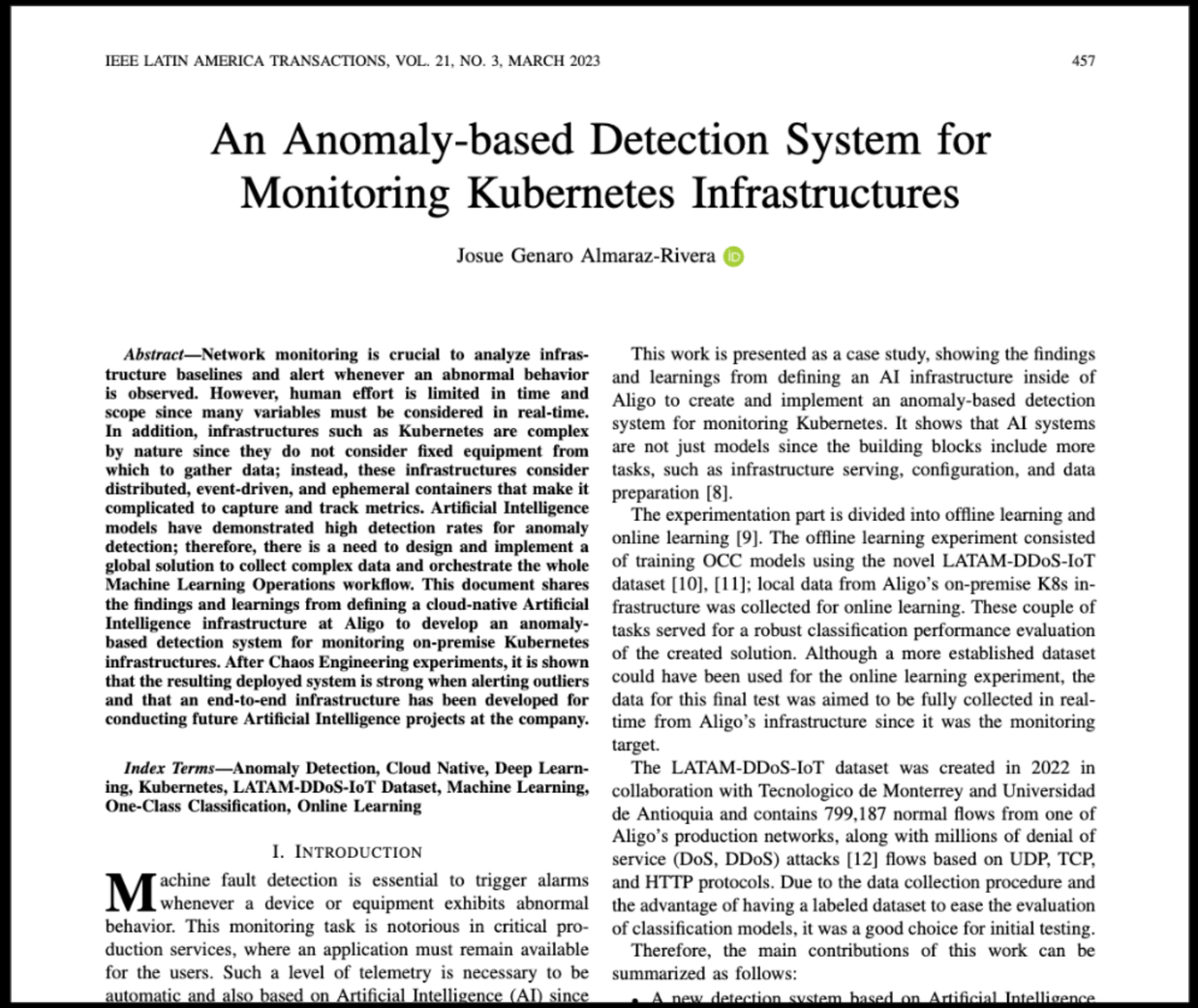
ALIGO

Informatic **Defensors**

An Anomaly-based Detection System for Monitoring Kubernetes Infrastructures

Journal paper

IEEE Latin America Transactions, March 2023
Citations: 7 (GS), 5 (Scopus)



