MARKETING ALGORITHM PREDICTIONS

CLASSIFICATION MODEL

Xgboost classifier is the machine learning model deployed,

- The train r2 score: 0.7349091834728909
- Test r2 score: -0.025773881989037628
- Train RMSE: 0.26
- Test RMSE: 0.32

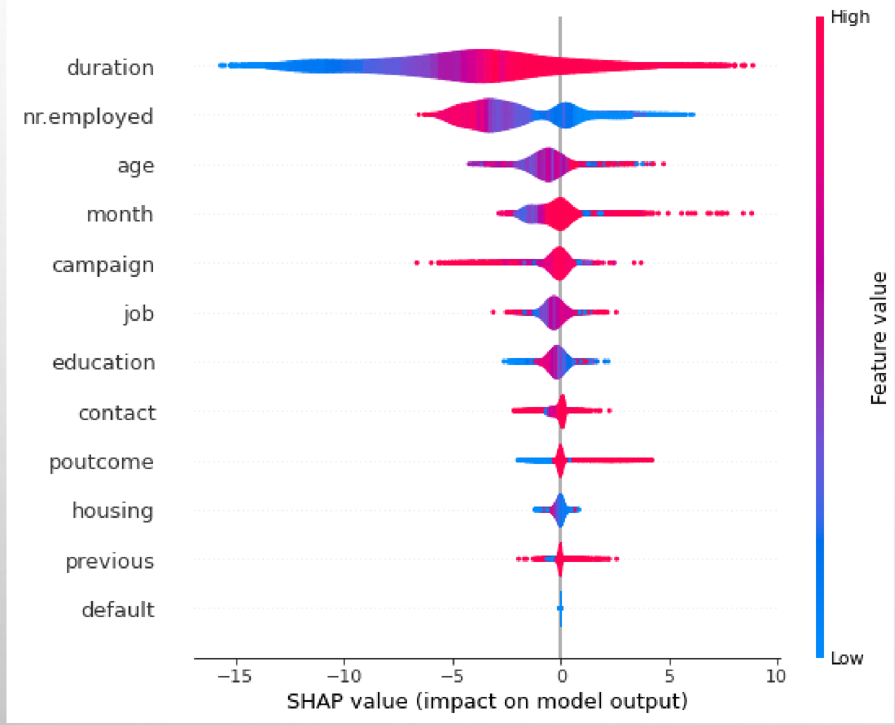
Hypertunining of xgboost machine learning model

- Train r2 score: 0.9946918746159615
- Test r2 score: 0.973367003850796
- Train RMSE: 0.0364
- Test RMSE: 0.0517

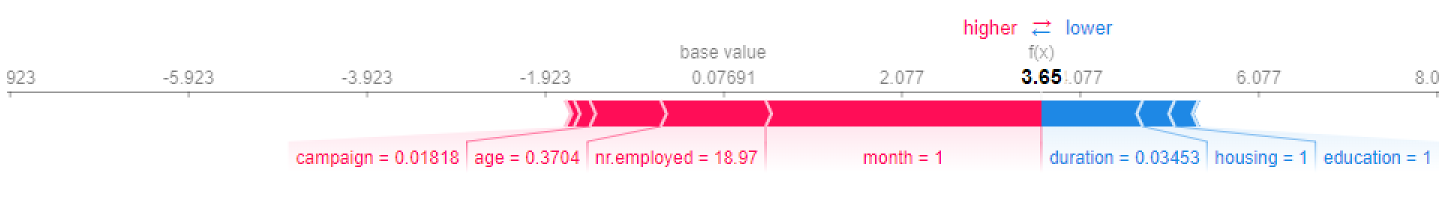
Two classification model was evaluated, Random Forest Classifier and Xgboost Classifier. The model. Shows that Xgboost Classifier hypertunning performs better in prediction, hence its deployment in machine learning model.

