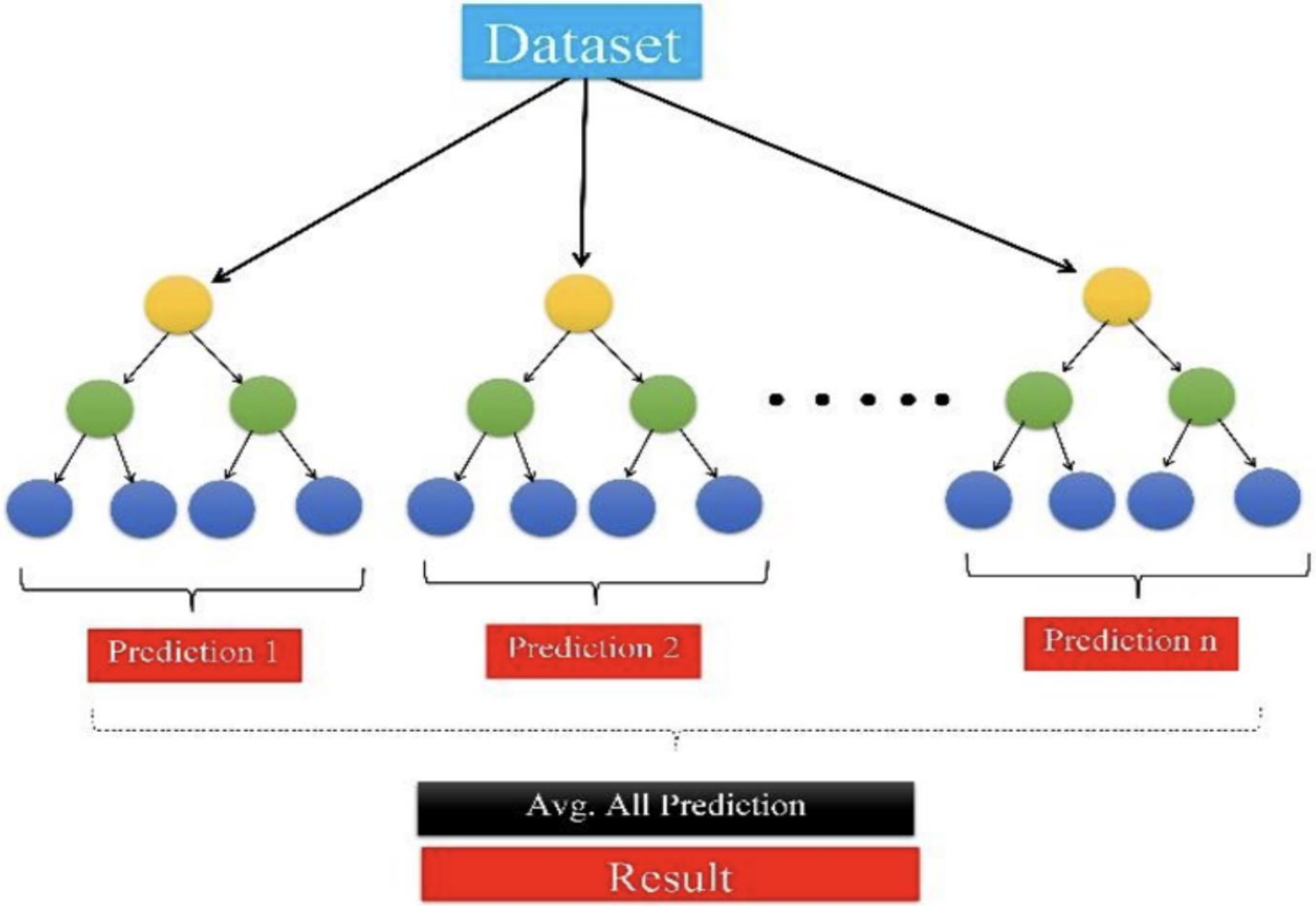
XG Boost - Regression

XGBoost, short for extreme Gradient Boosting, It works by building an ensemble of decision trees, where each tree is trained to make predictions based on a subset of the available data. The trees are grown sequentially, with each tree learning from the mistakes of the previous tree.

XG Boost Regression diagram



The diagram represents the XGBoost regression process, which builds multiple decision trees to make predictions. Here are the simple steps:

Dataset: Start with the input dataset, which contains the data to be analyzed.

Decision Trees: The dataset is split into multiple decision trees (shown as yellow nodes branching into green and blue nodes).

Predictions: Each tree generates a prediction (Prediction 1, Prediction 2, ..., Prediction n) based on its splits.

Average Prediction: The predictions from all trees are averaged to produce a final "Avg. All Prediction" value.

Result: This average is used as the final output or result of the regression model.

This process combines multiple weak models (trees) to create a stronger, more accurate prediction.

The formula for the final prediction in XGBoost regression is based on averaging the predictions from all individual trees. Mathematically, it can be expressed as:

$$\text{Final Prediction} = \frac{1}{n} \sum_{i=1}^{n} \text{Prediction}_i$$

Where:

- Prediction_i is the output of the i-th decision tree.
- n is the total number of trees.
- The sum of all tree predictions is divided by n to get the average.

This averaging process helps reduce overfitting and improve the model's accuracy.