Project Proposal

Filipe Sousa, fc52748
João Monteiro, fc49821

Goals

The goal of this proposal is to compare different tools/methodologies when it comes to model interpretability. For this, our problem definition is two datasets applying to two different types of algorithms and use LIME and SHAP to 1) between the algorithms compare how they learn and see if there´s any difference in this learning process and 2) how LIME and SHAP differ in the interpretation of this models.

Datasets

For our project, we will use two different datasets, the first one contains an airline passenger satisfaction survey , with the goal of understanding the factors that impact satisfaction on the customer and the second one contains key indicators of heart disease, with the goal of predict if someone has heart disease based on the other indicators.

The airline passenger satisfaction dataset has 23 features, divided into the target , 4 continuos features and 18 categorical features. This dataset has 130 thousand observations.

The key indicators of heart disease dataset contains 18 features, divided into the target, 4 continuos features and 14 categorical features. This dataset has 320 thousand observations.

The links of the two datasets are:

https://www.kaggle.com/datasets/teejmahal20/airline-passenger-satisfaction

https://www.kaggle.com/datasets/kamilpytlak/personal-key-indicators-of-heart-disease

Algorithms / ML Techniques

For this project we will use two different types of machine learning algorithms: 1) SVM (Support Vector Machine) and 2) Gradient Boost.

The first one it's a clear solution as a baseline model and should be interesting to see the interpretation of this type of algorithm.

The second one is a stronger algorithm with more computational costs. Due to it´s complexity it should be interesting to apply interpretability tools to it´s results.

For the SVM algorithm we will use the scikit learn implementation (https://scikit-learn.org/stable/modules/classes.html#module-sklearn.svm) and for the Gradient Boost algorithm we are still unsure if we are using CatBoost (https://catboost.ai) or XgBoost (https://xgboost.readthedocs.io/en/stable).

Interpretability

LIME and SHAP are techniques that can explain the output of any machine learning model.

Like we said we are going to use LIME and SHAP to interpretate and explain the difference between the two chosen algorithms, and why the data is classified the way it is, opening the door to explain the results of these models.

We will use the last updated implementation of LIME (https://github.com/marcotcr/lime) and the library SHAP (https://github.com/slundberg/shap).