# 3.2.3 SHAP Analysis to Investigate Relative Impact of Features on Target

Four kinds of classifier are used to predict the x value:

* Linear Regression
* Random Forest
* XG Boost
* LightGBM

This waterfall graph shows how the features contributing the output of the prediction of selected model. That is, the average model output based on each feature over the dataset we passed. The features are both in blue, which mean that, they give a lower expectation of the output of the model, and in red, which mean that, they give a higher expectation of the output of the model. Both features are in blue because our model is kind of overestimated.

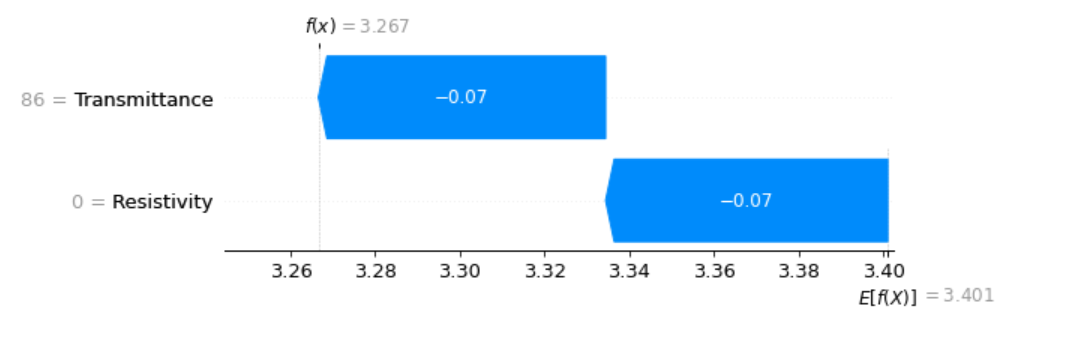


Fig 1 Waterfall graph – XGB Classifier

This is the plot show the base value of each feature and how it will react on the model.

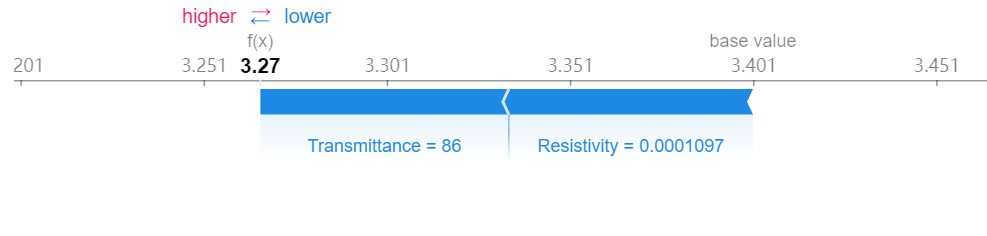


Fig 2 Force Plot – XGB Classifier – Visualize the prediction’s explanation with a force plot

The force graph only shows part of the model. Here, by rotating 90 angle and stack them horizontally, we will have a more detailed sight for the entire dataset.

