

Counterfeit Medicines Sales Prediction

Counterfeit medicines are fake medicines which are either contaminated or contain 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 health. 10% of the world's medicine is counterfeit and the 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 don't 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 a predictive model for predicting sales figures given other information related to counterfeit medicine selling operations. You need to build your model on the train dataset. Test dataset does not have a response column; you need to predict those values and submit it in a csv format.

Evaluation Criterion

Part 1: