Departamento de Ingeniería Informática
Universidad de Santiago de Chile

DEPARTAMENTO DE INGENIERÍA INFORMÁTICA

Evaluación de la eficacia de algoritmos de clustering en la detección de anomalías en entornos de hogares inteligentes

Matías Bozo-Pizarro, Juan Iturbe-Araya (profesor guía)
matias.bozo.p@usach.cl, juan.iturbe@usach.cl
Laboratorio/Grupo de Investigación - https://csl.diinf.usach.cl/

1- Introducción y problema

- Los hogares inteligentes integran múltiples dispositivos IoT, ampliando la superficie de ataque y los riesgos de ciberseguridad [1].
- La detección de anomalías en este contexto es compleja, debido a la alta variabilidad del tráfico, la generación continua de datos y la falta de etiquetas que distingan tráfico normal y anómalo [2].
- Esto motiva el uso de algoritmos de clustering, como alternativa no supervisada para analizar tráfico IoT en escenarios con alta dimensionalidad y desbalance de clases.

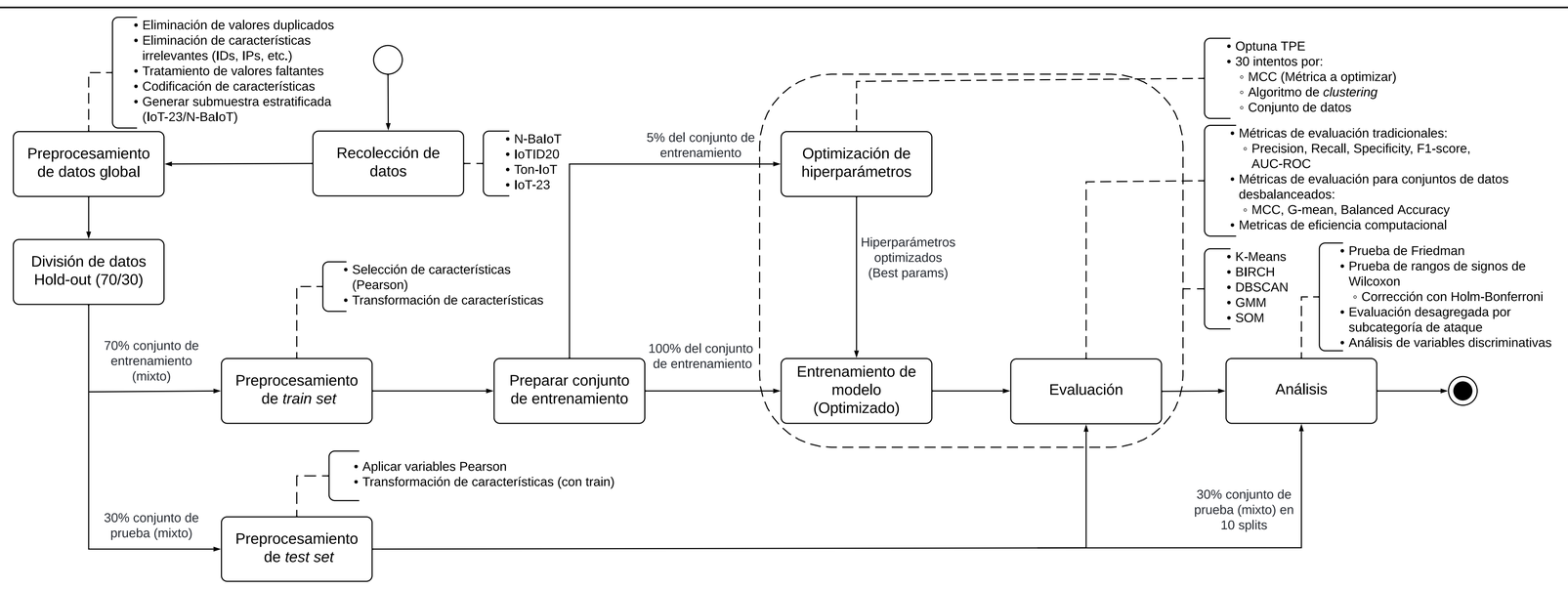
2- Objetivo y alcance

Objetivo: Evaluar la eficacia de distintos algoritmos de clustering en la detección de anomalías en datos de tráfico IoT de hogares inteligentes, comparando su desempeño en escenarios de alta dimensionalidad y datos desbalanceados.

