



ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL

Facultad de Ciencias Naturales y Matemáticas

**Análisis de la calidad del papel en una empresa  
ecuatoriana mediante un visualizador  
interactivo**

Proyecto Integrador

Previo a la obtención del Título de:  
Ingeniero en Estadística

**Presentado por:**

KATHERINE MARCELA CARCHI ORTEGA  
MELANY DENISSE GRANADOS CORTEZ

Guayaquil - Ecuador

2026

# Agradecimientos

Agradezco profundamente a mi familia, en especial a mi mamá, por su amor y apoyo constante a lo largo de este proceso; sin ella, este logro no habría sido posible.

*Katherine Carchi Ortega*

# Agradecimientos

A Dios y a mi familia por siempre estar y confiar en mí.  
Mamá y Papá, gracias por su apoyo y amor.

*Melany Granados Cortez*

## Declaración Expresa

---

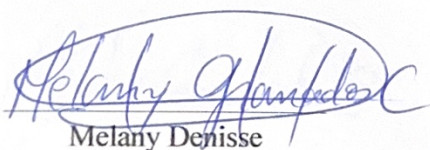
Nosotras, Melany Denisse Granados Cortez y Katherine Marcela Carchi Ortega, acordamos y reconocemos que:

La titularidad de los derechos patrimoniales de autor (derechos de autor) del proyecto de graduación corresponderá al autor o autores, sin perjuicio de lo cual la ESPOL recibe en este acto una licencia gratuita de plazo indefinido para el uso no comercial y comercial de la obra con facultad de sublicenciar, incluyendo la autorización para su divulgación, así como para la creación y uso de obras derivadas. En el caso de usos comerciales se respetará el porcentaje de participación en beneficios que corresponda a favor del autor o autores.

La titularidad total y exclusiva sobre los derechos patrimoniales de patente de invención, modelo de utilidad, diseño industrial, secreto industrial, software o información no divulgada que corresponda o pueda corresponder respecto de cualquier investigación, desarrollo tecnológico o invención realizada por nosotras durante el desarrollo del proyecto de graduación, pertenecerán de forma total, exclusiva e indivisible a la ESPOL, sin perjuicio del porcentaje que nos corresponda de los beneficios económicos que la ESPOL reciba por la explotación de nuestra innovación, de ser el caso.

En los casos donde la Oficina de Transferencia de Resultados de Investigación (OTRI) de la ESPOL comunique a las autoras que existe una innovación potencialmente patentable sobre los resultados del proyecto de graduación, no se realizará publicación o divulgación alguna, sin la autorización expresa y previa de la ESPOL.

Guayaquil, 17 de octubre del 2025



Melany Denisse

Granados Cortez



Katherine Marcela

Carchi Ortega

# Evaluadores

---

M.Sc. Roa López Heydi Mariana  
**Profesor de Materia**

---

Ph.D. Looz Valeriano Katherine  
**Tutor de proyecto**

# Resumen

El análisis y la integración de datos industriales generados de manera continua representan un gran desafío en la industria papelera, particularmente en la evaluación de la calidad del producto y en el estudio de eventos de rotura, los cuales se traducen en pérdidas de tiempo productivo y eficiencia operativa. Por ello, en este estudio se toman como unidad de investigación las bobinas de papel producidas por una empresa ecuatoriana, incorporando una variable binaria denominada *rotura* como criterio de segmentación, con el objetivo de caracterizar la calidad del proceso en torno a la ocurrencia de dicho evento.

Inicialmente, se realizó la reestructuración de la base de datos extraída del Sistema de Supervisión, Control y Adquisición de Datos (SCADA). Posteriormente, se llevó a cabo un análisis estadístico descriptivo correspondiente a 45 días de producción, empleando medidas de frecuencia, histogramas y diagramas de caja, con el fin de evaluar el comportamiento de las variables del proceso en relación con el evento de rotura. Además, se desarrolló un visualizador que incorpora gráficos de Control Estadístico de Procesos (SPC) para evaluar el comportamiento de las variables, filtradas según parámetros operativos específicos.

Los resultados evidenciaron que determinadas variables asociadas a las etapas de vacío y vapor presentan variaciones significativas en torno al evento de rotura, especialmente ante incrementos en temperatura y presión, dado que estas etapas intervienen directamente en el proceso de extracción de humedad del papel. Para contrastar estas diferencias, se compararon bobinas con y sin evento de rotura mediante la prueba no paramétrica de Mann–Whitney U y el tamaño del efecto de Cliff, permitiendo identificar tanto diferencias estadísticamente significativas como su magnitud práctica. Asimismo, se ajustó un modelo de regresión logística que alcanzó un valor de AUC-ROC del 77.64%. Estos hallazgos destacan el potencial de las herramientas estadísticas para el análisis de la calidad del papel en presencia de eventos anómalos, así como la relevancia de su aplicación en entornos industriales mediante enfoques proactivos apoyados en visualizadores interactivos desarrollados en Power BI.

**Palabras claves:** Calidad, Industria papelera, Papel, Procesos continuos, Procesos industriales, Rotura, Visualizador.

# Abstract

The analysis and integration of continuously generated industrial data represent a major challenge in the paper industry, particularly in the assessment of product quality and in the study of paper break events, which translate into losses of productive time and operational efficiency. Therefore, this study considers paper reels produced by an Ecuadorian company as the unit of analysis, incorporating a binary variable referred to as *break* as a segmentation criterion, with the aim of characterizing process quality around the occurrence of this event.

Initially, the database extracted from the Supervisory Control and Data Acquisition (SCADA) system was restructured and processed. Subsequently, a descriptive statistical analysis was conducted over 45 days of production, using frequency measures, histograms, and boxplots to evaluate the behavior of process variables in relation to the break event. In addition, a visualization tool incorporating Statistical Process Control (SPC) charts was developed to assess the behavior of the variables filtered according to specific operational parameters.

The results showed that certain variables associated with the vacuum and steam stages exhibit significant variations around the break event, particularly under increases in temperature and pressure, since these stages are directly involved in the moisture removal process of the paper. To assess these differences, reels with and without break events were compared using the nonparametric Mann-Whitney U test and Cliff's delta as a measure of effect size, allowing the identification of both statistically significant differences and their practical magnitude. Furthermore, a logistic regression model was fitted, achieving an AUC-ROC value of 77.64%. These findings highlight the potential of statistical tools for analyzing paper quality in the presence of anomalous events, as well as the relevance of their application in industrial environments through proactive approaches supported by interactive dashboards developed in Power BI.

**Keywords:** Quality, Paper Industry, Paper, Continuous Processes, Industrial Processes, Visualization Tool, Break.

# Índice de figuras

|   |    |
|---|----|
| Figura 1 – Diagrama de flujo del proceso de fabricación de papel. . . . .   | 18 |
| Figura 2 – Porcentaje de registros faltantes por variable. . . . .  | 24 |
| Figura 3 – Visualizador . . . . .   | 37 |
| Figura 4 – Distribución de roturas por día del proceso de fabricación. . . . .  | 41 |
| Figura 5 – Distribución de roturas por turno. . . . .   | 42 |
| Figura 6 – Distribución de roturas en variables claves. . . . .   | 44 |
| Figura 7 – Evolución temporal del torque de vacío en couch en las bobinas 232<br>y 135 respectivamente. . . . .                 | 45 |
| Figura 8 – Evolución temporal de la variable Control_Vapor.PIC_405_PV en<br>las bobinas 232 y 135 respectivamente. . . . .      | 47 |
| Figura 9 – Evolución temporal de la variable Nipcoflex_Com.P56_PV en las<br>bobinas 232 y 135 respectivamente. . . . .          | 48 |
| Figura 10 – Evolución temporal de variables de calidad en las bobinas 232 y 135. . . . .  | 56 |
| Figura 11 – Evolución temporal de la variable Aproximacion.SP_CHORRO_TELA<br>en las bobinas 232 y 135 respectivamente. . . . .  | 56 |
| Figura 12 – Evolución temporal de la variable Presecado_Couch.CORRIENTE_PV<br>en las bobinas 232 y 135 respectivamente. . . . . | 57 |

# Índice de cuadros

|  |    |
|--|----|
| Cuadro 1 – Cantidad de variables por etapa del proceso de fabricación . . . . .                          | 38 |
| Cuadro 2 – Distribución de evento de rotura . . . . .  | 39 |
| Cuadro 3 – Distribución de bobinas por roturas . . . . .   | 39 |
| Cuadro 4 – Bobinas que registraron al menos una rotura. . . . .  | 40 |
| Cuadro 5 – Estadísticas descriptivas globales de ciertas variables por etapa de<br>fabricación. . . . .  | 40 |
| Cuadro 6 – Roturas por hora del día . . . . .  | 42 |
| Cuadro 7 – Distribución porcentual de eventos de rotura asociados por etapa<br>del proceso . . . . .     | 42 |
| Cuadro 8 – Gramaje con mayor número de roturas . . . . .   | 43 |
| Cuadro 9 – SCT con mayor número de roturas . . . . .   | 43 |
| Cuadro 10 – Humedad promedio con mayor número de roturas . . . . .                                       | 43 |
| Cuadro 11 – Peso promedio con mayor número de roturas . . . . .  | 43 |
| Cuadro 12 – Chorro Tela con mayor número de roturas . . . . .  | 43 |
| Cuadro 13 – Resumen por mediana global de la estabilidad dinámica del sistema<br>de vacío . . . . .      | 46 |
| Cuadro 14 – Estabilidad dinámica del sistema de vacío en bobinas emparejadas                             | 46 |
| Cuadro 15 – Cliff's Delta . . . . .  | 50 |
| Cuadro 16 – Resultados del tamaño de efecto medido mediante Cliff's Delta . .                            | 50 |
| Cuadro 17 – Resultados del modelo de regresión logística para el evento de rotura<br>de bobina . . . . . | 51 |
| Cuadro 18 – Odds Ratios e intervalos de confianza al 95% para el evento de<br>rotura de bobina . . . . . | 54 |

# Lista de siglas y abreviaturas

|       |  |
|-------|--|
| SCADA | Sistema de Supervisión, Control y Adquisición de Datos |
| PLC   | Controlador Lógico Programable                         |
| SCP   | Sistema de Control de Proceso                          |
| SCT   | Resistencia a la Compresión de Fibras                  |
| CMT   | Resistencia al Aplastamiento                           |
| OCC   | Contenedores Corrugados Usados                         |
| DKL   | Doble Capa de papel Kraft Liner                        |
| LOCF  | Última Observación Arrastrada                          |
| MCAR  | Datos Faltantes Completamente Aleatorios               |
| PDIC  | Controlador Indicador de Presión Diferencial           |
| PV    | Variable de Proceso                                    |
| TIC   | Controlador Indicador de Temperatura                   |
| PIC   | Controlador Indicador de Presión                       |
| LIC   | Controlador Indicador de Nivel                         |
| LM    | Medición de Carga                                      |

# Índice general

|          |   |           |
|----------|---|-----------|
| <b>1</b> | <b>Introducción</b>                               | <b>12</b> |
| 1.1      | Descripción del problema                          | 13        |
| 1.2      | Justificación del problema                        | 13        |
| 1.3      | Objetivos   | 14        |
| 1.3.1    | Objetivo General                                  | 14        |
| 1.3.2    | Objetivos Específicos                             | 14        |
| 1.4      | Marco Teórico                                     | 14        |
| <b>2</b> | <b>Metodología</b>                                | <b>17</b> |
| 2.1      | Descripción del proceso de fabricación del papel  | 17        |
| 2.2      | Formulación de alternativas de solución           | 19        |
| 2.3      | Metodología de diseño                             | 20        |
| 2.3.1    | Fuente de datos                                   | 20        |
| 2.3.2    | Tratamiento de datos faltantes                    | 23        |
| 2.3.3    | Selección y codificación de variables             | 25        |
| 2.3.4    | Análisis exploratorio                             | 30        |
| 2.3.5    | Construcción del dataset de análisis preventivo   | 30        |
| 2.3.6    | Selección de variables para el enfoque preventivo | 32        |
| 2.3.7    | Modelos preventivos                               | 33        |
| 2.3.8    | Diseño del visualizador                           | 35        |
| 2.3.8.1  | Estructura funcional del visualizador             | 36        |
| <b>3</b> | <b>Resultados</b>                                 | <b>38</b> |
| 3.0.1    | Tamaño de efecto mediante Cliff's Delta           | 50        |
| 3.1      | Modelo de regresión logística                     | 50        |
| 3.1.1    | Odds Ratios del modelo logístico                  | 53        |
| <b>4</b> | <b>Conclusiones y Recomendaciones</b>             | <b>58</b> |
| 4.1      | Conclusiones                                      | 58        |
| 4.2      | Recomendaciones                                   | 59        |
|          | <b>Bibliografía</b>                               | <b>60</b> |

# 1 Introducción

La calidad del papel constituye un elemento esencial para el desarrollo de productos que satisfagan las expectativas del cliente y garanticen la funcionalidad del material en sus diferentes aplicaciones industriales. En particular, cuando el papel se destina a la producción de cartones, su resistencia física y estructural influye directamente en la capacidad de los empaques para proteger mercancías principalmente como banano, camarones, flores, además de otro tipo de alimentos procesados durante su transporte y almacenamiento, impactando de manera directa en la satisfacción del consumidor final e incidiendo en las exportación de estos productos en el país.

En la actualidad, las empresas papeleras se enfrentan al desafío de transformar la gran cantidad de datos generados en sus procesos productivos en información útil que permita optimizar la operación, reducir desperdicios, mejorar la eficiencia energética y fortalecer la toma de decisiones estratégicas dentro de planta (Area, 2008). El uso de metodologías de análisis estadístico y de inteligencia de datos ha demostrado ser una vía efectiva para identificar relaciones ocultas entre variables, modelar procesos y anticipar comportamientos que inciden en la calidad del producto final (Devi *et al.*, 2023).

En el contexto ecuatoriano, la industria papelera se encuentra en una fase de transición hacia la digitalización y la automatización de sus procesos, en respuesta a la necesidad de mejorar la competitividad y la sostenibilidad ambiental. Sin embargo, aún persisten limitaciones en la integración y estandarización de los datos disponibles, lo que dificulta el análisis integral de los datos generados a lo largo de la cadena productiva.

El presente proyecto se desarrolla en una empresa papelera ecuatoriana que posee un proceso de fabricación conformado por múltiples etapas y una gran cantidad de variables interrelacionadas. Aunque la empresa cuenta con registros históricos de producción, dichos datos no han sido sistemáticamente tratados ni analizados con técnicas estadísticas que permitan determinar los factores que influyen en la calidad del papel. Por tal motivo, se propone el desarrollo de un visualizador interactivo que integre y analice la información generada en planta, facilitando la interpretación de los resultados y el seguimiento de los indicadores de calidad de manera dinámica y accesible para los responsables del proceso productivo.

## 1.1 Descripción del problema

La empresa papelera objeto de estudio enfrenta dificultades para integrar y analizar los datos que se generan de manera continua en sus líneas de producción, y como consecuencia el manejo de tiempos improductivos en sus procesos industriales. En un día promedio, los tiempos improductivos asociados a eventos de rotura alcanzan aproximadamente 2 horas y 16 minutos, lo que se traduce en pérdidas económicas cercanas a USD 18,000.00. Esta situación limita la capacidad de analizar con precisión los factores operativos que afectan la calidad del producto final como eventos de roturas, impidiendo establecer estrategias de mejora sustentadas en evidencia.

Esta problemática es común en la industria papelera, donde los sistemas de registro de información se implementan de forma aislada y sin una estructura unificada, lo que ocasiona bases de datos fragmentadas, redundantes o incompletas. En consecuencia, la empresa carece de herramientas que le permitan evaluar de manera integral el desempeño de sus procesos productivos y establecer relaciones entre las variables del proceso, principalmente cuando aparecen daños a lo largo de la fabricación del papel, lo que se traduce en pérdidas económicas directas.

## 1.2 Justificación del problema

Las industrias papeleras deben cumplir con estándares de calidad cada vez más exigentes para mantener su competitividad y alinearse con los principios de sostenibilidad y eficiencia productiva (FAO, 2021; CEPI, 2021). En este marco, resulta indispensable desarrollar sistemas de análisis que permitan analizar la calidad del papel, minimizando las pérdidas de materia prima, tiempo de producción y optimizando el desempeño del proceso productivo.

La empresa objeto de estudio enfrenta desafíos vinculados con la variabilidad en la calidad del papel y la ausencia de herramientas analíticas que faciliten la detección temprana de desviaciones en el proceso productivo. En este contexto, el desarrollo de un visualizador interactivo permitirá a la organización monitorear de manera continua sus operaciones, identificar patrones anómalos y aplicar acciones correctivas fundamentadas en evidencia empírica. En consecuencia, resulta fundamental analizar la calidad del papel por medio de los factores que inciden significativamente a lo largo de la fabricación del papel de forma proactiva en torno al evento de la rotura y no rotura del papel, con el objetivo de determinar lo que está sucediendo con la variabilidad del proceso, reducir la brecha entre la calidad óptima y los productos no conformes, y fortalecer la eficiencia global del sistema productivo.

Este proyecto no solo beneficiará a la empresa participante, sino también a otras industrias papeleras ecuatorianas que trabajen con papel kraft, cartón corrugado, generación de grandes volúmenes de datos en sus procesos productivos continuos en sistemas como SCADA (Sistema de Supervisión, Control y Adquisición de Datos) y busquen mejorar su gestión de calidad. Los resultados esperados se traducen en impactos económicos (reducción de desperdicios y costos), ambientales (menor consumo de recursos) y sociales (mayor productividad y estabilidad laboral), consolidando un enfoque de mejora continua y gestión basada en evidencia (Gómez & Jiménez, 2019).

## 1.3 Objetivos

### 1.3.1 Objetivo General

Analizar la calidad del papel en una empresa ecuatoriana mediante la aplicación de técnicas estadísticas descriptivas y el desarrollo de un visualizador interactivo que facilite la interpretación de los resultados y apoye la toma de decisiones informadas.

