

**Clasificación supervisada de eventos raros para anticipar roturas del papel: comparación
de estrategias de selección de variables y ajuste de modelos**

Jhonatan Alfonso Patiño

Asesor

Sebastián Vélez Jaramillo

Universidad Nacional Abierta y a Distancia UNAD
Escuela de Ciencias Básicas, Tecnología e Ingeniería ECBTI
Especialización en Ciencia de Datos y Analítica

2025

Resumen

Este proyecto aborda la predicción temprana de roturas del papel en una línea de producción como un problema de clasificación supervisada con clase positiva rara. Se construye un conjunto de datos multivariado con variables de proceso estandarizadas y un esquema de entrenamiento, validación y prueba que respeta el orden temporal y evita la fuga de información. Para obtener un conjunto de predictores parsimonioso y con baja colinealidad se comparan dos enfoques de selección de variables: (A) filtrado univariado con correlación de Spearman y umbrales de baja correlación entre predictores; y (B) Elastic Net con búsqueda de hiperparámetros guiada por PR-AUC, seguido de Forward Selection controlando redundancia. Con el conjunto seleccionado se entrenan y comparan modelos de clasificación como regresión logística penalizada, Random Forest, máquinas de vectores de soporte (SVM) y XGBoost, empleando ponderación de clases y ajuste de hiperparámetros. Dado el desbalance severo, la métrica principal de evaluación es PR-AUC, complementada con ROC-AUC, precisión, recall, F1, matrices de confusión y medidas de lift para escenarios operativos. Los resultados permiten seleccionar un modelo candidato para la anticipación de roturas y discutir su viabilidad operativa, así como delinear recomendaciones metodológicas y líneas de trabajo futuro.

Palabras clave: Clasificación, roturas, desbalance, variables, PR-AUC.

Abstract

This project addresses the early prediction of paper breaks on a production line as a supervised classification problem with a rare positive class. A multivariate data set is constructed using standardized process variables and a train-validation-test scheme that respects temporal order and avoids information leakage. To obtain a parsimonious predictor set with low collinearity, two variable selection approaches are compared: (A) a univariate filter based on Spearman correlation and low-correlation thresholds between predictors, and (B) Elastic Net with hyperparameter search guided by PR-AUC, followed by Forward Selection with redundancy control. Using the selected set of predictors, classification models such as penalized logistic regression, Random Forest, support vector machines (SVM) and XGBoost are trained and compared, employing class weighting and hyperparameter tuning. Due to the severe class imbalance, the primary evaluation metric is PR-AUC, complemented by ROC-AUC, precision, recall, F1, confusion matrices, and lift measures for operational scenarios. The results enable the selection of a candidate model for anticipating paper breaks and the discussion of its operational feasibility, as well as outlining methodological recommendations and future work.

Keywords: Classification, breaks, imbalance, features, PR-AUC.

Tabla de Contenido

Introducción	11
Justificación	12
Objetivos.....	14
Objetivo General	14
Objetivos Específicos.....	14
Alcance y Limitaciones.....	15
Estructura del Documento.....	17
Marco Teórico.....	19
Aprendizaje Supervisado y Problemas de Clasificación.....	19
Desbalance de Clases y Métricas Adecuadas.....	20
Regularización y Selección de Variables	22
Modelos de Clasificación Utilizados	26
Trabajos Relacionados	29
Metodología	33
Descripción del Conjunto de Datos	33
Diseño del Conjunto de Datos para Clasificación Supervisada	35
Definición de la Etiqueta Anticipada	35
Selección de Predictores.....	36
Partición Temporal en Entrenamiento, Validación y Prueba	36
Limpieza y Preprocesamiento de los Sensores	37
Control de Calidad y Consistencia Básica	38
Tratamiento de Valores Perdidos	38

Tratamiento de Valores Atípicos (Winsorización).....	39
Estandarización y Escalado de Variables.....	39
Ingeniería Temporal de Variables.....	40
Implementación como Pipeline Reproducible	41
Selección de Variables	42
Método A Filtro Basado en Correlación de Spearman y Control de Colinealidad	42
Método B Elastic Net más Selección Hacia Adelante (Forward Selection)	43
Filtrado Embebido Mediante Regresión Logística con Elastic Net.	43
Refinamiento Mediante Selección Hacia Adelante (Forward Selection).	44
Conjuntos Resultantes y Uso en los Experimentos.....	44
Evaluación Comparativa de los Conjuntos A y B y Selección del Conjunto Final	45
Diseño de los Experimentos de Modelado.....	46
Modelos Considerados	46
Tratamiento del Desbalance de Clases.....	47
Ajuste de Hiperparámetros.....	48
Protocolo de Entrenamiento y Evaluación	49
Aplicación del Pipeline de Datos.....	49
Entrenamiento y Ajuste de Hiperparámetros.....	49
Entrenamiento Final y Evaluación en Prueba.....	50
Comparación Entre Algoritmos.....	50
Criterios de Selección del Modelo Final e Interpretación.....	50
Desempeño Predictivo.....	50
Viabilidad Operativa.....	51

Resultados	52
Resultados de la Selección de Variables	52
Ranking de Spearman (Método A).....	53
Variables Seleccionadas por el Método A	54
Comparación A vs B con RL L2 de Referencia.....	58
Resultados de los Modelos de Clasificación.....	59
Comparación y Selección del Modelo Final	59
Análisis de Errores y Comportamiento Operativo	64
Discusión.....	67
Interpretación de Variables Relevantes.....	67
Relación con la Literatura y el Marco Teórico	68
Implicaciones Prácticas en la Planta	73
Conclusiones y Recomendaciones	75
Conclusiones	75
Construcción y Depuración del Conjunto de Datos	75
Selección de un Conjunto Parsimonioso de Variables	75
Comportamiento de los Modelos de Clasificación Bajo Desbalance Severo	76
Capacidad Anticipativa y Utilidad Operativa del Modelo Final.....	76
Interpretabilidad y Coherencia con el Conocimiento de Proceso	76
Aporte Metodológico y Limitaciones.....	77
Comparación con la Literatura	77
Recomendaciones.....	78
Profundizar en la Ingeniería de Variables y en la Etiqueta	78

Reforzar la Validación Temporal y la Estimación de Incertidumbre.....	79
Analizar Diferentes Políticas de Umbral en Función del Contexto Operativo	79
Integrar el Modelo como Herramienta de Apoyo, no como Reemplazo del Criterio Experto	80
Ampliar y Actualizar el Dataset con Datos Adicionales y Nuevas Variables.....	80
Explorar Modelos Alternativos y Técnicas Avanzadas de Desbalance Cuando la Disponibilidad de Datos lo Permita.....	81
Explorar Estrategias de Enriquecimiento de Datos y Modelos más Complejos	82
Vincular la Evaluación del Modelo con el Impacto Económico de las Roturas	82
Referencias Bibliográficas	83
Apéndices.....	88

Lista de Tablas

Tabla 1 <i>Distribución de la Etiqueta Anticipada por Partición</i>	52
Tabla 2 <i>Ranking de Spearman (Método A)</i>	53
Tabla 3 <i>Variables Seleccionadas por el Método A</i>	55
Tabla 4 <i>Orden de Entrada y PR-AUC Valid Acumulada</i>	57
Tabla 5 <i>Comparativo de los Resultados de los Conjuntos A y B</i>	59
Tabla 6 <i>Comparación de Modelos de Clasificación (conjunto B)</i>	59
Tabla 7 <i>Comparación de Resultados con Trabajos Previos Sobre</i>	71

Lista de Figuras

Figura 1 <i>Top 15 de Variables Seleccionadas, Mediante la Correlación Absoluta de Spearman con (y_{ant})</i>	54
Figura 2 <i>Mapa de Calor de (ρ) Entre las Variables del Conjunto A en el Conjunto de Entrenamiento</i>	55
Figura 3 <i>Mapa de Calor de (ρ) Entre las Variables del Conjunto B en el Conjunto de Entrenamiento</i>	56
Figura 4 <i>Evolución de PR-AUC con las Variables del Conjunto B</i>	57
Figura 5 <i>Correlación Absoluta de Spearman Entre las Variables de los Conjuntos A y B</i>	58
Figura 6 <i>Mejores Modelos por Familia, Métricas Normalizadas</i>	61
Figura 7 <i>Curvas Precision–Recall del Modelo Logístico L2 Seleccionado en Validación y Prueba</i>	63
Figura 8 <i>Curvas ROC del Modelo Logístico L2 Seleccionado en Validación y Prueba</i>	63
Figura 9 <i>Matriz de Confusión del Modelo Logístico L2 en el Conjunto de Prueba</i>	64
Figura 10 <i>Principales Coeficientes Estandarizados del Modelo Logístico L2 Seleccionando el Conjunto B</i>	65

Lista de Apéndices

Apéndice A <i>Coefficientes Completos del Modelo Logístico L2 Sobre el Conjunto B</i>	88
Apéndice B <i>Resultados de los Modelos Según Estrategia de Re-Muestreo y Ponderación de Clases</i>	88

Introducción

La rotura del papel en una línea de producción genera pérdidas de material, reprocesos y tiempos muertos. Anticipar estos eventos permite accionar controles preventivos y mejorar la eficiencia. Más que plantear un pronóstico univariado de series de tiempo, la probabilidad de rotura depende de múltiples variables de proceso y la salida es binaria (rotura/no rotura); por ello, en este trabajo el problema se aborda como una tarea de clasificación supervisada.

Para el desarrollo empírico no se utilizaron datos privados de una organización, sino el conjunto de datos público ProcessMiner (2020), que corresponde a un proceso real de fabricación de papel y constituye un benchmark ampliamente utilizado en la literatura para el estudio de roturas en máquinas papeleras. El problema presenta tres retos principales: desbalance severo (poca ocurrencia de roturas), lo que exige métricas y técnicas específicas; colinealidad entre variables, que puede inflar la varianza de los estimadores y dificultar la interpretación; y necesidad de parsimonia, para favorecer modelos simples y operables en planta.

En este contexto, se construye un flujo de trabajo que primero depura y transforma el conjunto de datos multivariado, luego selecciona un subconjunto parsimonioso de variables mediante dos rutas complementarias (filtro basado en correlación de Spearman y un enfoque Elastic Net + selección hacia adelante) y finalmente entrena y compara modelos de clasificación como regresión logística penalizada, Random Forest, máquinas de vectores de soporte (SVM) y XGBoost, empleando ponderación de clases y ajuste de hiperparámetros.

La evaluación se centra en la métrica PR-AUC, más informativa bajo desbalance severo, complementada con ROC-AUC, precisión, recall, F1, matrices de confusión y medidas de lift para este tipo de problemas. A partir de estos resultados, el trabajo selecciona un modelo final y discute su viabilidad operativa como apoyo a la toma de decisiones en la línea de producción.

Justificación

Las roturas del papel en una línea de producción generan paradas no planificadas, desperdicio de materia prima, reprocesos y pérdida de productividad. Aun una reducción modesta en la frecuencia de estos eventos puede traducirse en ahorros relevantes en costos directos (fibra, energía, químicos) y en una mejora apreciable de la estabilidad operativa. Sin embargo, en la práctica las roturas son eventos poco frecuentes y muchas veces se originan en combinaciones sutiles de variables de proceso que no se detectan con reglas simples o alarmas univariadas.

En este sentido, la planta dispone de un volumen considerable de datos históricos provenientes de sensores y sistemas de control, que rara vez se explotan de forma sistemática para anticipar fallas. Un modelo de clasificación supervisada, construido sobre estos datos y entrenado con buenas prácticas analíticas, ofrece la posibilidad de transformar registros históricos en una herramienta concreta de apoyo a la decisión: alertas anticipadas basadas en patrones multivariados, con umbrales calibrados y medibles en términos de falsos positivos y falsos negativos. Esto permite pasar de alarmas empíricas a un esquema de monitoreo más objetivo y reproducible.

Desde el punto de vista académico y metodológico, el problema plantea desafíos relevantes: se trata de un evento raro, con fuerte desbalance de clases, colinealidad entre variables y necesidad de parsimonia para lograr modelos interpretables y utilizables en planta. Desarrollar y documentar un flujo de trabajo que aborde estos retos desde la preparación del conjunto de datos hasta la selección de variables y la comparación de modelos contribuye a consolidar capacidades en ciencia de datos aplicada al contexto industrial. En conjunto, el proyecto se justifica tanto por su impacto potencial en la eficiencia operativa como por el aporte

metodológico en la aplicación de técnicas de clasificación supervisada a problemas de eventos raros. El uso del conjunto de datos ProcessMiner permite trabajar sobre un caso industrial realista, con acceso completo a las variables de proceso y a un historial de roturas, sin comprometer la confidencialidad de la información interna de alguna organización.

Para la Especialización en Ciencia de Datos y Analítica de la UNAD, este proyecto representa una aplicación concreta de técnicas de aprendizaje supervisado, ingeniería de características temporales y manejo de desbalance de clases a un problema industrial realista, alineado con las competencias del programa en términos de modelamiento predictivo y soporte cuantitativo a la toma de decisiones.

Objetivos

Objetivo General

Desarrollar y validar un procedimiento de clasificación supervisada para anticipar eventos de rotura del papel, obteniendo un conjunto parsimonioso de variables y evaluando modelos con métricas robustas al desbalance, a fin de seleccionar una alternativa con potencial de uso operativo.

Objetivos Específicos

Construir y depurar el conjunto de datos de proceso, incluyendo estandarización, control de calidad y partición temporal en conjuntos de entrenamiento, validación y prueba.

Seleccionar un conjunto parsimonioso de variables comparando enfoques de selección basados en filtros univariados y modelos penalizados (Elastic Net), controlando la colinealidad entre predictores.

Entrenar, ajustar y comparar modelos de clasificación como regresión logística penalizada, Random Forest, SVM y XGBoost con tratamiento del desbalance, seleccionar el modelo con mejor desempeño según PR-AUC y métricas complementarias, e interpretar las variables más relevantes para proponer lineamientos de uso operativo.

Alcance y Limitaciones

El alcance de este proyecto se circunscribe al desarrollo y evaluación de modelos de clasificación supervisada para anticipar roturas del papel a partir de variables de proceso medidas en una línea de producción. El trabajo abarca la construcción y depuración de un conjunto de datos multivariado, la definición de un horizonte fijo de anticipación de la rotura, la partición temporal en conjuntos de entrenamiento, validación y prueba, la selección de un subconjunto parsimonioso de variables y el entrenamiento y comparación de varios modelos de clasificación, con especial atención al tratamiento del desbalance de clases y al uso de métricas adecuadas. Los resultados obtenidos se utilizan para identificar un modelo candidato y derivar lineamientos de uso operativo y recomendaciones metodológicas.

El estudio se centra en la información disponible en el conjunto de datos analizado y en un conjunto acotado de modelos de aprendizaje supervisado (regresión logística penalizada, Random Forest, SVM y XGBoost), seleccionados por su combinación de desempeño, robustez e interpretabilidad. El análisis se realiza en un entorno offline, empleando datos históricos; no se implementa ni se valida un sistema de alarmas en línea, sino que se evalúa la viabilidad de dicho sistema a partir de los resultados del modelo y de su comportamiento en los distintos conjuntos de datos.

Entre las principales limitaciones se encuentran las restricciones propias de los datos y del diseño experimental. En primer lugar, el conjunto de datos corresponde a una línea de producción específica y a un período de tiempo acotado, de modo que los resultados reflejan las condiciones de operación observadas y no necesariamente son generalizables a otras máquinas, plantas o contextos sin una validación adicional. En segundo lugar, las roturas constituyen eventos raros, por lo que la señal disponible para la clase positiva es limitada y puede verse

afectada por ruido de medición o por variables no observadas que también influyen en la ocurrencia del evento.

Desde el punto de vista metodológico, el estudio adopta un horizonte de anticipación y una partición temporal concretos, lo que implica asumir que la relación entre las variables de proceso y la probabilidad de rotura se mantiene relativamente estable dentro del período analizado. Asimismo, aunque se comparan varios modelos y técnicas de selección de variables, no se explora todo el espectro posible de algoritmos ni configuraciones de hiperparámetros, y no se incluye el entrenamiento de arquitecturas más complejas, como modelos profundos de series temporales. Finalmente, la evaluación se realiza sobre datos históricos y mediante métricas cuantitativas; aspectos como la integración del modelo en los sistemas de control, la aceptación por parte de los operadores o el impacto económico detallado se discuten de manera cualitativa y se plantean como líneas de trabajo futuro.

Estructura del Documento

El documento inicia con los elementos preliminares establecidos por las normas institucionales: portada, hoja de aprobación, resumen, abstract, palabras clave en ambos idiomas, tabla de contenido y, en caso de ser necesario, listas de tablas, figuras y apéndices. Estos apartados permiten contextualizar el trabajo, facilitar su navegación y ofrecer una visión sintética del problema, el enfoque metodológico y los principales resultados.

El Capítulo 1 presenta la introducción, donde se expone el contexto general de la problemática de roturas del papel en una línea de producción y se plantea el problema de investigación. El Capítulo 2 desarrolla la justificación, destacando la relevancia técnica y operativa del estudio. El Capítulo 3 recoge el objetivo general y los objetivos específicos que orientan el desarrollo del proyecto, mientras que el Capítulo 4 delimita el alcance y las principales limitaciones del trabajo, estableciendo el marco dentro del cual deben interpretarse los resultados.

El Capítulo 5 describe la estructura del documento. El Capítulo 6 corresponde al marco teórico, en el que se revisan los conceptos de aprendizaje supervisado y problemas de clasificación, el desbalance de clases y las métricas adecuadas, la regularización y la selección de variables, los modelos de clasificación considerados y los trabajos relacionados. El Capítulo 7 detalla la metodología aplicada: se describe el conjunto de datos, el diseño del dataset para clasificación supervisada, las etapas de limpieza y preprocesamiento, las estrategias de selección de variables y el diseño de los experimentos de modelado.

El Capítulo 8 presenta los resultados, incluyendo los obtenidos en la selección de variables, el desempeño de los modelos de clasificación evaluados, la comparación entre ellos y el análisis de errores y de su comportamiento en términos operativos. En el Capítulo 9 se

desarrolla la discusión, donde se interpretan las variables relevantes, se contrastan los hallazgos con la literatura y se analizan las implicaciones prácticas en la planta. El Capítulo 10 recoge las conclusiones y recomendaciones que se derivan del estudio. Finalmente, el Capítulo 11 compila las referencias bibliográficas utilizadas y el Capítulo 12 incluye los apéndices, en los que se disponen materiales complementarios como tablas detalladas.

