



**INFORMATICS  
INSTITUTE OF  
TECHNOLOGY**

**INFORMATICS INSTITUTE OF TECHNOLOGY**

In Collaboration with

**UNIVERSITY OF WESTMINSTER**

## **Detectic**

Automobile Insurance Claim Fraud Detection using  
Machine Learning Techniques.

Final Project Report by

Mr. Savidya Perera

Supervised by

Mr. Torin Wirasingha

Submitted in partial fulfillment of the requirements for the BEng Software Engineering  
degree at the University of Westminster.

**May 2022.**

## Abstract

As of today, Automobile insurance companies currently use a manual process to investigate fraudulent claims. Recent studies have shown that manual process has failed, and the companies are losing billions of profit every year. The manual investigation involves a lot of people with expertise and money too. The investigation can be time-consuming for the insurance company and the insured and can affect the loyal customer base when a legitimate claim has been investigated as fraud. Even if the investigation is done thoroughly, the accuracy of the process is very low.

The research goal is to design and develop an Automobile Claim Fraud Detection System to address the problem to improve the accuracy and simplify fraud investigations for both customers and insurers.

The proposed solution has been done using the best available machine learning techniques reviewed in the literature. An ensemble learning technique has been utilized to combine the results of 4 selected classification algorithms to achieve a better performance than each individual algorithm. As discussed in the literature review, XG Boost, Cat Boost, Ada Boost, and Logistic Regression algorithms were chosen as the most performing classifiers with a voting classifier, a well-known ensemble technique used.

Random Oversampling has been used to solve the skewed data problem, and the Chi-squared test has been used to identify the best subset of features to get a meaningful claim fraud prediction model, which has been an untouched part in recent years.

The proposed system was tested using several metrics to cover all areas of the study. Domain experts assessed the research's scope and importance, while technical experts assessed the system's design and implementation. According to the evaluation results, the proposed system outperformed the existing systems significantly in accuracy and recall metrics.

**Keywords** – Machine Learning, Automobile Insurance, Ensemble Learning, Classification, Fraud Detection, Feature Selection