### 1.3.2 Objetivos Específicos

- Realizar la limpieza y estructuración de los datos históricos de producción, garantizando la calidad y consistencia para su aplicación en el análisis de los procesos de la empresa.
- Examinar el comportamiento y la frecuencia de ocurrencia de la rotura del papel mediante técnicas estadísticas descriptivas y exploratorias.
- Implementar un visualizador interactivo en Power BI para el análisis dinámico de la ocurrencia de rotura del papel, facilitando la interpretación de los resultados por parte del personal técnico.

## 1.4 Marco Teórico

La industria papelerera se caracteriza por la operación de procesos continuos, complejos y altamente sensibles a variaciones en parámetros operativos como la temperatura, la presión, la humedad y la velocidad de la máquina. El control adecuado de estos parámetros resulta fundamental para garantizar la uniformidad del producto final y sus propiedades mecánicas, tales como el gramaje, el espesor, la resistencia y la capacidad de absorción del papel (Smook, 2016).

No obstante, el tratamiento de grandes volúmenes de datos generados de forma continua y a alta frecuencia en procesos industriales puede introducir niveles significativos de ruido, dificultando la identificación del comportamiento real del proceso. En este contexto, [Kuhn & Johnson \(2013\)](#) plantean la conveniencia de transformar series temporales de alta resolución hacia escalas temporales más consolidadas, mediante técnicas de agregación o re-muestreo, preservando la consistencia estadística de las variables y facilitando su análisis. Además, los sistemas de adquisición como SCADA (Sistema de Supervisión, Control y Adquisición de Datos), suelen operar bajo tiempos adaptables de muestreo pudiendo estar configurados a frecuencia de 1 minuto bajo cierto muestreo a las variables implicadas en el proceso industrial.

Según [Devi et al. \(2023\)](#), la fabricación de papel consta de tres etapas importantes como las secciones de alambre, prensa y secado. Además de ocho pasos que siguen un orden específico operando entre lo húmedo y lo seco de extremo a extremo, con parámetros medidos en cada una de ellas secuencialmente, además mencionan los tiempos de improductividad que son paradas dentro del proceso debido a eventos como posibles roturas. Su estudio estaba orientado al control del vapor para evitar daños en el papel, argumentando que si la humedad disminuye en un 45 % pierde estabilidad y suele dañarse, además de la implicación energética, ya que la sección de secado consume alrededor del 80 % de la energía del proceso.

Las técnicas estadísticas descriptivas son un buen soporte a la hora de analizar la calidad en general. El Control Estadístico de Procesos (SPC) permite monitorear y controlar procesos asegurando consistencia, ya que dentro de las industrias existen estándares. Según [Çiçekler & Tutus \(2023\)](#) existen dos enfoques, uno reactivo y otro proactivo, donde el proceso se estudia una vez terminado, examinando el producto final y en el otro enfoque se sigue durante la continuidad del proceso de fabricación anticipando fallos, respectivamente. Por otro lado, implementar estos controles fomenta la sostenibilidad, reduce desperdicios y costos.

Un enfoque proactivo requiere el estudio del proceso de fabricación del papel en tiempo real. Según [Reason \(2000\)](#) la gestión de riesgos proactiva consiste en el estudio de las condiciones previas al evento, siendo capaz de anticipar desde el “enfoque sistema” la ocurrencia del evento, donde la entidad cumple un rol importante analizando a detalle, entendiendo, alertando, anticipando escenario en donde podría darse el evento, además del soporte estadístico, ya que el autor asegura que los humanos toman mejores decisiones preventivas cuando se cuentan con predicciones basadas en datos.

[Kuhn & Johnson \(2013\)](#) mencionan que el uso de ventanas temporales ayuda a estudiar las condiciones previas al evento, permitiendo que la fabricación de algún

modelo detecte señales no vistas con los datos originales, teniendo la capacidad de tratar una gran cantidad de datos y a la vez tomando eventos pasados para formar estadísticos como la media, la desviación estándar, el recuento simplificando datos y ayudando a la estabilidad de los parámetros frente a la variabilidad de proceso. Paralelamente, la transformación digital de la industria, enmarcada en la Industria 4.0, impulsa el uso de herramientas de visualización y monitoreo de datos que facilitan la toma de decisiones en tiempo real, como la construcción de dashboards dinámicos, con el fin de analizar y tomar decisiones informadas a tiempo desde un enfoque proactivo, en lugar de reactivo. Además este tipo de herramientas van a la par con el avance de la tecnología y ayudan a mantener la rentabilidad y competitividad de la empresa. (Çiçekler & Tutuş, 2024; Othen *et al.*, 2025).

A nivel nacional, la industria papelera ecuatoriana ha experimentado un crecimiento sostenido en los últimos años; no obstante, enfrenta desafíos importantes vinculados con la modernización tecnológica, la gestión eficiente de los recursos y la incorporación de herramientas analíticas avanzadas en sus procesos productivos. En este contexto, el desarrollo e implementación de sistemas de análisis basados en datos, representan una oportunidad estratégica para potenciar la competitividad, promover la sostenibilidad y optimizar la gestión de la calidad del papel.

Se han identificado algunos estudios de la calidad del papel por medio del proceso de fabricación en Ecuador desde un punto de vista descriptivo con medidas usuales de resumen, frecuencias y descripción de variables de calidad, además de comparaciones con normas de acuerdo a lo establecidas por la ISO (Organización Internacional de Normalización) y uso de las cartas de control en variables de calidad.

## 2 Metodología

En una empresa papelería ecuatoriana surgió la necesidad de analizar la calidad del papel por medio de factores que inciden en el proceso de fabricación, entendida en este estudio como la capacidad del material para no presentar rupturas a lo largo de las distintas etapas del proceso productivo. La calidad del papel se evalúa a partir de un conjunto de propiedades físicas y mecánicas, entre las que se incluyen el gramaje, la humedad, el peso, la SCT (Resistencia a la Compresión de Fibras), el Mullen (Resistencia al Estallido) y la CMT (Resistencia al Aplastamiento), así como mediante variables registradas por sensores distribuidos a lo largo de la línea de producción.

Adicionalmente, se consideran variables operativas provenientes de distintas secciones del proceso, incluyendo la prensa de la máquina, el sistema de presecado, la sección de secadores y el sistema de vacío. El análisis integrado de estas variables tiene como finalidad generar una base técnica y cuantitativa que contribuya a mejorar la gestión del proceso productivo, mediante la reducción de desperdicios y costos operativos, promoviendo al mismo tiempo prácticas sostenibles en la producción de papel.

### 2.1 Descripción del proceso de fabricación del papel

El proceso de fabricación del papel comprende una serie de etapas interrelacionadas que transforman la materia prima fibrosa en un producto final con propiedades físicas y mecánicas específicas. De manera general, este proceso incluye las fases de preparación de la pulpa, formación de la hoja, prensado, secado y acabado.

En la etapa de preparación de la pulpa se definen variables asociadas al tipo de fibra, la consistencia y el uso de aditivos químicos, las cuales inciden directamente en propiedades como el gramaje y la resistencia mecánica del papel. Posteriormente, durante las etapas de formación de la hoja y prensado, intervienen variables operativas fundamentales, entre ellas la velocidad de la máquina, la presión aplicada y el contenido de humedad, que condicionan la estructura interna del papel y su comportamiento mecánico. Finalmente, las etapas de secado y acabado determinan las características finales del producto, tales como la estabilidad dimensional y la calidad superficial.

En este contexto, el presente proyecto se enfoca exclusivamente en el análisis de las variables asociadas a las etapas de formación de la hoja y prensado, delimitando el alcance del estudio y excluyendo la variabilidad atribuida a la materia prima y a los

compuestos químicos utilizados en el proceso de fabricación.

La Figura 1 muestra el esquema general del proceso de fabricación del papel en la empresa objeto de estudio. El proceso se inicia con el uso de materias primas recicladas, principalmente OCC (Contenedores Corrugados Usados) y DKL (Doble Capa de papel Kraft Liner), las cuales alimentan el sistema de preparación de la pasta. Esta es conducida hacia las licuadoras P1 y P2, donde se realiza la mezcla, para luego pasar por la plancha denominada “lise” y la refinadora, donde se ajustan las características físicas de las fibras. Posteriormente, la suspensión fibrosa atraviesa el depurador centrífugo y la mesa formadora, donde se inicia la conformación de la hoja de papel.

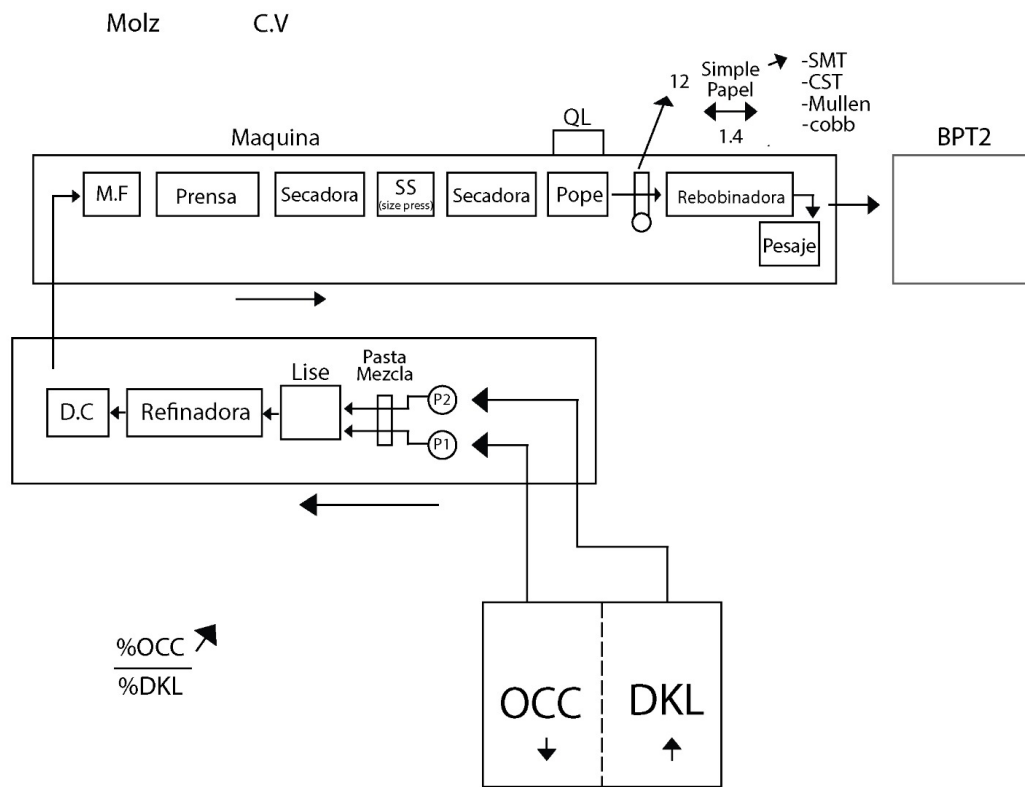


Figura 1 – Diagrama de flujo del proceso de fabricación de papel.

Una vez formada la hoja, el papel continúa su recorrido por la prensa, seguida de la secadora, la prensa encoladora y una segunda etapa de secado. Posteriormente, el material pasa por el Pope, donde se realiza el bobinado inicial, y por la rebobinadora, en la que se ajustan las dimensiones finales del producto. El proceso concluye con el pesaje del papel y su despacho hacia la etapa posterior.

Adicionalmente, el diagrama identifica puntos de control de calidad, en

los cuales se evalúan propiedades como CMT (Resistencia al Aplastamiento), SCT (Resistencia a la Compresión de Fibras), Mullen (Resistencia al Estallido) y Cobb, permitiendo verificar el cumplimiento de los estándares de calidad establecidos.

## 2.2 Formulación de alternativas de solución

Las soluciones consideradas difieren en la forma en que tratan los datos previos y posterior a la rotura. En particular, se analizaron enfoques basados en estadísticas descriptivas, modelos de clasificación y detección de anomalías, así como estrategias de segmentación temporal que utilizan exclusivamente información previa al evento.

1. Una alternativa metodológica al análisis puramente basado en variables individuales, se establece un análisis descriptivo que indique el comportamiento del proceso de cada bobina. Con el fin de identificar diferencias entre procesos normales y condiciones asociadas al evento de rotura.
2. Utilizar los datos previos a la ocurrencia de la rotura, y los datos completos para la no rotura, redefiniendo la unidad de análisis a nivel de bobina, incluyendo medidas de tendencia central, dispersión y variabilidad de corto plazo. Esta formulación permite incorporar un modelo de regresión logística, debido explícitamente a la independencia temporal y las interacciones entre variables, superando las limitaciones de los enfoques tradicionales.
3. Una tercera alternativa evaluada consistió en la aplicación de modelos de clasificación supervisados y técnicas de detección de anomalías. Utilizando técnicas como regresión logística, árboles de decisión o métodos de aprendizaje automático. De manera complementaria, se consideró el uso de modelos de detección de anomalías no supervisados, orientados a identificar desviaciones respecto a patrones normales de operación.

En base a esto, el estudio de alternativas condujo a realizar un análisis exploratorio mediante gráficos que evalúan el comportamiento del proceso, así como identificar sesgos respecto a condiciones normales y eventos de rotura. Adicional se integro un modelo de regresión logística usando información previa al evento de rotura, usando así un enfoque preventivo, basado en la estabilidad o variación del proceso.

## 2.3 Metodología de diseño

El desarrollo del proyecto siguió una metodología estructurada en seis etapas:

1. Fuente de datos
2. Tratamiento de datos faltantes
3. Selección y codificación de variables
4. Análisis exploratorio
5. Elaboración de modelo predictivo
6. Diseño de un visualizador

Cada una de estas etapas es descrita de manera detallada a continuación.

### 2.3.1 Fuente de datos

La presente investigación se sustenta en datos históricos obtenidos del sistema de automatización industrial de una planta papelera, el cual opera bajo una arquitectura integrada compuesta por Controladores Lógicos Programables (PLC) y un sistema de Supervisión, Control y Adquisición de Datos (SCADA). Esta infraestructura constituye el estándar en entornos industriales modernos, al permitir la adquisición continua y en tiempo real de variables críticas del proceso. Asimismo, el sistema posibilita el almacenamiento sistemático de los datos generados, lo que facilita su posterior análisis estadístico, el diagnóstico de fallas operativas y el soporte a la toma de decisiones tanto a nivel operativo como estratégico (Bailey & Wright, 2003; Seborg *et al.*, 2010).

El sistema PLC (Controladores Lógicos Programables)/SCADA (Sistema de Supervisión, Control y Adquisición de Datos) supervisa las distintas etapas del proceso de fabricación de papel, registrando variables asociadas a las condiciones mecánicas, operativas y de control del proceso productivo. La adquisición de datos se realiza mediante un esquema de registro por cambio (on-change logging), en el cual los valores de cada variable son almacenados únicamente cuando se detecta una variación respecto al último valor registrado. Como consecuencia de este esquema, durante períodos de operación estables, las variables mantienen su valor previo sin generar nuevos registros. Esto implica que el sistema no realiza muestreos periódicos ni registros asociados a cada rollo producido, sino que prioriza la eficiencia en el uso del almacenamiento y del ancho de banda de la infraestructura de automatización.

Si bien este enfoque resulta altamente eficiente desde el punto de vista operativo, introduce desafíos relevantes para el análisis de datos, ya que genera series temporales irregulares, asincrónicas y desalineadas entre variables. En primer lugar, los registros segundo a segundo exhiben una alta autocorrelación temporal, que implica que las observaciones no son independientes entre sí. Lo que afecta los supuestos fundamentales requeridos por diversos métodos estadísticos y de inferencia, llevando a estimaciones sesgadas, tal y como se lo discute en el análisis de datos longitudinales (Diggle *et al.*, 2002). Segundo, se genera una pseudo-replicación, debido a la presencia de miles de registros por bobinas, en el cual el modelo se ve forzado a aprender sobre las características específicas de una bobina, en lugar de capturar relaciones del proceso productivo, induciendo así a falsas significancias estadísticas (Harrell, 2015). Adicionalmente, el uso directo de los datos a nivel de segundos produce un desbalance extremo entre el evento de interés, dificultando la detección de patrones asociados al fallo. Estas características dificultan la aplicación directa de técnicas de análisis multivariado, modelado predictivo o detección de eventos sin un preprocesamiento adecuado, tal como se discute en la literatura especializada (Kadlec *et al.*, 2009).

Así se inició el proceso de re-estructuración temporal, mediante el cual los registros segundo a segundo son transformados a una base de datos minuto a minuto, con el objetivo de reducir la autocorrelación temporal y el ruido de alta frecuencia propio de las señales industriales (Seborg *et al.*, 2010; Box *et al.*, 2015). Esta agregación permite además mitigar el problema de pseudo-replicación, al evitar que miles de observaciones altamente correlacionadas por bobina induzcan falsas significancias estadísticas en los modelos (Harrell, 2015). De tal manera la representación agregada facilita una caracterización adecuada del estado operativo del proceso, alineada con la dinámica real del sistema y con los enfoques modernos de análisis de procesos y detección de fallos, particularmente en contextos preventivos (Kadlec *et al.*, 2009; Chiang *et al.*, 2001).

Los datos fueron extraídos del sistema en archivos de texto con extensión .txt, los cuales presentan una estructura particular basada en grupos repetitivos de tres columnas por cada variable del proceso. Cada una de estas tripletas está compuesta por:

- UTC-5: marca de tiempo que indica el instante exacto en el que ocurrió el cambio de valor, expresada en la zona horaria local.
- Sec: contador o referencia temporal auxiliar generada por el sistema de adquisición, utilizada como apoyo para el ordenamiento de los registros.

- Variable: valor numérico correspondiente a la variable de proceso registrada en dicho instante.

Como consecuencia directa del esquema de on-change logging, el conjunto de datos presenta las siguientes características estructurales:

- Marcas de tiempo no coincidentes entre variables.
- Ausencia de registros durante intervalos de estabilidad del proceso.
- Imposibilidad de realizar análisis multivariados sin una alineación temporal previa.
- Necesidad de aplicar técnicas de interpolación o propagación de valores.

