**Problem Statement:**

The goal of this hackathon is to develop algorithms for predicting accurately the possibility of disease or not based on features which have been provided in the dataset.

Here target label is either 0 or 1

The inputs are 100x5 matrices and target is the possibility of occurrence of disease or not.

**About Data:**

1. train.zip -Training set.

It consists of two files: x\_train.csv (training data features) and y\_train.csv (Target variable)

1. test.zip - Test set.

It consists of one file: x\_test.csv.

There is no label information in the test set .The participants need to predict the target on the test data.

**File descriptions:**

***x\_train.csv*** - It contains six columns separated by comma. The first column is the UniqueId (masked). The column headers beginning with H3K\* are the reads in each bin on five core histone modification marks. Each UniqueId has 100 bins.

***y\_train.csv*** - It contains two columns separated by comma. The first column is the UniqueId. The second column is its corresponding target label (0 or 1) which means either high or low

***x\_test.csv*** - The test set, similarly formatted as the x\_train.csv

**Evaluation:**

The participants will be ranked based on the accuracy of their predictions.

The evaluation metric for this competition is

***F1 score***

Submission files should contain two columns: Id and Prediction which could be 0 or 1.

The first column of the csv file is the Id (1,2,3,...,3870,3871). The second column is prediction (0,1, 0....,0,1)

The file should contain a header and have the following format as shared in the sample output.