- Evaluación sobre cuatro conjuntos de datos de tráfico IoT representativos (N-BaIoT, IoTID20, ToN-IoT e IoT-23).
- Comparación de cinco algoritmos de *clustering* (K-Means, BIRCH, GMM, DBSCAN y SOM).
- Detección de anomalías binaria: tráfico normal vs. anómalo.

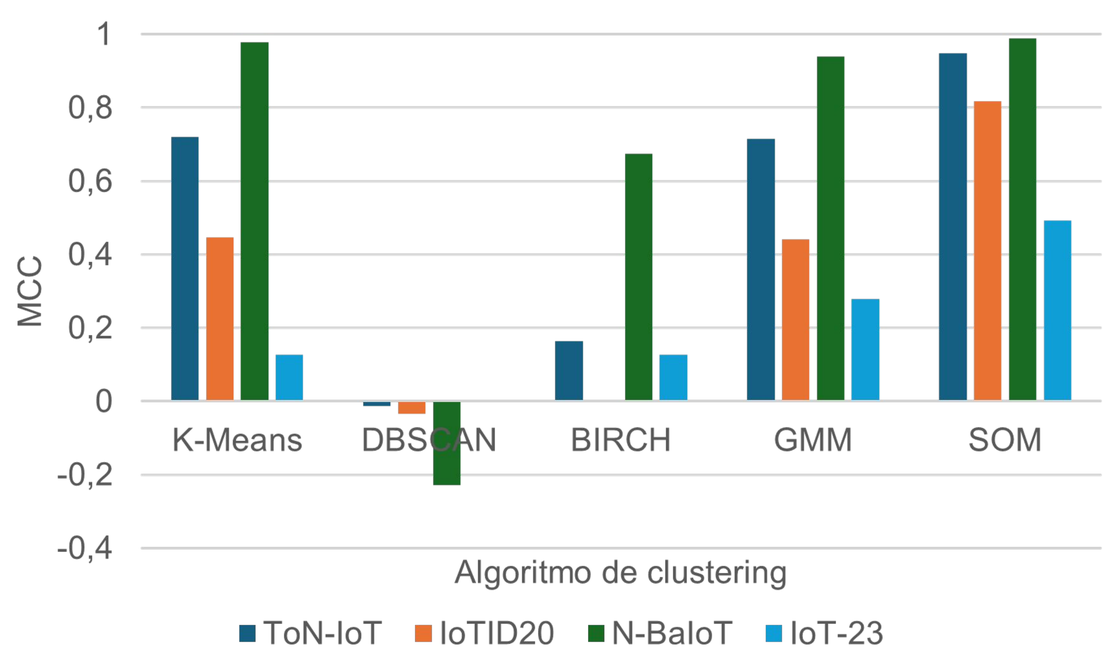
3- Metodología

- Preprocesamiento estandarizado de cuatro conjuntos de datos de tráfico IoT representativos de hogares inteligentes.
- Aplicación de cinco algoritmos de *clustering* sobre los datos procesados para identificar patrones normales y comportamientos anómalos en el tráfico de red.
- Evaluación comparativa del desempeño mediante métricas de clasificación robustas, considerando capacidad de detección y costo computacional de cada enfoque.
- Análisis comparativo de resultados mediante pruebas estadísticas y técnicas *post-hoc*.