Adicionalmente, se identificó que los archivos de datos contenían encabezados repetidos, principalmente en las columnas correspondientes a la marca de tiempo (UTC-5) y al contador auxiliar (Sec). Durante la importación de los datos, estas columnas eran renombradas automáticamente por las herramientas de análisis (por ejemplo, UTC-5.1, UTC-5.2), lo que hacía necesaria una correcta identificación y asociación de cada conjunto de columnas con su respectiva variable. Asimismo, la ausencia de una línea de tiempo común imposibilitaba la comparación directa entre señales.

Con el fin de resolver estas limitaciones y preparar los datos para su análisis, se desarrolló un proceso automatizado de preprocesamiento mediante un script en Python, el cual implementa de manera sistemática las siguientes etapas:

1. Identificación y detección automática de tripletas válidas (marca de tiempo, contador y valor), incluso en presencia de encabezados duplicados o renombrados.
2. Filtrado de registros para conservar únicamente observaciones con valores numéricos válidos.
3. Construcción de una serie temporal independiente para cada variable del proceso, preservando su dinámica original.
4. Integración horizontal (merge) de todas las series temporales en una estructura de datos unificada.
5. Definición de una línea de tiempo uniforme mediante un proceso de reindexado con una resolución temporal de 1 segundo.

Luego de esto se construyó una base de datos minuto a minuto, donde cada fila representa el estado operativo del proceso durante un minuto específico. Lo que permite obtener un representación estable y adecuada del comportamiento del sistema.

La agregación temporal se realizó bajo los siguientes criterios:

1. Se seleccionó el último valor medido como representación del estado real del proceso al cierre del intervalo, consistente con la lógica de control industrial.
2. Relleno de valores faltantes utilizando la técnica de propagación hacia adelante LOCF (Última Observación Arrastrada) dentro de cada minuto, la cual resulta apropiada en contextos industriales donde se asume que el último valor registrado permanece vigente hasta que se produce un nuevo cambio en la señal (Montgomery, 2019; Kadlec, 2009).

Este enfoque está alineado con las recomendaciones de Box & Jenkins (1976) quienes indican que la agregación temporal previa reduce la influencia del ruido y estabiliza la estructura de dependencia, y con Montgomery (2019); Kadlec (2009), quienes enfatizan que el tratamiento de datos debe realizarse a una escala temporal coherente con la dinámica de control del proceso. Adicionalmente se generó un archivo complementario de métricas de cambios, el cual contabiliza el número de variaciones registradas por cada variable dentro de cada intervalo de un minuto. Debido a que excluir valores por minuto puede ocultar comportamientos relevantes para la prevención de fallos.

La relevancia de este enfoque está respaldada por la literatura, que destaca que la variabilidad y la dinámica de corto plazo suelen ser más informativas que los niveles promedio para la detección temprana de fallos (Qin, 2003; Hastie *et al.*, 2009).

Tras la aplicación de los procesos de limpieza, alineación y sincronización, se obtuvo como resultado un conjunto de datos compuesto por 95 variables de proceso, todas ellas alineadas temporalmente con una resolución uniforme de un minuto. Este conjunto de datos procesado constituye la base sobre la cual se desarrollan los análisis estadísticos, exploratorios y preventivos presentados en los capítulos siguientes.

### 2.3.2 Tratamiento de datos faltantes

En esta fase se cuantificó la presencia de datos faltantes por variable, tanto en términos absolutos como porcentuales, con el objetivo de definir una estrategia de tratamiento adecuada para cada caso. La Figura 2 muestra el porcentaje de datos faltantes para algunas variables del proceso.

Como se observa en la Figura 2, ninguna de las variables de proceso presenta un porcentaje de datos faltantes superior al (20%), por lo que no fue necesaria la eliminación de variables del conjunto de datos. Las variables con mayor proporción

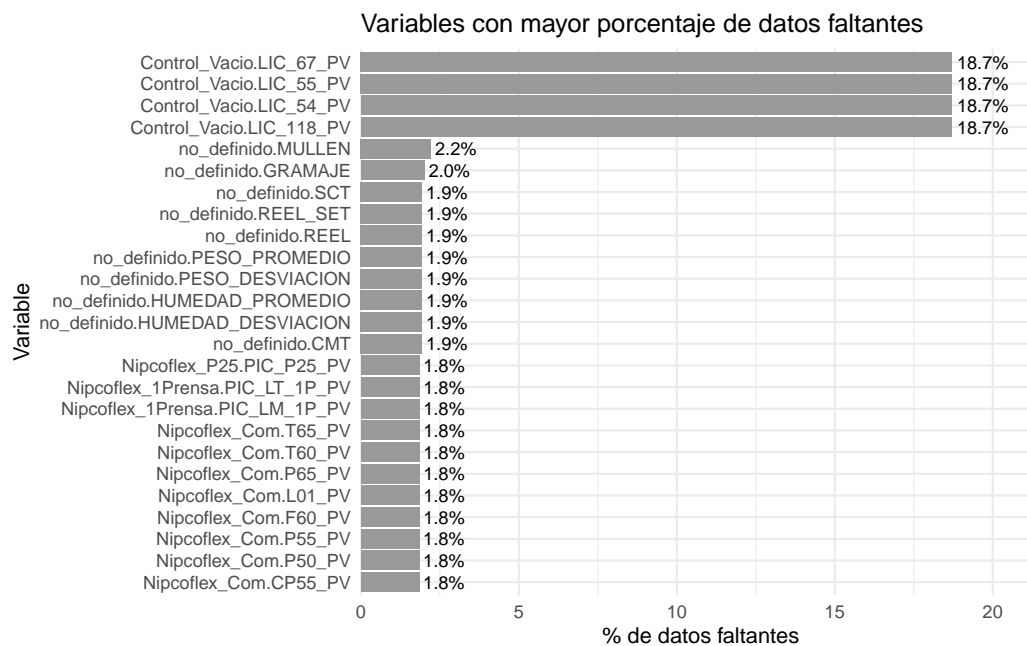


Figura 2 – Porcentaje de registros faltantes por variable.

de valores ausentes alcanzan aproximadamente un (18.7%), nivel que se considera aceptable en estudios de procesos industriales continuos, cuando el patrón de ausencia responde a características estructurales del sistema de adquisición de datos y no a fallas de medición o interrupciones del proceso (Montgomery, 2019).

Este análisis permitió identificar que los valores faltantes no se distribuyen a lo largo de la serie temporal, sino que corresponden exclusivamente a los primeros registros del conjunto de datos. Lo que está directamente asociado a la lógica del sistema industrial, el cual realiza el registro a partir de presenciar cambios, por lo que, al realizar el corte y la extracción de la base usada en este estudio, los estados previos al inicio del registro no quedan reflejados. Este tipo de ausencia de información es común en sistemas de adquisición de datos industriales y responde a un fenómeno de truncamiento temporal al inicio de la observación (Seborg *et al.*, 2010).

Bajo este contexto, el mecanismo de datos faltantes se clasifica como MCAR (Datos Faltantes Completamente Aleatorios) condicionado al inicio del registro, ya que la ausencia de información depende únicamente del momento de arranque del sistema y no de los valores observados ni de los valores no observados de las variables de proceso. De acuerdo con Little & Rubin (2002), cuando los datos faltantes se deben a truncamientos sistemáticos en los extremos de una serie temporal, la estrategia más apropiada consiste en eliminar los registros afectados, en lugar de aplicar técnicas de imputación que podrían introducir valores artificiales sin significado físico. En la misma línea, Kuhn & Johnson (2013) señalan que, en problemas de modelado predictivo con

series temporales de procesos industriales, los registros iniciales incompletos deben excluirse para garantizar la coherencia temporal del conjunto de datos y evitar sesgos derivados de imputaciones innecesarias en etapas transitorias del proceso.

En consecuencia, se optó por eliminar únicamente los primeros registros del conjunto de datos, correspondientes al período en el que el sistema de adquisición aún no había registrado la totalidad de las variables de proceso. Esta decisión permitió preservar la integridad estadística de los datos, mantener la consistencia física entre variables y asegurar que los análisis posteriores se realicen exclusivamente sobre datos representativos del proceso productivo (Montgomery, 2019; Seborg *et al.*, 2010).

Tras la eliminación de estos registros iniciales con valores faltantes, se obtuvo un conjunto de datos final compuesto por 63834 observaciones correspondientes a 95 variables de proceso, el cual constituye la base definitiva para el desarrollo de los análisis estadísticos y predictivos que se presentarán en el próximo capítulo.

### 2.3.3 Selección y codificación de variables

La selección de las variables analizadas en este estudio se fundamentó tanto en el conocimiento del proceso de fabricación de papel kraft como en la evidencia reportada en la literatura especializada, la cual analiza la relación entre las condiciones operativas, la calidad del papel y la ocurrencia de eventos de rotura de banda. Diversos autores coinciden en que las roturas de hoja están fuertemente influenciados por la interacción entre la formación de la hoja, la eficiencia en la extracción de agua, el perfil de secado, la resistencia mecánica del papel y las condiciones de operación de la máquina papelera (Paulapuro, 2000; Smook, 2016; Seborg *et al.*, 2010).

Bajo este enfoque, se seleccionaron variables asociadas a la calidad del papel, la mesa de formación, el sistema de prensas, el sistema de vacío, el control de vapor y las condiciones generales del proceso productivo. Todas estas variables fueron extraídas directamente del sistema SCADA de la planta, garantizando su representatividad y coherencia con la operación real del proceso.

Previo al análisis estadístico y al desarrollo de modelos preventivos, se aplicaron los siguientes procedimientos de codificación y transformación de variables:

- La variable temporal fue convertida de tipo fecha a tipo texto, con el objetivo de evitar pérdidas de información durante los procesos de alineación temporal y fusión de series, una práctica recomendada en el tratamiento de datos históricos provenientes de sistemas industriales (Montgomery, 2019).
- Se construyó una variable binaria denominada Corte\_Evento, definida a escala

de minuto. Esta variable toma el valor de 1 cuando ocurre al menos un evento de rotura dentro del minuto correspondiente, y 0 en caso contrario. La definición del target es:

0 : operación normal.

1 : evento de rotura de hoja en ese minuto.

La construcción de esta variable se realizó a partir de la señal original `Aproximacion.CORTE_HOJA`, considerando como evento todo valor distinto de cero registrado dentro del minuto. Cabe destacar que `Corte_Evento` no se utiliza como variable objetivo para el entrenamiento de modelos predictivos, sino exclusivamente como una herramienta metodológica para el control temporal del conjunto de datos. En particular, esta variable fue empleada para:

- Identificar el inicio del primer evento de rotura por bobina.
- Excluir ventanas temporales contaminadas por el evento.

Esta distinción es fundamental, ya que toda la información registrada a partir del momento en que ocurre la primera rotura corresponde a consecuencias del evento y no a sus causas. Incluir dichos registros en el análisis introduciría variables reactivas y generaría un sesgo significativo en los resultados.

Las variables asociadas a la calidad del papel consideran las principales propiedades físicas y mecánicas del papel kraft. La literatura especializada establece que variaciones en estas propiedades incrementan de forma significativa la probabilidad de roturas de la hoja, particularmente en la sección seca de la máquina papelera (Smook, 2016; Karlsson, 2006). Por lo tanto, a continuación se describirán las variables con las cuales se trabajará en este estudio.

- **Gramaje ( $\text{g}/\text{m}^2$ ):** representa la masa por unidad de área del papel. Variaciones locales de gramaje generan zonas estructuralmente débiles en la hoja, favoreciendo la ocurrencia de roturas (Gullichsen & Paulapuro, 2000).
- **Humedad Desviación (%)**: denota la desviación del contenido de humedad respecto al valor objetivo del proceso, asociada a inestabilidad dimensional del papel.
- **Humedad Promedio (%)**: denota el contenido promedio de humedad del papel, directamente relacionado con la resistencia mecánica y la capacidad de transferencia de esfuerzos.

- **Peso Desviación ( $\text{g}/\text{m}^2$ ):** es un indicador de variabilidad en la formación de la hoja y de irregularidades en la distribución de fibras.
- **Peso Promedio ( $\text{g}/\text{m}^2$ ):** es el peso base promedio del papel producido, parámetro clave para el control de calidad del producto final.
- **SCT ( $\text{kN}/\text{m}$ ):** representa la resistencia a la compresión en dirección de la fibra (Resistencia a la Compresión de Fibras), utilizada para evaluar la resistencia estructural del papel.
- **Mullen ( $\text{kPa}$ ):** denota la resistencia al estallido del papel, asociada al grado de enlace entre fibras y a la integridad de la estructura del material.
- **CMT ( $\text{N}$ ):** representa la resistencia a la compresión en columna (Resistencia al Aplastamiento), relevante para evaluar la rigidez y desempeño del papel en aplicaciones estructurales (Karlsson, 2006).

Los sensores generales de proceso influyen directamente en la estabilidad hidráulica y mecánica de la máquina papelera, siendo determinantes en la generación de perturbaciones que pueden derivar en eventos de rotura de hoja (Seborg *et al.*, 2010).

- **Corte Hoja:** variable de respuesta binaria que indica la ocurrencia del evento de rotura durante la operación de la máquina.
- **LIC P14 ABB PV (%)**: es el nivel de tanque o licor del sistema, relacionado con la consistencia de la pulpa y la estabilidad del proceso.
- **Visual PDIF Mult PV ( $\text{kPa}$ ):** denota la presión diferencial del sistema, asociada al balance hidráulico y a las condiciones de flujo del proceso.
- **SP Chorro Tela ( $\text{m}/\text{s}$  o %):** valor de referencia del chorro hacia la tela de formación, determinante en la orientación de fibras y uniformidad del papel.
- **PIC 407 PV ( $\text{kPa}$ ):** presión de proceso medida en un punto crítico del sistema.
- **LI Trivac PV (%)**: denota el nivel asociado al sistema de vacío, utilizado para el monitoreo operativo y la prevención de condiciones inestables.
- **FI 331 PV ( $\text{m}^3/\text{h}$ ):** caudal del proceso, relacionado con la alimentación de pulpa y la estabilidad hidráulica de la máquina.

La sección de prensas es crítica para la extracción de agua y la compactación del papel, y su influencia sobre la resistencia mecánica del producto está ampliamente documentada en la literatura técnica (Paulapuro, 2008).

- **Nipcoflex Com CP55 PV, F60 PV, L01 PV (bar / kN / mm):** variables asociadas a presiones, fuerzas y posiciones de los elementos de prensado.
- **Nipcoflex Com P50 PV, P55 PV, P56 PV, P65 PV (bar):** presiones aplicadas en las zapatas de la prensa, determinantes en la densificación del papel.
- **Nipcoflex Com T01 PV, T55 PV, T60 PV, T65 PV (°C):** temperaturas de operación de la prensa, influyentes en la eficiencia de extracción de agua.
- **Nipcoflex 1 Prensa PIC LM 1P PV (bar):** presión longitudinal en la primera prensa.
- **Nipcoflex 1 Prensa PIC LT 1P PV (bar):** presión lateral en la primera prensa.
- **Nipcoflex P25 PIC P25 PV (bar):** presión aplicada en la zapata P25.

Las siguientes variables controlan la reducción inicial de humedad y la estabilidad mecánica del papel antes del ingreso a la sección de secado final ([Smook, 2016](#)).

- **Presecado Couch Corriente PV (A):** corriente eléctrica consumida por los equipos del couch, asociada a la carga mecánica del sistema.
- **Presecado Couch Temperatura PV (°C):** temperatura de operación en la zona de couch.
- **Presecado Couch Torque PV (Nm):** torque mecánico del sistema, indicador de resistencia al movimiento.
- **Presecado Couch Velocidad PV (m/min):** velocidad de operación del couch, influyente en la estabilidad de la hoja.
- **Presecado 1 Prensa Corriente PV (A):** corriente eléctrica de la primera prensa húmeda.
- **Presecado 1 Prensa Temperatura PV (°C):** temperatura de operación de la prensa.
- **Presecado 1 Prensa Velocidad PV (m/min):** velocidad de operación de la primera prensa.

El sistema de vacío es determinante para la formación uniforme del papel y la eliminación de agua libre en la mesa húmeda, ya que controla la extracción de agua

durante la etapa de formación, influyendo directamente en la distribución de fibras, el gramaje y la estabilidad mecánica de la hoja (Gullichsen & Paulapuro, 2000).

- **Control Vacío I Filtro Inferior 1 PV, Control Vacío I Filtro Superior 1 PV (A):** corriente eléctrica consumida por las bombas de vacío asociadas a los filtros inferior y superior, utilizada como indicador indirecto de carga operativa y eficiencia del sistema de vacío.
- **Control Vacío Potencia Filtro Inferior PV, Control Vacío Potencia Filtro Superior PV (kW):** potencia absorbida por los motores de vacío, relacionada con la intensidad de extracción de agua y el estado mecánico del sistema.
- **Control Vacío Presión PIC 103 PV a PIC 353 PV, PIC 57 PV (kPa):** presiones de vacío medidas en diferentes puntos de la mesa húmeda y zona de couch, las cuales determinan el perfil de drenaje y la uniformidad transversal del papel.
- **Control Vacío Torque Vac Couch PV (Nm):** torque del sistema de vacío en la sección couch, asociado a la resistencia mecánica durante la extracción de agua y a la carga dinámica del equipo.
- **Control Vacío Velocidad Vac Couch PV, Control Vacío Velocidad Vacío Mesa PV (m/min):** velocidad de operación de los elementos de vacío, influyente en la capacidad de drenaje y en la estabilidad de la hoja durante la formación.
- **Control Vacío Nivel LIC 116 PV, LIC 117 PV, LIC 118 PV (%):** nivel de los tanques asociados al sistema de vacío, utilizado para monitorear el balance hidráulico y prevenir condiciones inestables que puedan afectar la continuidad del proceso.

El control térmico en la sección de secado regula la evaporación final del agua contenida en la hoja de papel y condiciona el contenido de humedad residual. Este factor es determinante en la generación de tensiones internas, deformaciones y roturas del papel, especialmente en condiciones de operación inestables (Paulapuro, 2008).

- **PDIC 313 PV, PDIC 400 PV, PDIC 401 PV, PDIC 406 PV (kPa):** controladores de presión diferencial del sistema de vapor en la sección de secado, asociados a la distribución y estabilidad del flujo de vapor en los cilindros secadores.

