Maestría en

Ciencia de Datos y Máquinas de Aprendizaje con mención en Inteligencia Artificial

Trabajo previo a la obtención de título de Magister en Ciencia de Datos y Máquinas de Aprendizaje con mención en Inteligencia Artificial

AUTORES:

José Alberto Rivadeneira Romero

Edwin Santiago Remache Males

Jhonatan Rafael Valencia Ramírez

TUTORES:

Karla Estefanía Mora Cajas

Pablo Renato Escobar Vallejos

Sistema para la detección de fraude Bypass en operadores de telefonía móvil del Ecuador mediante un análisis comparativo de modelos supervisados de aprendizaje automático utilizando registros CDR

Quito - Ecuador

2026

REINVENTEMOS EL FUTURO

RESUMEN

El fraude tipo Bypass constituye una de las modalidades de fraude en telecomunicaciones con mayor impacto económico a nivel mundial, y en Ecuador afecta tanto a los operadores autorizados como al Estado por la evasión de ingresos y obligaciones tributarias. Este trabajo desarrolla y evalúa un sistema de detección automática de fraude Bypass a partir de registros de llamadas de voz CDR de un operador de telefonía móvil ecuatoriano, mediante un análisis comparativo de modelos supervisados de aprendizaje automático. A partir de un gran volumen de registros correspondientes a varios períodos consecutivos, se construyó una matriz de variables de comportamiento por número que captura patrones que son característicos de este tipo de fraude. Se entrenaron y compararon distintos modelos de clasificación bajo varios escenarios operativos y una función de costo asimétrica para posteriormente evaluar su desempeño sobre un conjunto de prueba ciega que simula condiciones reales de producción. El modelo Ensemble que resulta de combinar los algoritmos de mejor desempeño individual, obtuvo los mejores resultados según los criterios establecidos y fue integrado en una interfaz web demostrativa que permite cargar registros y visualizar las clasificaciones generadas por el sistema.

Palabras Clave: Fraude Bypass, SIMBox, registros CDR, aprendizaje automático supervisado, detección de fraude en telecomunicaciones, Gradient Boosting, Ensemble, PR-AUC, función de costo asimétrica.

ABSTRACT

Bypass fraud is one of the telecommunications fraud modalities with the greatest economic impact worldwide, and in Ecuador it affects both authorized operators and the State through the evasion of revenue and tax obligations. This work develops and evaluates a system for the automatic detection of Bypass fraud based on voice call detail records (CDR) from an Ecuadorian mobile phone operator, through a comparative analysis of supervised machine learning models. Based on a large volume of records spanning several consecutive periods, a matrix of behavioral variables per phone number was built to capture patterns that are characteristic of this type of fraud. Different classification models were trained and compared under various operating scenarios and an asymmetric cost function, in order to subsequently evaluate their performance on a blind test set that simulates real production conditions. The Ensemble model, resulting from combining the best-performing individual algorithms, achieved the best results according to the established criteria and was integrated into a demonstrative web interface that allows uploading records and viewing the classifications generated by the system.

Keywords: Bypass fraud, SIMBox, CDR records, supervised machine learning, telecom fraud detection, Gradient Boosting, Ensemble, PR-AUC, asymmetric cost function.