MSc. in Computing
Practicum Approval Form

Section 1: Student Details

Project Title:	Fake Ad Click Detection by identifying anomalies
Student ID:	18210455
Student name:	Gautam Shanbhag
Student email	gautam.shanbhag2@mail.dcu.ie
Chosen major:	MCM – Data Analytics
Supervisor	Mark Roantree
Date of Submission	09-Nov-2018

Section 2: About your Practicum

Click fraud is a black-hat technique of falsely inflating the number of clicks on a pay-per-click ad. Click fraud is usually driven by one of two incentives:

- Advertisers are trying to sabotage their competitors by driving up their costs and meeting their budget caps early on in the day
- Ad publishers are clicking on the ads displayed on their own sites to generate more revenue for themselves.

The topic of this practicum is **Fake Ad Click Detection by Identifying Anomalies.**

By Clustering, Classifying and Using ML Modelling, I intend to predict whether a user will download an app after clicking a mobile app advertisement by measuring the journey of a user's click across their portfolio, and to flag IP addresses who produce many clicks, but never end up installing apps. With this information, we can built an IP blacklist and device blacklist.

Section 3: Papers referred

1. Quantifying Online Advertising Fraud: Ad-Click Bots vs Humans
   https://oxford-biochron.com/downloads/OxfordBioChron_Quantifying-Online-Advertising-Fraud_Report.pdf

2. Detecting Crowdsourcing Click Fraud in Search Advertising Based on Clustering Analysis
   https://ieeexplore.ieee.org/document/7518351

3. FCFraud: Fighting Click-Fraud from the User Side
   https://ieeexplore.ieee.org/document/7423147

4. Exposing click-fraud using a burst detection algorithm

https://ieeexplore.ieee.org/document/5983854

5. Cracking The Smart ClickBot
https://www.researchgate.net/publication/221272814_Cracking_the_Smart_ClickBot

Section 4: Practicum Abstract

For this project, I intend to perform clustering analysis of the data set into informative groups depending upon various parametrical values present. Identifying anomalies to distinguish between users & bots, to identify regions, time of attack, devices used or os versions ideally preferred which are best suited for click frauds.

The data set, which I will be using, is available on Kaggle by the name TalkingData AdTracking Fraud Detection Challenge (https://www.kaggle.com/c/talkingdata-adtracking-fraud-detection#Timeline).

Techniques like Hierarchical Clustering, Centroid based clustering (K Means Clustering), Distribution based clustering, and Density based clustering or a combination of some would be shortlisted to identify and segregate the clicks by humans or bots.

Analysis would be done using Python Language.