- **PIC 301 PV, PIC 306 PV, PIC 307 PV, PIC 314 PV, PIC 324 PV, PIC 402 PV, PIC 405 PV (kPa):** presiones absolutas del sistema de vapor, utilizadas para garantizar condiciones térmicas homogéneas y evitar gradientes de temperatura en la hoja.
- **TIC 302 PV, TIC 305 PV, TIC 325 PV, TIC 405 PV (°C):** temperaturas de operación en diferentes zonas de la sección de secado, directamente relacionadas con la tasa de evaporación del agua y el secado uniforme del papel.
- **TIC TQ A Filtrad PV (°C):** temperatura del tanque de filtrado, relevante para el control térmico del sistema auxiliar de vapor.

### 2.3.4 Análisis exploratorio

Se inicia el estudio con un análisis exploratorio de datos (EAD), como etapa previa a la selección de variables y al desarrollo de modelos predictivos, con el objetivo de comprender la estructura, distribución y comportamiento conjunto de las variables del proceso productivo.

El análisis se fundamenta en la aplicación de técnicas de estadística descriptiva donde se evalúan medidas de tendencia central, tales como la media, la mediana, incluyendo la desviación estándar. Estas métricas permiten caracterizar el comportamiento típico de las variables de calidad del papel, las condiciones operativas y las variables de control del proceso, así como identificar asimetrías, posibles sesgos y desviaciones respecto a condiciones nominales de operación (Montgomery, 2019). Este análisis, aunque parece simple, permite descubrir las etapas del proceso productivo donde las roturas de papel son recurrentes.

Adicionalmente, se emplean diagramas de caja (boxplots) para un subconjunto de variables, entre ellas el gramaje, la humedad promedio, las presiones del sistema de vacío, las presiones de vapor y las variables asociadas a la sección de prensas. Estos gráficos se utilizan para la detección de valores atípicos y la identificación de rangos operativos anómalos, los cuales pueden estar asociados tanto a perturbaciones transitorias del proceso como a condiciones precursoras de eventos de rotura de hoja (Gullichsen & Paulapuro, 2000).

### 2.3.5 Construcción del dataset de análisis preventivo

La literatura de modelado predictivo, indican la importancia de evitar la fuga de información temporal en problemas de series de tiempo. Por otro lado, Kuhn & Johnson (2013) señalan que la inclusión de observaciones posteriores al evento

objetivo conduce a un optimismo artificial del desempeño del modelo, mientras que Shmueli (2010) destaca la necesidad de separar claramente los objetivos de explicación y predicción, especialmente en contextos donde se busca un enfoque preventivo. Cuando la variable de interés está asociada a un evento puntual dentro de un proceso continuo, los predictores deben construirse utilizando exclusivamente información previa al evento. De lo contrario, el análisis se vuelve reactivo y pierde utilidad operativa, al incorporar efectos que son consecuencia directa del fallo (Kuhn & Johnson, 2013).

En este estudio, el objetivo no es predecir el minuto exacto en que ocurre una rotura de hoja, sino entender y prevenir su ocurrencia, identificando las condiciones operativas que caracterizan a bobinas de alto riesgo. En consecuencia, se definió como unidad de análisis final la bobina, construyendo un dataset compuesto por una fila por bobina. De tal manera, se identificó para cada bobina el primer minuto en el que ocurre un evento de rotura, conservando datos antes del evento y eliminando aquellos registros posteriores. Para el caso de las bobinas en donde no ocurrió una rotura se conservaron en su totalidad. Este procedimiento garantiza que los datos utilizados para el análisis contengan información previa al evento, evitando así la casualidad temporal.

A partir de la base minuto a minuto, se construyó el dataset final de análisis, en el cual cada fila representa un resumen del comportamiento operativo de una bobina, y no un instante aislado de tiempo. Cada fila contiene métricas agregadas que describen el comportamiento del proceso previo al primer evento de rotura, incluyendo medidas de tendencia central, dispersión y estabilidad operacional. En el caso de bobinas que presentaron uno o más eventos de rotura, se identificó el primer inicio de rotura, y se utilizaron únicamente los minutos anteriores a dicho evento. Los minutos correspondientes al evento de rotura y los posteriores fueron excluidos del análisis, dado que contienen información reactiva asociada a las consecuencias del fallo. Las bobinas que no presentaron roturas fueron conservadas en su totalidad, excluyendo únicamente los períodos de cambio de bobina cuando existieron.

Bajo esta lógica, se definió una variable indicadora a nivel bobina, donde  $Corte\_Bobina = 1$  si la bobina presentó al menos un evento de rotura, y  $Corte\_Bobina = 0$  en caso contrario. En bobinas con múltiples roturas, el interés no radica en la cantidad de eventos, sino en caracterizar las condiciones operativas que hacen a la bobina riesgosa. En consecuencia, la construcción de una fila por bobina proporciona un dataset equilibrado, interpretable y coherente con el objetivo preventivo del estudio, permitiendo identificar patrones operativos y variables críticas asociadas al riesgo de rotura, y constituyendo la base para los análisis estadísticos (Kuhn & Johnson, 2013; Bishop, 2006).

### 2.3.6 Selección de variables para el enfoque preventivo

La selección de variables se realizó siguiendo un enfoque jerárquico que prioriza la coherencia física y la utilidad operativa, antes de aplicar criterios puramente estadísticos. Este procedimiento es fundamental en modelos preventivos, donde el objetivo es identificar condiciones del proceso que anteceden al fallo.

En una primera etapa, se realizó una selección conceptual de variables basada en el conocimiento del proceso productivo y en criterios de control operativo. Para este estudio, se consideraron únicamente variables que cumplen simultáneamente las siguientes condiciones:

- Son manipulables o ajustables por el operador.
- Están directamente relacionadas con el proceso físico de fabricación.
- Corresponden a etapas previas a la ocurrencia del evento de rotura.

Bajo estos criterios, se seleccionaron variables pertenecientes a las etapas de:

- Aproximación,
- Control de vacío,
- Control de vapor,
- Nipcoflex,
- Presecado (Couch).

Por el contrario, se excluyeron explícitamente variables correspondientes a la etapa denominada *no\_definido*, las cuales representan características de calidad medidas posterior a la finalización del proceso de fabricación (por ejemplo, gramaje, humedad, Mullen). Estas variables incorporan información del futuro respecto al evento de interés y, por lo tanto, violan el principio de causalidad temporal (Pearl, 2009). Finalmente, no se consideró la variable *Aproximacion.CORTE\_HOJA* ni la variable temporal *minuto*, dado que el evento se utiliza únicamente como marcador temporal y no como predictor.

Una vez definida la base conceptual de variables, se aplicaron criterios estadísticos para depurar señales con bajo aporte informativo. En primer lugar, se

eliminaron variables con desviación estándar cercana a cero, dado que no presentan variabilidad suficiente para discriminar entre condiciones operativas y suelen corresponder a sensores constantes o saturados.

Posteriormente, se aplicó una selección basada en pruebas no paramétricas, utilizando la prueba estadística no paramétrica de Mann Whitney U que consiste en comparar si existen diferencias entre dos grupos independientes, junto con el tamaño de efecto Cliff's Delta. Este enfoque permite identificar diferencias relevantes entre bobinas con y sin rotura, sin asumir normalidad y es robusto frente a valores atípicos (Romano *et al.*, 2006). Se priorizaron variables con valores absolutos de Cliff's Delta mayores o iguales a 0.25 y valores de  $p$  ajustados, evitando la dependencia exclusiva en significancia estadística (Wasserstein *et al.*, 2019).

### 2.3.7 Modelos preventivos

Una vez definidas las variables, se desarrolló un modelo preventivo que ayude a caracterizar el riesgo operativo de una bobina. El objetivo de este modelo no es predecir el instante exacto de ocurrencia del fallo, sino estimar un *indicador de riesgo* asociado a la bobina, que permita identificar configuraciones operativas propensas a la aparición del evento. Por tal motivo se formuló como un problema de clasificación binaria.

Como modelo base se utilizará la regresión logística, debido a su idoneidad para problemas de clasificación binaria y a su amplia utilización en el análisis de procesos industriales y sistemas de control (Hastie *et al.*, 2009). También permite modelar una variable binaria a nivel de bobina, definida como:

$$Corte\_Bobina = \begin{cases} 1, & \text{si la bobina presentó al menos un evento de rotura} \\ 0, & \text{en caso contrario} \end{cases}$$

Este modelo permite estimar la probabilidad condicional de ocurrencia de un evento de rotura en función del comportamiento de las variables de proceso, manteniendo al mismo tiempo un elevado nivel de interpretabilidad de los coeficientes, lo cual resulta clave para la comprensión física del fenómeno y la toma de decisiones operativas.

El modelo fue entrenado utilizando exclusivamente información previa al primer evento de rotura de cada bobina, garantizando que todas las variables incluidas representan condiciones causales y no reacciones al fallo. De esta manera, la probabilidad estimada por el modelo debe interpretarse como un indicador de que una bobina opera

bajo condiciones que incrementan su susceptibilidad a la rotura, y no como una predicción temporal del evento.

El conjunto final de variables incluidas en el modelo fue seleccionado priorizando aquellas con mayor coherencia física, tamaño de efecto relevante y estabilidad estadística, resultando en un conjunto reducido de variables críticas para la interpretación operativa. En particular, se seleccionaron las once variables con mayor contribución relativa al modelo, permitiendo un balance adecuado entre capacidad explicativa e interpretabilidad.

Es importante destacar que la evaluación del modelo no se abordó desde una perspectiva de desempeño predictivo clásico, sino como una verificación de consistencia interna y estabilidad del indicador de riesgo. En este sentido, las métricas utilizadas se emplearon únicamente como herramientas de diagnóstico, evitando su interpretación como capacidad de predicción futura, en concordancia con el enfoque preventivo del estudio (Shmueli, 2010).

En conjunto, el modelo preventivo desarrollado constituye una herramienta de apoyo a la toma de decisiones operativas, permitiendo identificar patrones de comportamiento del proceso y variables críticas asociadas a bobinas de alto riesgo, proporcionando así una base analítica sólida para la implementación de estrategias de control orientadas a la prevención del evento de rotura.

Para la evaluación del desempeño de los modelos, el conjunto de datos se dividió en subconjuntos de entrenamiento y prueba, garantizando que información a una misma bobina no se superpongan entre ambos conjuntos. Esta separación se hizo a nivel de bobina, con el fin de preservar la independencia de las observaciones. Posteriormente, dado que el objetivo del modelo no es la predicción puntual del evento de rotura, sino la estimación de un indicador continuo de riesgo asociado a la bobina, la evaluación del modelo se realizó exclusivamente mediante el área bajo la curva ROC (AUC). Esta métrica permite cuantificar la capacidad del modelo para discriminar entre bobinas que presentan al menos un evento de rotura y aquellas que no lo presentan, independientemente de la selección de un umbral específico de decisión. La curva ROC (Característica Operativa del Receptor) se construye gráficamente la Tasa de Verdaderos Positivos (TPR o Recall) en el eje vertical frente a la Tasa de Falsos Positivos (FPR) en el eje horizontal, donde:

$$\text{TPR} = \frac{TP}{TP + FN}, \quad \text{FPR} = \frac{FP}{FP + TN}$$

donde  $TP$ ,  $FP$ ,  $TN$ , y  $FN$  representan los verdaderos positivos, falsos positivos, verdaderos negativos y falsos negativos, respectivamente. Por otro lado, el

área bajo la curva ROC (AUC) se determina como:

$$\text{AUC} = \int_0^1 \text{TPR}(u) du$$

donde,  $u = FPR$

El uso del AUC resulta especialmente adecuado en este contexto por tres razones principales:

- Es robusto frente al desbalance entre clases, condición frecuente en procesos industriales donde los eventos de fallo son poco frecuentes.
- Evalúa la capacidad de ordenamiento del modelo en términos de riesgo relativo, lo cual es coherente con un enfoque preventivo
- Evita la necesidad de fijar umbrales operativos arbitrarios, que no forman parte del objetivo del presente estudio (Hastie *et al.*, 2009; Shmueli, 2010).

En consecuencia, métricas dependientes de un umbral de clasificación, tales como precisión, *recall* o F1, no fueron consideradas, dado que su interpretación está orientada a escenarios de predicción binaria y no a la evaluación de un indicador de riesgo continuo. Para mitigar el efecto del desbalance existente entre bobinas con y sin eventos de rotura, se aplicó una ponderación de clases durante el entrenamiento del modelo, asignando pesos inversamente proporcionales a la frecuencia de cada clase. Este procedimiento permitió estabilizar la estimación de los parámetros sin alterar la interpretación del indicador de riesgo generado por el modelo.

### 2.3.8 Diseño del visualizador

Se implementa un visualizador en Power BI, aprovechando sus capacidades de interactividad y visualización dinámica para facilitar la exploración, análisis e interpretación de la información generada a partir de los datos del proceso productivo. La herramienta permite al usuario interactuar con los datos mediante filtros temporales, selección de bobinas pasadas y comparar entre condiciones de operatividad de la bobina fabricada en tiempo real. Esta interactividad resulta fundamental para el análisis del proceso y para la validación operativa de los resultados obtenidos a partir de los modelos preventivos, en un entorno industrial real.

El diseño del visualizador se estructuró con base en las variables significativas asociadas a la ocurrencia de eventos de rotura, previamente identificadas mediante el análisis exploratorio de datos y en el modelo de clasificación desarrollado. Dichas variables provienen de los principales subsistemas del proceso de fabricación del papel,

incluyendo la sección de formación, las prensas, el sistema de vacío, el presecado y el control de vapor, lo que permite una visión integral del estado operativo de la máquina.

### 2.3.8.1 Estructura funcional del visualizador

El panel principal del visualizador se organizó en los siguientes componentes:

- **Panel de indicador de riesgo:** este panel presenta una métrica agregada y de interpretación directa, diseñadas para apoyar el diagnóstico rápido del estado del proceso y la toma de decisiones operativas.
  - **Indicador de riesgo de rotura de hoja,** expresado como una probabilidad estimada a nivel de bobina, obtenida a partir del modelo preventivo. Este indicador no representa una predicción temporal del evento, sino un score relativo que permite identificar condiciones operativas asociadas a mayor susceptibilidad de la rotura.
  - **Estado operativo del sistema,** clasificado en categorías cualitativas (operación normal / condición de riesgo), derivadas del comportamiento conjunto de las variables críticas del proceso.

Estas medidas permiten al personal operativo identificar de manera temprana desviaciones respecto a condiciones normales de operación y priorizar acciones preventivas antes de que el proceso alcance estados críticos.

- **Visualizaciones temporales de variables de proceso:** el panel incluye gráficos de series temporales correspondientes a etapas de Control de Vacío, Control de Vapor, Presecado y Nipcoflex. Estas visualizaciones permiten comparar el comportamiento del proceso, facilitando la identificación de tendencias, incrementos de variabilidad y desviaciones progresivas que comprometen la integridad de la hoja.

En conjunto, el visualizador integra los resultados del análisis estadístico y del modelado predictivo en una plataforma interactiva, permitiendo transformar grandes volúmenes de datos industriales en información accionable para la mejora continua de la estabilidad del proceso y la reducción de eventos de rotura de hoja.

En la Figura 3 se muestra el resultado del visualizador con las series de cada una de las etapas identificadas con los eventos de rotura a lo largo del proceso. En el se puede filtra por las variables más significativas y ver el comportamiento de la última bobina en producción, si en caso la variable esta generando condiciones que podrían provocar un rotura, el indicador muestra una alerta y una acción sugerida.

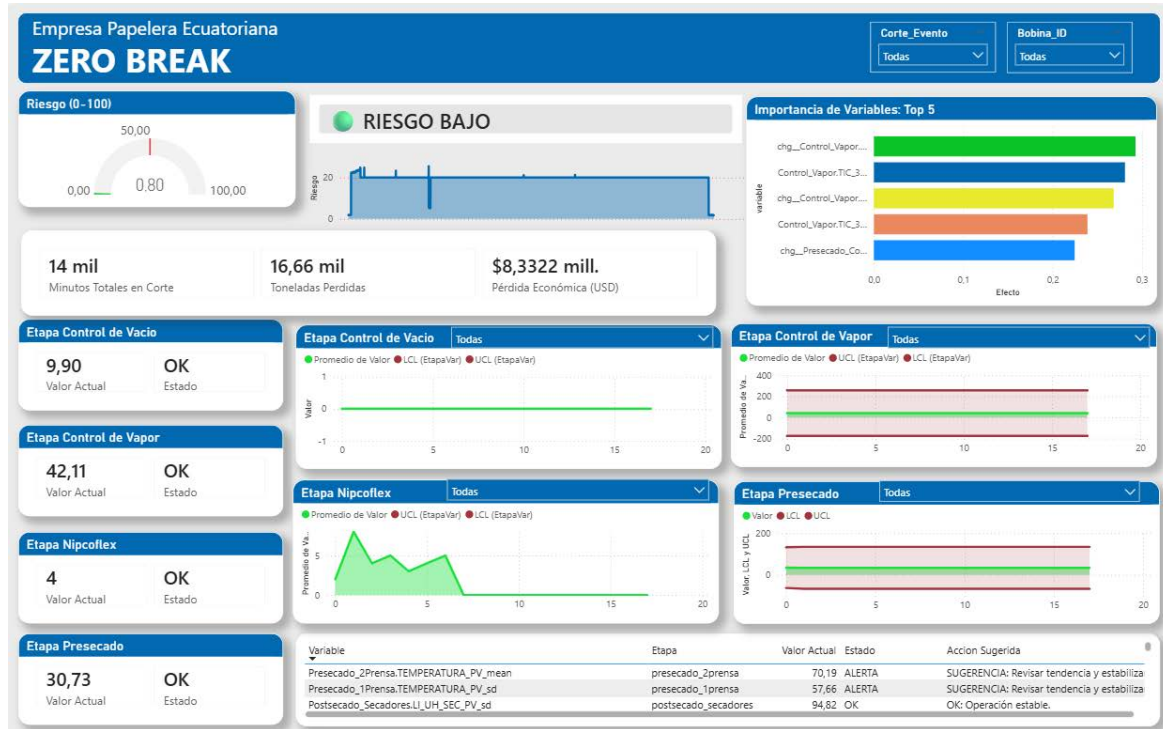


Figura 3 – Visualizador

### 3 Resultados

En esta sección se presentan los resultados obtenidos a partir de la aplicación de técnicas de estadística descriptiva, así como del uso de Power BI para el desarrollo de un visualizador interactivo orientado al análisis de la calidad del papel en una empresa ecuatoriana. El análisis se fundamenta en una base de datos previamente sometida a procesos de limpieza, depuración y estructuración, con el fin de garantizar la consistencia y confiabilidad de la información obtenida.