SHAPELY- GLOBAL INTERPRETATION

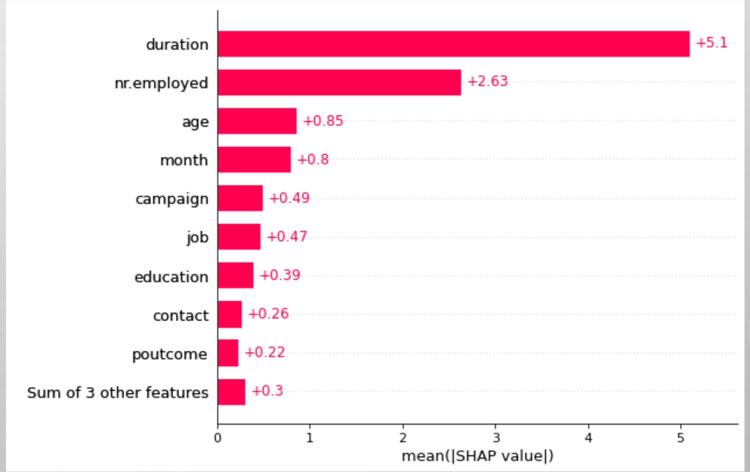
- SHAP- global variable specifically identify the important features that contribute to/impact a particular outcome - overall based
- With the duration feature, used as a benchmark the outcome is:



- Features on top contributed so much to the model's output. On a global scale, the important features as shown on the chart are ***duration, nr.employed*** (observation#20), ***age, month, campaign, job and education*** (observation#4)



- The duration has high importance impact in our model, has seen in the shapely explainer above.
- **observation #20**, nr.employed (number of employee)feature has positively low impact on the output of the XGBoost hypertuning classification model, as to whether a client will subscribe to a term deposit
- **observation #4**, education feature has a moderately high impact on the output of the XGBoost classification model on the global level, as to whether a client will subscribe to a term deposit.



SHAPELY- GLOBAL INTERPRETATION

- SHAP- global variable specifically identify the important features that contribute to/impact a particular outcome - overall based

