# Explainable AI (XAI) Implementation

LIME requires you to convert H2O models' data and predictions to formats it understands, which can be tricky.

So we used a RandomClassifier to explain LIME

LIME, or **Local Interpretable Model-agnostic Explanations**, is a powerful technique in the field of machine learning interpretability. It aims to provide insights into the predictions made by complex models by approximating their behavior locally using simpler, interpretable models. Here are key notes on LIME:

Overview of LIME

* **Model Agnostic**: LIME can be applied to any machine learning model, regardless of its complexity or architecture. This flexibility makes it a valuable tool for interpreting a wide range of models, from linear regressions to deep neural networks.
* **Local Interpretability**: Unlike some interpretability methods that provide global explanations, LIME focuses on explaining individual predictions. It does this by examining how small changes in the input features affect the model's output, allowing for a detailed understanding of the model's behavior in the vicinity of a specific instance.

# Visualizing the explanation

• left most section displays prediction probabilities

•The middle section returns 5 most important features. For the binary classification task, it would be in 2 colors orange/blue. Attributes in orange support class 1(Survived) and those in blue support class 0 (Not Survived). Float point numbers on the horizontal bars represent the relative importance of these features.

•The color-coding is consistent across sections. It contains the actual values of the top 5 variables.

A screenshot of a computer

Description automatically generated

The above explanation was for instance 0 of the test data.

The left most grapg shows the probabilities of “Survived” and “Not Survived”

The middle graph shows 5 most important features

It is seen that Sex contributes most (0.41) to “Not survived”

This is followed by Pclass at (010)

Lime also provides a range between which the contribution is seen

For example age between 20 and 28 contributes to the “Not survived”

A screenshot of a computer

Description automatically generated