

Predicción del corte de la hoja de celulosa mediante el uso de Machine Learning

Fredy Humberto Troncoso Espinosa

<https://orcid.org/0000-0002-9972-3123>

ftroncos@ubiobio.cl

Departamento de Ingeniería Industrial, Facultad de
Ingeniería, Universidad del Bío-Bío
Concepción, Chile

Yamil Gerard Avello Betancur

<https://orcid.org/0000-0002-4916-5784>

yavello@gmail.com

Superintendencia de Ingeniería de Celulosa Arauco y
Constitución S.A.,
Planta de Celulosa Nueva Aldea
Concepción, Chile

Luis Andrés Martínez Flores

<https://orcid.org/0000-0002-4199-465X>

lamartinez@ubiobio.cl

Departamento de Ingeniería Industrial, Facultad de
Ingeniería, Universidad del Bío-Bío
Concepción, Chile

Recibido (12/04/21) Aceptado (09/06/21)

Resumen: La celulosa es la principal materia prima para la producción de papel. Empresas que la producen presentan en su línea de producción el corte de la hoja de celulosa. Esta falla es esporádica y de alto impacto económico dado a que paraliza por varias horas la línea de producción, incurriéndose en horas improductivas y un gran despliegue de recursos humanos y financieros. En esta investigación se propone uso de Minería de Datos para definir un algoritmo de machine learning que permita predecir el corte de la hoja de celulosa en una línea de producción de una planta de celulosa en Chile. Los resultados muestran que mediante la aplicación de esta técnica es posible predecir el corte de la hoja de celulosa con la suficiente antelación como para tomar acciones correctivas que permitan evitar el corte y así minimizar el impacto económico asociado a la falla.

Palabras Clave: Minería de Datos, Machine Learning, Celulosa, Productividad.

Prediction of cellulose sheet cutting using Machine Learning

Abstract: Cellulose is the main raw material for the production of paper. Companies that produce it present in their production line the cutting of the cellulose sheet. This failure is sporadic and has a high economic impact since it paralyzes the production line for several hours, incurring unproductive hours and a large deployment of human and financial resources. In this research, the use of Data Mining is proposed to define a machine learning algorithm that allows predicting the cutting of the cellulose sheet in a production line of a cellulose plant in Chile. The results show that by applying this technique it is possible to predict the cutting of the cellulose sheet well in advance to take corrective actions to avoid cutting and thus minimize the economic impact associated with the failure.

Keywords: Data Mining, Machine Learning, Cellulose, Productivity.



I. INTRODUCCIÓN

El proceso de producción de celulosa está sujeto a varios requisitos que son determinantes para tener una producción de calidad óptima. En tal contexto, existe una falla sin solución y de alto costo económico para las empresas relacionadas a esta industria: el corte de hoja de celulosa. Esta falla afecta considerablemente el proceso productivo en términos de tiempo, dinero, materiales, entre otros. A raíz de lo anterior, es que se hace latente la necesidad contar con un sistema de predicción de fallas para este proceso [1].

La predicción de fallas es una temática que ha ido cobrando mayor relevancia para las empresas durante las últimas décadas. Si la predicción es confiable y cumple con los objetivos planteados por parte de los interesados, es posible emplear medidas correctivas para evitar potenciales problemas, haciendo del proceso de producción un sistema más tolerante a las fallas. Entre las principales consecuencias de estas fallas se encuentran: Disminución de los indicadores claves de rendimiento, pérdidas de producción y económicas, e incumplimiento en el plan operativo de producción de las empresas [2].

Para el análisis de fallas, las empresas productivas almacenan grandes volúmenes de datos obtenidos desde el proceso productivo, junto a otros datos característicos propios de los productos. Estos datos son útiles para la gestión de sus operaciones e inspección de sus productos. En este contexto, las herramientas tradicionales de análisis de datos han sido utilizadas con cierto éxito para predecir fallas en este tipo de sistemas. De-

bido al crecimiento masivo de los datos y complejidad de los procesos, los analistas han requerido técnicas más sofisticadas para enfrentar la predicción de fallas [3] [4]. Frente a este escenario de crecimiento de datos y complejidad de los procesos, la minería de datos se presenta como una alternativa útil para el apoyo a la toma de decisiones y una herramienta crucial en la predicción de fallas de sistemas productivos [5]. En el caso específico de la producción de celulosa, el proceso es complejo e involucra un volumen considerable de datos que representan muchas variables de entrada y salida. Además, se presenta cierto grado de dificultad para rescatar los datos del proceso, para que estos puedan ser aplicados a modelos predictivos y de optimización.