La base de datos está compuesta por un total de 63834 observaciones y 95 variables asociadas al proceso de fabricación del papel, registradas de manera continua a intervalos de un minuto durante un período de 45 días de producción. Adicionalmente, se dispone de 95 variables complementarias que capturan la variabilidad intraminuto de los principales parámetros del proceso.

Asimismo, el conjunto de datos incluye variables de carácter operativo y de control, tales como “minuto”, “Cambio.de.Bobina”, “CORTE\_ANY\_1MIN”, “CORTE\_END\_1MIN”, “CORTE\_DUR\_SEG\_1MIN”, “CAMBIO\_ANY\_1MIN”, “CAMBIO\_END\_1MIN”, “CAMBIO\_DUR\_SEG\_1MIN”, “Bobina\_ID” y “Corte\_Evento”, las cuales permiten identificar eventos relevantes dentro del proceso productivo y segmentar el análisis en función de la ocurrencia de roturas y cambios de bobina.

A continuación, se describen las distintas etapas del proceso de fabricación del papel, junto con la cantidad de variables asociadas a cada una de ellas.

Cuadro 1 – Cantidad de variables por etapa del proceso de fabricación

| <b>Etapa del proceso</b>       | <b>Número de variables</b> |
|--------------------------------|----------------------------|
| Control de Vacío               | 35                         |
| Control de Vapor               | 16                         |
| Prensas (Nipcoflex)            | 14                         |
| Calidad (peso/humedad/ensayos) | 12                         |
| Otros / No clasificado         | 10                         |
| Presecado                      | 9                          |
| Formación / Aproximación       | 6                          |
| Postsecado (Secadores)         | 2                          |
| Marcadores de Corte            | 1                          |

Como se observó en el Cuadro 1 existe una mayor concentración de variables en la etapa de Control de Vacío, Control de Valor y Prensas (Nipcoflex), lo

cual evidenció un mayor monitoreo de estas etapas a diferencia de las etapas de Pre-secado, Formación / Aproximación y Postsecado (Secadores). Por otro lado, existen dos clasificaciones que no entran dentro de las etapas del proceso, pero contienen variables de identificación y control como Marcadores de Corte que contiene únicamente la variable Aproximación.de.CORTE que cuenta los segundos que transcurrieron por cada evento de rotura y la clasificación Otros/ No clasificado que contiene Bobina\_ID, Corte\_Evento, minuto y Cambio.de.Bobina.

Entre las variables de identificación y control, destacan aquellas de vital importancia operativa, como la variable *Corte\_Evento* definida previamente como una variable binaria (0 = no ocurre rotura, 1= ocurre rotura) y *Bobina\_ID* que permitió la identificación de cada una de las bobinas sobre las cuales se llevó a cabo el proceso de la fabricación del papel.

Cuadro 2 – Distribución de evento de rotura

| Evento de Rotura | Frecuencia | Porcentaje (%) |
|------------------|------------|----------------|
| 0                | 63721      | 99.82 %        |
| 1                | 113        | 0.18 %         |

La naturaleza de los datos actualmente presenta independencia temporal minuto a minuto, ya que provienen de un proceso continuo originalmente registrado segundo a segundo, en el cual existía dependencia temporal. Las bobinas constituyen la unidad de observación independiente; sin embargo, es importante analizar la relación entre las variables. A continuación, se presenta la distribución de la variable “Bobina\_ID” según la ocurrencia de roturas:

Cuadro 3 – Distribución de bobinas por roturas

| Tipo      | Cantidad de bobinas | Porcentaje (%) |
|-----------|---------------------|----------------|
| No rotura | 1300                | 93.73          |
| Rotura    | 87                  | 6.27           |

En total, se identificaron 1387 bobinas de las cuales 87 presentaban al menos un evento de rotura, lo cual evidencia un desbalance entre las bobinas con respecto al evento de la rotura. Como se observa en los Cuadros 2 y 3, aunque la ocurrencia del evento de rotura represente un 0.18 % de presencia en la base de datos, dicho evento se presenta en el 6.27 % de las bobinas, es decir en promedio 1 de cada 16 bobinas presentan al menos una rotura, es decir, existe un impacto significativo a nivel de unidad productiva.

En el Cuadro 4 se presenta el número de roturas que ocurrieron en algunas de las bobinas. Por ejemplo, la bobina 803 presentó el mayor número de roturas (4), mientras que la bobina 506, 587, 1031 y 1032 presentaron tres roturas a lo largo del proceso de producción.

Cuadro 4 – Bobinas que registraron al menos una rotura.

| ID   | Cantidad de Roturas | Observaciones por bobina |
|------|---------------------|--------------------------|
| 803  | 4                   | 7517                     |
| 506  | 3                   | 64                       |
| 587  | 3                   | 70                       |
| 1031 | 3                   | 150                      |
| 1032 | 3                   | 74                       |
| 44   | 2                   | 38                       |
| 60   | 2                   | 43                       |
| 248  | 2                   | 54                       |
| 388  | 2                   | 326                      |
| 405  | 2                   | 58                       |

A continuación se muestran estadísticas descriptivas de un grupo de variables identificadas en cada una de las etapas del proceso de fabricación:

Cuadro 5 – Estadísticas descriptivas globales de ciertas variables por etapa de fabricación.

| Variable                          | Media  | SD     | P05   | Mediana | P95    |
|-----------------------------------|--------|--------|-------|---------|--------|
| Control_Vacio.LIC_TQ_TORRE_ENF_PV | 76.6   | 36.0   | -0.32 | 93.8    | 94.9   |
| Control_Vacio.PIC.103_PV          | 15.4   | 6.3    | 0.08  | 18.1    | 19.6   |
| Control_Vacio.PIC_338_PV          | -3.5   | 1.8    | -4.9  | -4.4    | 0.06   |
| Control_Vacio.TORQ_VAC_COUCH_PV   | 60.7   | 26.7   | 0.0   | 69.8    | 85.0   |
| Control_Vacio.VEL_VACIO_MESA_PV   | 61.3   | 36.3   | 0.0   | 70.7    | 100    |
| Control_Vapor.PIC_301_PV          | 3.3    | 0.43   | 2.99  | 3.40    | 3.62   |
| Control_Vapor.PIC_405_PV          | 2.56   | 0.53   | 2.00  | 2.50    | 3.60   |
| Nipcoflex_Com.L01_PV              | 96.5   | 5.0    | 94.9  | 96.2    | 101.8  |
| Nipcoflex_Com.P56_PV              | 38.9   | 18.9   | -0.08 | 49.5    | 52.1   |
| Nipcoflex_Com.T01_PV              | 44.7   | 5.7    | 29.3  | 46.6    | 48.09  |
| Nipcoflex_Com.T65_PV              | 49.2   | 9.07   | 26.6  | 52.1    | 55.4   |
| HUMEDAD_DESVIACION                | 0.30   | 0.44   | 0.13  | 0.29    | 0.59   |
| HUMEDAD_PROMEDIO                  | 7.36   | 0.53   | 6.68  | 7.42    | 7.92   |
| PESO_DESVIACION                   | 1.51   | 1.8    | 0.9   | 1.38    | 2.30   |
| Presecado_1Prensa.VELOCIDAD_PV    | 322.57 | 137.5  | 0.00  | 350.7   | 435.87 |
| Presecado_Couch.CORRIENTE_PV      | 480.74 | 199.51 | 0.00  | 556.29  | 632.26 |
| Presecado_Couch.TORQUE_PV         | 39.06  | 18.4   | 0.00  | 46.64   | 54.02  |
| Aproximacion.SP_CHORRO_TELA       | -41.57 | 7.4    | -52.0 | -42.0   | -29.0  |

El Cuadro 5 muestra las estadísticas descriptivas de las variables pertenecientes a la etapa de Control\_Vacio (PIC, LIC, TORQ, VELOCIDAD, VEL). Aquí observamos que variables como TORQ\_VAC\_COUCH, VEL\_VACIO\_MESA y LIC\_TQ\_TORRE\_ENF\_PV presentan alta dispersión. El ancho de los percentiles muestra distintos comportamientos que pueden afectar la estabilidad de la hoja, por ejemplo el 90 % del tiempo TORQ\_VAC\_COUCH\_PV opera entre 0.00 y 85.00, por otro lado VEL\_VACIO\_MESA\_PV opera entre 0.0 y 100, lo cual indica la existencia de distintos regímenes. La variables pertenecientes a la etapa Prensas (Nipcoflex), Nipcoflex (L01\_PV, T01\_PV) mostraron baja variabilidad con percentiles que el 90 % del tiempo mantienen el sistema estable. Por otro lado, T65\_PV presenta mayor dispersión que L01\_PV, pero moderadamente sin entrar en inestabilidad a diferencia de P56\_PV que opera entre -0.08 y 52.1 lo que indica que el sistema opera en variabilidad normalmente, además de tener distintos regímenes.

Además, se analizó el comportamiento de las roturas por días y turnos, dado que dentro de la empresa existen tres turnos donde se supervisa el proceso, además de la carga horaria.

En la Figura 4 se evidenció que la mayor cantidad de roturas ocurrieron el día martes representando el 22 % de evento, además del día lunes y miércoles representando el 20 %, acumulando entre estos tres día aproximadamente el 62 % de ocurrencias asociadas al evento rotura. Por otro lado los días con menos ocurrencias fueron el día domingo y el día viernes.

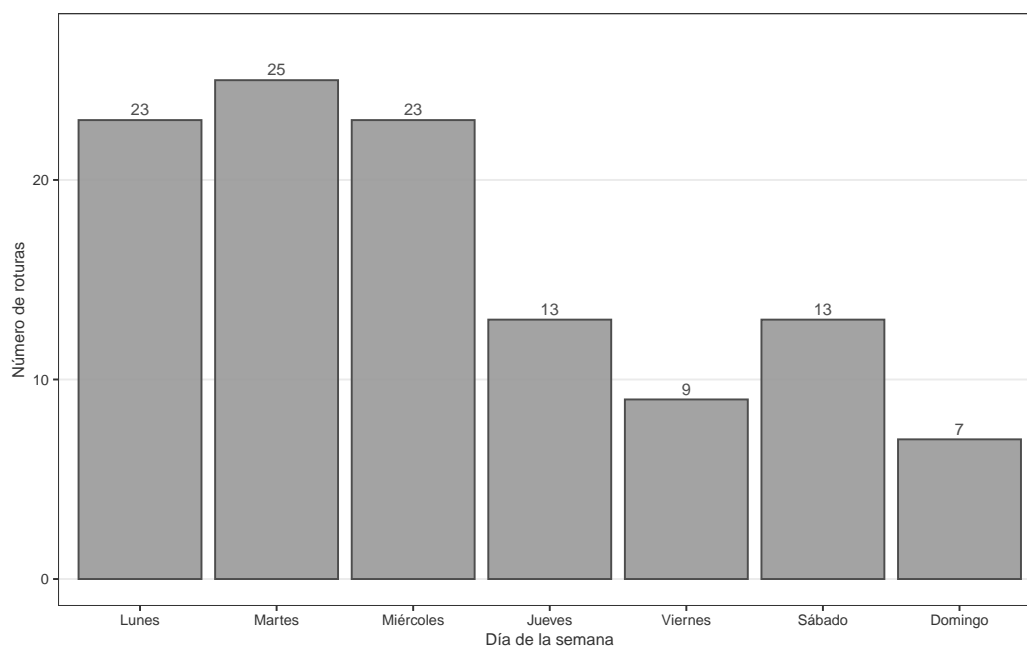


Figura 4 – Distribución de roturas por día del proceso de fabricación.

En la Figura 5 se observó que la mayor cantidad de ocurrencias en roturas sucedió en el turno 3, comprendido entre las 23h00 y las 7h00 con el 38 % de las roturas en su totalidad, y esto se evidencia en el Cuadro 6 donde las mayores ocurrencias están en el rango horario perteneciente al turno 3, siendo el mayor número de roturas 10 registradas a las 23h00 seguido de 9 roturas a las 00h00, finalmente se observó una concentración de ocurrencias entre las 23h00 a las 1h00.

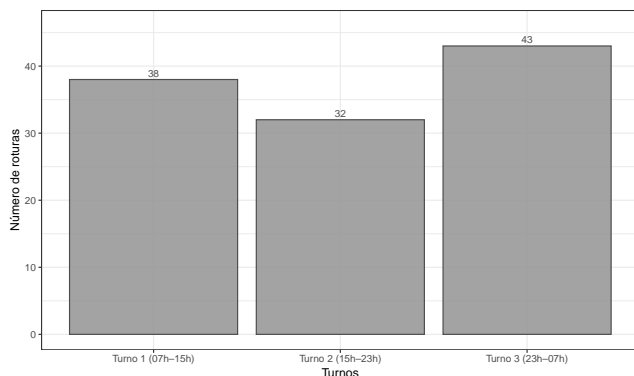


Figura 5 – Distribución de roturas por turno.

Cuadro 6 – Roturas por hora del día

| Hora | Cantidad de roturas |
|------|---------------------|
| 23   | 10                  |
| 0    | 9                   |
| 1    | 7                   |
| 8    | 7                   |
| 6    | 6                   |
| 14   | 6                   |
| 15   | 6                   |
| 19   | 6                   |
| 9    | 5                   |
| 11   | 5                   |

En el cuadro 7 muestra la distribución porcentual de los eventos de rotura asociados a las diferentes etapas del proceso productivo. Se pudo evidenciar que la etapa activa en la que ocurren más eventos de rotura es la etapa de Vacío encargada de eliminar el agua, seguida de la Etapa de Vapor con una presencia de 18 % aproximadamente, y seguida de la etapa de Nipcoflex con el 17.78 %.

Cuadro 7 – Distribución porcentual de eventos de rotura asociados por etapa del proceso

| Etapa del proceso | Porcentaje de eventos (%) |
|-------------------|---------------------------|
| Vacío             | 27.62                     |
| Vapor             | 18.42                     |
| Nipcoflex         | 17.78                     |
| Aproximación      | 9.21                      |
| No Clasificada    | 9.21                      |
| Calidad           | 8.89                      |
| Presecado         | 8.89                      |
| <b>Total</b>      | <b>100.00</b>             |

Además, se cuantificó el número de roturas por cada una de las variables de calidad. A continuación se muestran los resultados.

Cuadro 8 – Gramaje con mayor número de roturas

| Gramaje | Cant. de roturas |
|---------|------------------|
| 146     | 71               |
| 230     | 25               |
| 205     | 10               |
| 170     | 5                |

Cuadro 9 – SCT con mayor número de roturas

| SCT  | Cant. de roturas |
|------|------------------|
| 3.40 | 18               |
| 3.16 | 5                |
| 3.07 | 4                |
| 3.25 | 4                |

Cuadro 10 – Humedad promedio con mayor número de roturas

| Humedad prom. | Cant. de roturas |
|---------------|------------------|
| 6.82          | 12               |
| 7.67          | 8                |
| 6.70          | 6                |
| 7.47          | 5                |

Cuadro 11 – Peso promedio con mayor número de roturas

| Peso prom. | Cant. de roturas |
|------------|------------------|
| 149.25     | 10               |
| 148.08     | 9                |
| 146.75     | 5                |
| 148.17     | 5                |

Cuadro 12 – Chorro Tela con mayor número de roturas

| Chorro Tela | Cant. de roturas |
|-------------|------------------|
| -50         | 11               |
| -40         | 11               |
| -42         | 10               |
| -36         | 10               |
| -52         | 8                |

Estas variables fueron exploradas con el objetivo de evaluar su impacto sobre la calidad del papel al final del proceso productivo. Cabe señalar que las variables de calidad no influyen de manera directa en la ocurrencia de roturas, dado que sus

mediciones se registran únicamente en la etapa final del proceso, es decir, sobre el producto terminado, una vez que las bobinas han sido conformadas.

Según los Cuadros mostrados 8 9, 10, 11 , se observó que aproximadamente el 63 % de las roturas ocurrieron cuando el gramaje fue de 146g/m2 que suele ser un gramaje bastante demandado, el 16 % cuando la resistencia SCT fue de 3.40, el 10.6 % de las roturas se dieron cuando la humedad promedio fue de 6.82 %, y finalmente el peso promedio 149.25 con aproximadamente el 8.8 %. A nivel descriptivo esto indica que podría ser de interés estudiar el comportamiento de estas variables en relación a la ocurrencia de roturas. Por otro lado, en el Cuadro 12 se muestran los valores setpoint de chorro de tela, donde se observa que la mayor ocurrencia de roturas ocurrió alrededor del valor -50 y -40 aproximadamente (9 %). Esta variable está asociada con la formación de la bobina y sus fibras por lo cual es importante para la ocurrencia del evento.

En la Figura 6 se muestra el diagrama de cajas de algunas variables operativas y de calidad en torno al evento de las roturas, donde la medida central es la mediana, la cual no es influenciada por valores extremos que puedan ocurrir en el proceso.

