**TalkingData AdTracking Fraud Detection - Can you detect fraudulent click traffic for mobile app ads?**

***Fraud risk is everywhere, but for companies that advertise online, click fraud can happen at an overwhelming volume, resulting in misleading click data and wasted money. Ad channels can drive up costs by simply clicking on the ad at a large scale. With over 1 billion smart mobile devices in active use every month, China is the largest mobile market in the world and therefore suffers from huge volumes of fradulent traffic.***

***TalkingData, China’s largest independent big data service platform, covers over 70% of active mobile devices nationwide. They handle 3 billion clicks per day, of which 90% are potentially fraudulent. Their current approach to prevent click fraud for app developers is to measure the journey of a user’s click across their portfolio, and flag IP addresses who produce lots of clicks, but never end up installing apps. With this information, they've built an IP blacklist and device blacklist.***

***While successful, they want to always be one step ahead of fraudsters and have turned to the Kaggle community for help in further developing their solution. In their 2nd competition with Kaggle, you’re challenged to build an algorithm that predicts whether a user will download an app after clicking a mobile app ad. To support your modeling, they have provided a generous dataset covering approximately 200 million clicks over 4 days!***

1. What is the problem you want to solve?

Solving the fact that the we can decipher immediately whether a certain user/device/IP will be more prone to downloading the clicked app or not.

1. Who is your client and why do they care about this problem? In other words, what will your client DO or DECIDE based on your analysis that they wouldn’t have otherwise?

The client needs to decipher immediately whether a certain user will be more prone to downloading the clicked app or not. Decision is made whether we discount and/or interrupt such clicks or create another blacklist for such users. As an extreme measure the client can create a firewall rule against the origin of such clicks.

If clicking and not downloading the app may be means of fraudulent activities a more rigorous method of security verification such as 3-step verification maybe in order against such fraud-based activities.

On the other hand a reward maybe just the right solution that the client want to pass on to its users that actually download their app. For repeated downloads the client may go to the extent of starting a loyalty-based program for their users.

1. What data are you going to use for this? How will you acquire this data?

This data has been made available by TalkingData, a Chinese big data service platform for the Kaggle competition that they have sponsored.

This data is acquired from the Kaggle website.

1. In brief, outline your approach to solving this problem (knowing that this might change later).

Exploring the data and cleaning using visualization to understand the data will be the first goal. Next run statistical tests to get a sense of which features are stronger than others by quantifying and using the p-values. Lastly Machine Learning modeling techniques will follow to be used for making actual predictions.

1. What are your deliverables? Typically, this would include code, along with a paper and/or a slide deck.

The deliverables will be the code along with a paper or slide deck.