CKME136 - Capstone Project
Abstract

TalkingData AdTracking Fraud Detection Challenge
https://www.kaggle.com/c/talkingdata-adtracking-fraud-detection
Phi Huynh - 500777278

Click fraud is the act of generating fraudulent clicks on pay-per-click advertisements which results in artificially inflating web traffic and potential loss of revenue to the advertiser. TalkingData, a mobile big data platform, aims to reduce this fraud by creating blacklists for known IP addresses and devices that contribute to click fraud. TalkingData provided a dataset of around 200 million clicks spanning over 4 days for a Kaggle challenge with the aim of building an algorithm for determination of users that will download a mobile application after clicking on a relevant advertisement. The challenge proposed by TalkingData will be the focus of this project with the following research questions posed to assist with this challenge:

1. Are there particular times in the days where a user will download an application?
2. Do fraudulent clicks have any relations to the types of mobile devices used?
3. Are there particular types of an application that a user will be more likely to download?

Techniques for classification analysis will be employed on the R platform (RStudio) to tackle this challenge.