Esta investigación utilizó minería de datos para entrenar un conjunto de algoritmos de machine learning y se selecciona el de mejor desempeño para la predicción de fallas para el corte de hoja de celulosa en la Planta de Celulosa Nueva Aldea (PCNA) perteneciente al Holding Arauco, ubicada en la Región del Biobío, Chile. Específicamente, se analizó la parte del proceso asociada a la formación y secado de hoja de celulosa de la Línea 2 en el área de máquinas (ver figura 1). El periodo de estudio fue de 17 meses, desde enero de 2017 a mayo de 2018. Según lo observado durante este periodo, las constantes fallas disminuyeron los niveles de productividad, donde las pérdidas promediaron hasta en un 25% en los meses más críticos. Las pérdidas totales del periodo equivalen a 16.944 ADt (Air dry ton), que corresponden a 20,4 MM US\$.

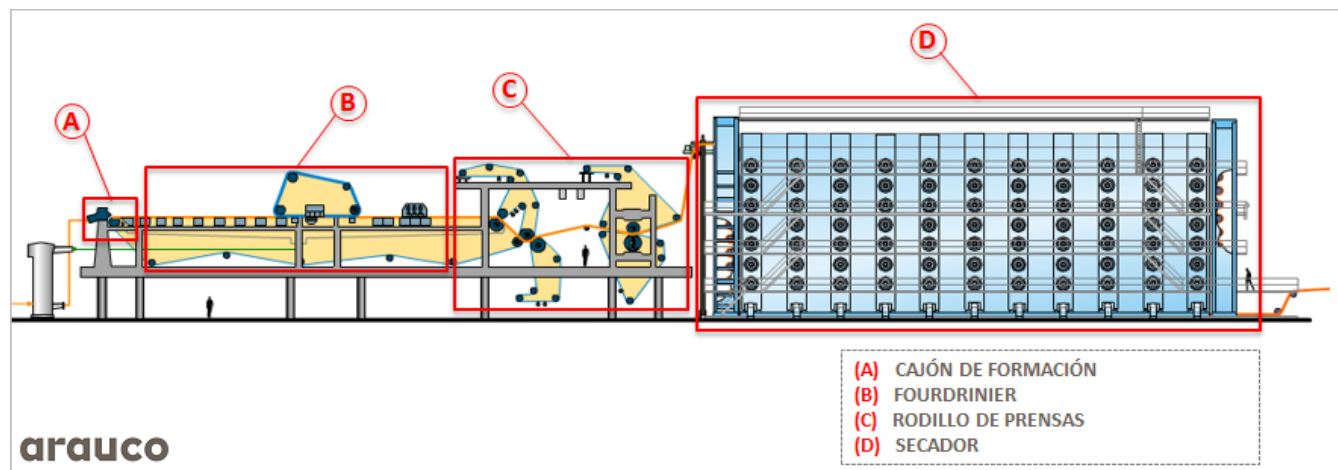


Fig. 1. Área de máquinas en PCNA, proceso de formación y secado de hoja de celulosa en línea final del proceso.

II.MACHINE LEARNING EN LA PREDICCIÓN DE FALLAS

La minería de datos es la parte del proceso Knowledge Discovery in Databases (KDD) donde se aplican los algoritmos necesarios para la identificación de patrones válidos, comprensibles, novedosos y potencialmente útiles en los datos [6]. Los algoritmos utilizados son conocidos como machine learning. La minería de datos ha sido aplicada con éxito a datos derivados de procesos productivos en la industria manufacturera para la prevención de fallas en maquinarias o partes del proceso productivo.