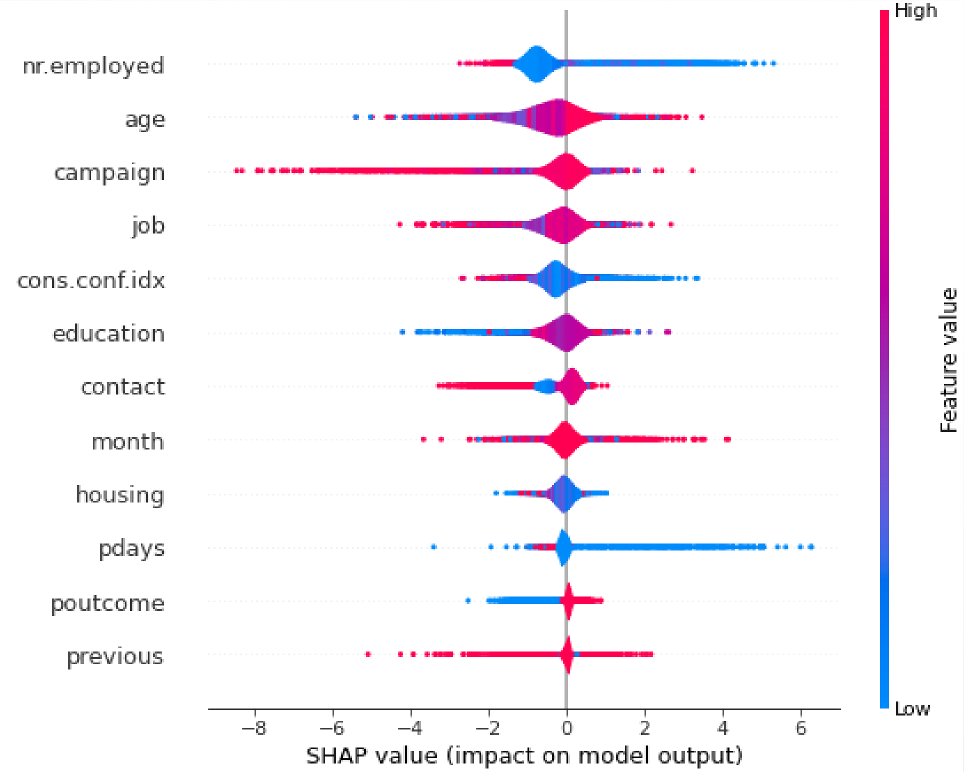


Figure 1 : important features summary plot chart

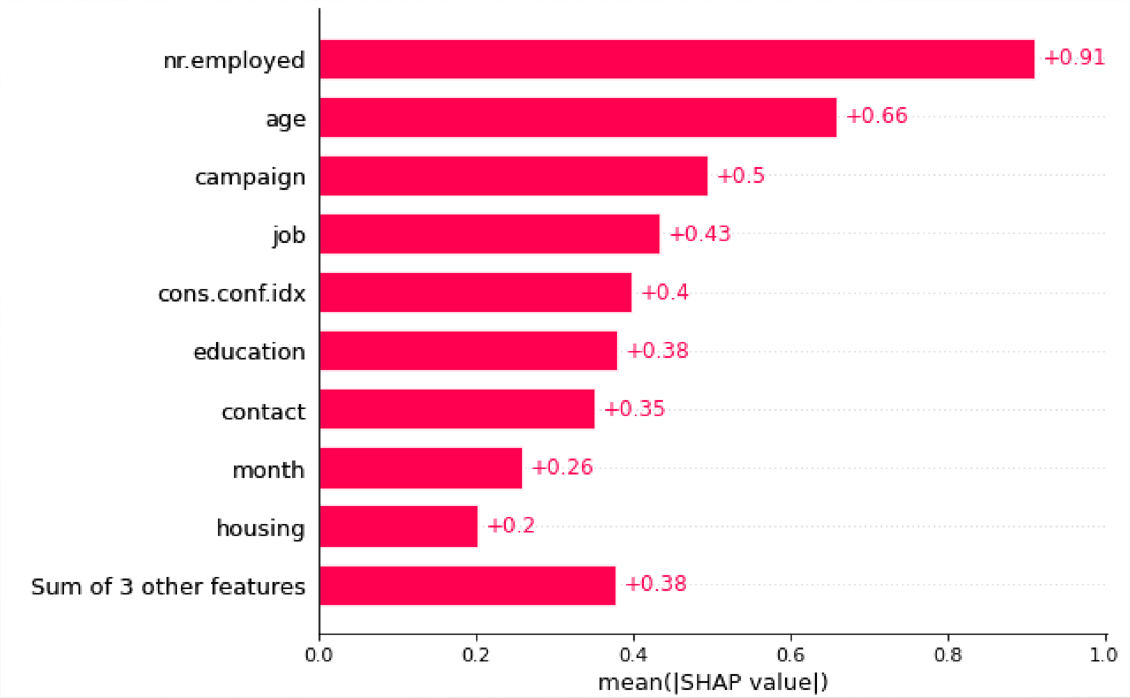


Figure 2 : important features summary barchart

- Features on top contributed so much to the model's output. On a global scale, the important features as shown on the Figure 1 chart are ***nr.employed*** (observation#20), ***age, campaign, job, cons.conf.idx and education*** (observation#4)…. etc

- The **observation #20**, nr.employed (number of employee)feature has positively low impact on the output of the XGBoost hypertuning classification model, as to whether a client will subscribe to a term deposit as shown in the shapely explainer above.
- **observation #4**, education feature has a moderately high impact on the output of the XGBoost classification model on the global level, as to whether a client will subscribe to a term deposit.