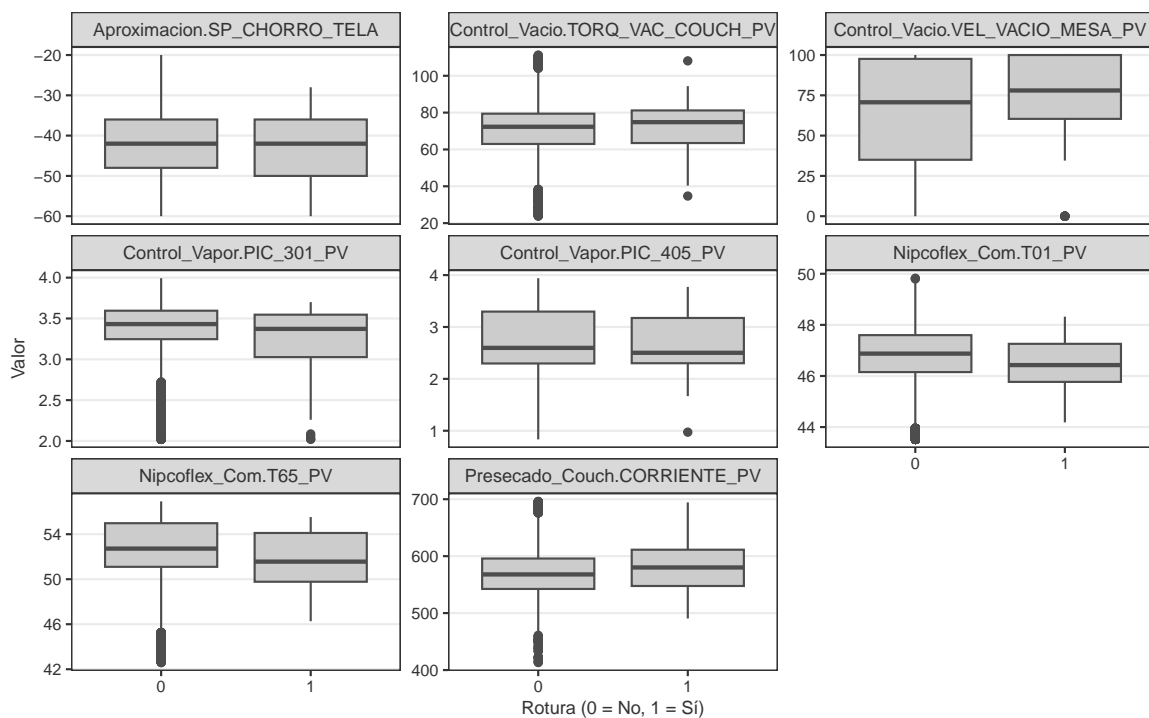


Figura 6 – Distribución de roturas en variables claves.

Si bien la mayoría de las variables observadas en el gráfico no presentan diferencias con respecto a la mediana entre bobinas con y sin rotura, siendo estas

similares, se observan diferencias con respecto a la dispersión, asimetría y valores atípicos, lo cual indica que la variabilidad del proceso podría estar directamente asociada al evento de rotura a nivel descriptivo.

Dado lo anterior, se hizo uso de ventanas móviles para estudiar la variabilidad, desviaciones, fluctuaciones de las variables del proceso en la rotura a nivel descriptivo en intervalos de 120 segundos cercanos a la ocurrencia de la rotura, es decir 2 minutos, con el fin de observar patrones. Se priorizaron variables por etapa, estudiando la variabilidad y la continuidad dado que las roturas se ven asociadas a variabilidad dentro del proceso de la fabricación del papel tomando de 7 variables para efectos gráficos.

A continuación, se presenta la evolución temporal de una de las variables pertenecientes a la etapa de control de vacío, `Control_Vacio.TORQ_VAC_COUCH_PV`, la cual fue identificada como una de las variables con mayor variabilidad en torno a los eventos de rotura dentro de esta etapa del proceso de fabricación.

La Figura 7 muestra que, para la bobina 232 que presenta un evento de rotura, ni en el valor instantáneo ni en la media móvil muestran quiebres significativos antes del evento, el cual está representado por la línea entrecortada roja, pero la desviación estándar móvil presenta ciertos picos de inestabilidad previos a la rotura. Por otro lado, la bobina 135 que no presenta eventos de rotura muestra estabilidad.

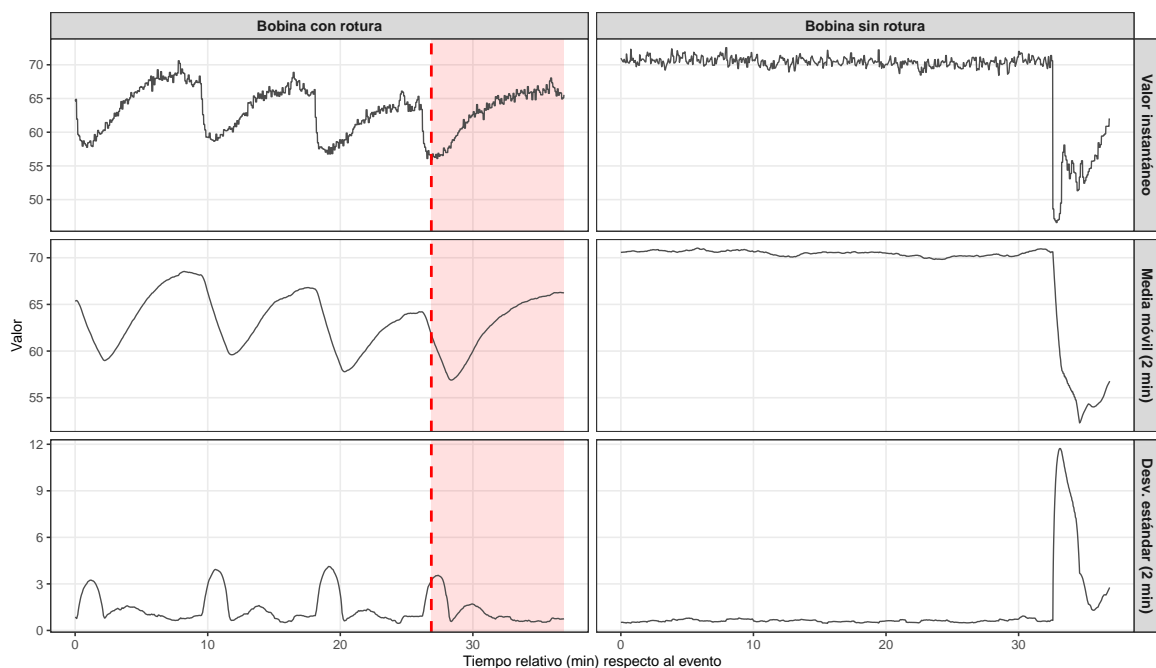


Figura 7 – Evolución temporal del torque de vacío en couch en las bobinas 232 y 135 respectivamente.

Las otras variables tomadas por su variabilidad en esta etapa fueron Con-

trol\_Vacio.PIC.103\_PV, Control\_Vacio.PIC\_338\_PV, Control\_Vacio.LIC\_TQ\_TORRE\_ENF\_PV y Control\_Vacio.VEL\_VACIO\_MESA\_PV. Para examinar lo que sucedía con estas variables se estudió la desviación estándar móvil. En el Cuadro 13 para PIC.103 se presenta una mayor estabilidad previa, a diferencia del PIC\_338 que presentó una pérdida de estabilidad, por otro lado la variable TORQ\_VAC\_COUCH presentó un cambio en el régimen dinámico del torque y la variable VEL\_VACIO\_MESA evidenció rigidez previo al evento de la rotura.

Cuadro 13 – Resumen por mediana global de la estabilidad dinámica del sistema de vacío

| Variable                  | No rotura | Rotura |
|---------------------------|-----------|--------|
| PIC.103_PV_lag1_sd        | 0.00      | 0.104  |
| PIC_338_PV_lag1_sd        | 0.00      | 0.04   |
| TORQ_VAC_COUCH_PV_lag1_sd | 1.10      | 0.83   |
| VEL_VACIO_MESA_PV_lag1_sd | 0.91      | 0.00   |

El Cuadro 14 muestra el comportamiento de las roturas bajo una bobina con presencia de rotura y una bobina de control, es decir sin presencia del evento tomando la mediana como medida robusta de la desviación estándar móvil de las variables PIC.103 y Vacio.TORQ\_VAC\_COUCH. Donde en la bobina 232 presentó una respuesta dinámica limitada del sistema previo al evento de rotura y a la vez una baja capacidad de amortiguación dinámica, es decir el sistema reacciona levemente frente a las roturas. Por otro lado, la bobina 105 evidenció una variabilidad controlada del sistema, donde responde moderadamente y a la vez el sistema responde moderadamente sin desestabilizar el proceso.

Cuadro 14 – Estabilidad dinámica del sistema de vacío en bobinas emparejadas

| Bobina | Estado     | Variable                                | Mediana sd |
|--------|------------|---|------------|
| 232    | Con rotura | Control_Vacio.PIC.103_PV_lag1_sd        | 0.00       |
| 232    | Con rotura | Control_Vacio.TORQ_VAC_COUCH_PV_lag1_sd | 0.97       |
| 105    | Sin rotura | Control_Vacio.PIC.103_PV_lag1_sd        | 0.13       |
| 105    | Sin rotura | Control_Vacio.TORQ_VAC_COUCH_PV_lag1_sd | 0.81       |

La Figura 8 presenta la comparación en relación a la variable Control\_Vapor.PIC\_405\_PV entre la bobina 232 que contiene el evento rotura y la bobina de control 135. En la bobina con rotura se observa un régimen estable previo al evento, con valores alrededor de 2.5. Sin embargo, a partir del instante de la rotura se produce un salto abrupto hasta valores cercanos a 2.7, seguido de una caída progresiva que parece estabilizarse alrededor de 2.42. En cambio, la bobina sin rotura presenta oscilaciones alrededor de 2.4 periódicamente, sin saltos abruptos. La media móvil no muestra señales tempranas previas al evento, ya que tiene un comportamiento similar al valor instantáneo

de forma suavizada. Por otro lado, el análisis de la desviación estándar móvil muestra un pico inmediatamente después de la rotura, reflejando un aumento significativo de la variabilidad asociada a la desestabilización del sistema. Por el contrario, la bobina sin rotura muestra oscilaciones constantes y moderadas a la variabilidad.

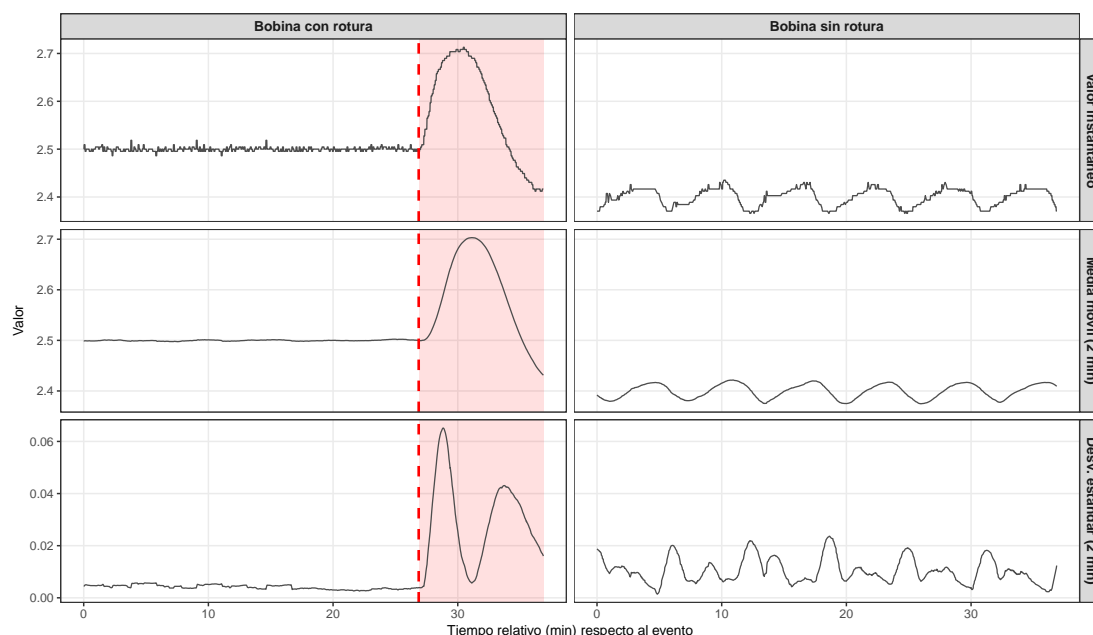


Figura 8 – Evolución temporal de la variable Control\_Vapor.PIC\_405\_PV en las bobinas 232 y 135 respectivamente.

Finalmente, estos resultados a nivel descriptivo indican que la rotura ocurre en un lapso de baja flexibilidad dinámica del sistema, no amortiguado de forma adecuada, manifestándose posteriormente como un pico de inestabilidad en la etapa control de vapor.

La Figura 9 muestra el comportamiento de una de las variables asociadas a la prensa Nipcoflex, donde la señal P56\_PV representa la presión aplicada por el sistema afectando directamente a la bobina de papel. En la bobina con rotura, el valor instantáneo presenta oscilaciones estables alrededor de los valores 42 y 45. Sin embargo, después del evento de rotura se observa una caída abrupta hasta aproximadamente 15, posteriormente manteniéndose estable. La media móvil presenta un comportamiento similar al valor instantáneo, pero suavizado, como consecuencia de esta caída abrupta, la desviación estándar móvil presenta un pico pronunciado posterior al evento de rotura.

Por otro lado, en la bobina sin rotura (bobina de control), el valor instante se mantiene cercano a cero, seguido de un incremento gradual a valores próximos a 15, donde no se producen caídas. Este comportamiento se refleja en la media móvil y en la

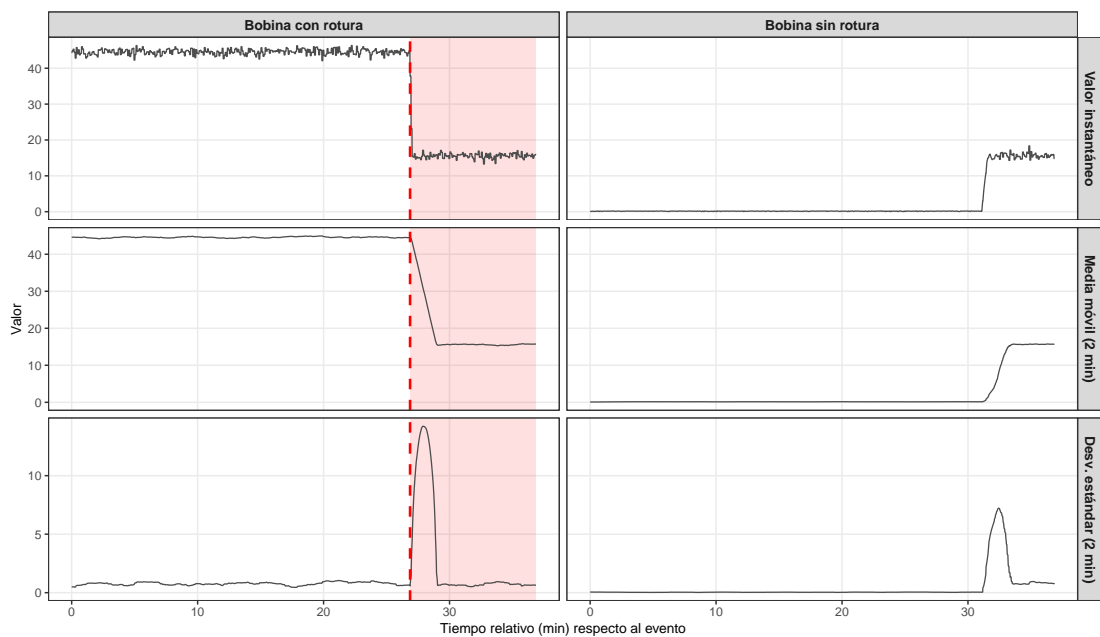


Figura 9 – Evolución temporal de la variable Nipcoflex\_Com.P56\_PV en las bobinas 232 y 135 respectivamente.

desviación estándar móvil donde se ve un pico moderado de variabilidad ante el cual el sistema tuvo control.

La Figura 11 muestra el comportamiento de la variable Aproximacion.SP\_CHO RRO\_TELA para dos bobinas, representando la etapa de formación. Esta variable controla la relación entre la velocidad del chorro de pasta y la velocidad de la tela formadora. En la bobina con evento de rotura se observa estabilidad alrededor del valor -36. Sin embargo, entre aproximadamente los minutos 25 y 27 se produce un salto abrupto hasta -38. Este comportamiento es similar en la media móvil de forma suavizada, dando soporte al cambio. Como consecuencia la desviación estándar móvil presenta un pico cercano a uno previo al evento de rotura, este comportamiento sugiere la existencia de un sistema rígido. Por otro lado, en la bobina de control se observa un valor contante en -34 tanto en el valor instantáneo como en la media móvil, dado lo anterior la desviación estándar móvil se mantiene en cero indicando estabilidad. Los resultados muestran que la rotura ocurre en un ambiente donde aparece un ajuste o perturbación introduciendo un pico antes del evento.

En la Figura 10 se presentan las series temporales de tres variables pertenecientes a la etapa de calidad, las cuales son HUMEDAD\_PROMEDIO, HUMEDAD\_DESVIACION, y PESO\_DESVIACION. Estas variables mostraron un comportamiento similar aproximadamente 25 minutos antes del evento de rotura en la bobina 232, donde previamente tenían un comportamiento constante. En estos minutos la varia-

ble HUMEDAD\_PROMEDIO presentó una subida de aproximadamente 0.05, mientras la HUMEDAD\_DESVIACION subió algunos puntos, y PESO\_DESVIACION bajó de manera escalonada. Por otro lado en la bobina de control, estas variables mostraron un comportamiento constante, es decir estable tanto en el valor instantáneo como en la media móvil. Como consecuencia la desviación estándar móvil tiene un valor de aproximadamente cero.

En la Figura 12 se presenta el comportamiento de la variable Presecado\_Couch.CORRIENTE\_PV correspondiente a la etapa del presecado. En la bobina con rotura se observa un comportamiento estable del valor instantáneo confirmado por la media móvil, incluso antes al evento de rotura. Asimismo, la desviación estándar móvil no presenta picos de variabilidad, lo que indica que el sistema se encontraba estable.

En la bobina de control el comportamiento también se muestra estable, aunque muestra un incremento de corriente desde valores cercanos a 350A hasta mucho más de 600A, valor aproximado de donde se mantenía constante la bobina con el evento de rotura, ya que estaba aproximadamente entre los valores de 550A y 620A. A pesar del incremento la bobina de control no presentó presencia del evento en esta variable.

Dado lo anterior, se hizo uso de ventanas preventivas para estudiar la variabilidad, desviaciones, fluctuaciones de las variables del proceso en la rotura, donde el tamaño del efecto fue cuantificado mediante Cliff's delta, una medida no paramétrica recomendada para comparar distribuciones asimétricas y con tamaños muestrales desbalanceados, características propias de los datos industriales analizados en este estudio (Cliff, 1993; Romano *et al.*, 2006).

Cliff's delta se define como:

$$\delta = P(X_1 > X_2) - P(X_1 < X_2)$$

donde  $X_1$  y  $X_2$  representan las observaciones correspondientes a ventanas previas al evento de rotura de hoja y a condiciones normales de operación, respectivamente. El estadístico  $\delta$  toma valores en el intervalo  $[-1, 1]$ , donde el signo indica la dirección del efecto y su magnitud absoluta refleja el grado de diferencia entre ambos grupos. Siguiendo los criterios propuestos en la literatura, valores de  $|\delta|$  entre 0.33 y 0.47 indican un efecto de magnitud mediana, mientras que valores inferiores corresponden a efectos pequeños o despreciables (Romano *et al.*, 2006).

Como se observa en el Cuadro 15, las variables asociadas al control de vacío, vapor y presecado presentan tamaños de efecto de magnitud pequeña a mediana, lo que indica diferencias sistemáticas entre las condiciones previas al evento de rotura de hoja y las condiciones normales de operación.

Cuadro 15 – Cliff’s Delta

| Valores de $ \delta $        | Nivel de efecto |
|------------------------------|-----------------|
| $ \delta  \geq 0.474$        | Grande          |
| $0.33 \leq  \delta  < 0.474$ | Mediano         |
| $0.147 \leq  \delta  < 0.33$ | Pequeño         |
| $ \delta  < 0.147$           | Despreciable    |

### 3.0.1 Tamaño de efecto mediante Cliff’s Delta

La Tabla 16 presenta los resultados de la prueba Mann Whitney y el estadístico de Cliff’s Delta. Con estos resultados se pudo identificar que dado sus p-value  $< 0.05$ , existen diferencias entre los eventos rotura y no rotura, y su efecto fue calculado mediante el estadístico de Cliff’s Delta, junto con sus respectivos valores de significancia estadística. Este indicador permite cuantificar la magnitud de la diferencia entre los estados con y sin rotura de bobina, independientemente del tamaño muestral.

Cuadro 16 – Resultados del tamaño de efecto medido mediante Cliff’s Delta

| Variable                                 | p-valor               | Cliff’s $\delta$ | $ \delta $ |
|--|-----------------------|------------------|------------|
| chg__Control_Vacio.I_VAC_COUCH1_PV_mean  | $5.44 \times 10^{-8}$ | 0.362            | 0.362      |
| chg__Control_Vacio.I_VAC_COUCH_PV_mean   | $7.24 \times 10^{-8}$ | 0.358            | 0.358      |
| chg__Control_Vapor.PIC_405_PV_mean       | $1.13 \times 10^{-7}$ | 0.353            | 0.353      |
| chg__Presecado_Couch.TEMPERATURA_PV_mean | $1.12 \times 10^{-6}$ | 0.324            | 0.324      |
| chg__Control_Vapor.TIC_305_PV_mean       | $1.86 \times 10^{-6}$ | 0.317            | 0.317      |
| chg__Nipcoflex_Com.T65_PV_mean           | $2.25 \times 10^{-6}$ | 0.315            | 0.315      |
| Control_Vapor.TIC_325_PV_sd              | $3.30 \times 10^{-6}$ | 0.309            | 0.309      |
| chg__Control_Vapor.PIC_324_PV_mean       | $1.81 \times 10^{-5}$ | 0.285            | 0.285      |
| Nipcoflex_Com.P50_PV_sd                  | $1.18 \times 10^{-4}$ | 0.256            | 0.256      |
| Control_Vapor.PDIC_406_PV_sd             | $1.21 \times 10^{-4}$ | 0.256            | 0.256      |
| chg__Control_Vapor.TIC_405_PV_mean       | $1.39 \times 10^{-4}$ | 0.253            | 0.253      |

## 3.1 Modelo de regresión logística

Basados en los resultados obtenidos a partir de la medida Cliff’s Delta, se considerarán las 11 variables con mayor efecto para el modelo de regresión logística, para explicar la ocurrencia del evento de rotura de bobina (*Corte\_Bobina*), utilizando como variables explicativas estadísticas resumidas (media, desviación estándar y cambios intra-minuto) de variables críticas del proceso. El modelo fue estimado mediante máxima verosimilitud usando un enlace logit e incorporando ponderaciones para manejar el desbalance de clases.

$$\log \left( \frac{P(\text{Corte\_Bobina} = 1)}{1 - P(\text{Corte\_Bobina} = 1)} \right) = \beta_0 + \sum_{i=1}^p \beta_i X_i \quad (3.1)$$

En el Cuadro 17 se presentan las estimaciones obtenidas para los coeficientes del modelo de regresión. El intercepto del modelo presenta un valor altamente negativo ( $\beta_0 = -47.54$ ,  $p < 0.001$ ), lo cual indica que, en ausencia de variaciones relevantes en las variables de proceso incluidas en el modelo, la probabilidad base de que una bobina presente un evento de rotura es prácticamente nula. En términos operativos, este resultado confirma que el proceso, bajo condiciones estables, no tiende a generar roturas de hoja, y que dichos eventos corresponden a situaciones asociadas a estados de inestabilidad del sistema.

Cuadro 17 – Resultados del modelo de regresión logística para el evento de rotura de bobina

| Variable                                 | Estimación | Error Std. | z      | p-valor    |
|--|------------|------------|--------|------------|
| Intercepto                               | -47.536    | 9.745      | -4.878 | < 0.001*** |
| chg__Control_Vapor.TIC_305_PV_mean       | 0.220      | 0.032      | 6.970  | < 0.001*** |
| Control_Vapor.TIC_305_PV_sd              | -0.189     | 0.041      | -4.551 | < 0.001*** |
| Control_Vapor.TIC_325_PV_sd              | 1.014      | 0.343      | 2.959  | 0.003**    |
| Control_Vapor.PIC_405_PV_sd              | 5.053      | 1.792      | 2.819  | 0.005**    |
| Control_Vapor.PDIC_400_PV_sd             | -7.066     | 4.717      | -1.498 | 0.134      |
| chg__Presecado_Couch.TEMPERATURA_PV_mean | 0.278      | 0.033      | 8.312  | < 0.001*** |
| chg__Control_Vacio.PIC_343_PV_mean       | 0.142      | 0.060      | 2.358  | 0.018*     |
| chg__Control_Vacio.PIC_347_PV_mean       | 0.928      | 0.222      | 4.175  | < 0.001*** |
| chg__Control_Vacio.PIC_352_PV_mean       | 0.011      | 0.049      | 0.218  | 0.827      |
| Control_Vacio.LIC_118_PV_mean            | 0.269      | 0.101      | 2.660  | 0.008**    |
| Nipcoflex_Com.F65_PV_sd                  | 5.205      | 1.503      | 3.458  | < 0.001*** |

Los coeficientes estimados representan el efecto marginal de cada variable sobre el logaritmo de las razones de momios (*log-odds*) de ocurrencia de una rotura. Para facilitar su interpretación, estos efectos se analizan mediante (*odds ratios*), calculadas como  $OR = e^\beta$ .

En el sistema de control de vapor, se observa que incrementos sostenidos en la temperatura promedio (*chg\_\_Control\_Vapor.TIC\_305\_PV\_mean*,  $OR \approx 1.25$ ,  $p < 0.001$ ) incrementan el riesgo de rotura, mientras que una menor variabilidad térmica en el mismo punto (*Control\_Vapor.TIC\_305\_PV\_sd*,  $OR \approx 0.83$ ,  $p < 0.001$ ) actúa como un factor protector. En contraste, aumentos en la variabilidad térmica en otros puntos del sistema de vapor, como *Control\_Vapor.TIC\_325\_PV\_sd* ( $OR \approx 2.76$ ,  $p = 0.003$ ), incrementan significativamente el riesgo, evidenciando que la estabilidad térmica es un elemento crítico para la integridad de la hoja.

Asimismo, la variabilidad en la presión de vapor (*Control\_Vapor.PIC\_405\_PV\_sd*,  $OR \approx 156$ ,  $p = 0.005$ ) muestra uno de los efectos más pronunciados del modelo, indicando que fluctuaciones en este subsistema multiplican de forma severa la probabilidad de rotura. Por otro lado, aunque la variable *Control\_Vapor.PDIC\_400\_PV\_sd* presenta un coeficiente negativo, su efecto no resulta estadísticamente significativo ( $p = 0.134$ ), por lo que no se considera prioritaria desde el punto de vista operativo.

En la etapa de presecado, los incrementos en la temperatura promedio del sistema Couch (*chg\_\_Presecado\_Couch.TEMPERATURA\_PV\_mean*,  $OR \approx 1.32$ ,  $p < 2 \times 10^{-16}$ ) constituyen uno de los predictores importantes del modelo, sugiriendo que cambios térmicos en esta etapa afectan directamente la resistencia inicial de la hoja y aumentan su susceptibilidad a la rotura.

Respecto al control de vacío, se observa que cambios sostenidos en las presiones de vacío (*chg\_\_Control\_Vacio.PIC\_343\_PV\_mean*,  $OR \approx 1.15$ ,  $p = 0.018$ ; *chg\_\_Control\_Vacio.PIC\_347\_PV\_mean*,  $OR \approx 2.53$ ,  $p < 0.001$ ) incrementan de manera significativa el riesgo, mientras que niveles elevados y sostenidos de vacío (*Control\_Vacio.LIC\_118\_PV\_mean*,  $OR \approx 1.31$ ,  $p = 0.008$ ) también se asocian a una mayor probabilidad de rotura. En contraste, la variable *chg\_\_Control\_Vacio.PIC\_352\_PV\_mean* no presenta un efecto significativo ( $p = 0.83$ ), por lo que no discrimina entre bobinas con y sin rotura.

Finalmente, en el sistema de compactación Nipcoflex, la variabilidad de la presión aplicada (*Nipcoflex\_Com.P65\_PV\_sd*,  $OR \approx 182$ ,  $p < 0.001$ ) emerge como uno de los factores de riesgo más críticos del modelo, indicando que una compactación irregular de la hoja incrementa de manera extrema la probabilidad de fallo.

- **Desviación nula:** 2853.0 (1103 grados de libertad)
- **Desviación residual:** 2389.1 (1092 grados de libertad)
- **Criterio de información de Akaike (AIC):** 2438.5
- **Número de iteraciones de Fisher Scoring:** 6

La deviance nula del modelo (2853.0 con 1103 grados de libertad) representa el ajuste obtenido al considerar únicamente el intercepto, es decir, sin incorporar información del proceso. Al incluir las variables explicativas, la deviance residual se reduce a 2389.1 con 1092 grados de libertad, lo que implica una reducción sustancial de 463.9 unidades. Esta disminución confirma que las variables seleccionadas aportan información relevante para caracterizar el riesgo de rotura a nivel bobina.

El valor del criterio de información de Akaike (AIC) obtenido es 2438.5. Este indicador refleja un balance adecuado entre calidad de ajuste y complejidad del modelo, y resulta útil para la comparación con modelos alternativos. En el contexto del presente estudio, un AIC reducido no se interpreta como mayor capacidad predictiva, sino como una descripción más parsimoniosa y coherente del riesgo operativo.

El modelo alcanzó convergencia tras seis iteraciones del algoritmo de Fisher Scoring, lo cual indica un ajuste estable y sin problemas numéricos, aun en presencia de desbalance entre las clases de bobinas con y sin rotura.

Dado el desbalance existente entre bobinas con eventos de rotura y bobinas sin fallos, se aplicó una ponderación de clases durante el entrenamiento del modelo, asignando pesos inversamente proporcionales a la frecuencia de cada clase. Esta estrategia permitió evitar que el modelo se sesgara hacia la clase mayoritaria y garantizó una estimación del indicador de riesgo, sin afectar la interpretación de los coeficientes ni de las razones de momios.

En conjunto, los resultados del modelo preventivo indican que el riesgo de rotura de hoja no está asociado principalmente a valores extremos de las variables de proceso, sino a la inestabilidad del sistema, reflejada en incrementos de la variabilidad y en cambios sostenidos en variables críticas de vapor, vacío, presecado y compactación. Esta evidencia respalda el enfoque preventivo adoptado, donde la identificación temprana de estados inestables del proceso permite intervenir antes de la ocurrencia del evento de rotura.

El modelo de regresión logística desarrollado estima un indicador de riesgo a nivel bobina, que representa la probabilidad de que una bobina opere bajo condiciones históricamente asociadas a la ocurrencia de al menos un evento de rotura de hoja. Dicho indicador se construye utilizando exclusivamente información previa al primer evento de rotura, por lo que su interpretación es de carácter preventivo y no predictivo.

### 3.1.1 Odds Ratios del modelo logístico

El Cuadro 18 presenta los Odds Ratios (OR) obtenidos a partir de la exponenciación de los coeficientes del modelo de regresión logística, junto con sus respectivos intervalos de confianza al 95%. Un OR mayor que 1 indica un incremento en la probabilidad de rotura de bobina, mientras que un OR menor que 1 indica un efecto protector sobre el evento.

Cuadro 18 – Odds Ratios e intervalos de confianza al 95 % para el evento de rotura de bobina

| Variable                                 | Odds Ratio             | IC 95 % (Inf.)         | IC 95 % (Sup.)         |
|--|------------------------|------------------------|------------------------|
| Intercepto                               | $2.27 \times 10^{-21}$ | $9.65 \times 10^{-30}$ | $3.33 \times 10^{-13}$ |
| chg__Control_Vapor.TIC_305_PV_mean       | 1.246                  | 1.172                  | 1.326                  |
| Control_Vapor.TIC_305_PV_sd              | 0.828                  | 0.763                  | 0.898                  |
| Control_Vapor.TIC_325_PV_sd              | 2.757                  | 1.550                  | 6.673                  |
| Control_Vapor.PIC_405_PV_sd              | 156.49                 | 5.78                   | 6237.04                |
| Control_Vapor.PDIC_400_PV_sd             | $8.54 \times 10^{-4}$  | $8.11 \times 10^{-8}$  | 8.98                   |
| chg__Presecado_Couch.TEMPERATURA_PV_mean | 1.320                  | 1.239                  | 1.413                  |
| chg__Control_Vacio.PIC_343_PV_mean       | 1.152                  | 1.026                  | 1.298                  |
| chg__Control_Vacio.PIC_347_PV_mean       | 2.528                  | 1.644                  | 3.928                  |
| chg__Control_Vacio.PIC_352_PV_mean       | 1.011                  | 0.921                  | 1.115                  |
| Control_Vacio.LIC_118_PV_mean            | 1.308                  | 1.074                  | 1.596                  |
| Nipcoflex_Com.P65_PV_sd                  | 182.17                 | 9.75                   | 3342.79                |

El intercepto del modelo presenta un OR extremadamente bajo ( $2.27 \times 10^{-21}$ ), lo cual indica que, en condiciones estables del proceso y en ausencia de variaciones relevantes en las variables incluidas, la probabilidad base de que una bobina presente un evento de rotura es prácticamente nula.

En el sistema de control de vapor, se observa que incrementos sostenidos en la temperatura promedio (*chg\_\_Control\_Vapor.TIC\_305\_PV\_mean*, OR = 1.25, IC<sub>95%</sub> [1.17, 1.33]) incrementan el riesgo de rotura en aproximadamente un 25 % por unidad. En contraste, una menor variabilidad térmica en el mismo punto (*Control\_Vapor.TIC\_305\_PV\_sd*, OR = 0.83, IC<sub>95%</sub> [0.76, 0.90]) actúa como un factor protector, reduciendo el riesgo en torno al 17 %. Estos resultados confirman que la estabilidad térmica resulta más relevante que el valor absoluto de la temperatura.

Asimismo, la variabilidad térmica en otros puntos del sistema de vapor, como *Control\_Vapor.TIC\_325\_PV\_sd* (OR = 2.76, IC<sub>95%</sub> [1.55, 6.67]), incrementa significativamente el riesgo, mientras que la variabilidad en la presión de vapor (*Control\_Vapor.PIC\_405\_PV\_sd*, OR  $\approx$  156, IC<sub>95%</sub> [5.78, 6237]) emerge como uno de los factores de riesgo más críticos del modelo. Aunque el intervalo de confianza en este último caso es amplio, la dirección del efecto es clara y consistente con una condición de fuerte inestabilidad del sistema.

Por otro lado, la variable *Control\_Vapor.PDIC\_400\_PV\_sd* presenta un OR menor que 1, pero con un intervalo de confianza que cruza la unidad, por lo que su efecto no se considera estadísticamente concluyente ni prioritario desde el punto de vista operativo.

En la etapa de presecado, los incrementos térmicos en el sistema Couch (*chg\_\_Presecado\_Couch.TEMPERATURA\_PV\_mean*, OR = 1.32, IC<sub>95%</sub> [1.24, 1.41]) muestran uno de los efectos más robustos del modelo, indicando que cambios térmicos

en esta etapa afectan directamente la resistencia inicial de la hoja y aumentan su susceptibilidad al rotura.

Respecto al sistema de control de vacío, se observa que cambios sostenidos en las presiones de vacío (*chg\_\_Control\_Vacio.PIC\_343\_PV\_mean*, OR = 1.15, IC<sub>95%</sub> [1.03, 1.30]; *chg\_\_Control\_Vacio.PIC\_347\_PV\_mean*, OR = 2.53, IC<sub>95%</sub> [1.64, 3.93]) incrementan de manera significativa el riesgo de rotura. Asimismo, niveles elevados y sostenidos de vacío (*Control\_Vacio.LIC\_118\_PV\_mean*, OR = 1.31, IC<sub>95%</sub> [1.07, 1.60]) se asocian a un mayor riesgo, mientras que la variable *chg\_\_Control\_Vacio.PIC\_352\_PV\_mean* no presenta un efecto discriminante relevante.

