# MACHINE LEARNING APPROACH FOR CLICK FRAUD DETECTION

**Devendrage Dilki Sasikala**

A dissertation submitted in partial fulfilment of the requirement for Bachelor of Engineering (Honours) degree in Software Engineering

**Department of Computing**

**Informatics Institute of Technology, Sri Lanka**
**in collaboration with**
**University of Westminster, UK**

**2021**

# Abstract

Machine learning has been incorporated into many industries in recent years. Capturing acts such as committing fraud has become a game-changer with the integration of machine learning. It is quick as well as efficient and can be used to solve many real-world problems. Even though machine learning can become an excellent choice for specific types of problem-solving, it is not always easy to implement. As the knowledge required to properly handle machine learning is quite extensive, experts are needed that has the knowledge whether if it's the domain in question or technology. Click fraud is one of the major issues in the digital advertisement industry. As advertisements are a factor for businesses to become successful, any form of fraudulent activity can taint that progress because this will eventually cause them to lose money. Depending on the different payout systems that are available the type of fraud may change. But at the end of the day, it can produce a major blow to the industry.

By considering the different factors such as the type of payments or the type of advertisement, there are solutions developed or being developed. Different learning models have been tested out with different sets of datasets. In this research, several of these classifiers like the Logistic Regression, KNN, Lenear Discriminant Analysis, Quadratic Discriminant Analysis and Extra trees classifier are put to the test for detecting click fraud where more than 90% performance results are acquired with the Extra Trees classifier. All the results that were produced are compared with each other to determine the classifier that has the potential to be further developed into an accurate fraud detection model.



**Keywords -** Click fraud, Logistic Regression, KNN, Lenear Discriminant Analysis, Quadratic Discriminant Analysis, Extra trees classifier