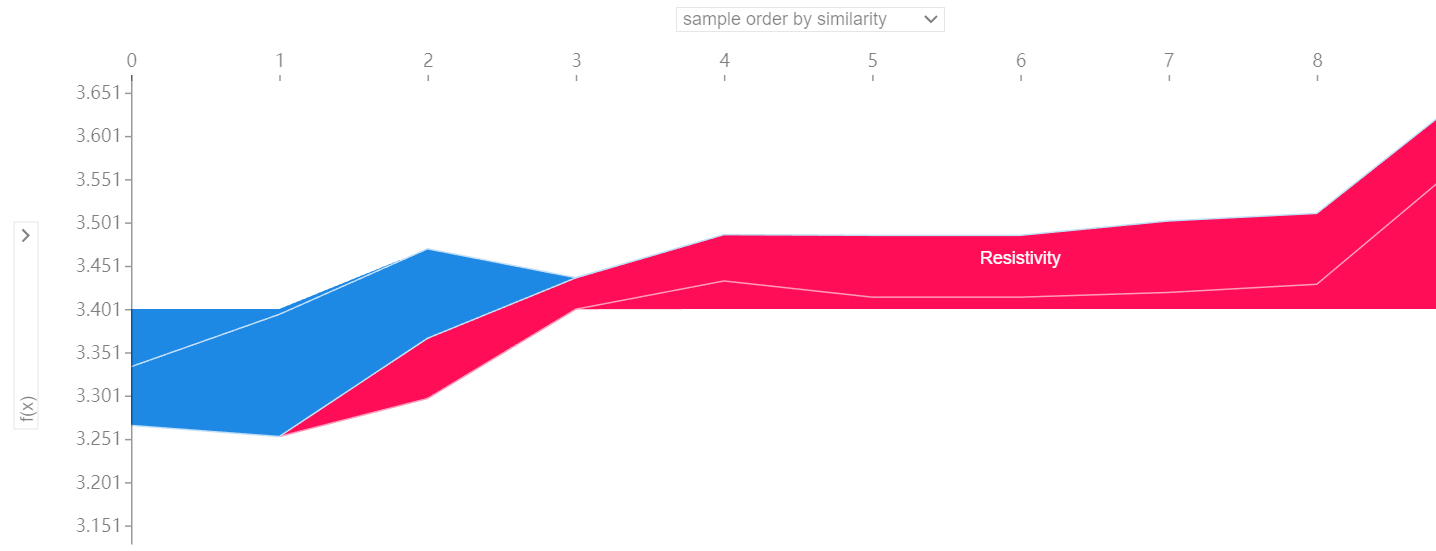


Fig 3 Force Plot – Entire Model

The bee swarm can give us an outlook of features involved in this model. It can give us an idea which feature has more influence on the model. The bee swarm plot sorts the features by their influence on the model, that is, the sum of SHAP value magnitudes of samples. We can see that the most important features are the ones with the highest SHAP values while the SHAP values show the distribution of the impacts of the features.

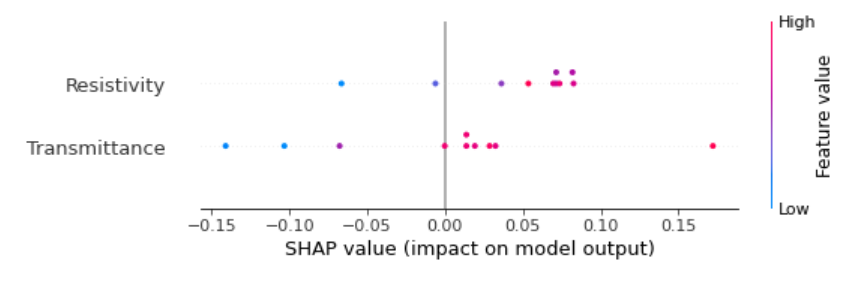


Fig 4 Bee Swarm – Outlook the features involved

This standard bar plot shows the mean absolute value of the SHAP value of both resistivity and transmittance.

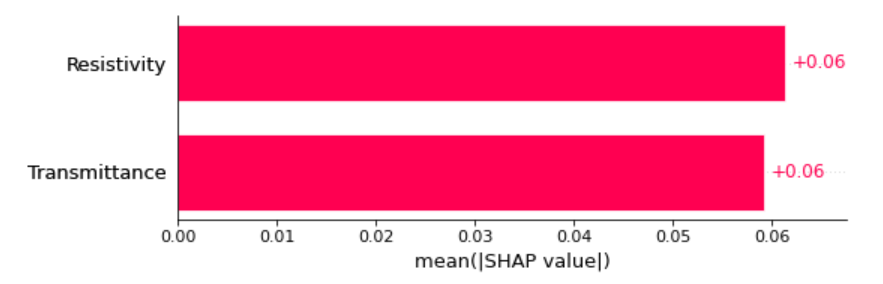
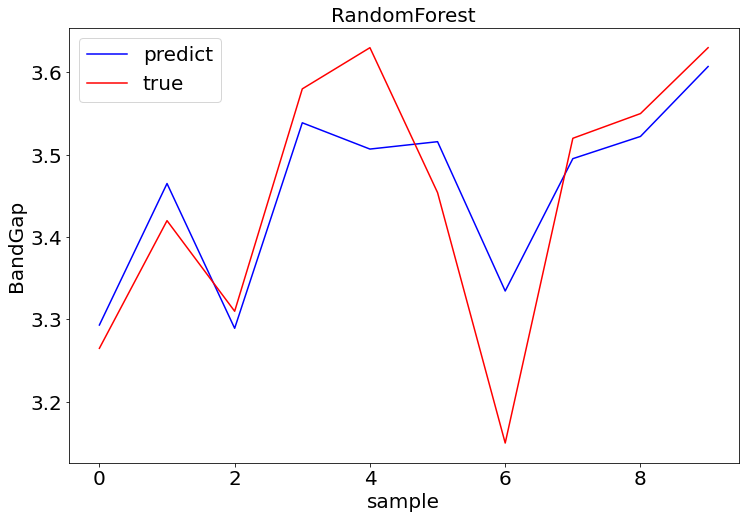
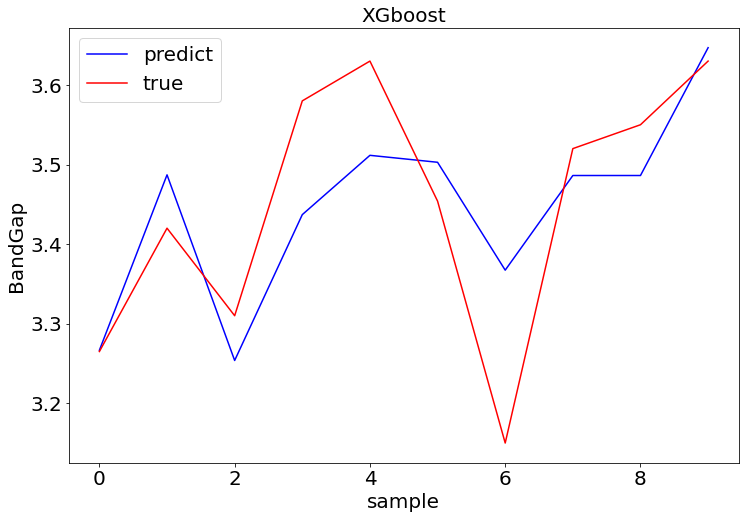