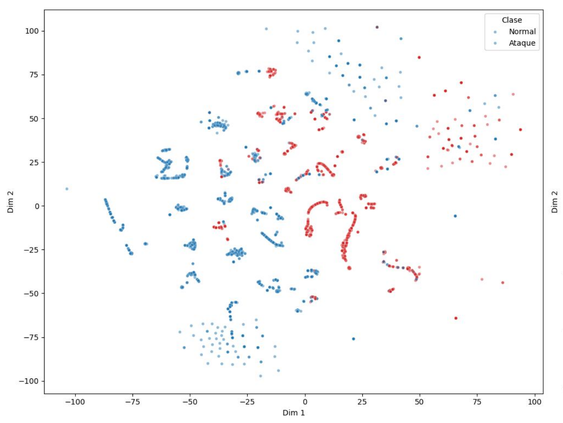
Resumen conceptual de metodología experimental aplicada

4- Resultados y análisis

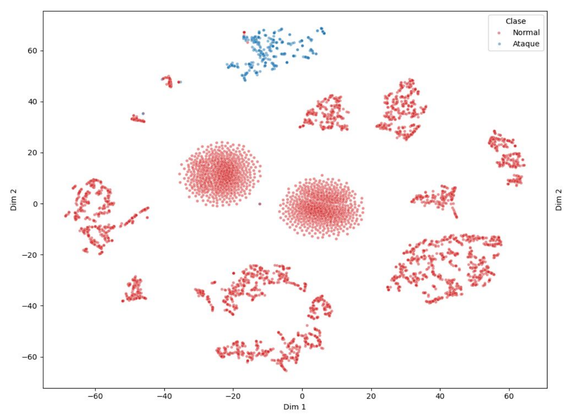


Comparación del desempeño de los algoritmos de clustering mediante el coeficiente de correlación de Matthews (MCC) en los conjuntos de datos evaluados

- **SOM obtuvo el mejor desempeño global**, destacando en ToN-IoT y N-BaIoT con alto MCC. Las pruebas de Friedman y Wilcoxon confirmaron la superioridad estadística de SOM frente a los otros algoritmos de clustering.
- Visualizaciones t-SNE sugieren que **los solapamientos dificultan la separación en escenarios altamente desbalanceados**.
- **IoT-23 fue el conjunto más desafiante**, con una caída general del desempeño debido a su mayor complejidad y desbalance.



Visualización del espacio de datos mediante t-SNE para el conjunto de datos ToN-IoT



Visualización del espacio de datos mediante t-SNE para el conjunto de datos N-BaIoT

5- Conclusiones

- SOM fue el algoritmo con mejor desempeño global, mostrando mayor robustez y mejor equilibrio entre sensibilidad y especificidad en los escenarios evaluados.
- La elección del algoritmo de clustering implica un compromiso entre precisión y costo computacional.
- La eficacia del algoritmo de clustering depende principalmente de la estructura del tráfico IoT, más que de la complejidad del algoritmo utilizado.

6- Trabajo futuro

- Evaluar el enfoque en nuevos conjuntos de datos IoT representativos de hogares inteligentes
- Validar el desempeño de algoritmos de clustering en dispositivos IoT de bajo consumo
- Ampliar la detección de anomalías desde un esquema binario a uno multiclase

Referencias

[1] Saadouni, R., Gherbi, C., Aliouat, Z., Harbi, Y., & Khacha, A. (2024). Intrusion detection systems for IoT based on bio-inspired and machine learning techniques: a systematic review of the literature. Cluster Computing, 27(7), 8655–8681. https://doi.org/10.1007/s10586-024-04388-5.
[2] Jalali, N., Sahu, K. S., Oetomo, A., & Morita, P. P. (2020). Understanding user behavior through the use of unsupervised anomaly detection: proof of concept using Internet of Things smart home thermostat data for improving public health surveillance. JMIR mHealth and uHealth, 8(11), e21209. https://doi.org/10.2196/21209