**Extreme Gradient Boosted Trees of Depth 2 (XGB2)**

* XGB2 refers to a specific variant of the Extreme Gradient Boosted (XGB) Forests model. XGB2 limits the depth of each tree to 2 splits, which allows for pairwise interactions between features.
* Formula: f(x) = ∑ fk(xi)
* Advantages:
  + Handles non-linear relationships
  + Distinguishes interactions between variables better than linear models
* Disadvantages:
  + Prone to overfitting

## **Explainability and Interpretability**

### *Global Explainability*

#### **Permutation Feature Importance (PFI)-Computes the change in prediction performance as the measure of feature importance.** Breaks the relationship between the feature and the target, thus the drop in the model score is indicative of how much the model depends on the feature

A picture containing text, screenshot, software, number

Description automatically generated

* Loan\_to \_value \_ratio\_ std, Property value std and debt to income ratio std are the top three features that play pivotal role in deciding high\_priced\_mortagage among other features.

#### Partial Dependency Plot (PDP) and Accumulated Local Effects (ALE)

* **Partial Dependence Plot (PDP): used to understand how the prediction varies as a function of variables of interest, by averaging over other variables.**
  + Not recommended if the features are correlated
* **Accumulated Local Effects (ALE): describes how features affect a model prediction**
  + Shares the same goal as PDP (Partial Dependence Plot)
  + Overcomes the features correlation problem by averaging and accumulating the difference in predictions across the conditional distribution, limiting the effects of specific features.

A screenshot of a graph

Description automatically generated with medium confidence

There is a positive relationship between loan\_to\_value\_ratio\_std and high\_priced\_mortgage. As loan\_to\_value\_ratio\_std increases, the probability of that individual of getting high\_priced\_mortgage increases.

A screenshot of a computer screen

Description automatically generated with low confidence

There is a negative relationship between property\_value\_std and high\_priced\_mortgage. As loan\_to\_value\_ratio\_std increases, the probability of that individual of getting high\_priced\_mortgage increases.

A screenshot of a graph

Description automatically generated with low confidence

There is a positive relationship between debt\_to\_income\_ratio\_std and high\_priced\_mortgage. As loan\_to\_value\_ratio\_std increases, the probability of that individual of getting high\_priced\_mortgage increases.

A picture containing line, diagram, plot, screenshot

Description automatically generated

There is a positive relationship between loan\_amount\_std and high\_priced\_mortgage. As loan\_to\_value\_ratio\_std increases, the probability of that individual of getting high\_priced\_mortgage increases.

A picture containing line, diagram, rectangle, text

Description automatically generated

………………………………….

A picture containing line, diagram, plot

Description automatically generated

…………………………..

A picture containing diagram, screenshot, plot, design

Description automatically generated

……………………………….

A picture containing screenshot, diagram, rectangle, plot

Description automatically generated

……………………………….

A picture containing screenshot, diagram, plot, line

Description automatically generated

……………………….

A picture containing plot, diagram, screenshot, line

Description automatically generated

…………………………….