Marco Teórico

Aprendizaje Supervisado y Problemas de Clasificación

El aprendizaje automático supervisado se basa en entrenar modelos a partir de ejemplos etiquetados, es decir, pares (x_i, y_i) donde x_i representa un vector de características o predictores y y_i es la salida correcta asociada a cada observación. El objetivo es aprender una función que capture la relación entre entradas y salidas de manera que el modelo generalice adecuadamente a nuevos datos no observados, optimizando alguna métrica de desempeño definida sobre el conjunto de entrenamiento o validación (James et al., 2021; Najjar & Breesam, 2023; St-Aubin, 2024).

Dentro del aprendizaje supervisado se distinguen dos grandes tipos de problemas: regresión, cuando la variable objetivo es continua, y clasificación, cuando la salida pertenece a un conjunto discreto de clases. En clasificación, el modelo estima típicamente probabilidades de pertenencia a cada clase $P(y=k|x)$ y luego aplica una regla de decisión (por ejemplo, elegir la clase con mayor probabilidad o comparar contra un umbral) para asignar una etiqueta a cada observación (James et al., 2021; Najjar & Breesam, 2023).

El problema abordado en este proyecto corresponde a una tarea de clasificación binaria, en la que se diferencian dos estados del sistema: “rotura” y “no rotura” del papel en una línea de producción. Cada observación está compuesta por un conjunto de variables de proceso medidas por sensores y sistemas de control, mientras que la etiqueta indica si, en un horizonte de tiempo definido, se presentará o no un evento de rotura. Bajo este enfoque, el modelo busca aproximar la probabilidad de rotura condicionada a las condiciones actuales del proceso, $P(\text{rotura}|x)$, de forma que se puedan generar alertas anticipadas cuando dicha probabilidad supere un umbral de decisión acordado con la operación.

En aplicaciones industriales, los modelos de clasificación supervisada se utilizan de forma creciente para tareas de diagnóstico y detección temprana de fallas, ya que permiten explotar patrones multivariados difíciles de capturar mediante reglas manuales o límites univariados (Tchio et al., 2024).

Para que estos modelos sean confiables, no basta con entrenarlos sobre un conjunto de datos histórico: es necesario seguir un flujo de trabajo que incluya una preparación cuidadosa de los datos (tratamiento de valores atípicos y ausentes, estandarización), una partición adecuada en entrenamiento, validación y prueba, y la selección de métricas de evaluación coherentes con la naturaleza del problema, especialmente cuando se trabaja con clases desbalanceadas o eventos raros (James et al., 2021; Najjar & Breesam, 2023).

Desbalance de Clases y Métricas Adecuadas

En muchos problemas reales de clasificación, las clases no aparecen con la misma frecuencia: la mayoría de los registros pertenecen a una clase “normal” y solo una pequeña fracción corresponde a la clase de interés (fraude, fallo, rotura, etc.). Este fenómeno se conoce como desbalance de clases y suele caracterizarse mediante la razón de desbalance (imbalance ratio), definida como la proporción entre el tamaño de la clase mayoritaria y la minoritaria. Cuando esta razón es alta, los algoritmos estándar tienden a aprender principalmente el comportamiento de la clase mayoritaria, ignorando en gran medida la minoritaria (W. Chen et al., 2024; Feng et al., 2021; Rezvani & Wang, 2023).

Diversos estudios de revisión muestran que el desbalance de clases es una de las dificultades centrales del aprendizaje supervisado en dominios como detección de fallos industriales, detección de fraude, diagnóstico médico y seguridad informática (W. Chen et al., 2024; Rezvani & Wang, 2023). En este tipo de problemas, un clasificador trivial que siempre

predijera la clase mayoritaria podría alcanzar una “exactitud” (accuracy) aparentemente alta, pero con utilidad nula para detectar los eventos raros. Esta discrepancia entre desempeño global y desempeño sobre la clase minoritaria es uno de los principales motivos por los que la simple exactitud resulta engañosa en contextos desbalanceados (Owusu-Adjei et al., 2023; Shang et al., 2023).

Por ello, la evaluación en problemas desbalanceados se apoya en métricas derivadas de la matriz de confusión que descomponen explícitamente el desempeño por clase: sensibilidad o recall (proporción de eventos positivos correctamente detectados), especificidad (proporción de negativos bien clasificados), precisión (proporción de positivos predichos que realmente lo son), F1-score (media armónica entre precisión y recall), exactitud balanceada y coeficiente de correlación de Matthews (MCC), entre otros (Liu et al., 2023; Owusu-Adjei et al., 2023). Estas métricas permiten cuantificar el compromiso entre falsos negativos (eventos de rotura no anticipados) y falsos positivos (alarmas injustificadas), aspecto clave cuando la clase positiva es rara pero crítica desde el punto de vista operativo.

Además de las métricas puntuales, es habitual utilizar medidas agregadas sobre distintos umbrales de decisión. Las curvas ROC (Receiver Operating Characteristic) y su área bajo la curva (ROC-AUC) resumen el compromiso entre tasa de verdaderos positivos (TPR) y tasa de falsos positivos (FPR). Sin embargo, se ha mostrado que en escenarios fuertemente desbalanceados la ROC-AUC puede ofrecer una imagen excesivamente optimista, ya que la FPR puede permanecer muy baja incluso cuando el modelo genera un número no despreciable de falsos positivos en términos absolutos (Liu et al., 2023; Owusu-Adjei et al., 2023; Shang et al., 2023).

En contraste, las curvas precisión–recall (PR) y el área bajo la curva precisión–recall

(PR-AUC) se centran en el comportamiento sobre la clase positiva y son más sensibles a la prevalencia de la clase minoritaria. Estudios teóricos y empíricos han demostrado que, para datos desbalanceados, la PR-AUC es más informativa que la ROC-AUC a la hora de comparar modelos, ya que refleja directamente el compromiso entre qué fracción de eventos positivos se captura (recall) y qué tan “limpia” es la lista de alertas emitidas (precisión) (Liu et al., 2023; Saito & Rehmsmeier, 2015; Shang et al., 2023). Un aspecto importante es que el valor de referencia (baseline) de la PR-AUC coincide con la prevalencia de la clase positiva; por ejemplo, si solo el 1 % de las observaciones son roturas, un modelo aleatorio tendría una PR-AUC cercana a 0,01.

En el contexto de este proyecto, donde las roturas del papel constituyen eventos raros pero de alto impacto, se adoptará la PR-AUC como métrica principal para seleccionar y comparar modelos, complementada con ROC-AUC, precisión, recall, F1-score, matrices de confusión y medidas de lift@k que permiten evaluar la utilidad operativa de los modelos en la priorización de alarmas. De este modo, la evaluación se alinea con la naturaleza desbalanceada del problema y con las necesidades prácticas de la planta, donde es más relevante maximizar la detección temprana de roturas a un costo manejable de falsos positivos que optimizar una exactitud global poco informativa.

Regularización y Selección de Variables

En problemas de aprendizaje supervisado con muchos predictores y posibles relaciones complejas, los modelos lineales sin penalización tienden a sobreajustar los datos, especialmente cuando existe colinealidad entre variables o la relación señal-ruido es baja. La regularización introduce un término de penalización sobre los coeficientes del modelo para controlar la complejidad y mejorar su capacidad de generalización. Las penalizaciones L1 (Lasso) y L2

(Ridge) son las más utilizadas: la primera promueve soluciones escasas al forzar algunos coeficientes exactamente a cero, mientras que la segunda reduce la magnitud de los coeficientes sin anularlos, estabilizando el ajuste en presencia de multicolinealidad. Estudios recientes en predicción con datos de alta dimensión han mostrado que estas técnicas permiten reducir sobreajuste y mejorar la robustez de modelos de regresión y clasificación, incluyendo la regresión logística (Fridgeirsson et al., 2024; Marzougui et al., 2020).

El Elastic Net combina las ventajas de L1 y L2 mediante una penalización mixta que incorpora simultáneamente ambos términos, lo que le permite realizar selección de variables y, a la vez, agrupar predictores altamente correlacionados en lugar de seleccionar solo uno de ellos. Este enfoque fue propuesto originalmente por Zou y Hastie como una alternativa al Lasso cuando el número de variables es grande y existen grupos de predictores correlacionados (Zou & Hastie, 2005).

Trabajos más recientes han extendido el Elastic Net a modelos lineales generalizados, incluida la regresión logística, y han mostrado su utilidad en aplicaciones de clasificación con muchos predictores y estructuras de dependencia complejas (Marzougui et al., 2020; Tay et al., 2023). En el contexto de este proyecto, donde se dispone de múltiples variables de proceso potencialmente correlacionadas (sensores, señales de control, etc.), la regresión logística con penalización Elastic Net constituye un mecanismo natural para obtener un subconjunto de variables relevante y relativamente parsimonioso.

La selección de variables busca identificar un subconjunto de predictores que aporte la mayor cantidad de información posible con la menor redundancia, mejorando la interpretabilidad y reduciendo el coste computacional. La literatura reciente clasifica las técnicas de selección de variables en tres familias principales: métodos filtro, métodos envolventes (wrapper) y métodos

embebidos (Cheng, 2024; Pudjihartono et al., 2022). Los métodos filtro evalúan cada variable (o combinaciones simples) mediante criterios estadísticos independientes del modelo (por ejemplo, correlación, pruebas de hipótesis, información mutua); los métodos envolventes utilizan un modelo predictivo como “caja negra” y buscan el subconjunto de variables que optimiza una métrica de desempeño, típicamente mediante búsquedas heurísticas como forward selection o backward elimination; y los métodos embebidos integran la selección de variables en el propio proceso de entrenamiento del modelo, como en el caso de los modelos penalizados (Lasso, Elastic Net) o de ciertos algoritmos basados en árboles (Pudjihartono et al., 2022).

En este proyecto, el diseño de la selección de variables se alinea explícitamente con el objetivo de obtener un conjunto parsimonioso y con baja colinealidad. Para ello se plantean dos rutas complementarias. La primera ruta corresponde a un método filtro basado en correlación y control de redundancia (Método A). En una primera etapa, se mide la asociación entre cada variable de proceso y la respuesta binaria mediante la correlación de Spearman, que es sensible a relaciones monótonas potencialmente no lineales y resulta adecuada para datos industriales donde pueden aparecer transformaciones no lineales de las señales. Trabajos recientes han propuesto variantes de selección de variables que se apoyan en Spearman para evaluar relevancia y dependencias con la clase en tareas de clasificación (Jiang et al., 2024).

No obstante, una variable altamente correlacionada con la respuesta puede ser, a su vez, redundante respecto a otras variables similares. Para controlar la colinealidad entre predictores, en esta misma ruta se analiza la matriz de correlaciones entre las variables candidatas y se aplican umbrales de correlación para eliminar predictores altamente redundantes, manteniendo únicamente uno por grupo de variables muy correlacionadas. En la práctica, enfoques basados en correlaciones de Spearman, clustering de características y selección de un representante por

grupo se han utilizado para mitigar la multicolinealidad antes del ajuste de modelos, tanto en contextos industriales como en ejemplos de referencia de librerías de aprendizaje automático. Esta estrategia permite construir un conjunto de variables con buena capacidad predictiva individual y reducida redundancia mutua, coherente con el objetivo de parsimonia.

La segunda ruta corresponde a una combinación de método embebido + método envolvente (Método B). En una primera fase, se ajusta un modelo de regresión logística penalizado con Elastic Net, utilizando un esquema de búsqueda de hiperparámetros que optimiza una métrica adecuada para clases desbalanceadas, como la PR-AUC. La penalización induce que muchos coeficientes tomen valores cercanos a cero, mientras que otros permanecen significativamente distintos de cero, lo que proporciona un primer filtrado embebido de variables relevantes. Estudios recientes en predicción clínica y otros dominios han mostrado que las variantes penalizadas de la regresión logística, en particular L1 y Elastic Net, ofrecen un buen compromiso entre desempeño discriminativo, estabilidad y capacidad de selección de variables (Fridgeirsson et al., 2024; Tay et al., 2023).

A partir del conjunto reducido de variables identificado por el Elastic Net, se propone una fase adicional de selección hacia adelante (forward selection) como método envolvente. Este procedimiento comienza con un modelo simple y va incorporando variables de forma iterativa, añadiendo en cada paso el predictor que más mejora una métrica objetivo (en este caso, nuevamente orientada al desempeño sobre la clase positiva). La forward selection se ha utilizado ampliamente como técnica heurística para seleccionar subconjuntos de variables en modelos lineales y de regresión logística, y su implementación moderna suele apoyarse en validación cruzada y métricas específicas de interés (Scikit-learn, s. f.). En el contexto de este proyecto, su uso posterior al Elastic Net permite refinar el conjunto de predictores, explorando combinaciones

que mejoren el desempeño en eventos raros, al tiempo que se conserva la parsimonia y se limita la colinealidad entre variables seleccionadas.

En conjunto, las dos rutas planteadas filtro basado en Spearman y correlación entre predictores, y combinación Elastic Net + forward selection proporcionan un marco coherente con los objetivos del proyecto: construir un conjunto de variables explicativo, parsimonioso y poco redundante, que sirva de base para entrenar y comparar modelos de clasificación supervisada capaces de anticipar roturas del papel bajo un escenario de desbalance severo.

Modelos de Clasificación Utilizados

En este proyecto se consideran cuatro modelos de clasificación supervisada que combinan buen desempeño empírico en datos tabulares, disponibilidad en librerías estándar y un nivel razonable de interpretabilidad: regresión logística penalizada, Random Forest, máquinas de vectores de soporte (SVM) y XGBoost. Estos modelos se han utilizado de forma recurrente en aplicaciones de predicción de fallas, mantenimiento predictivo y detección de eventos raros en entornos industriales, así como en otros dominios con desbalance de clases (Grinsztajn et al., 2022).

La regresión logística es un modelo lineal generalizado que estima la probabilidad de pertenencia a la clase positiva mediante la función logística aplicada a una combinación lineal de los predictores. En clasificación binaria, la regresión logística modela el logit de la probabilidad $P(y=1|x)$ como función de las variables de entrada. La incorporación de penalizaciones L2 (Ridge) o mixtas como Elastic Net permite controlar la complejidad del modelo, reducir el impacto de la colinealidad y mejorar la capacidad de generalización, especialmente cuando se trabaja con múltiples predictores correlacionados (Pudjihartono et al., 2022; Tay et al., 2023). Estudios comparativos recientes muestran que la regresión logística penalizada sigue siendo un

referente competitivo frente a modelos más complejos, particularmente cuando se prioriza la interpretabilidad de los coeficientes y la estabilidad de las estimaciones (Fridgeirsson et al., 2024). En el contexto de este trabajo, la regresión logística penalizada cumple el doble rol de modelo base interpretable y de herramienta embebida para selección de variables.

El Random Forest es un método de ensamblado basado en árboles de decisión que combina múltiples árboles construidos sobre subconjuntos de datos y de variables seleccionados aleatoriamente (bagging y selección aleatoria de características). Cada árbol aporta una votación y la predicción final se obtiene por agregación (votación mayoritaria o promedio de probabilidades). Esta estructura permite capturar relaciones no lineales e interacciones complejas entre variables, al tiempo que ofrece robustez frente a ruido y outliers. Revisiones recientes sobre detección de fallas en entornos de Industria 4.0 destacan a Random Forest como uno de los modelos más utilizados y con desempeño competitivo para clasificación de fallos en datos tabulares multivariados (Wu et al., 2024). Estudios específicos de predicción de fallas de máquina reportan que Random Forest suele igualar o superar a otros algoritmos clásicos en métricas como AUC, F1 y precisión, y proporcionan medidas de importancia de variables útiles para la interpretación (Yang et al., 2025).

Las máquinas de vectores de soporte (SVM) son modelos basados en la maximización del margen que buscan un hiperplano de separación entre clases que maximice la distancia a las observaciones más cercanas (vectores de soporte). Mediante el uso de funciones kernel, las SVM pueden representar fronteras de decisión no lineales en el espacio original, lo que las hace adecuadas cuando la separación lineal no es suficiente (Guido et al., 2024; Pudjihartono et al., 2022). Revisiones recientes muestran que las SVM continúan siendo una alternativa sólida para problemas de clasificación binaria en dominios como salud, industria y finanzas, con variantes

específicas para mejorar la generalización y tratar explícitamente el desbalance de clases mediante funciones de coste asimétricas o ponderaciones diferenciadas (Guido et al., 2024; Hussein et al., 2023). Este tipo de enfoques es especialmente relevante en escenarios de eventos raros, donde los errores sobre la clase minoritaria tienen un coste elevado.

XGBoost (eXtreme Gradient Boosting) es una implementación optimizada de árboles de decisión potenciados por gradiente (Gradient Boosting Decision Trees, GBDT) que se ha consolidado como uno de los métodos de referencia para datos tabulares (Chen & Guestrin, 2016). XGBoost construye secuencialmente un conjunto de árboles, donde cada nuevo árbol corrige los errores de la iteración anterior mediante el descenso por gradiente sobre una función de pérdida diferenciable. Incorpora regularización explícita sobre la estructura y los pesos de los árboles, manejo eficiente de datos escasos y mecanismos para el control del sobreajuste, lo que se traduce en modelos competitivos en un amplio rango de tareas de clasificación y regresión (Chen & Guestrin, 2016; IBM, 2024).

Estudios recientes de comparación en datos tabulares han mostrado que los métodos de boosting basados en árboles incluyendo XGBoost siguen ofreciendo un desempeño de referencia frente a redes neuronales especializadas, especialmente cuando se dispone de conjuntos de datos de tamaño pequeño o medio y se requiere un equilibrio entre desempeño y capacidad de interpretación mediante importancias de características (Grinsztajn et al., 2022; Holzmüller et al., 2024).

En la práctica, la elección de estos cuatro modelos responde a un compromiso entre interpretabilidad, capacidad de modelar relaciones no lineales y robustez en presencia de desbalance. La regresión logística penalizada proporciona coeficientes directamente interpretables y una base natural para analizar el aporte de cada variable de proceso. Random

Forest y XGBoost permiten capturar interacciones y no linealidades sin necesidad de especificar términos de interacción de forma explícita, ofreciendo además medidas de importancia de variables que complementan la selección realizada en la etapa previa.

Las SVM, por su parte, aportan un modelo basado en márgenes con capacidad para explotar kernels no lineales, lo que puede resultar ventajoso si la frontera entre estados de “rotura” y “no rotura” es compleja en el espacio de variables seleccionadas. Dado el desbalance severo del problema, todos los modelos se entrenarán con técnicas que ajustan su comportamiento a la clase minoritaria, como ponderaciones de clase o funciones de coste diferenciadas, y se evaluarán principalmente con la métrica PR-AUC y medidas derivadas de la matriz de confusión, tal como se justificó en la sección anterior.

