

# Counterfeit Medicines Sales Prediction

Counterfeit medicine are fake medicines which are either contaminated or contain the wrong or no active ingredient. They could have the right active ingredient but at the wrong dose. Counterfeit drugs are illegal and are harmful to the health. 10% of the world's medicine is counterfeit, problem is even worse in developing countries. Up to 30% of medicines in developing countries are counterfeit Millions of pills, bottles and sachets of counterfeit and illegal medicines are being traded across the world The World Health Organization (WHO) is working with International Criminal Police Organization (Interpol) to dislodge the criminal networks raking in billions of dollars from this cynical trade. Despite all these efforts, counterfeit medicine selling rackets do not seem to stop popping here and there. It has become a challenge to deploy resources to counter these; without spreading them too thin and eventually rendering them ineffective. Government has decided that they should focus on illegal operations of high net worth first instead of trying to control all of them. In order to do that they have collected data which will help them to predict sales figures given an illegal operation's characteristics.

## Data Files

Train Dataset = counterfeit\_train.csv

Test Dataset = counterfeit\_test.csv

## Formal Problem Statement

Variable names are self-explanatory. Your task here is to build predictive model for predicting sales figures given other information related to counterfeit medicine selling operations. You need to build your model on Train data. Test data does not have response column, you need to predict those values and submit it as a csv format.

Submission CSV should resemble the file: Sample Submission = 'sample\_submission.csv'

Column names, value types should exactly match. Also, number of rows in the submission csv should be exactly same as test data. If this is not taken care of, your submission will not be graded.

## Evaluation Criterion

Score will be calculated as:  $\text{Score} = 1 - (\text{MAE} / 1660)$

Where MAE is mean absolute error on test file. You need to score more than 0.5 in order to pass the project submission. Do not read too much into score formulation, it is just to scale MAE. You just need to focus on minimising MAE