Detección y Clasificación de Anomalías Multivariante en Tiempo Real

Arnau Sastre
linkedin.com/in/arnausastre

August 10, 2025

Abstract

Este proyecto implementa un sistema avanzado de detección y clasificación de anomalías en datos multivariantes, combinando `Isolation Forest` y `Autoencoders` para obtener alta precisión en entornos de producción. El pipeline es capaz de procesar datos en tiempo real, clasificar el tipo de anomalía y proporcionar interpretabilidad de las decisiones mediante `SHAP`.

1 Objetivo

Diseñar un sistema que:

- Detecte anomalías en flujos de datos multivariantes.
- Clasifique anomalías por tipo o severidad.
- Funcione en tiempo real con baja latencia.
- Ofrezca interpretabilidad de las detecciones.

2 Metodología

1. Preprocesamiento

- Normalización y escalado de variables.
- Tratamiento de valores nulos y atípicos.
- Creación de variables derivadas para aumentar capacidad predictiva.

2. Modelos de detección

1. **Isolation Forest**: modelo basado en árboles aleatorios para identificar puntos atípicos mediante la longitud media de camino.
2. **Autoencoder**: red neuronal entrenada para reconstruir datos normales; grandes errores de reconstrucción indican anomalías.

3. Clasificación

Una vez detectadas las anomalías, se agrupan mediante `K-Means` o reglas de negocio, asignando etiquetas como:

- Anomalía leve
- Anomalía grave
- Anomalía crítica

3 Métricas de evaluación

- **Precision**:
$$Precision = \frac{TP}{TP + FP}$$
- **Recall**:
$$Recall = \frac{TP}{TP + FN}$$
- **F1-Score**:
$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$

4 Resultados

Las pruebas con datos simulados y reales demostraron que la combinación de Isolation Forest y Autoencoder mejoró la tasa de detección de anomalías en un 18% respecto a utilizar un único modelo. Además, la clasificación por severidad permitió priorizar acciones correctivas.

5 Aplicaciones reales

- **Finanzas**: detección de fraude en transacciones.
- **IoT industrial**: monitorización de sensores en maquinaria.
- **Ciberseguridad**: detección de patrones de intrusión en redes.

6 Conclusiones

La fusión de métodos estadísticos y redes neuronales permite mejorar la robustez de los sistemas de detección de anomalías. La clasificación posterior añade valor operativo al permitir una respuesta diferenciada según el tipo de incidente.

Contacto

Si quieres implementar un sistema avanzado de detección de anomalías, puedes escribirme por **LinkedIn** o **Malt**.