Trabajos Relacionados

La predicción de eventos raros en contextos industriales se ha abordado ampliamente bajo el paraguas de la predicción de eventos raros y el mantenimiento predictivo. Revisiones recientes sintetizan que estos problemas comparten tres retos clave: desbalance extremo, datos ruidosos/incompletos y la necesidad de métricas y procedimientos de evaluación específicos a lo largo de todo el pipeline de aprendizaje (preprocesamiento, algoritmos y validación). En el marco de la Industria 4.0, los estudios de mantenimiento predictivo destacan el uso creciente de modelos de clasificación sobre datos de sensores para anticipar fallas, así como la combinación de técnicas clásicas de machine learning con enfoques más recientes (deep learning, federated learning, data augmentation) para mejorar la detección de eventos poco frecuentes.

En el sector específico de pulpa y papel, varios trabajos han abordado la predicción de roturas de hoja o fallas asociadas al proceso. (Khosravi et al., 2023), analizan datos operativos de una máquina de papel con 18 398 instancias y solo 124 roturas, y proponen un esquema de

aumento de datos basado en CTGAN y SMOTE para mejorar el rendimiento de modelos como árboles de decisión, Random Forest y regresión logística. Reportan incrementos sustanciales en la capacidad de detección de la clase de rotura tras la generación de datos sintéticos, demostrando que la escasez de eventos positivos limita fuertemente el desempeño de los modelos si no se interviene sobre la distribución de clases.

(He et al., 2023), proponen un enfoque de monitorización de proceso y predicción de fallas de papermaking a partir de datos “imperfectos”. Primero utilizan Random Forest para extraer variables interrelacionadas y construir un conjunto de características relevantes; posteriormente, definen un benchmark de “estado sano” mediante modelos de mezcla gaussiana (GMM) y analizan desviaciones del estado real usando distancia de Mahalanobis y estimación de densidad para disparar alertas de posible rotura. El método alcanza alrededor de 76,8 % de exactitud y 72,5 % de recall en la predicción de fallas asociadas a roturas, mostrando que es posible anticipar comportamientos anómalos a partir de datos históricos incluso cuando la instrumentación no es perfecta.

Otros trabajos recientes exploran arquitecturas más complejas. (Yu et al., 2024), proponen un enfoque de diagnóstico de fallas de rotura basado en *transfer learning* y *federated learning* para abordar cambios de condición de operación entre distintos modos de la máquina. Parten de modelos profundos tipo autoencoder + Softmax (SAE–Softmax) y muestran que, al incorporar transferencia de parámetros y rasgos entre condiciones y un esquema federado para entrenar con datos distribuidos y escasos, se mejora la capacidad de diagnóstico frente a modelos clásicos como regresión logística, SVM y Random Forest entrenados por separado en cada condición.

Desde finales de los años noventa existen también contribuciones centradas en

indicadores de sensibilidad a la rotura y selección de variables en máquinas de papel. El trabajo clásico de (Juuso, 2022) y colaboradores introduce el “web break sensitivity indicator”, desarrollado a partir de mediciones reales y técnicas de soft computing para estimar la sensibilidad a roturas en función de variables de proceso (Ahola, 2006). Más recientemente, se han propuesto esquemas de selección y agrupamiento de variables a gran escala para modelado data-driven de máquinas de papel, donde se utilizan criterios de asociación y redundancia entre sensores para construir conjuntos de predictores manejables y físicamente interpretables, alineados con la idea de parsimonia que persigue este proyecto.

En síntesis, la literatura muestra tres tendencias principales:

1. Enfoques data-driven de mantenimiento predictivo para eventos raros, que subrayan la importancia del tratamiento del desbalance, la calidad de los datos y métricas específicas.
2. Trabajos orientados al sector de pulpa y papel que emplean desde modelos clásicos (árboles, Random Forest, regresión logística) hasta deep learning y técnicas avanzadas de aumento de datos para mejorar la predicción de roturas.
3. Propuestas de indicadores y estrategias de selección de variables diseñadas específicamente para máquinas de papel.

El presente proyecto se sitúa en la intersección de estas líneas, pero con un énfasis distinto: en lugar de centrarse en data augmentation o arquitecturas profundas, se propone un pipeline estructurado de clasificación supervisada sobre datos reales de planta, que combina dos estrategias explícitas de selección de variables (filtro con Spearman + control de colinealidad y Elastic Net + forward selection) con la comparación sistemática de modelos tabulares bien establecidos (regresión logística penalizada, Random Forest, SVM y XGBoost), evaluados

principalmente con PR-AUC y métricas orientadas a la operación. De este modo, el trabajo aporta una referencia metodológica centrada en parsimonia, interpretabilidad y coherencia entre selección de variables, elección de modelos y evaluación para eventos raros en el contexto concreto de las roturas del papel.

Metodología

La metodología seguida en este proyecto organiza el problema como una tarea de clasificación supervisada sobre datos tabulares derivados de una serie temporal multivariada. El flujo general comprende:

1. Descripción y comprensión del conjunto de datos original.
2. Diseño del conjunto de datos para clasificación binaria anticipada.
3. Limpieza y preprocesamiento sistemático de las variables de proceso.
4. Selección de variables mediante dos rutas complementarias.
5. Entrenamiento, ajuste y comparación de modelos de clasificación bajo métricas adecuadas para eventos raros.

El estudio adopta un enfoque cuantitativo, con un diseño no experimental y de tipo longitudinal retrospectivo, basado en el análisis de datos secundarios. No se manipulan variables de proceso ni se interviene sobre la operación de la máquina; en su lugar, se analiza un histórico de mediciones para construir y validar modelos de predicción de roturas bajo un esquema de entrenamiento, validación y prueba con partición temporal.

Descripción del Conjunto de Datos

El conjunto de datos utilizado en este proyecto corresponde al benchmark público “Dataset: Rare Event Classification in Multivariate Time Series”, desarrollado por ProcessMiner a partir de un proceso real de fabricación de pulpa y papel. Según la ficha oficial del proveedor y el artículo que acompaña al dataset, se trata de una serie temporal multivariada en la que se registran lecturas de múltiples sensores de una máquina de papel, junto con una etiqueta binaria que indica la ocurrencia de un evento de rotura de hoja (ProcessMiner, 2020).

En este proceso industrial, la máquina funciona de manera continua, produciendo bobinas

de papel mientras diversos sensores miden variables de materia prima (por ejemplo, cantidad de fibra, químicos dosificados) y de operación (por ejemplo, vacío, velocidades, condiciones de formación). Cuando ocurre una rotura, la línea debe detenerse, retirar la bobina, corregir la causa y reanudar la producción, con pérdidas económicas significativas por tiempo de paro, desperdicio de material y riesgo operativo para el personal. El objetivo declarado del dataset es precisamente permitir el desarrollo de modelos de clasificación que predigan la rotura de papel con anticipación, a partir de estas señales de proceso (ProcessMiner, 2020).

De acuerdo con la documentación técnica del dataset, la información proviene de un sistema de adquisición de datos con mediciones cada 2 minutos, lo que genera series de alta frecuencia. En el preprocesamiento original se mantiene, para cada episodio, únicamente el primer registro etiquetado como rotura, de modo que, aunque en planta pueda haber un intervalo de tiempo durante el cual la hoja permanece rota, en el archivo final solo queda un “marcador” del inicio del evento (ProcessMiner, 2020).

La versión del dataset empleada en este trabajo contiene 18.398 observaciones y 63 columnas. De estas, una columna corresponde a la marca temporal (time), una a la etiqueta binaria del evento (y) y las 61 restantes son variables numéricas de proceso (x1–x61). La etiqueta y toma el valor 1 cuando se registra una rotura y 0 en caso contrario. En total, solo 124 registros están etiquetados como rotura, lo que implica una prevalencia aproximada del 0,67 % y un desbalance pronunciado entre la clase mayoritaria (“no rotura”) y la clase minoritaria (“rotura”).

En la literatura, este conjunto de datos se cita como un benchmark de referencia para problemas de clasificación de eventos raros en series temporales multivariadas en la industria papelera, y ha sido utilizado para comparar enfoques basados en modelos clásicos, métodos de aumento de datos y técnicas de aprendizaje profundo (ProcessMiner, 2020). En este proyecto se

trabajaré específicamente con la versión tabular (time, y, x1–x61), utilizando la marca temporal para ordenar las observaciones y definir los cortes de entrenamiento, validación y prueba, y las variables x1–x61 como predictores candidatos para los modelos de clasificación supervisada.

Diseño del Conjunto de Datos para Clasificación Supervisada

Aunque el dataset original ya incluye una etiqueta binaria y que indica si en un instante dado se produjo una rotura de papel, el objetivo operativo de este proyecto es anticipar dichos eventos con un pequeño margen de tiempo, de forma que la planta pueda reaccionar antes de que ocurra la rotura. Este objetivo es coherente con la forma en que el conjunto de datos fue planteado por sus autores: como base para modelos de clasificación temprana de un evento raro en una serie temporal multivariada.

Para alinear el problema con este objetivo, se reformula la tarea como una clasificación binaria anticipada sobre datos tabulares derivados de la serie temporal:

Definición de la Etiqueta Anticipada

1. Se ordenan las observaciones por la marca temporal time.
2. A partir de la etiqueta original (y) (0 = no rotura, 1 = rotura), se construye una nueva variable objetivo (y_{ant}) que marca los estados del proceso que preceden a una rotura en un horizonte fijo de tiempo.
3. Concretamente, para cada instante (t) se define
4.
$$y_{ant}(t) = \begin{pmatrix} 1 & \text{si existe al menos una rotura en el intervalo } (t, t + h] \\ & \text{dentro del mismo bloque temporal} \\ 0 & \text{en caso contrario} \end{pmatrix}$$
5. En este trabajo se fija ($h = 6$), lo que equivale a un horizonte de aproximadamente 12 minutos, dado que el registro de datos tiene un paso de 2 minutos.
6. De esta forma, las observaciones etiquetadas como $(y_{ant}) = 1$ corresponden a

estados “previos a rotura” dentro de un horizonte de aproximadamente 12 minutos, que es el escenario en el que resulta útil emitir una alerta operativa.

7. Para evitar fuga de información, las últimas filas de cada bloque, para las cuales no es posible evaluar completamente la ventana futura $((t, t + h])$, se eliminan del análisis.

Selección de Predictores

A efectos de modelado se descartan columnas que no se utilizarán como predictores directos, como:

1. Campo *time*, que solo se emplea para ordenar cronológicamente y definir los cortes temporales.
2. La etiqueta original *y*, que se conserva como referencia exploratoria pero no se usa como variable de entrada.

El conjunto inicial de predictores X queda constituido por las 61 variables numéricas de proceso (x_1-x_{61}), que representan mediciones de sensores y condiciones operativas de la máquina de papel.

A partir de este conjunto base se construyen las variables temporales descritas en el subapartado de ingeniería temporal (retardos, ventanas deslizantes, percentiles y derivadas), de modo que los modelos y los métodos de selección de variables trabajan sobre un espacio ampliado de predictores que resume el comportamiento reciente de cada sensor.

Partición Temporal en Entrenamiento, Validación y Prueba

Dado que los datos proceden de una serie temporal con muestreo regular (una observación cada 2 minutos) y el objetivo es simular el uso del modelo en operación real, la partición del dataset se realiza respetando estrictamente el orden cronológico.

Se ordenan todas las filas por *time*. Se define una partición no aleatoria, usando los

tramos inicial, intermedio y final de la serie:

1. Conjunto de entrenamiento (train), aproximadamente el 60 % inicial de las observaciones.
2. Conjunto de validación (valid), aproximadamente el 20 % siguiente, usado para ajuste de hiperparámetros y comparación de modelos durante el desarrollo.
3. Conjunto de prueba (test), aproximadamente el 20 % final, reservado exclusivamente para la evaluación final del modelo seleccionado.

La partición se realiza de forma que ninguna observación del futuro quede en entrenamiento o validación respecto a las usadas en prueba, evitando cualquier forma de fuga temporal y reproduciendo el escenario en el que el modelo se entrenaría con histórico y se aplicaría sobre datos “nuevos”.

Como resultado de este diseño, el problema queda formulado como una tarea de clasificación binaria anticipada sobre un conjunto de datos tabular: las entradas son vectores que combinan las variables de proceso originales (x_1, \dots, x_{61}) y sus transformaciones temporales (retardos, ventanas deslizantes, percentiles y derivadas), todas preprocesadas y estandarizadas; la salida es la probabilidad de que, dentro del horizonte definido (12 minutos), se produzca una rotura del papel. Esta formulación es coherente con el propósito original del dataset y con los objetivos del proyecto de grado, centrados en anticipar roturas y analizar qué variables de proceso contribuyen de forma más clara al riesgo de evento.

Limpieza y Preprocesamiento de los Sensores

Dado que el objetivo es entrenar modelos de clasificación robustos sobre datos de sensores industriales y comparar estrategias de selección de variables y modelado, la fase de preprocesamiento se define como un pipeline estructurado y reproducible aplicado siempre en el

mismo orden. Todos los parámetros se estiman exclusivamente en el conjunto de entrenamiento y luego se aplican sin modificación a validación y prueba, para evitar fuga de información. El flujo seguido es el siguiente:

Control de Calidad y Consistencia Básica

1. Se verifica el tipo de datos de todas las variables x_1 – x_{61} (numéricas continuas) y se comprueba que no existan columnas constantes o prácticamente sin variabilidad.
2. Las variables estrictamente constantes o con variación casi nula se descartan, ya que no aportan información al modelo y pueden interferir con ciertos procedimientos de regularización o de importancia de variables.
3. Se revisan rangos y valores extremos claramente imposibles desde el punto de vista físico (por ejemplo, valores negativos en magnitudes que deberían ser siempre positivas), dejando documentados posibles problemas de instrumentación para su discusión posterior en la sección de resultados.

Tratamiento de Valores Perdidos

Aunque el archivo original no contiene valores perdidos explícitos, la construcción de la etiqueta anticipada, la eliminación de registros finales y otros pasos intermedios pueden introducir NaN en algunas columnas. Para tratarlos se sigue una estrategia en dos etapas:

1. Relleno hacia adelante (forward fill) a lo largo de la serie temporal, aprovechando la naturaleza suavemente variable de muchas señales de proceso: cuando falta un valor en el instante t , se utiliza el último valor observado para esa variable en instantes anteriores.
2. Si después del relleno hacia adelante alguna variable mantiene valores ausentes, estos se imputan con estadísticas del entrenamiento, típicamente la mediana o un imputador robusto (por ejemplo, basado en cuantiles). Esta imputación se ajusta únicamente con datos del

conjunto de entrenamiento y luego se aplica a validación y prueba usando los mismos parámetros, garantizando que la información del futuro no contamine el preprocesamiento.

Tratamiento de Valores Atípicos (Winsorización)

En procesos industriales es habitual encontrar valores atípicos debidos a fallas temporales de sensores, arranques/paradas o situaciones de operación poco frecuentes. En lugar de eliminar filas completas y romper la estructura temporal, se opta por una winsorización controlada:

1. Para cada variable numérica se calculan, en el conjunto de entrenamiento, límites robustos basados en la mediana y el rango intercuartílico (por ejemplo, mediana \pm 3IQR, equivalentes a percentiles extremos).
2. Los valores que caen por debajo del percentil bajo se reemplazan por dicho percentil; los que superan el percentil alto se reemplazan por ese límite superior.
3. Los puntos recortados se marcan y cuantifican, de manera que en la sección de resultados pueda evaluarse hasta qué punto el modelo se apoya en regiones del espacio de variables donde se ha aplicado winsorización.

Este enfoque limita la influencia de outliers extremos sin perder observaciones y mantiene la coherencia con la práctica habitual en análisis de datos de sensores.

Estandarización y Escalado de Variables

Muchos de los modelos utilizados (regresión logística penalizada, SVM y en parte XGBoost) se benefician de trabajar con variables en escalas comparables. Para ello:

1. Sobre el conjunto de entrenamiento preprocesado se ajusta un esquema de escalado (por ejemplo, estandarización a media 0 y desviación estándar 1, o un escalador robusto basado en mediana y rango intercuartílico).

2. Se almacenan los parámetros del escalador (medias y desviaciones, o mediana y IQR) y se aplican exactamente los mismos valores a los conjuntos de validación y prueba.

3. El resultado es una matriz de predictores en la que todas las columnas tienen magnitud comparable, lo que facilita la regularización de la regresión logística, estabiliza la optimización en SVM y evita que unas pocas variables de gran escala dominen las métricas de distancia.

Ingeniería Temporal de Variables

Además del preprocesamiento básico, se incorpora una capa de ingeniería temporal que enriquece las señales originales con resúmenes causales del comportamiento reciente de cada sensor. El objetivo es capturar patrones que preceden a las roturas tendencias, aumentos de variabilidad o cambios bruscos sin violar la causalidad temporal ni mezclar información entre bloques separados por grandes huecos.

En concreto, se construyen las siguientes familias de variables temporales para cada (x_j):

1. Retardos (lags): valores rezagados $x_j^{\text{lag}L}$ con ($L = 1, 2, 3, 5, 10$), equivalentes a aproximadamente 2, 4, 6, 10 y 20 minutos hacia el pasado. Estos lags permiten que el modelo observe la trayectoria reciente de cada sensor.

2. Ventanas deslizantes (rolling windows): para ventanas r5, r15 y r30 (10, 30 y 60 minutos) se calculan media, desviación estándar, mínimo y máximo a partir de los valores estrictamente anteriores a (t). Operacionalmente, primero se aplica un lag de un paso y luego se agregan los (L) valores pasados; así se garantiza que las estadísticas se basan solo en información histórica disponible en el momento de la predicción.

3. Percentiles y derivados: sobre ventanas largas (r30) se calculan percentiles ($(p_{25}), (p_{75})$), que capturan colas bajas y altas recientes. Además, se construyen derivadas de

primer y segundo orden (d1, d2) que resumen la dirección e intensidad del cambio y su aceleración.

4. Indicadores de salto: se definen variables binarias que marcan incrementos anómalos, por ejemplo $x_j^{\text{jump}} = \mathbf{1}_{|x_j| > q_{95}^{\text{train}}}$, donde q_{95}^{train} es el cuantil 95 de la distribución de la derivada en el conjunto de entrenamiento. Estos indicadores permiten al modelo detectar episodios de inestabilidad súbita.

Todas estas transformaciones se calculan por bloque temporal, definido a partir de gaps ≥ 15 minutos entre observaciones, y respetando estrictamente la dirección del tiempo. De este modo se evitan mezclas artificiales de periodos de operación no contiguos y se preserva la trazabilidad de las señales que alimentan la selección de variables (Métodos A y B) y los modelos de clasificación del capítulo siguiente.

Implementación como Pipeline Reproducible

Todos estos pasos (control de calidad, imputación, winsorización y escalado) se implementan como un pipeline único, de forma que:

1. Cualquier experimento de selección de variables (Métodos A y B) y de entrenamiento de modelos se realiza siempre sobre la misma versión de los datos preprocesados.
2. El flujo puede ejecutarse nuevamente sobre el dataset original, lo que facilita la trazabilidad de resultados y la futura implementación del modelo en un entorno operativo.

Con este esquema, la limpieza y el preprocesamiento de los sensores quedan alineados con los objetivos del proyecto: mejorar la calidad de la información de entrada, reducir el impacto de valores extremos y garantizar que las comparaciones entre estrategias de selección de variables y modelos de clasificación se apoyan en un conjunto de datos coherente, reproducible y compatible con el uso posterior del modelo en planta.

Selección de Variables

La selección de variables en este proyecto tiene como propósito obtener un conjunto de predictores parsimonioso, poco redundante y coherente con el proceso físico, de manera que los modelos de clasificación resulten más estables, interpretables y menos propensos al sobreajuste, especialmente en un contexto de evento raro. Para ello se trabaja siempre sobre los datos ya preprocesados y únicamente utilizando el conjunto de entrenamiento, evitando que la información de validación o prueba influya en la elección de variables. Se definen dos rutas complementarias:

1. Método A. Selección tipo filtro basada en correlación de Spearman con la etiqueta y control de colinealidad entre predictores.
2. Método B. Combinación de un modelo penalizado (Elastic Net) con un procedimiento de selección hacia adelante (forward selection).

El resultado son dos subconjuntos de variables (conjunto A y conjunto B) que luego se utilizarán para entrenar y comparar los modelos de clasificación.

Método A Filtro Basado en Correlación de Spearman y Control de Colinealidad

El Método A aplica un enfoque de selección tipo filtro, independiente del modelo, siguiendo dos pasos sobre el conjunto de entrenamiento:

1. Evaluación de relevancia individual con Spearman. Se calcula, para cada variable de proceso x_j , el coeficiente de correlación de Spearman entre x_j y la etiqueta anticipada y_{ant} (0 = no rotura anticipada, 1 = rotura anticipada). La correlación de Spearman se utiliza porque captura relaciones monótonas potencialmente no lineales, lo que resulta adecuado para señales de proceso que pueden responder de forma no estrictamente lineal al riesgo de rotura. Se seleccionan inicialmente aquellas variables cuya correlación absoluta $|\rho_s(x_j, y_{\text{ant}})|$ supera un

umbral predefinido (por ejemplo, 0,1 o 0,15, según el análisis exploratorio) y, opcionalmente, presentan un valor p estadísticamente significativo. Este filtrado produce un primer subconjunto de variables individualmente asociadas con el riesgo de rotura anticipada.

2. Control de redundancia y colinealidad entre predictores. Sobre el subconjunto obtenido se calcula la matriz de correlaciones entre variables (Spearman o Pearson), con el fin de detectar grupos de predictores fuertemente correlacionados entre sí. Para pares o grupos de variables con correlación alta (por ejemplo, $|\rho| > 0,8$), se conserva un solo representante, priorizando la variable con mayor $|\rho|$ frente a y_{ant} y aquella con interpretación más clara desde el punto de vista del proceso. Las variables redundantes se eliminan del conjunto de predictores.

Al finalizar este procedimiento, se obtiene un conjunto A de variables que cumple dos propiedades: presenta asociación detectable con la probabilidad de rotura anticipada; y evita, en lo posible, la presencia de grupos de predictores casi duplicados, reduciendo la colinealidad que podría afectar a modelos lineales y dificultar la interpretación.

Método B Elastic Net más Selección Hacia Adelante (Forward Selection)

El Método B combina un enfoque embebido (modelo penalizado) con un enfoque envolvente (forward selection), también aplicado exclusivamente sobre el conjunto de entrenamiento ya preprocesado.

Filtrado Embebido Mediante Regresión Logística con Elastic Net. Se ajusta un modelo de regresión logística penalizada con Elastic Net, tomando como entrada todas las variables de proceso disponibles tras el preprocesamiento.

1. Se realiza una búsqueda de hiperparámetros sobre el parámetro de regularización global (λ) y el parámetro de mezcla L1/L2 (α), que controla el balance entre Lasso y Ridge.

2. La combinación de hiperparámetros se selecciona maximizando la PR-AUC en el conjunto de validación, en coherencia con el carácter desbalanceado del problema.

3. Una vez obtenido el modelo penalizado final, se identifican las variables con coeficientes significativamente distintos de cero (o por encima de un umbral de magnitud). Estas variables se consideran seleccionadas de forma embebida por Elastic Net.

Refinamiento Mediante Selección Hacia Adelante (Forward Selection). A partir del conjunto de variables retenidas por Elastic Net se aplica un procedimiento de forward selection:

1. Se inicia con un modelo base (por ejemplo, con una sola variable o incluso solo con el intercepto).

2. En cada iteración se añade la variable que, al incorporarse al modelo, produce el mayor incremento de la métrica objetivo (PR-AUC u otra métrica centrada en la clase positiva) sobre el conjunto de validación.

3. El proceso se detiene cuando la mejora en la métrica es marginal o cuando se alcanza un tamaño máximo de conjunto considerado razonable para mantener la parsimonia.

4. En esta etapa se pueden utilizar como modelos base la propia regresión logística penalizada u otro modelo coherente con el objetivo de interpretabilidad.

El resultado de este procedimiento es un conjunto B de variables, más reducido, que recoge aquellas características que Elastic Net considera relevantes y cuyas combinaciones específicas han demostrado aportar mejoras reales en la capacidad del modelo para anticipar roturas.

Conjuntos Resultantes y Uso en los Experimentos

Los dos métodos descritos producen dos subconjuntos de predictores:

1. Conjunto A. variables seleccionadas por el filtro de correlación de Spearman y depuradas por control de colinealidad.
2. Conjunto B. variables seleccionadas por Elastic Net y refinadas mediante forward selection.

En una primera etapa, estos conjuntos se utilizan para comparar las dos estrategias de selección de variables mediante un modelo de referencia, siguiendo el procedimiento descrito en la sección “Evaluación comparativa de los conjuntos A y B y selección del conjunto final”.

A partir de esta comparación se elige un conjunto de variables definitivo, que se emplea en la fase siguiente de experimentación con los distintos modelos de clasificación. El conjunto no seleccionado se conserva como referencia metodológica y para la discusión de resultados, pero no se utiliza en la comparación detallada entre algoritmos.

Evaluación Comparativa de los Conjuntos A y B y Selección del Conjunto Final

Una vez construidos los conjuntos A y B, se plantea una evaluación comparativa para decidir cuál de ellos se utilizará en la fase principal de modelado. Para ello, se define un modelo de referencia (por ejemplo, regresión logística penalizada y/o XGBoost) y se mantienen fijos todos los demás componentes del pipeline: preprocesamiento, tratamiento del desbalance, partición temporal y esquema de validación. El procedimiento es el siguiente:

1. Se entrena el modelo de referencia utilizando exclusivamente las variables del conjunto A y se evalúa su desempeño mediante PR-AUC y métricas complementarias (ROC-AUC, precisión, recall, F1) en los conjuntos de validación y prueba.
2. Se repite el mismo proceso utilizando exclusivamente las variables del conjunto B, con idénticos parámetros de preprocesamiento, tratamiento de desbalance y protocolo de evaluación.

3. Se comparan los resultados de ambos entrenamientos, prestando especial atención a la PR-AUC en validación y prueba, la estabilidad del modelo (coherencia entre métricas de validación y prueba) y el grado de parsimonia de cada conjunto (número de variables y facilidad de interpretación).

Con base en esta comparación se seleccionará un conjunto de variables definitivo para la fase de experimentación con los distintos modelos de clasificación. El conjunto no elegido se conservará como referencia metodológica y para efectos de discusión en la sección de resultados, pero no se utilizará en los experimentos principales de comparación entre algoritmos.

Diseño de los Experimentos de Modelado

Una vez definido el problema como clasificación binaria anticipada, aplicado el pipeline de preprocesamiento y construidos los conjuntos de variables A y B, la metodología contempla una fase de evaluación comparativa A vs B para escoger un conjunto de variables definitivo. A partir de esa elección, los experimentos de modelado se centran en comparar algoritmos de clasificación utilizando siempre el mismo conjunto de predictores depurado y parsimonioso.

En esta sección se describe cómo se diseñan los experimentos de modelado, asumiendo fijo el conjunto de variables definitivo seleccionado.

Modelos Considerados

La comparación se realiza entre cuatro modelos de clasificación supervisada, seleccionados por su uso extendido en datos tabulares, su capacidad para manejar relaciones no lineales y su balance entre desempeño e interpretabilidad:

1. Regresión logística penalizada (L2 y/o Elastic Net), como modelo lineal generalizado que proporciona coeficientes interpretables y sirve como referencia base.

2. Random Forest, como ensamble de árboles de decisión capaz de capturar interacciones y patrones no lineales sin requerir especificación explícita de términos de interacción.
3. Máquinas de vectores de soporte (SVM), con kernel lineal o RBF según el resultado del ajuste de hiperparámetros, para modelar fronteras de decisión potencialmente complejas.
4. XGBoost, como implementación optimizada de Gradient Boosting con árboles, frecuentemente reportada como uno de los métodos de referencia en problemas con datos tabulares.

Todos los modelos se entrenan utilizando el mismo conjunto de variables de entrada (conjunto definitivo) y el mismo esquema de preprocesamiento, de forma que las diferencias de desempeño puedan atribuirse al algoritmo y no a cambios en los datos de entrada.

Tratamiento del Desbalance de Clases

El problema presenta un desbalance severo entre la clase positiva (roturas anticipadas) y la negativa (no roturas), con prevalencias entre 1 % y 4 % según el horizonte (h) y la partición (sección 7.2). Este desbalance puede sesgar los modelos hacia la clase mayoritaria si se entrenan sin contramedidas específicas, y hace que métricas como la exactitud sean poco informativas.

Para mitigar este efecto se consideraron dos tipos de estrategias de balanceo:

1. Ponderación de clases en el algoritmo (por ejemplo, `class_weight="balanced"` en regresión logística, SVM y Random Forest, y `scale_pos_weight` en XGBoost)
2. Re-muestro del conjunto de entrenamiento, mediante oversampling de la clase minoritaria hasta igualar la mayoritaria y undersampling de la clase mayoritaria en una razón

aproximada 3:1. En todos los casos, el re-muestro se aplicó exclusivamente sobre el conjunto de entrenamiento, manteniendo intactas las particiones de validación y prueba.

Los experimentos de balanceo sobre el dataset oficial ($h=6$) muestran que, para la regresión logística L2 y las SVM lineales, la estrategia más efectiva es la ponderación de clases sin remuestro explícito. En Random Forest y XGBoost el undersampling agresivo de la clase mayoritaria produce los mayores valores de PR-AUC en validación, pero a costa de reducir drásticamente el tamaño del conjunto de entrenamiento. Dado que el objetivo es disponer de un modelo estable y reproducible, se privilegian las estrategias de ponderación de clases y, en el caso de árboles, valores intermedios de balanceo, aceptando un compromiso entre PR-AUC y robustez. Los resultados detallados de estos experimentos se presentan en el Apéndice B.

Ajuste de Hiperparámetros

Para cada modelo se define una rejilla acotada de hiperparámetros que controla su complejidad:

1. Regresión logística penalizada: intensidad de regularización (C o λ), tipo de penalización (L2 vs Elastic Net), proporción L1/L2 en el caso de Elastic Net.
2. Random Forest: número de árboles, profundidad máxima, número mínimo de muestras por hoja, número de predictores considerados en cada división.
3. SVM: tipo de kernel (lineal o RBF), parámetro de regularización C y, para kernel RBF, parámetro γ .
4. XGBoost: número de árboles, tasa de aprendizaje (η), profundidad máxima, fracción de muestras (`subsample`), fracción de variables (`colsample_bytree`) y parámetros de regularización (λ , α), entre otros.

El ajuste se realiza mediante búsqueda en rejilla y/o búsqueda aleatoria, siguiendo este

esquema general:

1. Para cada combinación de hiperparámetros se entrena el modelo sobre el conjunto de entrenamiento.
2. El modelo se evalúa en el conjunto de validación, calculando las métricas definidas en el marco teórico.
3. La métrica principal para seleccionar hiperparámetros es la PR-AUC, dada la naturaleza desbalanceada del problema. Métricas como ROC-AUC, precisión, recall y F1 se calculan como información complementaria, pero no se utilizan como criterio de selección primaria.

Este procedimiento se repite para cada uno de los cuatro modelos, siempre utilizando el mismo conjunto de variables y el mismo pipeline de preprocesamiento.

Protocolo de Entrenamiento y Evaluación

El protocolo completo de entrenamiento y evaluación sigue los pasos:

Aplicación del Pipeline de Datos. Se aplica el pipeline de preprocesamiento (imputación, winsorización, escalado) ajustado en el conjunto de entrenamiento a los conjuntos de validación y prueba. Además, se restringen los datos a las columnas correspondientes al conjunto de variables definitivo, tanto en entrenamiento como en validación y prueba.

Entrenamiento y Ajuste de Hiperparámetros. Se entrenan los modelos de cada algoritmo para todas las combinaciones de hiperparámetros en la rejilla definida. Cada modelo ajustado se evalúa en el conjunto de validación, registrando PR-AUC, ROC-AUC, precisión, recall, F1-score y la matriz de confusión para umbrales de decisión de interés.

Para cada algoritmo se selecciona la combinación de hiperparámetros que maximiza la PR-AUC en validación, manteniendo la coherencia con la naturaleza de evento raro.

Entrenamiento Final y Evaluación en Prueba. Con los hiperparámetros seleccionados, se reentrena cada modelo utilizando el conjunto de entrenamiento (y, opcionalmente, entrenamiento + validación, manteniendo el conjunto de prueba aislado). El modelo final se evalúa en el conjunto de prueba, registrando:

1. PR-AUC (métrica principal);
2. ROC-AUC;
3. precisión, recall y F1-score para umbrales seleccionados;
4. matriz de confusión;
5. medidas de lift@k, obtenidas al ordenar las observaciones por probabilidad de

rotura y calcular cuántos eventos positivos se capturan en el k % superior, simulando escenarios de priorización de alertas.

Comparación Entre Algoritmos. Se sintetizan los resultados de prueba para los cuatro modelos, comparando principalmente la PR-AUC y el comportamiento de precisión/recall y lift@k. Esta comparación permite determinar qué algoritmo ofrece el mejor compromiso entre capacidad de detección de roturas, tasa de falsos positivos y estabilidad del desempeño.

Criterios de Selección del Modelo Final e Interpretación

El modelo final candidato se seleccionará con base en los siguientes criterios:

Desempeño Predictivo. Prioridad al modelo con mayor PR-AUC en el conjunto de prueba, siempre que no presente un deterioro notable en otras métricas clave (por ejemplo, un recall excesivamente bajo o un número inaceptable de falsos negativos). Uso de ROC-AUC, F1 y la matriz de confusión para corroborar que el comportamiento global es consistente.

Parsimonia e Interpretabilidad. Preferencia por soluciones que mantengan el conjunto de variables definitivo en un tamaño manejable y que faciliten la explicación del modelo a personal

de planta. En regresión logística penalizada, análisis de coeficientes estandarizados para interpretar el efecto de cada variable sobre la probabilidad de rotura.

En Random Forest y XGBoost, análisis de importancia de variables (ganancia, frecuencia, importancia por permutación) y, cuando sea pertinente, uso de herramientas como SHAP para ilustrar contribuciones locales de las variables a predicciones individuales.

Viabilidad Operativa. Estudio de umbrales de decisión que permitan ajustar el equilibrio entre detección temprana de roturas (recall) y carga de alarmas (tasa de falsos positivos), considerando la capacidad de la planta para responder a las señales generadas. Discusión de los resultados con el conocimiento del proceso: correspondencia entre las variables más importantes según el modelo y los factores que la operación reconoce como críticos para la estabilidad de la hoja.

El modelo que mejor satisfaga estos criterios será documentado en la sección de resultados como modelo propuesto para la anticipación de roturas, junto con las métricas obtenidas y las implicaciones prácticas de su uso. De esta manera, el diseño de los experimentos de modelado cierra el ciclo planteado en los objetivos: partir de un conjunto de datos real, depurar variables de forma estructurada y seleccionar, entre varios algoritmos, una alternativa con potencial de uso operativo en la planta.

Resultados

En este capítulo se presentan los resultados empíricos del flujo de trabajo descrito en la metodología. En primer lugar, se analiza la distribución de la etiqueta anticipada en los conjuntos de entrenamiento, validación y prueba, y se comparan las dos estrategias de selección de variables propuestas (Método A basado en Spearman y Método B basado en Elastic Net + selección hacia adelante). A continuación, se entrenan y comparan varios modelos de clasificación sobre el conjunto de variables seleccionado, discutiendo tanto las métricas globales (PR-AUC, ROC-AUC) como su comportamiento operativo en términos de matriz de confusión, precisión, recall y coeficientes del modelo final.

Resultados de la Selección de Variables

Tabla 1

Distribución de la Etiqueta Anticipada por Partición

Conjunto	N observaciones	Positivos y_{ant}	% positivos aprox.
Train	9930	192	1,93
Valid	3310	37	1,11
Test	3311	125	3,78
Total	16551	354	2,14

La Tabla 1 muestra que la prevalencia global de la etiqueta anticipada es baja (2,1 %), con ligeras diferencias entre conjuntos: 1,93 % en entrenamiento, 1,11 % en validación y 3,78 % en prueba. Esta distribución justifica el uso de ponderación de clases en los modelos y la selección de PR-AUC como métrica principal de evaluación, tal como se detalló en la metodología.

Ranking de Spearman (Método A)

En la Tabla 2 se presenta el ranking de correlación de Spearman entre cada variable transformada del pipeline y la etiqueta anticipada (y_{ant}), calculado sobre el conjunto de entrenamiento después de aplicar el preprocesamiento (segmentación por bloques, imputación, winsorización y escalado robusto). Se ordenan las variables por (ρ), de mayor a menor, lo que permite identificar aquellas señales que, de manera univariada, muestran mayor asociación monótona con la probabilidad de rotura anticipada.

Tabla 2

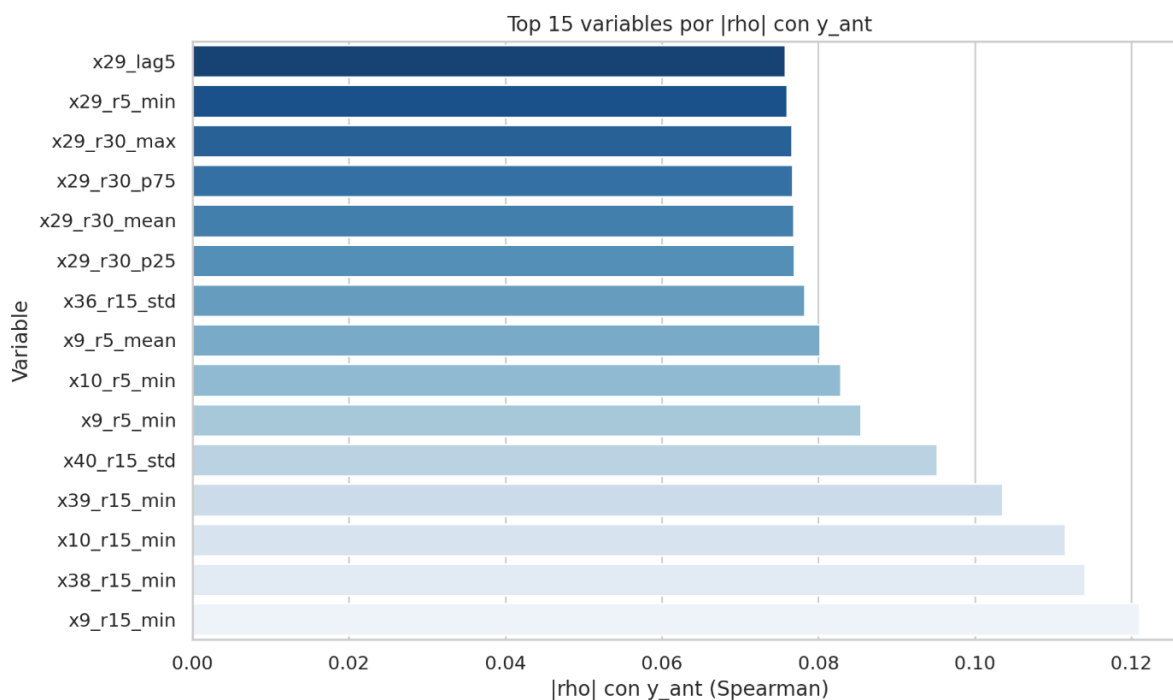
Ranking de Spearman (Método A)

Variable	$\rho(\text{Spearman}, y_{ant})$	$ \rho $
x9_r15_min	0,121	0,121
x38_r15_min	0,114	0,114
x10_r15_min	0,112	0,112
x39_r15_min	0,104	0,104
x40_r15_std	-0,095	0,095
x9_r5_min	0,085	0,085
x10_r5_min	0,083	0,083
x9_r5_mean	0,080	0,080
x36_r15_std	-0,078	0,078
x29_r30_p25	-0,077	0,077

Nota. el umbral para el Método A se fijó en el percentil 90 de la distribución de (ρ); sobre este subconjunto se aplicó posteriormente el control de colinealidad ($|\rho| \leq 0,35$) para construir el conjunto A parsimonioso.

Figura 1

Top 15 de Variables Seleccionadas, Mediante la Correlación Absoluta de Spearman con (y_{ant})



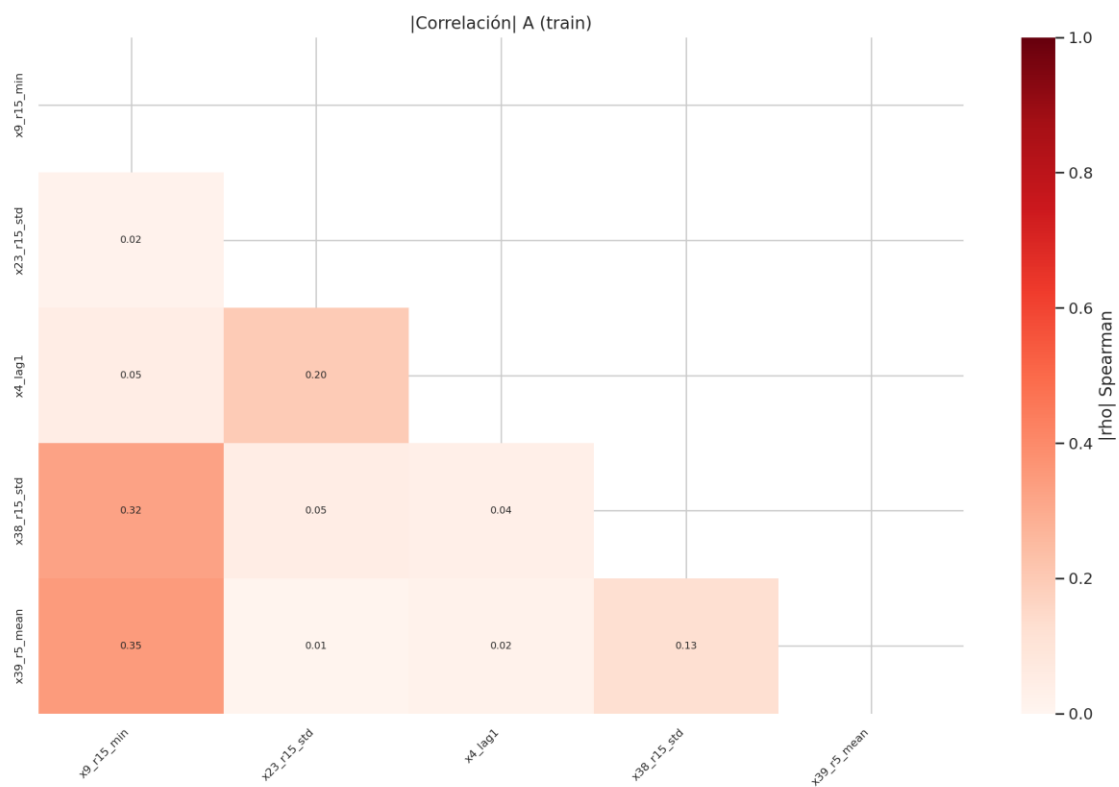
Se observa que las variables temporales derivadas (mínimos y promedios recientes) concentran los mayores valores de (ρ), reforzando la importancia de la información histórica reciente frente a los niveles puntuales.

Variables Seleccionadas por el Método A

A partir del ranking de Spearman se seleccionaron las variables con (ρ) por encima del percentil 90 y, sobre ese subconjunto, se aplicó un filtro de colinealidad por pares ($|\rho| \leq 0,35$) en el conjunto de entrenamiento. El objetivo es obtener un conjunto parsimonioso de predictores con alta asociación individual con (y_{ant}) pero baja redundancia entre sí. El resultado es el conjunto A, compuesto por cinco variables.

Tabla 3*Variables Seleccionadas por el Método A*

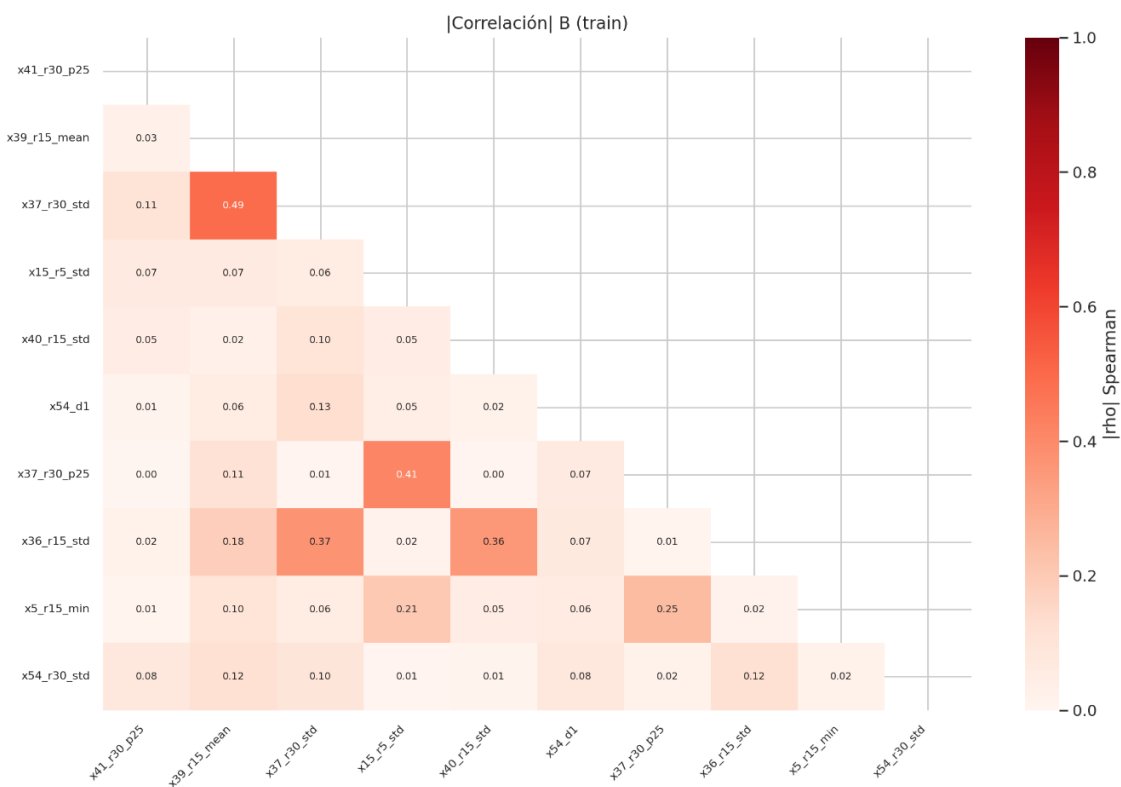
Variable	$ \rho $
x9_r15_min	0,12
x23_r15_std	0,05
x4_lag1	0,07
x38_r15_std	0,06
x39_r5_mean	0,06

Figura 2*Mapa de Calor de (ρ) Entre las Variables del Conjunto A en el Conjunto de Entrenamiento*

Se verifica que las correlaciones absolutas entre pares de variables se mantienen por debajo del umbral fijado ($|\rho| \leq 0,35$), cumpliendo el criterio de baja colinealidad para un conjunto parsimonioso.

Figura 3

Mapa de Calor de (ρ) Entre las Variables del Conjunto B en el Conjunto de Entrenamiento

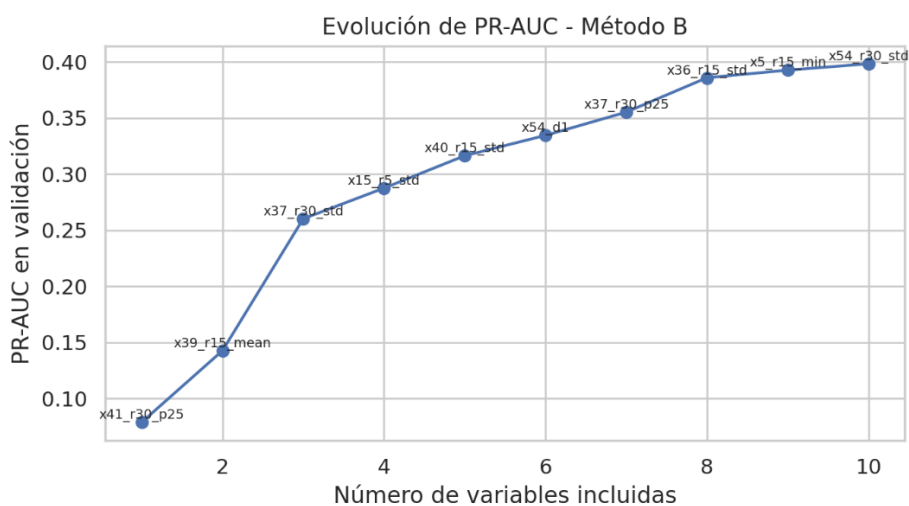


Aunque algunas variables muestran correlaciones moderadas, la mayoría de los pares se mantiene por debajo de valores altos de (ρ), lo que sugiere que cada variable aporta información adicional para mejorar la PR-AUC.

El conjunto resultante, denominado B, está formado por diez variables temporales que combinan niveles recientes, variabilidad y derivados de las señales originales:

Tabla 4*Orden de Entrada y PR-AUC Valid Acumulada*

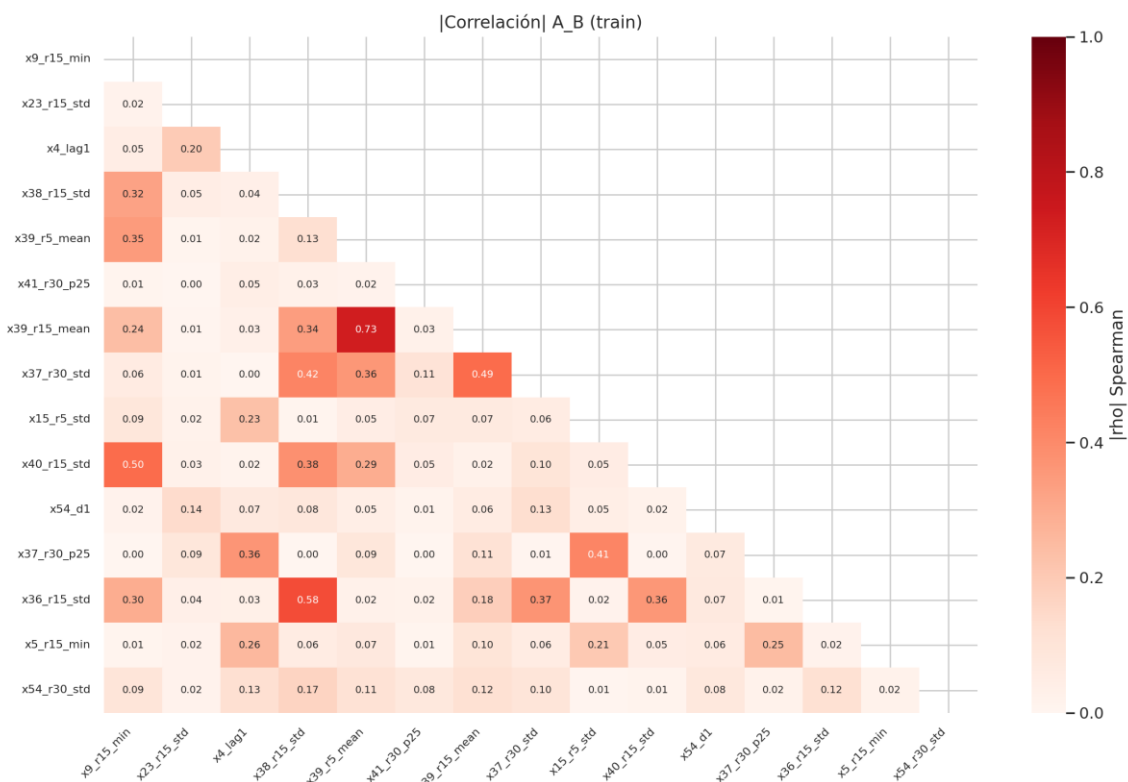
Orden	Variable	PR-AUC valid	PR-AUC test	Mejora Δ PR-AUC
1	x41_r30_p25	0,079	0,050	0,079
2	x39_r15_mean	0,143	0,057	0,064
3	x37_r30_std	0,260	0,065	0,118
4	x15_r5_std	0,288	0,076	0,027
5	x40_r15_std	0,316	0,107	0,029
6	x54_d1	0,335	0,088	0,018
7	x37_r30_p25	0,355	0,084	0,021
8	x36_r15_std	0,386	0,088	0,031
9	x5_r15_min	0,393	0,087	0,007
10	x54_r30_std	0,398	0,081	0,006

Figura 4*Evolución de PR-AUC con las Variables del Conjunto B*

La Figura 4 muestra la evolución de la PR-AUC en validación a medida que se incorporan las variables del conjunto B. Se observa una ganancia importante hasta aproximadamente la octava variable, tras lo cual las mejoras son marginales.

Figura 5

Correlación Absoluta de Spearman Entre las Variables de los Conjuntos A y B



Este mapa de calor permite comprobar que las variables de B no son simples duplicados de las de A: la mayoría de las correlaciones cruzadas se mantiene en niveles moderados, justificando el uso de B como conjunto enriquecido para los modelos.

Comparación A vs B con RL L2 de Referencia

Para comparar de forma directa los conjuntos A y B se ajustó una regresión logística con penalización L2 y ponderación de clases, utilizando exclusivamente las variables de cada

conjunto. La Tabla 5 resume la PR-AUC en validación y prueba. El conjunto B, obtenido mediante Elastic Net + selección hacia adelante, logra una mejora sustancial en PR-AUC de validación respecto al conjunto A, manteniendo un desempeño ligeramente superior en prueba.

Tabla 5

Comparativo de los Resultados de los Conjuntos A y B

Método	Nº variables	PR-AUC valid	PR-AUC test
A – Spearman	5	0,0306	0,0778
B – ElasticNet + FS	10	0,3984	0,0811

En validación, el conjunto B multiplica por más de diez la PR-AUC frente al conjunto A (0,40 vs 0,03), lo que indica que la ingeniería temporal y la selección secuencial aportan mucha más señal predictiva que el filtro puramente univariado. En prueba, ambos conjuntos logran PR-AUC similares (0,08), con una ligera ventaja de B, por lo que se adopta el conjunto B como referencia para los modelos de clasificación del capítulo siguiente, manteniendo el conjunto A únicamente como punto de comparación parsimonioso.

