Model Construction:

Before the model is constructed there are a lot of pre processing steps that need to be done in order for the model to be properly constructed.

The Machine learning algorithms used in model construction are:

1. Logistic Regression
2. Decision Tree Classifier
3. Random Forest Classifier
4. K-NN classifier
5. AdaBoost Classifier
6. Gradient Boosting Classifier
7. eXtreme Gradient Boosting classifier

The model construction can be divided into three major parts:

1. Splitting the Predictors(X) and the Target variable(y).

While splitting the features into Predictors and Target variables, we use the iloc method to split the features column wise. In our case, we have 10 features as X. Those features are:

* Credit Score
* Geography
* Gender
* Age
* Tenure
* Balance
* Number of products
* Whether the customer has credit card or not
* Whether the customer is an active member or not
* Estimated Salary of the Customer

1. Splitting the X and y further. That is the test-train split. Now, we have 4 variables. X\_train, X\_test ; y\_train, y\_test. The train data is 80% of the complete data whereas the test data is 20% of the complete data sampled randomly.

We have used the train\_test\_split python library to split the data and defined the random state as 42.

1. Fitting the Required machine learning model using the algorithms.

Model Training, Testing and Evaluation:

The algorithm we use are imported from the Scikit Learn library which is an open-source library.

We trained the models using the train data that we divided earlier using the train test split.

To predict the values, we use the test data from the train test split.

For the purpose of evaluation, we use the Confusion matrix and accuracy parameter derived from the confusion matrix parameters.

Accuracy score = (Number of Correct predictions) ÷ (Total Number of Predictions made)

The evaluation Metrics used were Train/Test split validation as well as K fold Cross validation techniques.

Since these two are different validation techniques, there will be a slight deference in the accuracy scores of the same model.