RESUMEN

Los delitos informáticos cada vez son más frecuentes en la actualidad, uno de dichos delitos constituye el acceso no consentido a un sistema informático el cual ha tenido un crecimiento notable de denuncias en los últimos años.

El objetivo del presente proyecto es predecir éste tipo de accesos inusuales a través del análisis de información de las pistas de auditoria de la plataforma SIIPNE3w de la Policía Nacional del Ecuador haciendo uso de herramientas de Aprendizaje Automático.

El propósito del estudio es comparar algoritmos de clasificación supervisado y no supervisado, su precisión/sensibilidad en la detección de anomalías (accesos no autorizados o inusuales) para determinar cuál otorga mejor resultado. Los algoritmos elegidos incluyen: árbol de decisiones y bosque de aislamiento. El etiquetado de datos necesario para el algoritmo supervisado se lo alcanzó utilizando clúster k-means.

Para una adecuada gestión se utilizan el estudio de caso como metodología de investigación y CRISP-DM para el desarrollo de los modelos de aprendizaje.

Como resultado se obtiene que la etiquetación de datos juega un papel importante al momento de la clasificación para los algoritmos supervisados, además que el algoritmo de bosque de aislamiento nos brinda mejor resultado de clasificación para el caso que se estudia.

Palabras clave:

- **DETECCIÓN DE INTRUSOS**
- **PISTAS DE AUDITORIA**
- **APRENDIZAJE AUTOMATICO**
- **APRENDIZAJE SUPERVISADO**

ABSTRACT

Nowadays cybercrime are more frequently every time, one of them is the informatics system unauthorized access who has been having a significant growing of complaints in the last years. The target of the project is to predict those unauthorized access through information's analysis from audit tracks produced by the SIIPNE3w platform of the Ecuador National Police using machine learning tools.

The study aim is to compare the supervised and not supervised classification algorithms, its accuracy/sensibility to detect outliers (unauthorized access) and determine which of them offer a better result. The used algorithms include: Decision Tree and Isolation Forest. Data annotation for supervised algorithm is performed with k-means clustering.

For an appropriate management of the project were used Study Case as a research methodology and CRISP-DM in the development of the machine learnings models.

As a result was obtained that annotation is an important factor for classification in supervised algorithms, also that the Isolation Forest algorithm give us a better classification result for the study case.

Key words:

- **INTRUSION DETECTION**
- **AUDIT TRAILS**
- **MACHINE LEARNING**
- **SUPERVISED LEARNING**