more interpretable can definitely be found. Therefore the Rashomon set is realistic in a way where it allows the researcher/ expert to have an idea of how complex the model would be without forgoing its accuracy, also keeping its generalizability and interpretability.

The Rashomon set could encourage researchers/ experts in looking for a highly accurate yet interpretable model instead of accepting black box model usage. It could also solve the problem of requiring information outside of the database when black box model prediction is in use. Biasness and poor decisions made through black box prediction can also be reduced.