En [7] se muestra el desarrollo de un enfoque híbrido del algoritmo de machine learning red Neuronal artificial y lógica difusa para predecir el ancho de ranura en el proceso de corte por rayo láser en láminas de acero delgadas. Los resultados permitieron minimizar potenciales fallas haciendo del sistema un proceso más tolerante a los errores. Por otro lado, en [8] se desarrolló un método para la preparación y transformación de datos de fallas de sistemas de control para la industria automotriz. Luego se aplicaron algoritmos de clústers para encontrar relaciones en los datos de fallas emergentes del sistema, detectando que el agrupamiento jerárquico produce agrupaciones de mejor calidad en la predicción de fallas. En [9] se emplearon técnicas de minería de datos para predecir y clasificar los modos de fallas de dos placas compuestas tipo sándwich fijadas/atornilladas. Específicamente, se utilizaron máquina de vectores de soporte, métodos bayesianos, k-vecinos más cercanos, regresión logística, entre otros, cuya exactitud en la predicción de fallas superó el 95% en algunos casos. Por otro lado, en [10] se aplicaron enfoques de clasificación y agrupamiento para el reconocimiento de patrones y la predicción de fallas en palas mineras. Lo anterior, mediante el algoritmo máquina de vectores de soporte (SVM) que demostró un nivel de asertividad de más del 75%. Esto proporcionó información valiosa para la toma de decisiones. En [11] se investigó la predicción de fallas en transformadores de potencia, el cual es un equipo crítico para la correcta distribución de electricidad. Para lo anterior se emplearon varias técnicas de minería de datos para una mejor predicción y comprensión de conjuntos de datos. Se realizaron pruebas con árboles de decisión, máquinas de vectores de soporte (SVM) y redes neuronales. Se obtuvo un desempeño de un 77% en SVM y una tasa de falsos positivos de 35%. Se señaló que los resultados podrían haber sido mejores si la recopilación de datos hubiese sido la adecuada. En [12] se aplicaron enfoques de machine learning para datos recopilados de una mina de carbón de Turquía, con el propósito de predecir fallas en dragalinas. Los re-

sultados permitieron reducir costos de mantenimiento, una mayor vida útil de las máquinas excavadoras y un aumento en la producción.

Por otra parte, en [13] se investigó la predicción de fallas en el proceso de mantenimiento de aeronaves. Para lo anterior, se empleó un modelo de agrupamiento de red de fallas bayesiano en tiempo real. Este modelo demostró una alta tasa de precisión y la investigación contribuyó considerablemente al proceso de mantenimiento, protección y conservación de los sistemas aeronáuticos. En [14] se desarrolló un modelo de predicción de fallas en los sistemas de nube y de computación de alto rendimiento. Para esto se aplicaron algoritmos de machine learning tales como máquina de vectores de soporte (SVM), k-vecinos más cercanos (KNN) y árboles de clasificación. Los resultados experimentales mostraron que el modelo SVM presentó una precisión en la predicción de fallas de un 90%, pronosticando eficazmente todas las posibles fallas futuras del sistema. En [15] se desarrolló un modelo de predicción de fallas en la profundidad del piso en la industria de la minería. Los experimentos mostraron una alta precisión del modelo, cuyas predicciones de fallas permitieron controlar factores de seguridad, profundidad de excavación, intervención humana, entre otros, asegurando el máximo beneficio económico de las faenas de extracción de carbón. Por otro lado, en [16] mediante la aplicación de un algoritmo de machine learning se mejoró la calidad de la soldadura en tubos. Los algoritmos permitieron predecir los tubos con soldaduras débiles durante un proceso de producción en tiempo real, con un nivel de asertividad significativa.

Dada esta revisión bibliográfica se determinaron los algoritmos generales de machine learning que se utilizaron en esta investigación. Estos algoritmos se describen a continuación:

Clasificador Bayesiano: Este es un clasificador estadístico que estima la probabilidad de pertenencia de un individuo a una clase o no, dado un conjunto de variables que lo caracteriza. Está basado en el teorema de Bayes y muestra un alto accuracy y rapidez cuando es aplicado a grandes bases de datos [9] [13].

K-Nearest Neighbor: Es un clasificador basado en una técnica simple pero eficiente de aprendizaje, más conocido como el método supervisado de los k vecinos más cercanos. Consiste básicamente en la clasificación de valores buscando los puntos de datos (vecindades) más similares [12] [14].

Árbol de Decisión: un árbol de decisión es un diagrama de flujo con una estructura de árbol en forma jerárquica [11] [17] donde:

•Cada nodo denota un atributo sobre el que se realiza

una prueba.

- Cada rama derivada de un nodo representa las categorías del atributo como resultado de la prueba.

- Cada hoja representa una clase que se asigna a un registro.

Support Vector Machine (SVM): Es un modelo de clasificación y regresión de la teoría de aprendizaje estadístico. La metodología se basa en minimizar el error cuadrático de la clasificación, construyendo un hiperplano que separa los datos de la forma más precisa posible [10] [16].