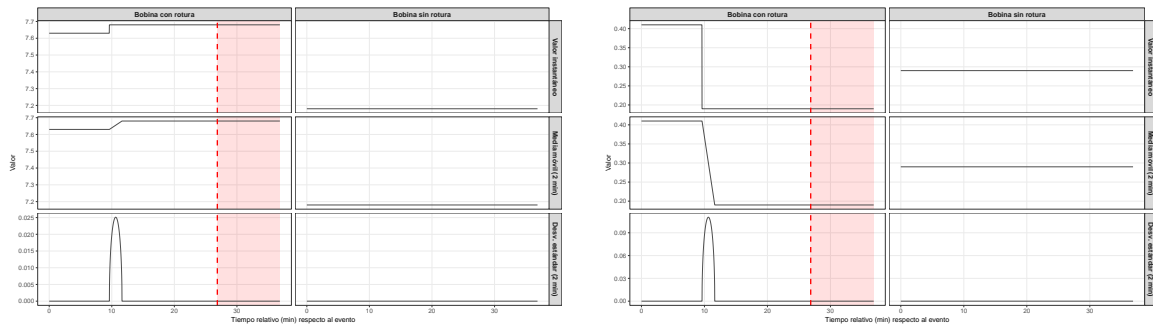
Finalmente, en el sistema de compactación Nipcoflex, la variabilidad de la presión aplicada (*Nipcoflex\_Com.P65\_PV\_sd*, OR  $\approx$  182, IC<sub>95%</sub> [9.75, 3343]) constituye uno de los principales factores de riesgo identificados, indicando que una compactación irregular de la hoja incrementa de forma severa la probabilidad de fallo.

La capacidad discriminativa del modelo preventivo se evaluó mediante el área bajo la curva ROC (AUC), obteniéndose un valor de

$$\text{AUC} = 0.7764.$$

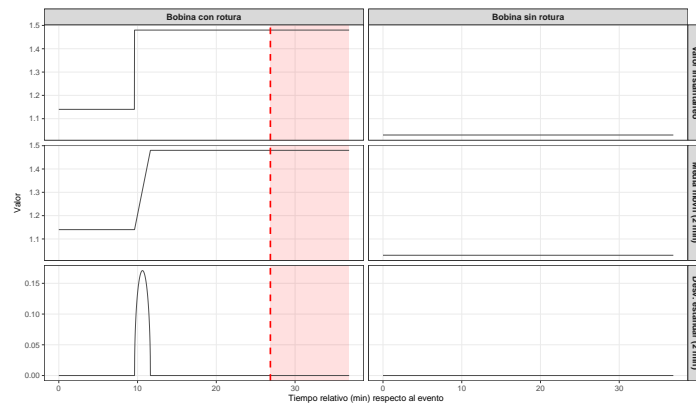
Este resultado indica que el modelo es capaz de ordenar correctamente bobinas con mayor riesgo por encima de bobinas con operación estable en aproximadamente el 77.6% de los casos. En el contexto del presente estudio, este valor se considera adecuado y coherente con un enfoque preventivo, dado que evalúa la capacidad global de discriminación del indicador de riesgo sin depender de un umbral específico de clasificación.

En síntesis, la evaluación mediante AUC confirma que el modelo posee una capacidad adecuada para discriminar entre distintos niveles de riesgo operativo a nivel de bobina. Los resultados obtenidos evidencian que el riesgo de rotura de hoja está asociado principalmente a la inestabilidad del proceso, reflejada en incrementos de la variabilidad y en cambios sostenidos en variables críticas de vapor, vacío, presecado y compactación. Esta evidencia respalda el enfoque preventivo adoptado y proporciona una base sólida para la identificación temprana de condiciones operativas que requieren intervención antes de la ocurrencia del evento de rotura.



(a) Evolución temporal de HUME-DAD\_PROMEDIO

(b) Evolución temporal de HUME-DAD\_DESVIACION



(c) Evolución temporal de PESO\_DESVIACION

Figura 10 – Evolución temporal de variables de calidad en las bobinas 232 y 135.

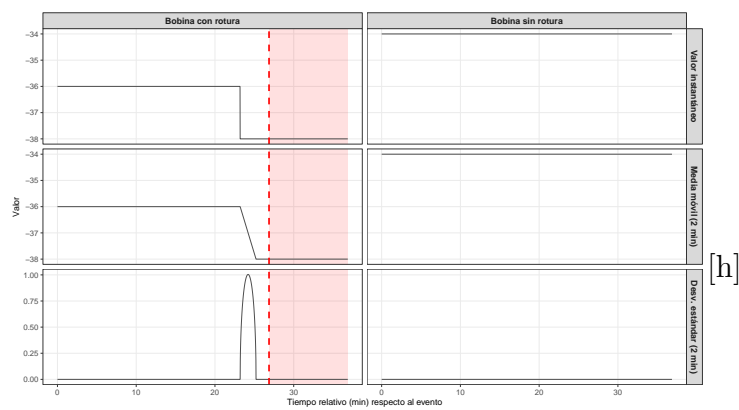


Figura 11 – Evolución temporal de la variable Aproximacion.SP\_CHORRO\_TELA en las bobinas 232 y 135 respectivamente.

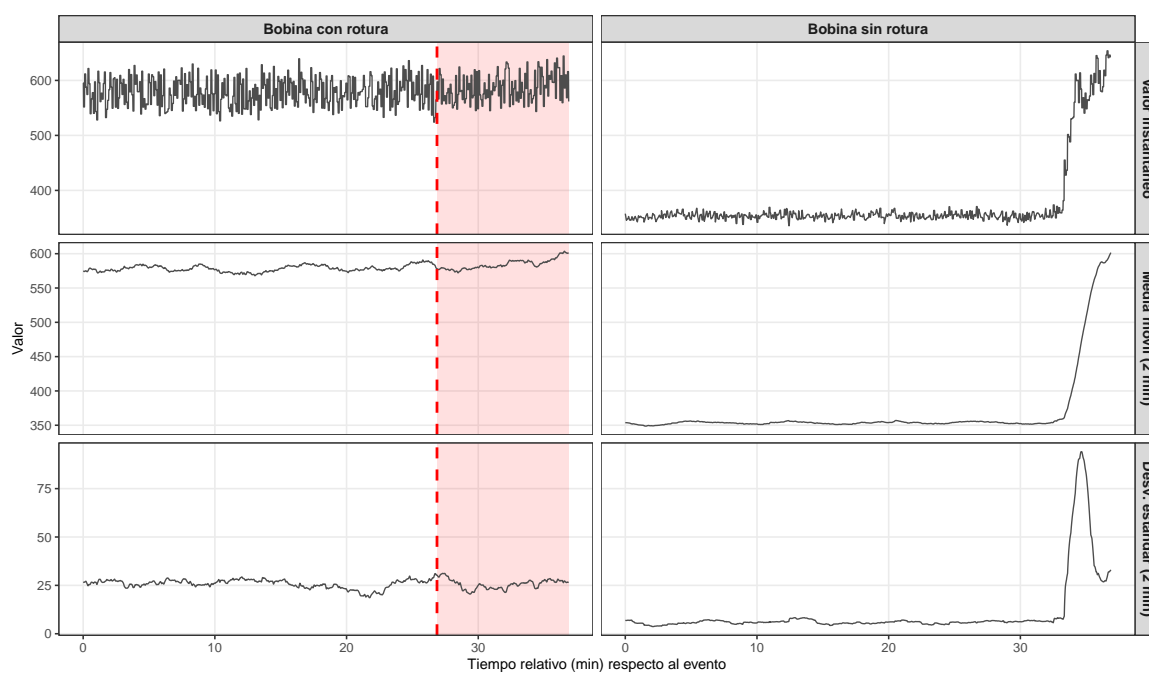


Figura 12 – Evolución temporal de la variable Presecado\_Couch.CORRIENTE\_PV en las bobinas 232 y 135 respectivamente.

## 4 Conclusiones y Recomendaciones

En el presente estudio se llevó a cabo la limpieza, integración y reestructuración de la base de datos proporcionada por una empresa papelería ecuatoriana, proveniente del sistema SCADA (Supervisory Control and Data Acquisition). Los datos originales, registrados a nivel de segundos, fueron consolidados a una resolución temporal de un minuto, con el objetivo de reducir el ruido inherente a los procesos industriales continuos y permitir una identificación fiel del comportamiento real de las variables del proceso. Asimismo, se realizó el tratamiento de datos faltantes bajo el supuesto de Missing Completely at Random (MCAR), junto con la recodificación de variables clave como Bobina\_ID y Corte\_Evento.

Se desarrolló un análisis exploratorio de las variables correspondientes a cada una de las etapas del proceso productivo mediante medidas descriptivas, análisis de frecuencias y evaluación de la estructura temporal de los datos, considerando la diferenciación por turnos, días y horas de operación. Este análisis incluyó el uso de diagramas de caja para variables asociadas tanto de forma reactiva como proactiva al evento de rotura, definido como una variable binaria. Adicionalmente, se construyeron ventanas temporales preventivas en torno al evento, y se aplicaron pruebas no paramétricas de Mann-Whitney U junto con la medida de tamaño del efecto Cliff's Delta, con el fin de evaluar la existencia y magnitud de diferencias entre los grupos definidos por la variable objetivo.

Las variables identificadas como relevantes fueron posteriormente incorporadas en un modelo de regresión logística, lo que permitió interpretar los resultados mediante razones de momios (Odds Ratios) y cuantificar el efecto de cada variable sobre el riesgo de ocurrencia del evento de rotura. Finalmente, se desarrolló un visualizador interactivo en Power BI, orientado a la toma de decisiones desde un enfoque proactivo, permitiendo el análisis del proceso por etapas y facilitando el monitoreo continuo de condiciones operativas críticas.

### 4.1 Conclusiones

- Se logró la limpieza, integración y estructuración de los datos históricos de producción, transformando registros de alta frecuencia (segundo a segundo) en una base de datos consolidada a nivel minuto, adecuada para el análisis estadístico y la visualización del proceso productivo.

- El análisis estadístico descriptivo y exploratorio permitió identificar patrones claros asociados a la ocurrencia de eventos de rotura, evidenciando que estos no se presentan de manera aleatoria, sino que están relacionados con variaciones específicas en variables críticas del proceso.
- Las variables pertenecientes a las etapas de vapor y vacío, encargadas de regular la remoción de humedad del papel, mostraron una asociación significativa con la ocurrencia de roturas, especialmente ante incrementos en temperatura y presión. Estos resultados fueron respaldados por el modelo de regresión logística, el cual permitió identificar tanto factores de riesgo como factores protectores frente al evento. Asimismo, se identificó la influencia de variables correspondientes a las etapas de presecado y Nipcoflex, reforzando la hipótesis de que la variabilidad del proceso juega un papel clave en la ocurrencia de fallos.
- La implementación del visualizador interactivo en Power BI facilitó la exploración dinámica de la información, permitiendo el monitoreo de la frecuencia y distribución temporal de las variables mediante gráficos de Control Estadístico de Procesos (CEP), constituyéndose como una herramienta de apoyo para la toma de decisiones informadas.
- El estudio permitió desarrollar un enfoque preventivo para el análisis de eventos de rotura en el proceso de fabricación del papel, integrando datos industriales de alta frecuencia con técnicas estadísticas y modelos explicativos. A diferencia de enfoques puramente predictivos, el objetivo se centró en la identificación de condiciones operativas que incrementan la probabilidad de que una bobina presente al menos un evento de fallo.

## 4.2 Recomendaciones

- Implementar como política de la empresa un monitoreo continuo y sistemático de las etapas de vapor, vacío y Nipcoflex, dado que presentan una mayor variabilidad y un impacto significativo sobre la estabilidad del proceso productivo.
- Se recomienda que futuras investigaciones incorporen técnicas estadísticas y de aprendizaje automático más avanzadas, tomando como base el enfoque descriptivo-analítico desarrollado en este estudio.
- Explorar el uso de métodos de detección de anomalías, tales como Isolation Forest y One-Class Support Vector Machine, con el fin de identificar comportamientos atípicos del proceso de manera temprana.

# Bibliografía

- Area, M. C. (2008). *Panorama de la industria de celulosa y papel en Iberoamérica 2008*. Red Iberoamericana de Docencia e Investigación en Celulosa y Papel-Riadicypp. Citado en la página 12.
- Bailey, J. E. & Wright, P. I. (2003). *Practical SCADA for Industry*. Newnes. Citado en la página 20.
- Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer, New York. Citado en la página 31.
- Box, G. E. P. & Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control*. Holden-Day, San Francisco, revised edition edition. Citado en la página 23.
- Box, G. E. P., Jenkins, G. M., Reinsel, G. C. & Ljung, G. M. (2015). *Time Series Analysis: Forecasting and Control*. Wiley. Citado en la página 21.
- CEPI (2021). Key statistics 2020: European pulp & paper industry. Technical report, Confederation of European Paper Industries, Brussels. Statistical booklet on performance of European pulp and paper industry. Citado en la página 13.
- Chiang, L. H., Russell, E. L. & Braatz, R. D. (2001). *Fault Detection and Diagnosis in Industrial Systems*. Springer. Citado en la página 21.
- Çiçekler, M. & Tutuş, A. (2024). Industry 4.0 in the paper industry: A review from past to present. *4th International Conference on Innovative Academic Studies*. Citado en la página 16.
- Cliff, N. (1993). Dominance statistics: Ordinal analyses to answer ordinal questions. *Psychological Bulletin*, **114**(3), 494–509. Citado en la página 49.
- Devi, T. K., Priyanka, E. & Sakthivel, P. (2023). Paper quality enhancement and model prediction using machine learning techniques. *Results in Engineering*, **17**, 100950. Citado 2 veces en las páginas 12 y 15.
- Diggle, P., Heagerty, P., Liang, K.-Y. & Zeger, S. (2002). *Analysis of Longitudinal Data*. Oxford University Press. Citado en la página 21.
- FAO (2021). Pulp and paper capacities — survey 2020–2025. Technical report, Food and Agriculture Organization of the United Nations, Rome. Survey on global pulp

- and paper production capacity and utilization of recovered paper. Citado en la página 13.
- Gullichsen, J. & Paulapuro, H. (2000). *Papermaking Science and Technology, Book 8: Papermaking Part 1, Stock Preparation and Wet End*. Fapet Oy. Citado 3 veces en las páginas 26, 29 y 30.
- Gómez, L. & Jiménez, A. (2019). Gestión de la calidad y sostenibilidad en la industria manufacturera latinoamericana. *Revista Ingeniería Industrial*, **40**(2), 87–96. Citado en la página 14.
- Harrell, F. E. (2015). *Regression Modeling Strategies*. Springer. Citado en la página 21.
- Hastie, T., Tibshirani, R. & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, second edition. Citado 3 veces en las páginas 23, 33 y 35.
- Kadlec, P. (2009). *Soft Sensors for Process Control*. Ph.D. thesis, University of Nottingham. Citado en la página 23.
- Kadlec, P., Gabrys, B. & Strandt, S. (2009). Data-driven soft sensors in the process industry. *Computers & Chemical Engineering*, **33**(4), 795–814. Citado en la página 21.
- Karlsson, M. (2006). *Paper and Board Grades*. Elsevier. Citado 2 veces en las páginas 26 y 27.
- Kuhn, M. & Johnson, K. (2013). *Applied Predictive Modeling*. Springer. Citado 4 veces en las páginas 15, 24, 30 y 31.
- Little, R. J. A. & Rubin, D. B. (2002). *Statistical Analysis with Missing Data*. Wiley, second edition. Citado en la página 24.
- Montgomery, D. C. (2019). *Introduction to Statistical Quality Control*. John Wiley & Sons, 8th edition. Citado 4 veces en las páginas 23, 24, 25 y 30.
- Othen, R., Padberg, J., Möbitz, C. & Gries, T. (2025). Digital transformation in the paper industry: Assessing maturity, challenges, and opportunities. *Sustainability*, **17**(2), 770. Citado en la página 16.
- Paulapuro, H., editor (2000). *Papermaking Science and Technology, Book 8: Papermaking Part 1 — Stock Preparation and Wet End*. Fapet Oy. Citado en la página 25.

- Paulapuro, H. (2008). *Papermaking Science and Technology, Book 9: Papermaking Part 2, Drying*. Fapet Oy. Citado 2 veces en las páginas 27 y 29.
- Pearl, J. (2009). *Causality: Models, Reasoning, and Inference*. Cambridge University Press, Cambridge, second edition. Citado en la página 32.
- Qin, S. J. (2003). Statistical process monitoring: basics and beyond. *Journal of Chemometrics*, **17**(8–9), 480–502. Citado en la página 23.
- Reason, J. (2000). Human error: models and management. *BMJ*, **320**(7237), 768–770. Citado en la página 15.
- Romano, J., Kromrey, J. D., Coraggio, J. & Skowronek, J. (2006). Appropriate statistics for ordinal level data: Should we really be using t-test and cohen's d for evaluating group differences? In *Proceedings of the Florida Association of Institutional Research*. Citado 2 veces en las páginas 33 y 49.
- Seborg, D. E., Edgar, T. F., Mellichamp, D. A. & Doyle, F. J. (2010). *Process Dynamics and Control*. John Wiley & Sons, third edition. Citado 5 veces en las páginas 20, 21, 24, 25 y 27.
- Shmueli, G. (2010). To explain or to predict? *Statistical Science*, **25**(3), 289–310. Citado 3 veces en las páginas 31, 34 y 35.
- Smook, G. A. (2016). *Handbook for Pulp and Paper Technologists*. TAPPI Press, fourth edition. Citado 4 veces en las páginas 14, 25, 26 y 28.
- Wasserstein, R. L., Schirm, A. L. & Lazar, N. A. (2019). Moving to a world beyond “ $p < 0.05$ ”. *The American Statistician*, **73**(sup1), 1–19. Citado en la página 33.
- Çiçekler, M. & Tutus, A. (2023). Overview of quality control in the paper industry. volume 1, pages 100–110. Citado en la página 15.