Fig 5 Bar Plot – Mean Absolute Value of the SHAP value

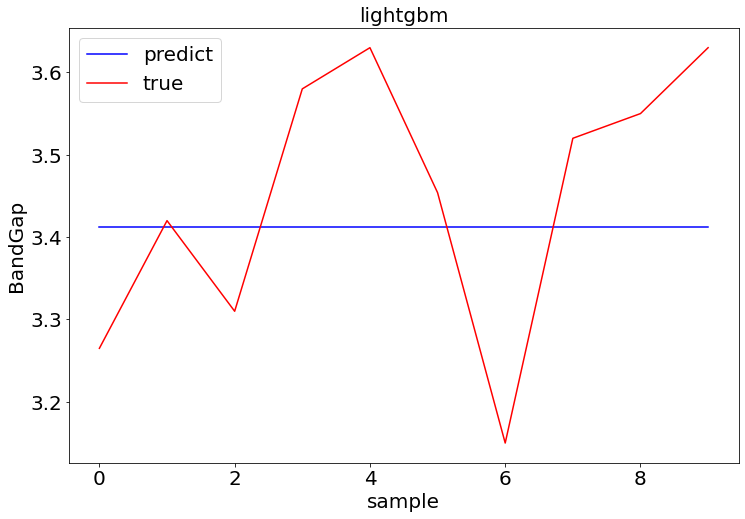
Possibly used Image:



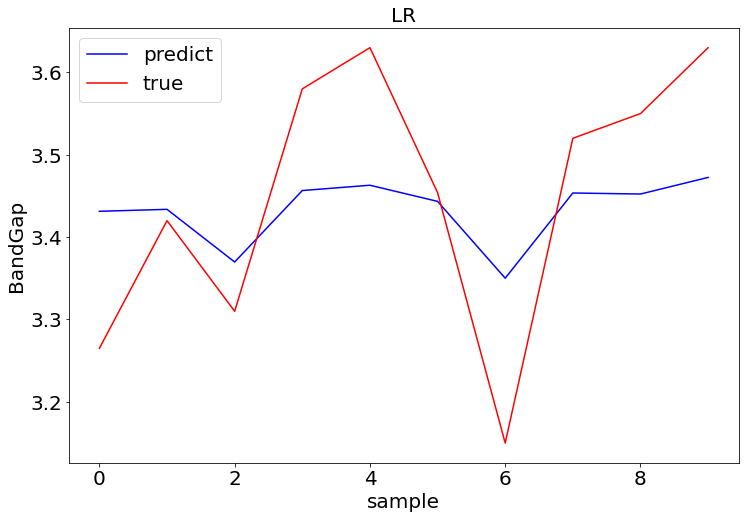
The difference between predicted and true value with random forest



The difference between predict and true value with XGboost



The difference between predict and true value with lightgbm



The difference between predict and true value with linear regression