Rede Neuronal: Este algoritmo imita el cerebro humano en la transformación de entradas en una o más salidas como un conjunto de neuronas [7] [18]. Una red neuronal está compuesta por una capa de entrada y de salida que están conectadas a través de un conjunto de nodos interconectados, conocido como capa oculta. La capa oculta procesa la información utilizando un conjunto de ponderaciones, asignadas a cada conexión. El proceso de aprendizaje de la red neuronal consiste en asignar estas ponderaciones a través del conjunto de entrenamiento [19].

Para el entrenamiento de estos algoritmos de machine learning se aplicó la técnica de entrenamiento y prueba llamada validación cruzada. Esta técnica consiste en separar el conjunto de datos en k subconjuntos de la misma cantidad de registros, utilizando $k-1$ para el entrenamiento y uno para la prueba. Este proceso se repite k veces, de manera que cada subconjunto pueda ser utilizado como prueba [15] [20]. El resultado de esta técnica es una medida de desempeño predictivo promedio y su desviación estándar.

El resultado de este proceso de entrenamiento y prueba fue representado mediante una matriz llamada Matriz de Confusión [21]. La Matriz de Confusión clasifica la predicción de un modelo en cuatro categorías: verdaderos positivos (VP) que son los elementos de la clase positiva (Corte de hoja) correctamente predichos por el modelo, los falso negativo FN que representa los elementos de la clase positiva incorrectamente predichos por el modelo, los verdaderos negativos VN que representa los elementos de la clase negativa (No corte de la hoja) correctamente predichos por el modelo y la tasa falso positivo FP que representa los elementos de la clase negativa incorrectamente predichos. Mediante estos cuatro valores se definen las siguientes medidas de desempeño predictivo:

Accuracy: mide el desempeño general del modelo y representa la proporción total de predicciones que fueron correctamente clasificadas y se obtiene la suma de VP y VN dividido por el total de datos en la matriz.

Recall: representa la tasa de elementos pertenecien-

te a la clase positiva que fueron clasificadas correctamente y se obtiene al dividir VP entre la suma de VP y FN.

Precision: representa la tasa de elementos de la clase positiva entre el total de elementos predichos como clase 1. Se obtiene al dividir VP entre la suma de VP y FP.

A continuación, se muestra la metodología utilizada para búsqueda de patrones de corte de la hoja de celulosa mediante minería de datos y su posterior aplicación a la predicción y prevención del corte.

III. METODOLOGÍA

Como se mencionó anteriormente, la minería de datos es la parte del proceso Knowledge Discovery in Databases (KDD) [22]. Este proceso es el fundamento metodológico mediante el cual se desarrolla esta investigación. Esta metodología o proceso explora de forma iterativa grandes volúmenes de datos y está compuesta por cinco etapas:

- Selección de datos, donde se determinan las fuentes de datos y el tipo de información con la cual se trabajará [23] [24].

- Pre procesamiento y transformación de los datos, con el fin de tener información más confiable y que aporte mayor valor a la predicción. En esta etapa se preparan y limpian los datos extraídos para aplicar de manera correcta las siguientes etapas. Se utiliza el análisis de datos faltantes, el análisis de datos inconsistentes, y el análisis de datos fuera de rango [25]. También, si es necesario, se modifica la naturaleza de los datos y se generan nuevas variables para obtener patrones de mejor calidad [8] [26].

- Selección de variables, donde se identifican las variables redundantes y las de mayor poder predictivo sobre la variable dependiente o variable a predecir. Esta selección permite la obtención de modelos más sencillos y explicables. También permite obtener patrones más robustos y de mejor calidad [8] [26].

- Minería de datos, etapa en la cual se aplican los algoritmos de machine learning que permiten extraer patrones relevantes desde los datos [19] [27].

- Evaluación de los algoritmos e interpretación de los resultados, donde se evalúan los algoritmos respecto a la búsqueda de patrones. Para esto se realiza una comparación de sus desempeños. Luego se interpretan los patrones obtenidos y se evalúa el impacto de la implementación del algoritmo seleccionado [28].

Tabla 1. Síntesis extracción de datos.

N° Total de variables	500
N° de variables consideradas	53
Periodo de estudio	17 meses
Extracción de datos	Extracción de datos cada 1 minuto
Registros Línea 2	77.800
Registros asociados a 47 Cortes	5.640

B. Identificación de variables con mayor poder predictivo

En esta etapa, a cada una de las variables seleccionadas se le midió la capacidad predictiva respecto a la variable dependiente. Esta variable dependiente es conocida como "label" y se asocia a cada registro. Esta variable toma la categoría "si sucedió" o la categoría

"no sucedió" el evento de corte de la hoja de celulosa. Esta capacidad predictiva fue medida mediante la Ganancia de Información. Esta técnica mide la cantidad de información contenida en una variable y que explica el label. La tabla 2 muestra los 20 atributos con mayor poder predictivo según esta técnica.

