Detección Distribuida de Anomalías Transaccionales en SAP Business One para Auditoría Minera mediante Técnicas Híbridas de Big Data y Machine Learning

Universidad Nacional de Ingeniería – Maestría en Inteligencia Artificial
Curso: Big Data
Modalidad propuesta: Artículo científico / Informe técnico

Este proyecto propone el diseño e implementación de una arquitectura Big Data para la detección inteligente de anomalías transaccionales sobre datos históricos reales extraídos de SAP Business One de una empresa minera peruana.

Problema

La auditoría manual sobre sistemas ERP empresariales presenta limitaciones importantes frente al crecimiento del volumen transaccional y la complejidad operacional de las empresas mineras. Detectar pagos duplicados, transacciones fraccionadas, movimientos sospechosos y comportamientos anómalos mediante reglas tradicionales resulta costoso y poco escalable.

Objetivo General

Desarrollar una arquitectura Big Data capaz de procesar y analizar datos históricos de SAP Business One utilizando técnicas híbridas de Machine Learning para identificar anomalías financieras y operativas.

Dataset

• Base de datos SAP Business One SQL Server.
• Aproximadamente 30 GB de información histórica.
• Período: 2017 – 2026.
• Tablas principales: OINV, OPCH, OPOR, OJDT, OCRD, OITM, INV1, PCH1, entre otras.

Arquitectura Propuesta

SQL Server SAP B1 → PySpark → Amazon S3 (Raw/Curated) → Procesamiento Spark → Modelos ML → DynamoDB → Tableau Dashboard.

Tecnologías Utilizadas

• Apache Spark (PySpark)
• Amazon EC2
• Amazon S3
• DynamoDB
• Tableau

• TensorFlow / Keras
• Scikit-learn
• SQL Server

Modelos de Detección

1. Ley de Benford para análisis estadístico.
2. Isolation Forest para detección no supervisada.
3. Autoencoder basado en Deep Learning.
4. Ensemble híbrido para priorización de anomalías.

Anomalías Objetivo

• Pagos duplicados.
• Fraccionamiento de órdenes de compra.
• Backdating de documentos.
• Movimientos financieros atípicos.
• Activos fantasma o inconsistencias operativas.

Estrategia de Anonimización

Los datos sensibles serán protegidos mediante hashing SHA-256 sobre identificadores críticos y técnicas de preservación distribucional para ocultar montos reales manteniendo propiedades estadísticas.

Resultados Esperados

• Comparación de métricas Precision, Recall y F1-score.
• Benchmark de rendimiento Spark.
• Dashboard interactivo de auditoría.
• Pipeline Big Data reproducible.
• Identificación automatizada de anomalías empresariales.

Valor Diferencial

El principal valor del proyecto es el uso de datos empresariales reales provenientes de un ERP minero peruano, permitiendo construir una solución aplicada, escalable y alineada con escenarios reales de auditoría financiera y analítica empresarial.

Stack Tecnológico

Capa	Tecnología	Función
Storage	Amazon S3	Raw y Curated Zone
Compute	EC2 + PySpark	Procesamiento paralelo

NoSQL	DynamoDB	Top-K anomalías
Visualización	Tableau	Dashboard interactivo
Origen	SAP Business One	ERP empresarial

Conclusión

El proyecto busca demostrar cómo una arquitectura Big Data bien diseñada puede mejorar significativamente los procesos de auditoría empresarial mediante técnicas modernas de procesamiento distribuido y Machine Learning aplicado sobre datos ERP reales.