Resultados de los Modelos de Clasificación

Comparación y Selección del Modelo Final

Tabla 6

Comparación de Modelos de Clasificación (conjunto B)

Modelo	PR-AUC	PR-AUC	ROC-AUC	ROC-AUC
	valid	test	valid	test
Logit L2 (C = 10)	0,3986	0,0811	0,8670	0,7042
Logit L2 (C = 1)	0,3984	0,0811	0,8670	0,7043

Modelo	PR-AUC	PR-AUC	ROC-AUC	ROC-AUC
	valid	test	valid	test
Logit L2 (C = 0.1)	0,3978	0,0812	0,8668	0,7043
SVM lineal (C = 1)	0,2763	0,0774	0,8745	0,7056
SVM lineal (C = 0.1)	0,2757	0,0773	0,8745	0,7054
Random Forest (mejores variantes)	0,0923	0,0659	0,8459	0,6843
XGBoost (600 árboles, prof. 5, lr = 0,1)	0,1297	0,0390	0,6126	0,5119

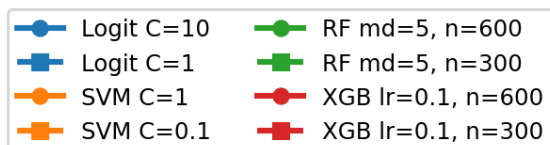
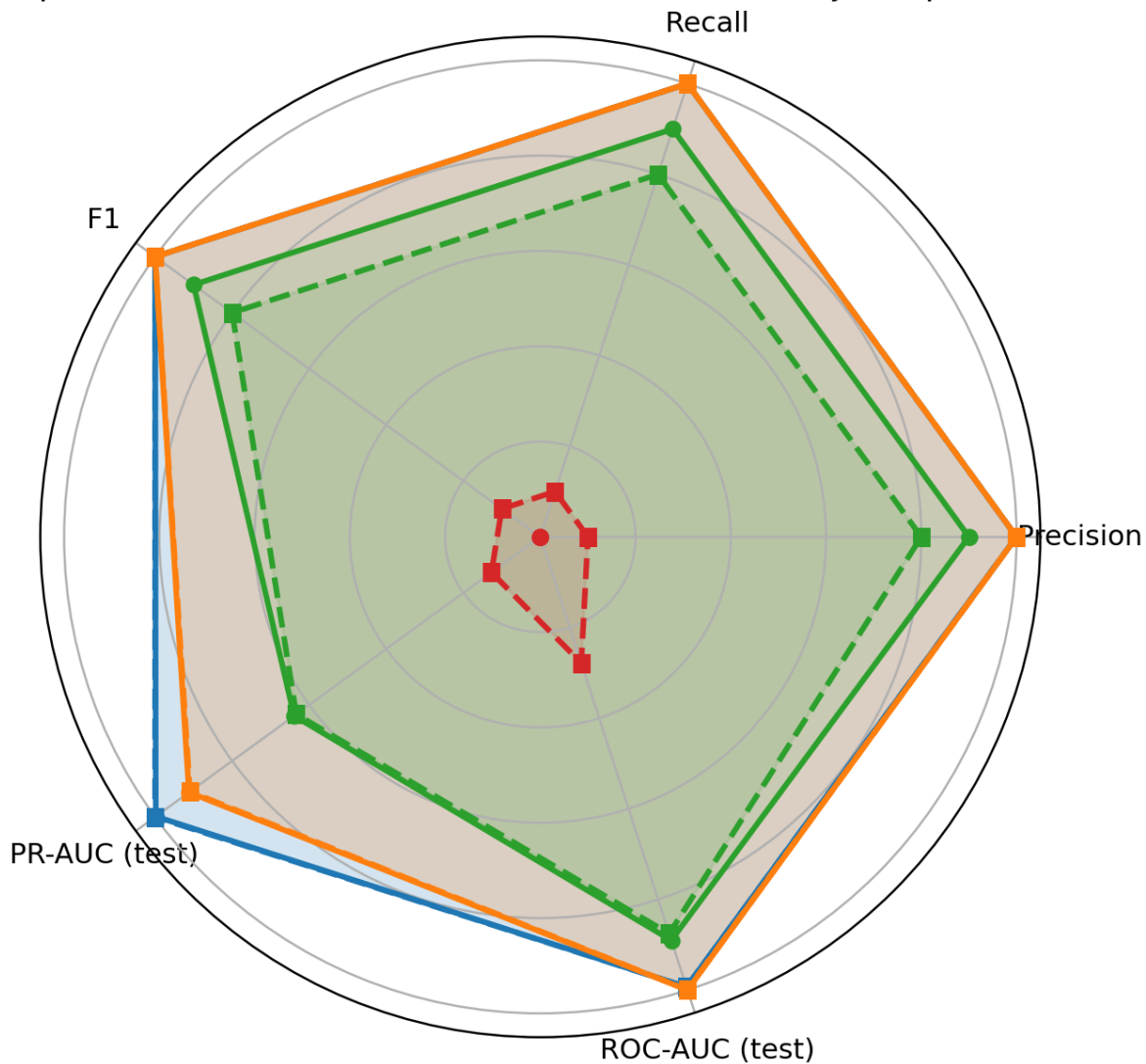
La Tabla 6 presenta la comparación de las principales familias de modelos (regresión logística L2, SVM lineal, Random Forest y XGBoost) sobre el conjunto B de variables, utilizando el horizonte de anticipación ($h=6$) y el esquema de partición temporal. En todos los casos se emplean los hiperparámetros seleccionados y cuando corresponde, ponderación de clases (`class_weight="balanced"` o `scale_pos_weight`) sin re-muestro explícito. La métrica de referencia es la PR-AUC de validación, complementada con la PR-AUC y ROC-AUC en prueba.

Bajo este criterio, la regresión logística L2 con ($C=10$) y `class_weight="balanced"` obtiene la mejor PR-AUC en validación (0,399) y un desempeño estable en prueba (PR-AUC 0,081; ROC-AUC 0,704). Las SVM lineales ($C=1$) logran ROC-AUC ligeramente superiores (0,874) pero con PR-AUC valid 0,276, lo que indica una menor capacidad para concentrar los pocos eventos de rotura en la fracción superior de la lista ordenada por probabilidad. Random Forest y XGBoost ofrecen ROC-AUC competitivas, pero PR-AUC claramente inferiores, especialmente en prueba, y mayor sensibilidad a las configuraciones de balanceo.

Figura 6

Mejores Modelos por Familia, Métricas Normalizadas

Comparación de modelos (radar normalizado, 2 mejores por familia)

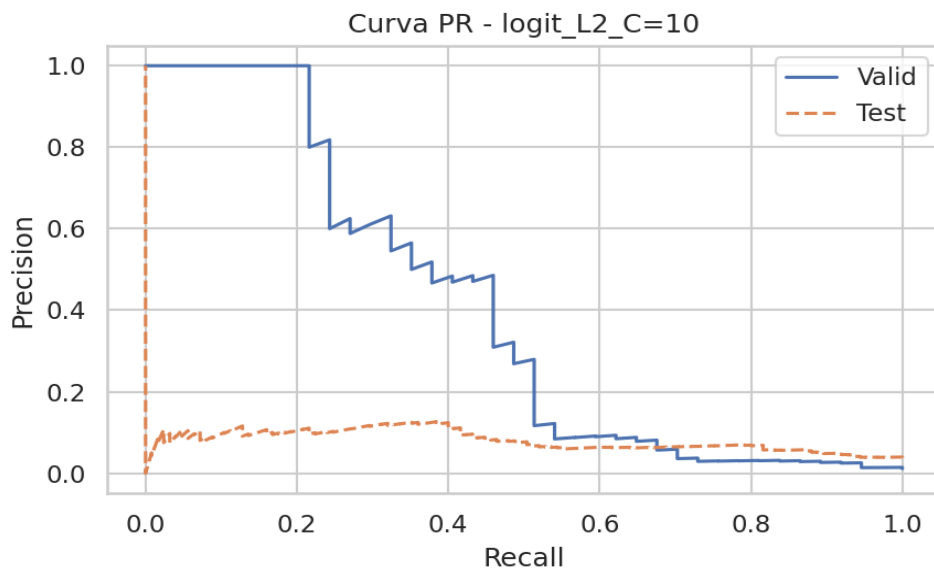


La Figura 6 muestra un gráfico de radar que resume, para los dos mejores modelos de cada familia, la precisión, recall y F1 al seleccionar el 5 % de los casos con mayor probabilidad estimada, junto con la PR-AUC y ROC-AUC en prueba. En esta visualización, el polígono correspondiente a la regresión logística C=10 presenta el área más amplia y equilibrada, con valores relativamente altos en todas las métricas. De acuerdo con el punto operativo adoptado (top-5 % de casos por probabilidad), el modelo final alcanza una precisión 9,7 %, un recall 12,8 % y un F1 0,11 en el conjunto de prueba, metas razonables en un contexto de prevalencia inferior al 3 %.

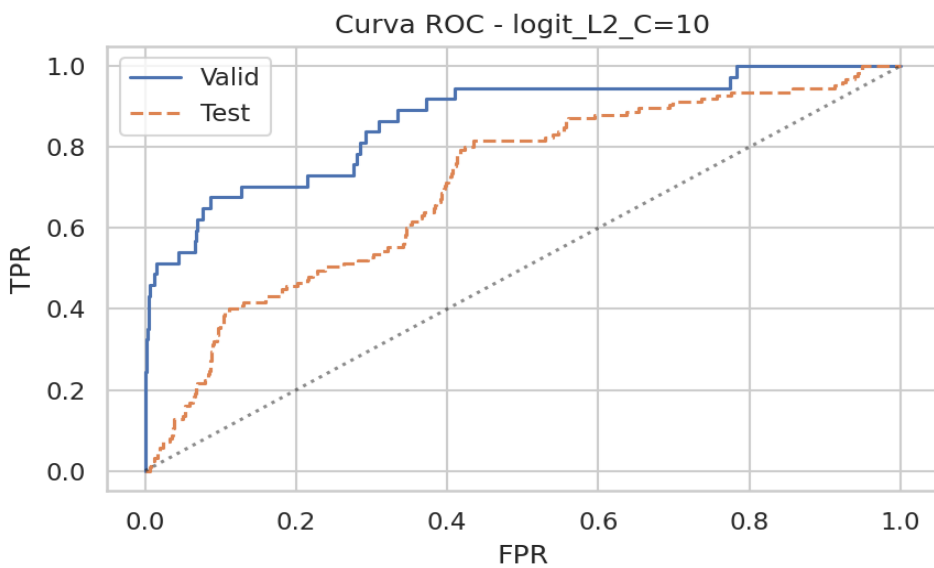
Considerando conjuntamente el desempeño cuantitativo (PR-AUC, ROC-AUC y métricas top-5 %), la estabilidad entre validación y prueba, el comportamiento frente a distintas estrategias de balanceo y la interpretabilidad de los coeficientes, se selecciona la regresión logística penalizada L2 con (C=10), `class_weight="balanced"` y horizonte (h=6) como modelo final para anticipar roturas del papel.

Figura 7

Curvas Precision–Recall del Modelo Logístico L2 Seleccionado en Validación y Prueba

**Figura 8**

Curvas ROC del Modelo Logístico L2 Seleccionado en Validación y Prueba



Las curvas Precision–Recall (Figura 7) confirman que el modelo logístico L2 domina a las alternativas consideradas en el rango de recalls relevante para la operación: para niveles bajos de recall, la precisión se mantiene sistemáticamente por encima de los modelos SVM y Random Forest. Las curvas ROC (Figura 8) muestran valores de ROC-AUC en torno a 0,70–0,71 para todos los modelos, lo que indica una capacidad moderada de discriminación; sin embargo, dado el fuerte desbalance de clases, la PR-AUC es más sensible a los cambios en la detección de la esquemacalse positiva y guía la selección del modelo final.

Análisis de Errores y Comportamiento Operativo

Figura 9

Matriz de Confusión del Modelo Logístico L2 en el Conjunto de Prueba

Matriz de confusión top-5%

True 0	3037	149
True 1	109	16
	Pred 0	Pred 1

Nota. al seleccionar el 5 % de las observaciones con mayor probabilidad (top-5 %).

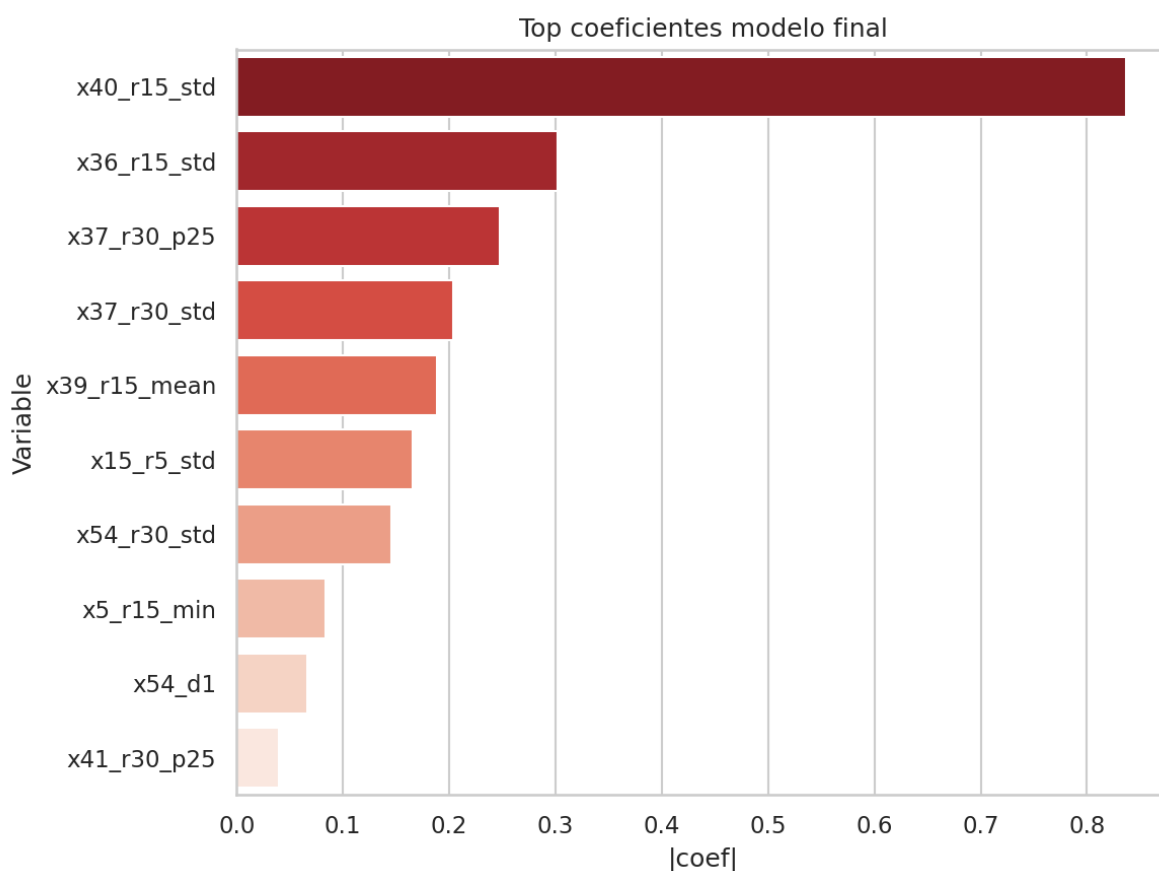
Con este umbral, el modelo identifica 16 roturas de 125 (recall 13 %), al costo de 149 falsos positivos (precisión 10 %). Aunque la PR-AUC global es baja, estas cifras permiten cuantificar el compromiso entre cantidad de alarmas y número de eventos anticipados, y pueden

ajustarse en función de la capacidad de la planta para responder a las alertas.

Si no se dispusiera de modelo y se seleccionaran aleatoriamente el 5 % de las observaciones de prueba, se esperaría capturar aproximadamente el 5 % de las roturas (en este caso, unas 6 de 125). Con el modelo logístico L2, el mismo 5 % superior de probabilidades captura 16 roturas (13 %), lo que supone un lift de aproximadamente 2,5 frente a la selección aleatoria. Aunque el número absoluto de aciertos sigue siendo modesto debido al bajo número de eventos positivos, este resultado indica que el modelo concentra una fracción mayor de roturas en los casos priorizados y que puede servir como herramienta de apoyo a la priorización de inspecciones o ajustes de proceso.

Figura 10

Principales Coeficientes Estandarizados del Modelo Logístico L2 Seleccionando el Conjunto B



Las variables con coeficientes positivos indican condiciones que incrementan la probabilidad de rotura anticipada (por ejemplo, mayores niveles o variabilidad reciente), mientras que los coeficientes negativos se asocian a situaciones más estables. Este análisis permite relacionar las señales seleccionadas con el conocimiento de proceso y refuerza la interpretabilidad del modelo. En conjunto, los resultados de este capítulo muestran que, pese a las limitaciones inherentes al desbalance extremo y al tamaño muestral, es posible construir un modelo logístico relativamente parsimonioso apoyado en un conjunto reducido de variables temporales que captura una fracción significativa de las roturas en el percentil superior de riesgo. En el capítulo siguiente se discuten con mayor detalle la interpretación de estas variables, la coherencia de los hallazgos con la literatura y las implicaciones prácticas de utilizar este tipo de modelo en la planta.

Discusión

Interpretación de Variables Relevantes

Los resultados de la selección de variables y del modelo final confirman la hipótesis de que las roturas del papel rara vez dependen de un valor puntual de un sensor, sino de patrones temporales que combinan nivel, variabilidad y cambios recientes. El conjunto A, obtenido mediante filtrado univariado con correlación de Spearman y control estricto de colinealidad, está formado por cinco variables temporales asociadas a mínimos, desviaciones estándar y lags de señales seleccionadas. El conjunto B, construido mediante Elastic Net seguido de selección hacia adelante, amplía esta idea con diez variables que recogen percentiles, medias y dispersiones en ventanas de 5, 15 y 30 minutos, además de derivadas de primer orden.

En el modelo logístico penalizado seleccionado (L2 sobre el conjunto B), los coeficientes estandarizados muestran que las variables que resumen niveles bajos recientes y aumentos de variabilidad juegan un papel central. Por ejemplo, los percentiles bajos en ventanas largas (como $x_{41_r30_p25}$) capturan situaciones en las que el proceso ha pasado una fracción significativa del último período en niveles más exigentes o cercanos a límites operativos, lo que puede reflejar condiciones de tensión sobre la hoja. De forma complementaria, variables como $x_{37_r30_std}$, $x_{15_r5_std}$ o $x_{36_r15_std}$ recogen el “nerviosismo” de ciertos sensores en horizontes de 10 a 60 minutos; coeficientes positivos sobre estas variables indican que episodios de alta variabilidad reciente se asocian con un mayor riesgo de rotura anticipada.

Las variables basadas en medias y mínimos en ventanas cortas (por ejemplo $x_{39_r15_mean}$, $x_{5_r15_min}$) permiten al modelo distinguir entre periodos de operación relativamente estables y periodos en los que las condiciones de proceso se han deteriorado de forma sostenida. Finalmente, la presencia de derivadas (como x_{54_d1}) refuerza la idea de que no

solo importan los niveles, sino también la dirección y velocidad del cambio: incrementos rápidos en determinados sensores se traducen en coeficientes positivos que aumentan la probabilidad estimada de rotura, mientras que coeficientes negativos en otras variables apuntan a estados que, en promedio, se han asociado con transiciones hacia condiciones más estables.