SHAPELY- GLOBAL INTERPRETATION

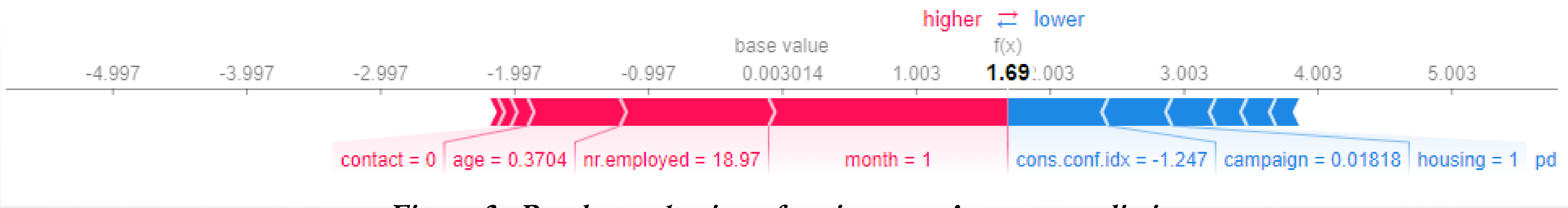


Figure 3 : Random selection of an instances' output prediction

- The output value 1.69 marked in bold in Figure 3 is the prediction probability for the instance being predicted as client subscription to term deposit. Xgboost model voted for the subscription to term deposit class, so the instance was predicted as subscription to term deposit

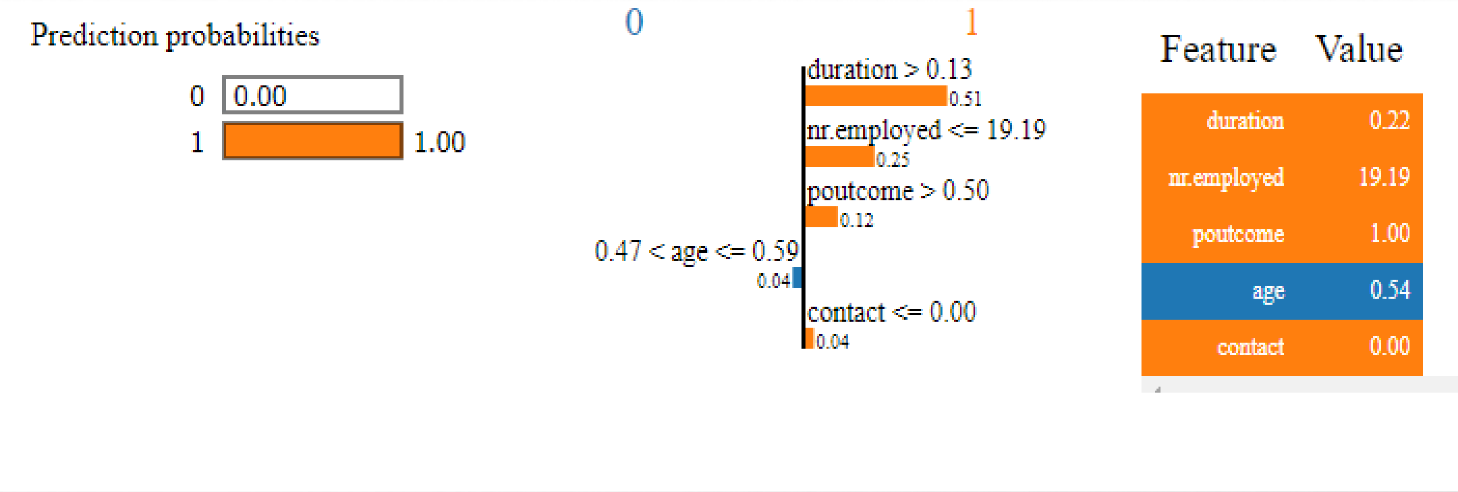
- Selecting a random instance from the test set, SHAP's force_plot function in Figure 3 shows how each feature contributes to the model's prediction for this specific instance, with positive values pushing the prediction towards the positive class and negative values pushing it towards the negative class

