**KDD-CUP-98**

**Introduction KDD Cup 1998: Direct marketing for profit optimization**

**The competition task is a regression problem where the goal is to estimate the return from a direct mailing in order to maximize donation profits.**

[**http://www.kdd.org/kdd-cup/view/kdd-cup-1998/Intro**](http://www.kdd.org/kdd-cup/view/kdd-cup-1998/Intro)

<http://kdd.org/cupfiles/KDDCupData/1998/readme.txt>

<https://www.kdd.org/cupfiles/KDDCupData/1998/instruct.txt>

Problem Statement and Idea:

The goal of the problem is to maximize net profit from a marketing campaign for fundraising. Since the response rate is low (about 5 percent), it is necessary to devise a strategy to identify profitable candidates. Given a candidate-member x, to predict:

1. whether x will donate? (Clasification)

2. what amount will x donate? (Regresion, not implemented yet)

Then calculate revenue earned from x is:



where prob[.] denotes probability and E[.] denotes expectation

Files description:

**1l. Data ingestion and preprocessing:**

DataExplAndPrep.ipynb

**2. Clasification Models:**

ClassificationModel.ipynb

**3. Test set scoring:**

TestSetScoreCalc.ipynb