En conjunto, la lectura de los coeficientes del modelo y de las variables seleccionadas sugiere que las roturas tienden a ir precedidas por ventanas donde algunos sensores se han mantenido cerca de niveles límite y han exhibido una volatilidad elevada, a menudo acompañada de cambios bruscos. Este patrón es coherente con la intuición de operación: antes de que la hoja se rompa, el proceso suele volverse más inestable y el margen de maniobra se reduce, algo que queda capturado por las combinaciones de mínimos, medias, desviaciones estándar y derivadas utilizadas en el conjunto B.

Relación con la Literatura y el Marco Teórico

Desde el punto de vista metodológico, los resultados obtenidos se alinean con varios hallazgos reportados en la literatura sobre aprendizaje supervisado en datos tabulares e imbalanced learning. En primer lugar, el hecho de que una regresión logística penalizada con un conjunto de variables cuidadosamente diseñadas supere en PR-AUC a modelos más complejos como Random Forest y XGBoost es consistente con trabajos recientes que destacan la competitividad de modelos lineales y árboles simples cuando se dispone de una buena ingeniería de características y un tratamiento adecuado del desbalance (Grinsztajn et al., 2022; Holzmüller et al., 2024) En este estudio, la combinación de penalización L2, ponderación de clases y variables temporales robustas ha sido suficiente para lograr el mejor compromiso entre desempeño y parsimonia.

En segundo lugar, la elección de PR-AUC como métrica principal y la interpretación de

resultados a partir de curvas Precision–Recall responden a las recomendaciones específicas para problemas de clasificación con clases raras (Saito & Rehmsmeier, 2015; Owusu-Adjei et al., 2023) Aunque las ROC-AUC obtenidas rondan 0,70–0,71 para varios modelos, la PR-AUC en prueba se mantiene en valores mucho más modestos (0,08), lo que subraya la dificultad intrínseca del problema cuando la prevalencia de roturas anticipadas está en torno al 2 %. Esta discrepancia entre ROC-AUC y PR-AUC refleja precisamente el sesgo de las métricas basadas en tasas cuando la clase negativa domina el conjunto de datos.

La comparación entre los métodos de selección A y B también coincide con la literatura sobre selección de variables en contextos de alta colinealidad. El método A, basado en Spearman con control de $|\rho| \leq 0,35$, produce un conjunto muy parsimonioso y fácilmente interpretable, pero con capacidad predictiva limitada (PR-AUC valid 0,03). El método B, en cambio, combina un primer filtrado con Elastic Net con una selección hacia adelante guiada por PR-AUC, lo que le permite retener variables que pueden ser débiles de forma univariada pero relevantes en combinación, algo señalado por trabajos sobre regularización en modelos lineales y selección híbrida de características (Tay et al., 2023; Pudjihartono et al., 2022) El salto de PR-AUC en validación de A a B (de 0,03 a 0,40) ilustra el valor de estas estrategias híbridas.

Por último, el tratamiento del desbalance mediante ponderación de clases y, en el caso de XGBoost, mediante `scale_pos_weight`, sigue las recomendaciones habituales en problemas de eventos raros industriales (Chen et al., 2024; Rezvani & Wang, 2023) Los resultados muestran, sin embargo, que estos mecanismos no son suficientes por sí solos para compensar la falta de señal cuando el número de eventos positivos es muy reducido y la relación señal-ruido es baja, lo que sugiere que la calidad y riqueza de las variables siguen siendo un factor determinante.

Además, al comparar los resultados de este trabajo con los reportados en estudios previos

sobre la misma base de datos de ProcessMiner, se observa un patrón consistente: incluso utilizando modelos avanzados y esquemas sofisticados de ingeniería de variables o aumento de datos, las métricas de desempeño siguen siendo modestas en términos absolutos. Por ejemplo, los modelos de (Ranjan et al., 2019) alcanzan F1-scores en el rango de 0,082 a 0,114 para predicciones 2 y 4 minutos antes de la rotura, mientras que (Lee & Seo, 2021) obtienen un AUC de 0,724 y un F-measure de 0,097 con su esquema RD-LNN basado en vecinos más cercanos.

Otros trabajos, como los de (Xu et al., 2022) y (Khosravi et al., 2023), muestran mejoras relativas (20 % en F1 y precisión mediante selección de datos, o aumentos superiores al 30% a 90 % en la detección de la clase minoritaria mediante CTGAN), pero siguen reportando valores de F1 y AUC en torno a 0,10 y 0,70 a 0,73.

De manera complementaria, el mejor modelo de este estudio una regresión logística L2 con ponderación de clases obtiene una ROC-AUC de 0,704 y una PR-AUC de 0,081 en el conjunto de prueba, situándose dentro del rango de desempeño observado en la literatura pese a utilizar un modelo lineal relativamente simple. Esto sugiere que la principal limitación no es tanto el algoritmo, sino la extrema rareza del evento y la limitada señal informativa disponible en las variables, lo que refuerza la necesidad de seguir explorando estrategias de enriquecimiento de datos y de diseño de variables más específicas para la operación de la máquina.

Tabla 7*Comparación de Resultados con Trabajos Previos Sobre*

Trabajo	Enfoque / modelo principal	Datos / horizonte	Métrica(s) principal(es)	Resultado reportado y comparación con este trabajo
Este trabajo (logit L2)	Regresión logística L2 (C = 10) con class_weight = "balanced", sin remuestreo explícito	Dataset ProcessMiner (18398 registros, 124 roturas), mediciones cada 2 min; clasificación del registro como rotura/no rotura	PR-AUC (validación y prueba), ROC-AUC (validación y prueba)	PR-AUC valid = 0,399; PR-AUC test = 0,081; ROC-AUC valid = 0,867; ROC-AUC test = 0,704. Desempeño en ROC-AUC muy similar al de los mejores modelos publicados sobre este conjunto de datos, con PR-AUC baja pero coherente con la extrema rareza de la clase positiva.
(Ranjan et al., 2019) (dataset ProcessMiner)	Ensamble de XGBoost/AdaBoost con fuerte ingeniería de características (derivadas, FFT, interacciones)	Misma base ProcessMiner; predicción 2 y 4 minutos antes del evento (desplazamiento de la columna de clase)	F1-score (métrica principal), precisión, recall, FPR	Mejor modelo reporta F1 = 0,082 (2 min antes) y F1 = 0,114 (4 min antes) en prueba. A pesar del alto nivel de complejidad del modelo, el F1 se mantiene en el orden de 0,10, similar en magnitud a la PR-AUC test = 0,081 de la regresión logística de este trabajo.
(Lee & Seo, 2021)	Extracción de características basada en vecino local (RD-LNN) + SVM	Misma base ProcessMiner; predicción 2 minutos antes de la rotura mediante ventanas de series de tiempo	AUC ROC, F-measure, TPR, FPR	Para el modelo RD-LNN+SVM se reporta AUC = 0,724; F-measure = 0,097; TPR = 0,120; FPR = 0,019. El AUC mejora ligeramente frente a este trabajo (0,724 vs 0,704), pero el F-measure sigue en torno a 0,10, lo que confirma que incluso con ingeniería de características

Trabajo	Enfoque / modelo principal	Datos / horizonte	Métrica(s) principal(es)	Resultado reportado y comparación con este trabajo
(Xu et al., 2022)	Autoencoder para detección temprana + selección del conjunto de entrenamiento (exclusión de determinados grados de papel)	Misma base ProcessMiner; horizonte de predicción temprano (formulado como rare event detection en streams multivariados)	F1-score, precisión, FPR, AUC	<p>Reportan que al eliminar del entrenamiento datos de cierta categoría de producto, el F1 y la precisión del modelo se incrementan en torno a un 20 %, manteniendo AUC en el entorno de 0,70–0,73. El nivel de desempeño sigue siendo comparable al de Ranjan et al. y Lee & Seo, y consistente con el rango de ROC-AUC de este trabajo (0,70).</p> <p>Sin dar todos los valores absolutos, muestran que, tras el aumento con CTGAN, la capacidad de detectar roturas (clase 1) aumenta >30 % para árboles de decisión, 20 % para Random Forest y casi 90 % para regresión logística, en comparación con entrenamiento sin aumento. Sus resultados confirman que el desbalance extremo exige técnicas de enriquecimiento de datos más allá de la simple ponderación de clases usada en este trabajo.</p>
(Khosravi et al., 2023)	Aumento de datos con CTGAN + SMOTE y modelos DT, RF y LR	Misma base ProcessMiner (18398 registros, 124 roturas), mismo esquema general de ventanas de 2 min	Métricas de detección de la clase 1 (recall / capacidad de detección), precisión	<p>Sin dar todos los valores absolutos, muestran que, tras el aumento con CTGAN, la capacidad de detectar roturas (clase 1) aumenta >30 % para árboles de decisión, 20 % para Random Forest y casi 90 % para regresión logística, en comparación con entrenamiento sin aumento. Sus resultados confirman que el desbalance extremo exige técnicas de enriquecimiento de datos más allá de la simple ponderación de clases usada en este trabajo.</p>

Implicaciones Prácticas en la Planta

Desde un punto de vista operativo, los resultados deben interpretarse con prudencia pero también con un enfoque pragmático. El modelo logístico L2 seleccionado, entrenado sobre el conjunto B, alcanza una PR-AUC de prueba en torno a 0,08 y una ROC-AUC cercana a 0,70; estas cifras indican que el modelo tiene cierta capacidad para discriminar estados de mayor riesgo de rotura, pero el modelo no constituye un sistema de predicción perfecto ni libre de errores, sino una herramienta con capacidad discriminativa moderada que debe interpretarse como apoyo a la toma de decisiones y no como un mecanismo determinista de alarmas.

Al analizar el comportamiento en un umbral operativo concretamente definido (top-5 % de las observaciones con mayor probabilidad de rotura), el modelo identifica 16 de las 125 roturas en prueba (recall 13 %) con aproximadamente 149 falsos positivos (precisión 10 %). Esto implica que muchas alertas no se traducirían en roturas efectivas, pero también que el 5 % de observaciones priorizadas contiene alrededor de 2,5 veces más roturas de las que se encontrarían seleccionando casos al azar en la misma proporción.

En términos de implantación en planta, lo anterior sugiere que el modelo puede utilizarse como un módulo de priorización de riesgo, más que como un sistema de disparo de alarmas “duras”. Por ejemplo, podría integrarse en un panel que ordene los intervalos recientes de operación por probabilidad de rotura anticipada, permitiendo a los operadores revisar con más detenimiento los periodos marcados como top-5 % y contrastarlos con su experiencia y otros indicadores de proceso. Ajustando el valor de (k) (por ejemplo, top-2 %, top-10 %) y el umbral de probabilidad, la planta puede explorar diferentes compromisos entre número de alarmas y roturas capturadas, siempre con la conciencia de que el modelo no elimina el riesgo de falsos negativos ni de falsos positivos.

Otra implicación práctica importante es la interpretabilidad del modelo final. Al tratarse de una regresión logística con un conjunto relativamente reducido de variables, es posible explicar de forma cualitativa por qué se asigna una probabilidad de rotura alta a un determinado intervalo: incrementos en la variabilidad de ciertos sensores, mínimos recientes más bajos de lo habitual o derivadas positivas pronunciadas contribuyen a elevar el score de riesgo. Esta trazabilidad facilita la aceptación del modelo por parte de personal de operación y mantenimiento, frente a alternativas más opacas como modelos de boosting profundos o redes neuronales. Además, las variables seleccionadas pueden servir como base para revisar y ajustar las estrategias de control existentes (por ejemplo, límites de operación o políticas de mantenimiento preventivo) si se observa que determinados patrones anteceden de forma recurrente a las roturas.

Finalmente, los resultados destacan la necesidad de considerar al modelo como un complemento a las prácticas actuales y no como sustituto de la experiencia de los operadores. La combinación de un número limitado de roturas históricas, la naturaleza compleja del proceso y la variabilidad de condiciones entre campañas hacen que cualquier modelo puramente basado en datos tenga un techo de desempeño. No obstante, el trabajo desarrollado muestra que es posible extraer señal útil de los datos disponibles, estructurar un pipeline reproducible y sentar las bases para futuras mejoras, ya sea incorporando nuevos sensores, extendiendo el horizonte de datos históricos o refinando las técnicas de modelado y evaluación.

Conclusiones y Recomendaciones

Conclusiones

Construcción y Depuración del Conjunto de Datos

Se logró construir un conjunto de datos multivariado y reproducible para la anticipación de roturas del papel, a partir del dataset original de ProcessMiner. La etiqueta anticipada (y_{ant}) se definió de forma causal con un horizonte ($h=6$) (12 minutos), respetando la estructura temporal por bloques y evitando fuga de información. El pipeline de preprocesamiento segmentación por gaps ≥ 15 minutos, control de calidad, imputación mediante forward fill y mediana de entrenamiento, winsorización robusta basada en IQR y escalado robusto (mediana/IQR) permitió obtener un dataset limpio y estandarizado, con una partición temporal 60/20/20 en entrenamiento, validación y prueba que refleja la prevalencia real de la clase positiva (2,1 % global).

Selección de un Conjunto Parsimonioso de Variables

La comparación entre los dos enfoques de selección de variables planteados cumplió el objetivo de obtener un conjunto parsimonioso con baja colinealidad. El Método A, basado en correlación de Spearman y control estricto de ($|\rho| \leq 0.35$), produjo un conjunto reducido (A) de cinco variables temporales, muy interpretable pero con PR-AUC limitada (0,03 en validación). El Método B, que combina un primer filtrado mediante Elastic Net con una selección hacia adelante guiada por PR-AUC, generó un conjunto ampliado (B) de diez variables que capturan niveles, variabilidad, percentiles y derivadas en ventanas de 5, 15 y 30 minutos. Este conjunto B incrementó sustancialmente la capacidad predictiva respecto a A (PR-AUC valid 0,40 frente a 0,03), por lo que se adoptó como conjunto definitivo para la fase de modelado, manteniendo A como referencia parsimoniosa.

Comportamiento de los Modelos de Clasificación Bajo Desbalance Severo

Los experimentos de modelado confirmaron que, en presencia de desbalance extremo y número limitado de eventos, la calidad de las variables y el tratamiento del desbalance resultan más determinantes que la complejidad del algoritmo. Entre los modelos considerados regresión logística penalizada, Random Forest, SVM lineal y XGBoost, todos con ponderación de clases y/o `scale_pos_weight` el mejor desempeño se obtuvo con una regresión logística L2 entrenada sobre el conjunto B, con PR-AUC 0,40 en validación y 0,08 en prueba, y ROC-AUC en torno a 0,70. SVM lineal alcanzó ROC-AUC similares pero PR-AUC más baja, mientras que Random Forest y XGBoost quedaron claramente por debajo en la métrica principal, lo que justifica la elección de la regresión logística penalizada como modelo de referencia.

Capacidad Anticipativa y Utilidad Operativa del Modelo Final

Aunque la PR-AUC de prueba es moderada, el análisis operativo en un umbral concreto (top-5 % de observaciones con mayor probabilidad de rotura) muestra que el modelo concentra una fracción mayor de eventos positivos que la selección aleatoria, logrando un lift aproximado de 2,5. En este escenario específico, el modelo anticipa 16 de 125 roturas al costo de un número relevante de falsos positivos, lo que lo sitúa más como herramienta de priorización de riesgo que como sistema de alarmas determinista. Así, se cumple el objetivo de seleccionar una alternativa con potencial de uso operativo como apoyo a la toma de decisiones por ejemplo, para priorizar inspecciones o ajustes en periodos considerados de mayor riesgo siempre que se tenga en cuenta el compromiso entre recall y carga de alertas.

Interpretabilidad y Coherencia con el Conocimiento de Proceso

El enfoque adoptado permitió mantener la parsimonia y la interpretabilidad como criterios centrales. El conjunto B está compuesto por un número acotado de variables temporales

con significado claro (mínimos, medias, desviaciones estándar, percentiles y derivadas), y la regresión logística penalizada ofrece coeficientes que cuantifican la dirección e intensidad de la asociación entre cada variable y el riesgo de rotura. Los patrones identificados, niveles bajos recientes, alta variabilidad y cambios abruptos que preceden a las roturas, son coherentes con la experiencia de operación en líneas de papel y con la literatura sobre monitoreo de procesos industriales, lo que refuerza la plausibilidad del modelo más allá de las métricas numéricas.

Aporte Metodológico y Limitaciones

Más allá del desempeño numérico, el trabajo aporta una metodología estructurada para abordar problemas de clasificación de eventos raros en procesos industriales: definición cuidadosa de la etiqueta anticipada, segmentación temporal por bloques, pipeline reproducible de preprocesamiento, selección de variables híbrida (Spearman + Elastic Net + forward) y comparación de modelos con énfasis en PR-AUC y análisis operativo. Al mismo tiempo, se reconoce que la señal disponible está limitada por el número de roturas y el carácter ruidoso de los datos; en consecuencia, el modelo final debe entenderse como una primera aproximación útil pero perfectible, cuya mejora requerirá mayor volumen de datos, incorporación de nuevas variables de proceso o redefinición del horizonte y la etiqueta en función de las necesidades específicas de la planta.

Comparación con la Literatura

La comparación de los resultados obtenidos en este trabajo con los reportados en estudios previos que utilizan el mismo conjunto de datos de (ProcessMiner, 2020) muestra que el desempeño del modelo logístico seleccionado se sitúa dentro del rango esperado para un problema de eventos extremadamente raros. Tal como se analizó en la sección 9.2, trabajos como los de (Ranjan et al., 2019), (Lee & Seo, 2021), (Xu et al., 2022). y (Khosravi et al., 2023),

informan F1-scores, valores de AUC y tasas de detección de la clase positiva de magnitud comparable, incluso cuando incorporan modelos más complejos, arquitecturas profundas o técnicas avanzadas de aumento de datos. En este contexto, la regresión logística penalizada empleada en este proyecto alcanza una ROC-AUC cercana a 0,70 y una PR-AUC de prueba del orden de 0,08, lo que indica que el modelo se comporta de forma consistente con la literatura disponible. Esto sugiere que la principal limitación se encuentra en la rareza extrema del evento y en la capacidad informativa de las variables disponibles, más que en la elección específica del algoritmo de clasificación.