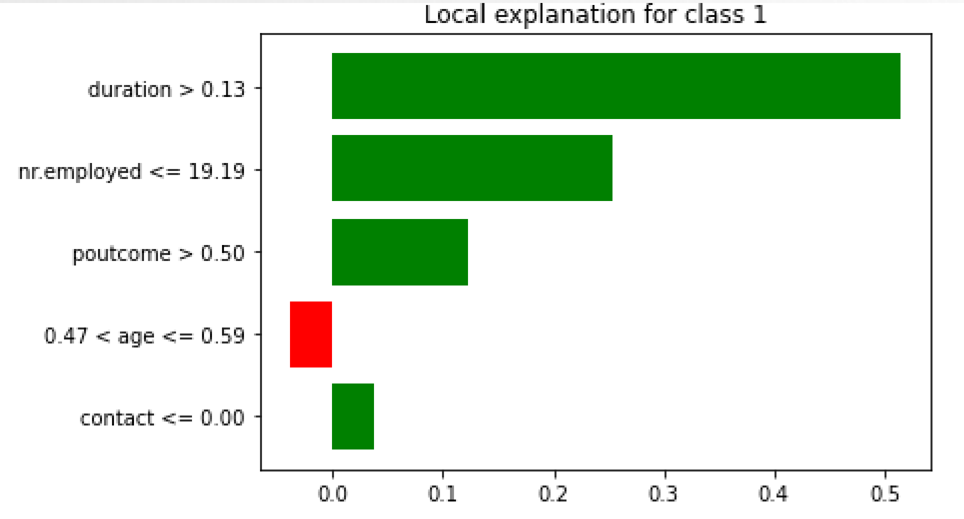
This plot shows the mean absolute SHAP values for each feature, providing an overview of the most influential features in the model.

LIME- LOCAL INTERPRETATION

- LIME - Local Interpretation variable It implies the features that contributes to a decision made - instance specific

- Features on top contributed so much to the model's output. On a global scale, the important features as shown on the chart are ***duration, nr.employed*** (observation#20), ***age, month, campaign, job and education*** (observation#4)





- The duration has high importance impact in our model, has seen in the shapely explainer above.
- **observation #20**, nr.employed (number of employee)feature has positively low impact on the output of the XGBoost hypertuning classification model, as to whether a client will subscribe to a term deposit
- **observation #4**, education feature has a moderately high impact on the output of the XGBoost classification model on the global level, as to whether a client will subscribe to a term deposit.

LIME- LOCAL INTERPRETATION

- LIME - Local Interpretation variable It implies the features that contributes to a decision made - instance specific
- This considered the top five (5) features that impact/contributed to the model's output., the important features as shown on the Figure 4 & 5 chart are ***nr.employed*** (observation#20), ***contact, previous, campaign and education*** (observation#4)

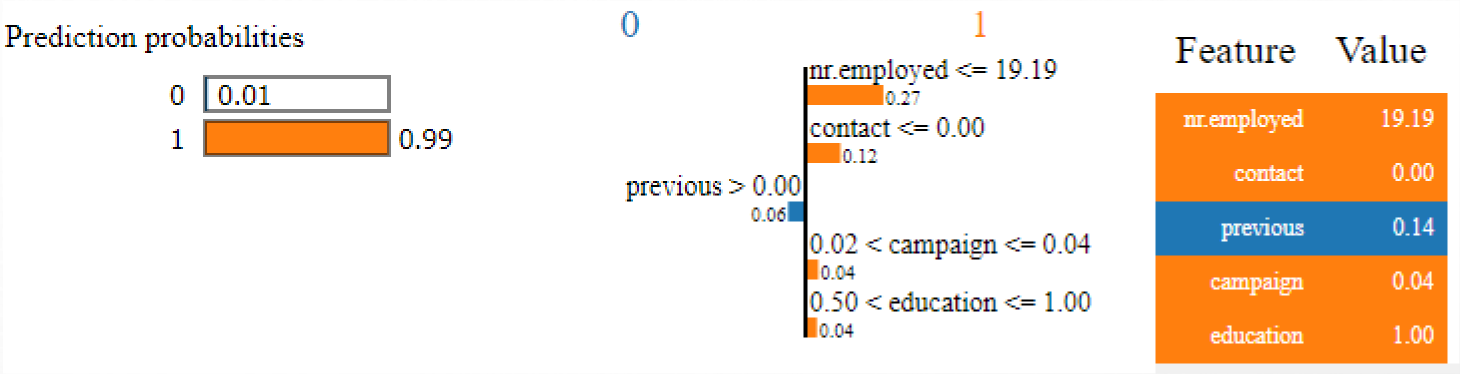


Figure 4 : Prediction Probabilities

This specific instance shows that this features contributes to the reason behind a client subscription to a term deposit.