An Anomaly-based Detection System for Monitoring Kubernetes Infrastructures

Josue Genaro Almaraz-Rivera

***Abstract*—Network monitoring is crucial to analyze infrastructure baselines and alert whenever an abnormal behavior is observed. However, human effort is limited in time and scope since many variables must be considered in real-time. In addition, infrastructures such as Kubernetes are complex by nature since they do not consider fixed equipment from which to gather data; instead, these infrastructures consider distributed, event-driven, and ephemeral containers that make it complicated to capture and track metrics. Artificial Intelligence models have demonstrated high detection rates for anomaly detection; therefore, there is a need to design and implement a global solution to collect complex data and orchestrate the whole Machine Learning Operations workflow. This document shares the findings and learnings from defining a cloud-native Artificial Intelligence infrastructure at Aligo to develop an anomaly-based detection system for monitoring on-premise Kubernetes infrastructures. After Chaos Engineering experiments, it is shown that the resulting deployed system is strong when alerting outliers and that an end-to-end infrastructure has been developed for conducting future Artificial Intelligence projects at the company.**

***Index Terms*—Anomaly Detection, Cloud Native, Deep Learning, Kubernetes, LATAM-DDoS-IoT Dataset, Machine Learning, One-Class Classification, Online Learning**

I. Introduction

Machine fault detection is essential to trigger alarms whenever a device or equipment exhibits abnormal behavior. This monitoring task is notorious in critical production services, where an application must remain available for the users. Such a level of telemetry is necessary to be automatic and also based on Artificial Intelligence (AI) since

This work is presented as a case study, showing the findings and learnings from defining an AI infrastructure inside of Aligo to create and implement an anomaly-based detection system for monitoring Kubernetes. It shows that AI systems are not just models since the building blocks include more tasks, such as infrastructure serving, configuration, and data preparation [8].

The experimentation part is divided into offline learning and online learning [9]. The offline learning experiment consisted of training OCC models using the novel LATAM-DDoS-IoT dataset [10], [11]; local data from Aligo's on-premise K8s infrastructure was collected for online learning. These couple of tasks served for a robust classification performance evaluation of the created solution. Although a more established dataset could have been used for the online learning experiment, the data for this final test was aimed to be fully collected in real-time from Aligo's infrastructure since it was the monitoring target.

The LATAM-DDoS-IoT dataset was created in 2022 in collaboration with Tecnologico de Monterrey and Universidad de Antioquia and contains 799,187 normal flows from one of Aligo's production networks, along with millions of denial of service (DoS, DDoS) attacks [12] flows based on UDP, TCP, and HTTP protocols. Due to the data collection procedure and the advantage of having a labeled dataset to ease the evaluation of classification models, it was a good choice for initial testing.

Therefore, the main contributions of this work can be summarized as follows:

- A new detection system based on Artificial Intelligence

Contributions

What did we contribute to the state of the art?

01

A new detection system based on AI models tested on the novel LATAM-DDoS-IoT dataset and local data to identify outliers on on-premise Kubernetes infrastructures.