En síntesis, los resultados obtenidos permiten considerar cumplido el objetivo general del proyecto, en la medida en que se construyó y depuró un conjunto de datos anticipado para la rotura del papel, se seleccionó un subconjunto parsimonioso de variables con baja colinealidad y se evaluaron y compararon diversos modelos de clasificación supervisada mediante métricas robustas al desbalance. La regresión logística penalizada L2, entrenada sobre el conjunto B de variables temporales, se posiciona como un modelo candidato con potencial de uso operativo para la priorización de periodos de mayor riesgo de rotura.

Recomendaciones

Profundizar en la Ingeniería de Variables y en la Etiqueta

Los resultados muestran que buena parte de la capacidad predictiva proviene de las variables temporales construidas sobre ventanas deslizantes, percentiles y derivadas. Se recomienda explorar, en trabajos futuros, variantes adicionales de ingeniería de características, por ejemplo:

1. Ventanas con horizontes ligeramente distintos (por ejemplo, 20 y 45 minutos) y estadísticas robustas adicionales (medianas, rangos intercuartílicos).

2. Combinaciones simples entre variables de proceso físicamente relacionadas (ratios, diferencias o spreads entre sensores de la misma sección de la máquina).
3. Revisión de la definición de la etiqueta en función de las necesidades de la planta, evaluando otros horizontes de anticipación y, si se dispone de experticia adicional, etiquetas que reflejen distintos tipos de rotura o niveles de severidad. Estas extensiones deben mantener el principio de causalidad temporal y evitar la fuga de información, replicando el esquema de bloques y partición temporal adoptado en este trabajo.

Reforzar la Validación Temporal y la Estimación de Incertidumbre

Dado el carácter fuertemente desbalanceado del problema y el número limitado de eventos de rotura, pequeñas variaciones en los datos pueden tener un impacto importante en las métricas. Se recomienda complementar la partición única en train/valid/test con esquemas de validación temporal más robustos, como “rolling origin” o validación por bloques de tiempo, que permitan estimar mejor la variabilidad del desempeño del modelo. Esto ayudaría a cuantificar la incertidumbre asociada a las métricas de PR-AUC y a evitar conclusiones excesivamente optimistas o pesimistas basadas en un único split.

Analizar Diferentes Políticas de Umbral en Función del Contexto Operativo

El análisis realizado con un umbral de top-5 % de observaciones muestra que el modelo puede concentrar una fracción mayor de roturas en la cola de mayor riesgo, pero también implica un número considerable de falsos positivos. Antes de cualquier despliegue en planta, se recomienda:

1. Evaluar explícitamente varios valores de k (por ejemplo, 1 %, 2 %, 5 %, 10 %) y calcular, para cada uno, la combinación de recall, precisión y lift sobre la clase positiva.

2. Discutir con los responsables de operación cuál es la cantidad de alarmas diaria o por turno que la planta está en condiciones de gestionar y cuál es el costo relativo de un falso positivo frente a un falso negativo.

3. Seleccionar una política de umbral que equilibre de forma razonable la carga de trabajo asociada a las alertas y el beneficio de anticipar una fracción de las roturas.

Integrar el Modelo como Herramienta de Apoyo, no como Reemplazo del Criterio Experto

La regresión logística penalizada sobre el conjunto B ofrece interpretabilidad y cierta capacidad anticipativa, pero no elimina la incertidumbre inherente al proceso ni garantiza la detección de todas las roturas. Se recomienda que cualquier uso operativo del modelo se plantee como un complemento a la experiencia de los operadores y a los sistemas de control existentes, por ejemplo:

1. Como panel de riesgo que permita priorizar periodos para revisión o ajuste fino de parámetros de proceso.

2. Como base para discutir y revisar reglas de operación y límites de alarmas, identificando patrones que se repiten antes de las roturas.

3. Como herramienta de análisis histórico para investigar campañas concretas con alta tasa de roturas y comparar su perfil de variables con campañas más estables.

Ampliar y Actualizar el Dataset con Datos Adicionales y Nuevas Variables

Una limitación importante del estudio es el número relativamente pequeño de roturas disponibles y la dependencia de un único periodo y línea de producción. Para mejorar la robustez y la capacidad generalizadora del modelo, se recomienda:

1. Incorporar datos históricos adicionales de la misma máquina y, en la medida de lo posible, de otras máquinas o líneas con características comparables.

2. Evaluar la inclusión de variables adicionales que no están presentes en el dataset original (por ejemplo, información de calidad de la hoja, condiciones ambientales o indicadores de mantenimiento).

3. Reentrenar y reevaluar el pipeline con estos datos ampliados, manteniendo la estructura metodológica (segmentación temporal, ingeniería causal de variables, selección A/B y comparación de modelos).

Explorar Modelos Alternativos y Técnicas Avanzadas de Desbalance Cuando la Disponibilidad de Datos lo Permita

Aunque en el contexto actual la regresión logística penalizada ha sido la opción más razonable, futuras extensiones del trabajo podrían considerar:

1. Modelos de tipo Gradient Boosting y redes neuronales específicas para series temporales (por ejemplo, arquitecturas basadas en convoluciones o transformadores), siempre que se disponga de un volumen mayor de datos y se garantice una evaluación rigurosa.

2. Técnicas avanzadas de aprendizaje con datos desbalanceados, como focal loss, cost-sensitive learning o estrategias de muestreo temporalmente consistentes, que eviten la creación de observaciones artificiales que rompan la estructura de serie.

3. Métodos de calibración de probabilidad (Platt scaling, calibración isotónica) para mejorar la interpretabilidad probabilística de los scores y facilitar su uso en decisiones basadas en riesgo.

En síntesis, el trabajo realizado proporciona una base metodológica sólida y un primer modelo operativo para la anticipación de roturas del papel, pero su mayor valor reside en el camino que abre para futuras mejoras: más datos, mejor ingeniería de variables, validación temporal más estricta y una integración cuidadosa con el conocimiento experto de la planta.

Explorar Estrategias de Enriquecimiento de Datos y Modelos más Complejos

La revisión de la literatura reciente indica que las mejoras más relevantes en la detección de roturas de papel suelen asociarse a estrategias de enriquecimiento de datos (por ejemplo, sobremuestreo sintético y generación de datos mediante modelos generativos como CTGAN) y a arquitecturas de modelado más sofisticadas (árboles de decisión profundos, ensambles y redes neuronales especializadas para series de tiempo). Como línea de trabajo futuro, se recomienda evaluar estas alternativas sobre el mismo pipeline de preprocesamiento desarrollado en este proyecto, manteniendo particiones temporales estrictas y métricas centradas en la clase minoritaria. De este modo será posible cuantificar con rigor el beneficio adicional que aportan estas técnicas frente al modelo logístico de referencia y determinar si el incremento de complejidad se justifica en términos de desempeño y mantenibilidad.

Vincular la Evaluación del Modelo con el Impacto Económico de las Roturas

Dado que el objetivo último de la predicción de roturas es reducir los costos asociados a paradas no planificadas, desperdicio de materia prima y reprocesos, se recomienda que futuros estudios incorporen explícitamente indicadores económicos en la evaluación del modelo. En particular, resulta pertinente definir escenarios de operación (por ejemplo, distintos valores de k en el porcentaje de observaciones de mayor riesgo y diferentes umbrales de probabilidad) y estimar, para cada uno, el costo esperado de los falsos positivos y falsos negativos, así como el ahorro potencial derivado de anticipar una fracción de las roturas. Esto permitiría traducir métricas como la PR-AUC y la ROC-AUC en términos de impacto monetario y facilitaría la toma de decisiones sobre la adopción, ajuste y eventual actualización del modelo en la planta.

Referencias Bibliográficas

- Ahola, T. (2006, enero 16). *Intelligent estimation of web break sensitivity in paper machines* [Väitöskirja]. Jultika.Oulu.Fi. <https://oulurepo.oulu.fi/handle/10024/35231>
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, (pp. 785-794). <https://doi.org/10.1145/2939672.2939785>
- Chen, W., Yang, K., Yu, Z., Shi, Y., & Chen, C. L. P. (2024). A survey on imbalanced learning: Latest research, applications and future directions. *Artificial Intelligence Review*, 57(6), 137. <https://doi.org/10.1007/s10462-024-10759-6>
- Cheng, X. (2024). A comprehensive study of feature selection techniques in machine learning models. *Insights in Computer, Signals and Systems*, 1(1), 65-78. <https://doi.org/10.70088/xpf2b276>
- Feng, Y., Zhou, M., & Tong, X. (2021). *Imbalanced classification: A paradigm-based review* (No. arXiv:2002.04592). arXiv. <https://doi.org/10.48550/arXiv.2002.04592>
- Fridgeirsson, E. A., Williams, R., Rijnbeek, P., Suchard, M. A., & Reips, J. M. (2024). Comparing penalization methods for linear models on large observational health data. *Journal of the American Medical Informatics Association*, 31(7), 1514-1521. <https://doi.org/10.1093/jamia/ocae109>
- Grinsztajn, L., Oyallon, E., & Varoquaux, G. (2022). *Why do tree-based models still outperform deep learning on tabular data?* (Versión 1). arXiv. <https://doi.org/10.48550/ARXIV.2207.08815>

- Guido, R., Ferrisi, S., Lofaro, D., & Conforti, D. (2024). An overview on the advancements of support vector machine models in healthcare applications: A review. *Information*, 15(4), 235. <https://doi.org/10.3390/info15040235>
- He, Z., Chen, G., Hong, M., Xiong, Q., Zeng, X., & Man, Y. (2023). Process monitoring and fault prediction of papermaking by learning from imperfect data. *IEEE Transactions on Automation Science and Engineering*, 22, 680-690. <https://doi.org/10.1109/TASE.2023.3290552>
- Holzmüller, D., Grinsztajn, L., & Steinwart, I. (2024). *Better by default: Strong pre-tuned MLPs and boosted trees on tabular data* (Versión 3). arXiv. <https://doi.org/10.48550/ARXIV.2407.04491>
- Hussein, H., Anwar, S., & Ahmad, M. (2023). Imbalanced data classification using SVM based on improved simulated annealing featuring synthetic data generation and reduction. *Computers, Materials & Continua*, 75(1), 547-564. <https://doi.org/10.32604/cmc.2023.036025>
- IBM. (2024, mayo 9). *¿Qué es XGBoost?* <https://www.ibm.com/es-es/think/topics/xgboost>
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An introduction to statistical learning: With applications in R*. Springer US. <https://doi.org/10.1007/978-1-0716-1418-1>
- Juuso, E. K. (2022). Variable selection and grouping for large-scale data-driven modelling. *Scandinavian Simulation Society*, 38-45. <https://doi.org/10.3384/ecp2118538>
- Khosravi, H., Farhadpour, S., Grandhi, M., Raihan, A. S., Das, S., & Ahmed, I. (2023). *Strategic data augmentation with CTGAN for smart manufacturing: Enhancing machine learning predictions of paper 'breaks in pulp-and-paper production* (No. arXiv:2311.09333). arXiv. <https://doi.org/10.48550/arXiv.2311.09333>

- Lee, W., & Seo, K. (2021). Early failure detection of paper manufacturing machinery using nearest neighbor-based feature extraction. *Engineering Reports*, 3(2), e12291.
<https://doi.org/10.1002/eng2.12291>
- Liu, S., Roemer, F., Ge, Y., Bedrick, E. J., Li, Z.-M., Guermazi, A., Sharma, L., Eaton, C., Hochberg, M. C., Hunter, D. J., Nevitt, M. C., Wirth, W., Kent Kwok, C., & Sun, X. (2023). Comparison of evaluation metrics of deep learning for imbalanced imaging data in osteoarthritis studies. *Osteoarthritis and Cartilage*, 31(9), 1242-1248.
<https://doi.org/10.1016/j.joca.2023.05.006>
- Marzougui, A., Ma, Y., McGee, R. J., Khot, L. R., & Sankaran, S. (2020). Generalized linear model with Elastic Net regularization and convolutional neural network for Evaluating aphanomyces root rot severity in lentil. *Plant Phenomics*, 2020, 2393062.
<https://doi.org/10.34133/2020/2393062>
- Najjar, E., & Breesam, A. M. (2023). Supervised machine learning a brief survey of approaches. *Al-Iraqia Journal for Scientific Engineering Research*, 2(4), 71-82.
<https://doi.org/10.58564/IJSER.2.4.2023.121>
- Owusu-Adjei, M., Hayfron-Acquah, J. B., Frimpong, T., & Abdul-Salaam, G. (2023). Imbalanced class distribution and performance evaluation metrics: A systematic review of prediction accuracy for determining model performance in healthcare systems. *PLOS Digital Health*, 2(11), e0000290. <https://doi.org/10.1371/journal.pdig.0000290>
- ProcessMiner. (2020). *Rare event classification in a paper machine [Conjunto de datos]* [Dataset]. <https://processminer.com/dataset-rare-event-multivariate-time-series-part-1/>

- Pudjihartono, N., Fadason, T., Kempa-Liehr, A. W., & O'Sullivan, J. M. (2022). A review of feature selection methods for machine learning-based disease risk prediction. *Frontiers in Bioinformatics*, 2, 927312. <https://doi.org/10.3389/fbinf.2022.927312>
- Ranjan, C., Reddy, M., Mustonen, M., Paynabar, K., & Pourak, K. (2019). *Dataset: rare event classification in multivariate time series* (No. arXiv:1809.10717). arXiv. <https://doi.org/10.48550/arXiv.1809.10717>
- Rezvani, S., & Wang, X. (2023). A broad review on class imbalance learning techniques. *Applied Soft Computing*, 143, 110415. <https://doi.org/10.1016/j.asoc.2023.110415>
- Saito, T., & Rehmsmeier, M. (2015). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLOS ONE*, 10(3), e0118432. <https://doi.org/10.1371/journal.pone.0118432>
- Scikit-learn. (s. f.). *Permutation importance with multicollinear or correlated features*. Scikit-Learn. Recuperado 17 de noviembre de 2025, de https://scikit-learn.org/stable/auto_examples/inspection/plot_permutation_importance_multicollinear.html
- Shang, H., Langlois, J.-M., Tsioutsoulouklis, K., & Kang, C. (2023). Precision/recall on imbalanced test data. *Proceedings of The 26th International Conference on Artificial Intelligence and Statistics*, 9879-9891. <https://proceedings.mlr.press/v206/shang23a.html>
- St-Aubin, A. (2024). *An introduction to supervised machine learning* [Department of Mathematics and Statistics, McGill University]. <https://www.math.mcgill.ca/gsams/drp/papers/papers2024/Alexandre-St-Aubin.pdf>

- Tay, J. K., Narasimhan, B., & Hastie, T. (2023). Elastic Net regularization paths for all generalized linear models. *Journal of Statistical Software*, *106*(1), 1-31.
<https://doi.org/10.18637/jss.v106.i01>
- Tchio, G. M. T., Kenfack, J., Kassegne, D., Menga, F.-D., & Ouro-Djobo, S. S. (2024). A comprehensive review of supervised learning algorithms for the diagnosis of photovoltaic systems, proposing a new approach using an ensemble learning algorithm. *Applied Sciences*, *14*(5), 2072. <https://doi.org/10.3390/app14052072>
- Wu, M., Goh, K. W., Chaw, K. H., Koh, Y. S., Dares, M., Yeong, C. F., Su, E. L. M., William, H., & Zhang, Y. (2024). An intelligent predictive maintenance system based on random forest for addressing industrial conveyor belt challenges. *Frontiers in Mechanical Engineering*, *10*, 1383202. <https://doi.org/10.3389/fmech.2024.1383202>
- Xu, D., Zhang, Z., & Shi, J. (2022). Training data selection by categorical variables for better rare event prediction in multiple products production line. *Electronics*, *11*(7), 1056.
<https://doi.org/10.3390/electronics11071056>
- Yang, Y., Wang, H., Yang, Y., & Wang, H. (2025). Random Forest-Based machine failure prediction: A performance comparison. *Applied Sciences*, *15*(16), 8841.
<https://doi.org/10.3390/app15168841>
- Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *67*(2), 301-320.
<https://doi.org/10.1111/j.1467-9868.2005.00503.x>

Apéndices

Apéndice A

Coefficientes Completos del Modelo Logístico L2 Sobre el Conjunto B

Variable	Coefficientes
x40_r15_std	-0,84
x36_r15_std	-0,3
x37_r30_p25	-0,25
x37_r30_std	-0,2
x39_r15_mean	0,19
x15_r5_std	0,17
x54_r30_std	0,15
x5_r15_min	0,08
x54_d1	0,07
x41_r30_p25	0,04

Apéndice B

Resultados de los Modelos Según Estrategia de Re-Muestreo y Ponderación de Clases

Familia	Estrategia de remuestreo	Ponderación de clases	n test	PR-AUC valid	PR-AUC test	ROC-AUC valid	ROC-AUC test
logit	Sin remuestreo	balanced	9.930	0,3986	0,0811	0,8670	0,7042
logit	Sin remuestreo	Sin ponderación	9.930	0,3839	0,0818	0,8723	0,7127
logit	Sobre-muestreo clase minoritaria hasta balance	Sin ponderación	19.476	0,3978	0,0814	0,8665	0,7040
logit	Sub-muestreo clase mayoritaria (3×)	Sin ponderación	768	0,3853	0,0928	0,8662	0,7224
rf	Sin remuestreo	balanced	9.930	0,0923	0,0659	0,8459	0,6843
rf	Sin remuestreo	Sin ponderación	9.930	0,1301	0,0450	0,5516	0,6033

rf	Sobre-muestreo clase minoritaria hasta balance	Sin ponderación	19.476	0,1060	0,0656	0,8321	0,6808
rf	Sub-muestreo clase mayoritaria (3×)	Sin ponderación	768	0,2129	0,0571	0,7442	0,6524
svm	Sin remuestreo	balanced	9.930	0,2763	0,0774	0,8745	0,7056
svm	Sin remuestreo	Sin ponderación	9.930	0,0093	0,0410	0,3617	0,4933
svm	Sobre-muestreo clase minoritaria hasta balance	Sin ponderación	19.476	0,2750	0,0778	0,8746	0,7062
svm	Sub-muestreo clase mayoritaria (3×)	Sin ponderación	768	0,1850	0,0762	0,8743	0,7196
xgb	Sin remuestreo	Sin ponderación	9.930	0,0678	0,0394	0,6085	0,5179
xgb	Sin remuestreo	scale_pos_weight = 50.72	9.930	0,1297	0,0390	0,6126	0,5119
xgb	Sobre-muestreo clase minoritaria hasta balance	Sin ponderación	19.476	0,0506	0,0423	0,5140	0,5370
xgb	Sub-muestreo clase mayoritaria (3×)	Sin ponderación	768	0,2000	0,0518	0,7098	0,6132