Tabla 2. Los 20 atributos con mejor capacidad predictiva según su Ganancia de Información.

ID	Variable	Pond.	Nombre Variable	Unidad de Medida	Proceso
1	Ce05_ip21_572pi2245	0,045	Combi Press Top Felt Suction Box	[kPa]	Rodillos Prensa
2	Ce05_l21_622ai581.0.8	0,044	Freeness Cajon Entrada L2	CSF [mm filtrado de "agua"]	Var. Laboratorio
3	Ce05_ip21_572aic2122	0,040	pH Pozo Fan Maquina 2	[pH]	Cajón
4	Ce05_ip21_572pi2247	0,034	Combi Press Top Felt Suction Box	[kPa]	Rodillos Prensa
5	Ce05_ip21_572zx2881e	0,033	Velocidad Foudrinier L2	[m/min]	Cajón
6	Ce05_ip21_572pi2204	0,033	Vacufoil Unit 2 Pressure	[kPa]	Fourdrinier
7	Ce05_ip21_572pi2214	0,033	Vacufoil Unit 6 Pressure	[kPa]	Fourdrinier
8	Ce05_ip21_572fi2292	0,033	Combi Press TF Water Separator 5	[L/min]	Rodillos Prensa
9	Ce05_ip21_572fi2293	0,032	Combi Press TF Water Separator 5	[L/min]	Rodillos Prensa
10	Ce05_ip21_572pic2904	0,030	Steam Pressure Controller	[kPa]	Secador
11	Ce05_ip21_572fi2901	0,029	Consumo vapor secador L2	[kg/s]	Secador
12	Ce05_ip21_572pi2233	0,027	Flat Suction Box 2 Pressure	[kPa]	Fourdrinier
13	Ce05_ip21_572qic2127	0,027	Razon Chorro Tela	[Veloc. Chorro Pulpa / Veloc. Tela]	Cajón
14	Ce05_ip21_572fi2297	0,025	Combi Press TF Water Separator 57	[L/min]	Rodillos Prensa
15	Ce05_ip21_572pi2202	0,025	Vacufoil Unit 1 Pressure	[kPa]	Fourdrinier
16	Ce05_ip21_572pi2207	0,023	Vacufoil Unit 3 Pressure	[kPa]	Fourdrinier
17	Ce05_ip21_572fi2295	0,023	Pickup 2nd Chamber Water Separat	[L/min]	Rodillos Prensa
18	Ce05_ip21_572pi2206	0,022	Vacufoil Unit Pressure	[kPa]	Fourdrinier
19	Ce05_ip21_572TI2905	0,022	Dryer Temperature	[°C]	Secador
20	Ce05_ip21_572pi2212	0,022	Vacufoil Unit 5 Pressure	[kPa]	Fourdrinier

Esta definición del poder predictivo permitió asignar criterios para el entrenamiento y prueba de los algoritmos de machine learning en la etapa de minería de datos.

Al comparar el poder predictivo de cada variable con el juicio de los expertos del proceso productivo, se validó su importancia. Un ejemplo de esto es la relación presión y flujo que se aprecia entre las variables Combi Press Top Felt Suction Box, Combi Press TF Water Separator 5 y Combi Press Top Felt Suction Box, Pickup 2nd Chamber Water Separat, las cuales coinciden en un punto del proceso inyectando presión y flujo a la forma-

ción de la hoja.

C. Minería de datos

En esta etapa se trabajó con el software RapidMiner [29] para la búsqueda del algoritmo de machine learning que identificara de mejor manera el patrón que permite predecir el corte de hoja de celulosa. Los nombres de los algoritmos específicos de RapidMiner utilizados, sus características principales y los parámetros ajustados luego de un proceso iterativo, se muestran en la tabla 3.

Tabla 3. Descripción de algoritmos de RapidMiner y ajuste de parámetros.