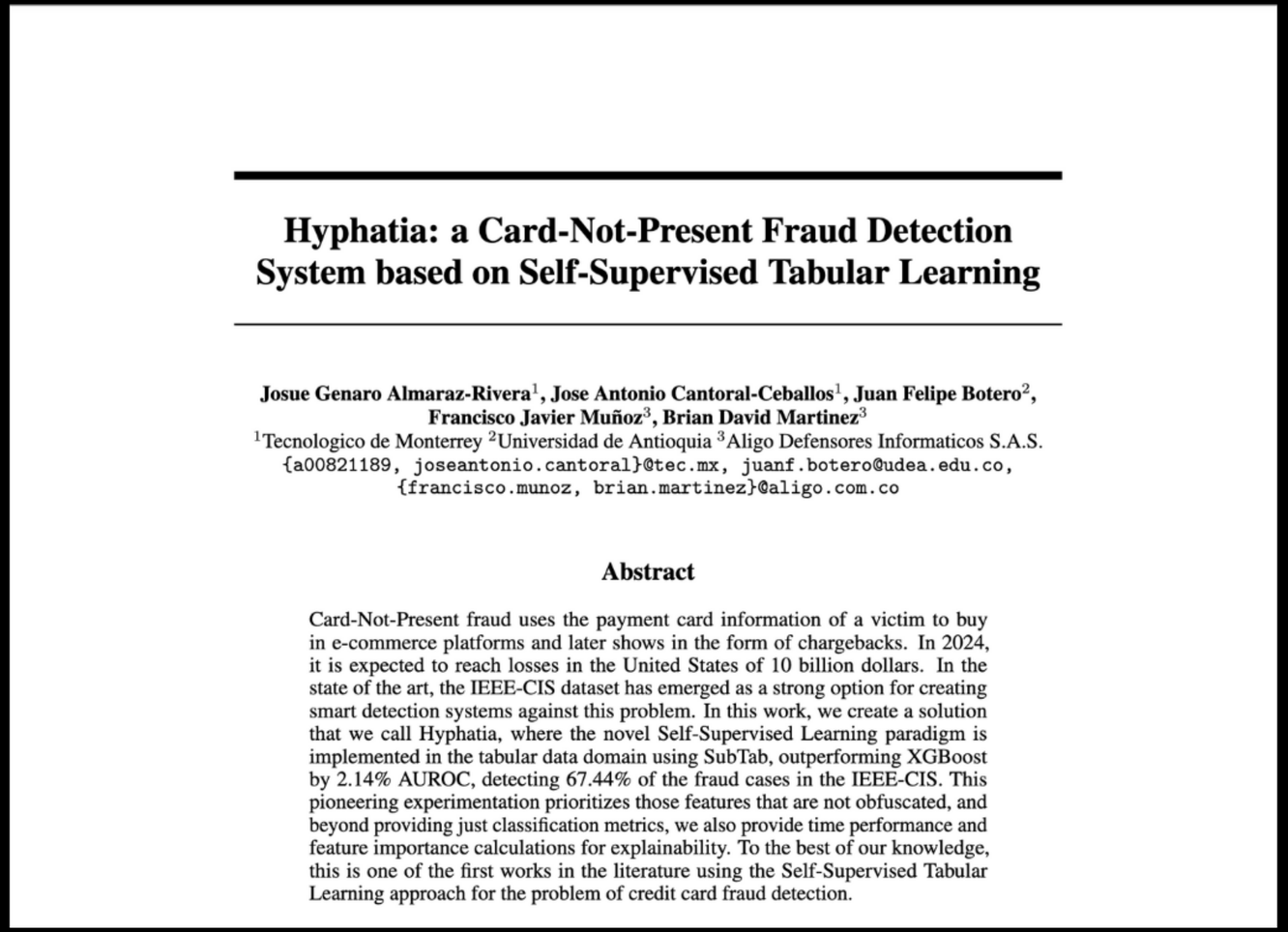
02

A tested cloud-native workflow proposal for conducting AI-based projects on Kubernetes environments.

HyphatIA
a Card-Not-Present Fraud Detection System based on Self-Supervised Tabular Learning

Conference Paper

LatinX in AI @ NeurIPS 2024



Hyphatia: a Card-Not-Present Fraud Detection System based on Self-Supervised Tabular Learning

Josue Genaro Almaraz-Rivera[1], **Jose Antonio Cantoral-Ceballos**[1], **Juan Felipe Botero**[2], **Francisco Javier Muñoz**[3], **Brian David Martinez**[3]
[1]Tecnologico de Monterrey [2]Universidad de Antioquia [3]Aligo Defensores Informaticos S.A.S.
{a00821189, joseantonio.cantoral}@tec.mx, juanf.botero@udea.edu.co,
{francisco.munoz, brian.martinez}@aligo.com.co

Abstract

Card-Not-Present fraud uses the payment card information of a victim to buy in e-commerce platforms and later shows in the form of chargebacks. In 2024, it is expected to reach losses in the United States of 10 billion dollars. In the state of the art, the IEEE-CIS dataset has emerged as a strong option for creating smart detection systems against this problem. In this work, we create a solution that we call Hyphatia, where the novel Self-Supervised Learning paradigm is implemented in the tabular data domain using SubTab, outperforming XGBoost by 2.14% AUROC, detecting 67.44% of the fraud cases in the IEEE-CIS. This pioneering experimentation prioritizes those features that are not obfuscated, and beyond providing just classification metrics, we also provide time performance and feature importance calculations for explainability. To the best of our knowledge, this is one of the first works in the literature using the Self-Supervised Tabular Learning approach for the problem of credit card fraud detection.

Contributions

What did we contribute to the state of the art?

01

Self-Supervised Tabular Learning models with performance improvements over state-of-the-art baselines including XGBoost, for the binary credit card fraud detection task, using SubTab and the IEEE-CIS dataset.

02

Non-linear evaluation head using MLP is added to the original SubTab architecture to capture more complicated boundaries in the customers' data behavior.

03

Pioneering experimentation in the S-SL and tabular learning domains, closing the existing performance gap between Machine Learning and Deep Learning supervised models, without requiring labeling of large amounts of input data.

16 years of experience

ALIGO

Contact Information

comercial@aligo.com.co
+57 310 418 78 20
Aligo - Defensores Informáticos