The prediction probabilities from our xgboost classifier has 0.99, that is 99% chances that a client will subscribe to the term deposit.

The Feature and the value that contributes to this prediction is given in Figure 4.

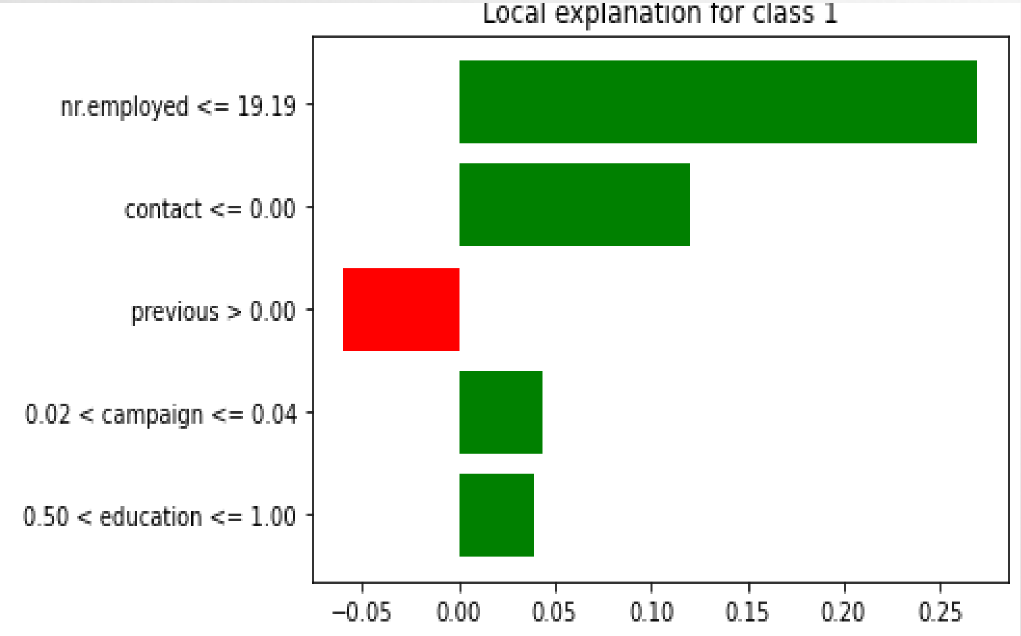


Figure 5 : Top Five(5) features on Local explanation

SHAP – GLOBAL INTERPRETATION

The following was carried out:

1. generate a synthetic binary classification dataset using scikit-learn's make_classification function with 1000 samples and 10 features, 5 of which are informative and 5 are redundant.
2. split the dataset into training and test sets and train an XGBoost classifier on the training data.
3. initialize a SHAP TreeExplainer with the trained XGBoost model. This explainer is specifically designed for tree-based models like XGBoost.
4. calculate SHAP values for the test set using the shap_values method of the explainer. These values represent the feature importances for each instance in the test set.
5. visualize the global feature importances using SHAP's summary_plot function with plot_type="bar". This plot shows the mean absolute SHAP values for each feature, providing an overview of the most influential features in the model.
6. interpret an individual prediction, we select a random instance from the test set (instance_idx = 42) and use

SHAP's force_plot function. This plot shows how each feature contributes to the model's prediction for this specific instance, with positive values pushing the prediction towards the positive class and negative values pushing it towards the negative class.

CONCLUSION

- **SHAP**

SHAP provides a powerful way to interpret XGBoost models by quantifying the impact of each feature on the model's predictions. The summary_plot gives a global view of feature importances, while the force_plot allows you to understand the factors driving a specific prediction. By combining these insights, you can gain a deeper understanding of your model's behavior and make more informed decisions based on its predictions.

- **LIME**

LIME explanations are local, meaning they are specific to the chosen instance and may not represent the model's overall behavior. It's essential to consider multiple instances and the global feature importances to get a comprehensive understanding of the model's decision-making process.