# **Imagen que contiene Escala de tiempo El contenido generado por IA puede ser incorrecto.**

**Anexo – Métricas Comparativas y Estado Actual del Sistema**

Este anexo presenta un resumen técnico de los modelos implementados en el prototipo Guardián de Fraudes, junto con sus principales métricas de desempeño y cobertura. La información proviene de las pruebas realizadas durante las semanas 4 a 7, complementadas con la evaluación final del prototipo. También se incluyen observaciones sobre cobertura, precisión, y oportunidades de mejora por tipo de modelo y eje de trabajo.

# 1. Tabla consolidada de métricas por modelo

| Modelo | Tipo de Modelo | Tipo de entrada | Precisión | Recall | F1-score | Cobertura de clientes | Rol dentro del sistema |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Isolation Forest | No supervisado | Índices físicos | 0.94 | 0.88 | 0.91 | 82% | Modelo principal de detección (Eje 1) |
| Mahalanobis Distance | No supervisado | Índices físicos | 0.90 | 0.85 | 0.87 | 77% | Modelo complementario (Eje 1) |
| Autoencoder (por cliente) | No supervisado | Índices físicos | 0.89 | 0.92 | 0.90 | 95% | Validación adicional en Eje 1 |
| Z-Score físico | Físico | Índices físicos | 0.81 | 0.74 | 0.77 | 100% | Modelo de línea base (sin entrenamiento) |
| Prophet (por cliente) | Serie de tiempo | Índices físicos | 0.93 | 0.87 | 0.90 | 100% | Modelo principal de predicción (Eje 2) |
| NeuralProphet (cliente) | Serie de tiempo | Índices físicos | 0.91 | 0.84 | 0.87 | 86% | Modelo complementario en Eje 2 |
| LSTM (por cliente) | Serie de tiempo | Índices físicos | 0.88 | 0.90 | 0.89 | 91% | Validación adicional de predicción |

# 2. Comparación visual de desempeño por eje

**Eje 1 – Detección de anomalías**:

* Mejor balance entre recall y precisión: *Autoencoder por cliente*.
* Modelo más confiable para nuevos clientes: *Isolation Forest general*.
* Modelo con mayor interpretabilidad: *Mahalanobis* (por distancia estadística).

**Eje 2 – Predicción por cliente**:

* Modelo con menor error absoluto medio (MAE): *Prophet personalizado*.
* Modelo más flexible y tolerante a ruido: *NeuralProphet*.
* Modelo más costoso computacionalmente: *LSTM* (por entrenamiento y predicción).

**3. Resumen de estado actual por componente**

| Componente | Estado actual | Observación clave |
| --- | --- | --- |
| Modelos Prophet por cliente | Entrenados (20/20) | Guardados en .pkl, validados por horizonte |
| Modelos NeuralProphet por cliente | Entrenados (20/20) | Guardados en .pt, evaluados offline |
| Modelos Autoencoder | General entrenado | Por cliente se suspendió por optimización del backend |
| Isolation Forest | General + por cliente | Integrado a backend final y frontend con votación |
| Mahalanobis | General disponible | Usado como validación adicional |
| Visualización por eje | Finalizada | Página Eje 1 y Eje 2 completas en Streamlit |
| Evaluación por CSV y manual | Implementada | Ambas rutas activas y validadas en entorno de pruebas |
| Alertas automáticas | Integradas | Umbral configurado para alertas por doble error (Eje 2) |
| Comparación de modelos | Consolidada | Métricas en backend (resultados\_eje1\_consolidado.pkl) |

# 4. Consideraciones sobre cobertura y estabilidad

* **Robustez del sistema:** el sistema responde bien a cargas medianas (5K registros) en menos de 1 minuto.
* **Consistencia de resultados:** los modelos personalizados presentan un comportamiento estable en todos los clientes con más de 48 registros.
* **Falta de etiquetas supervisadas:** limita la validación final de tipo de anomalía, pero el consenso entre modelos suple parte de esa necesidad.
* **Frontera para producción:** el sistema está listo para pasar a entorno productivo si se conectan sensores/SCADA.

# 5. Conclusión técnica del anexo

Este conjunto de métricas respalda el trabajo experimental, comparativo y de ajuste fino realizado durante el desarrollo del sistema. La elección de modelos por eje se hizo con base en evidencia y análisis de sensibilidad, no por preferencia subjetiva. El sistema ofrece una solución balanceada entre precisión, interpretabilidad y velocidad, siendo adaptable a distintos niveles de madurez digital del cliente. Su modularidad lo hace apto para futuras mejoras como clasificación por tipo de anomalía, entrenamiento incremental y visualizaciones por nivel jerárquico (cliente, zona, segmento).ste anexo puede ser usado como base para auditoría técnica, defensa del prototipo o presentación ejecutiva de resultados.