Nombre	Características principales	Parámetros
Support Vector Machine	Este algoritmo utiliza la implementación de Java de la máquina de vectores de soporte mySVM de Stefan Rueping	Kernel Type: Dot; Kernel Cache= 200; C= 0; Convergence epsilon = 0.001; Max iterations = 1000000
AutoMLP (Auto Multi-Layer Perceptron)	Es un algoritmo de red neuronal que ajusta la tasa de aprendizaje y el tamaño de las redes neuronales durante el entrenamiento. Combina ideas de algoritmos genéticos y optimización estocástica.	Training Cycles= 500 Numbers of generations 10 Numbers of esemble mlps = 4
Naïve Bayes	El supuesto fundamental del algoritmo es el de independencia entre los atributos, es decir que dado el valor del label, el valor de un atributo es independiente del valor de cualquier otro. Este supuesto simplifica enormemente los cálculos necesarios para construir el modelo de probabilidad.	Laplace correction (Evita que la probabilidad condicional se establezca en cero cuando dentro de los datos de entrenamiento un valor de atributo dado nunca ocurre en el contexto de una clase dada).
k-NN	Utiliza diferentes tipos de métricas para calcular la distancia entre el registro desconocido y los registros de entrenamiento.	K=5; Measure type= Numerical Measures; Numerical measures = Euclidean Distance; Weighted vote
Decision Tree	Este operador puede procesar bases de datos con atributos tanto nominales como numéricos.	Criterion = Gain Ratio; Maximal depth = 10; Confidence= 0.1; Minimal Gain = 0.01; Minimal leaf size = 2

Para el entrenamiento, ajuste y prueba de los algoritmos se utilizó el proceso que se muestra en la figura 2. Se observa que la base de datos pre procesada pasa por una etapa de normalización para ajustar los datos a una misma escala, luego se seleccionan las variables rele-

vantes para pasar a la etapa de entrenamiento y prueba mediante validación cruzada con k=10 folders. El detalle de la etapa de validación cruzada la muestra la flecha en color azul.

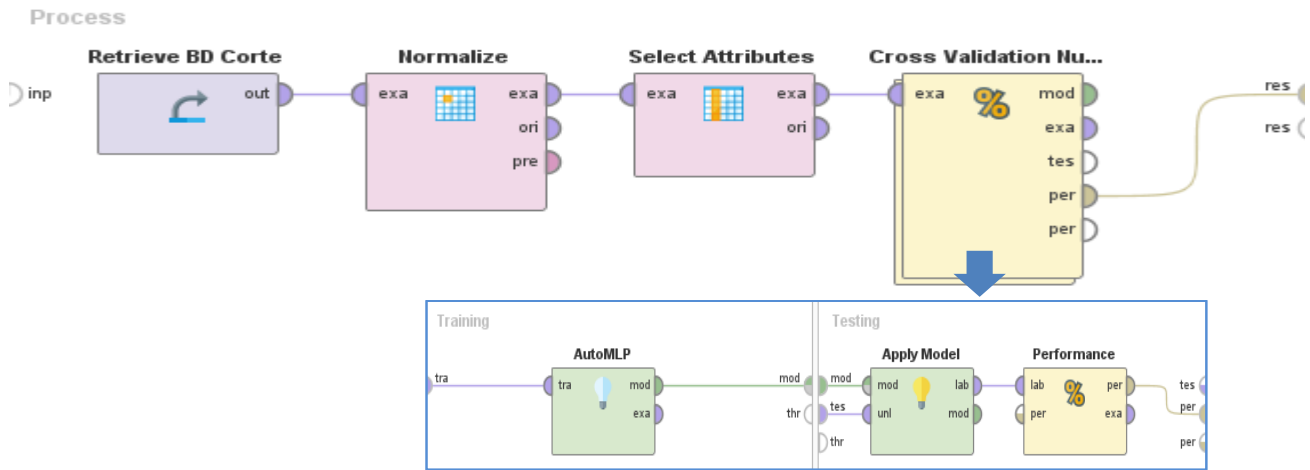


Fig. 2. Proceso de entrenamiento y prueba de algoritmos en RapidMiner.

D.Evaluación de los modelos e interpretación económica de los resultados

La tabla 4 muestra el desempeño general promedio y la desviación estándar de los algoritmos de machine

learning entrenados con las 53 variables consideradas. El de mayor accuracy fue AutoMLP, por lo que fue el elegido para predecir el corte de la hoja de celulosa.

Tabla 4. Desempeño predictivo de los algoritmos de RapidMiner para 53 atributos.

Algoritmo	Accuracy
Support Vector Machine	69,81% [+/-1,70%]
AutoMLP	99,87% [+/-0,15%]
Naïve Bayes	57,93% [+/- 2,40%]
k-NN	95,90% [+/- 1,09%]
Decision Tree	95,77% [+/-0,66%]

La tabla 5 muestra el desempeño del algoritmo AutoMLP en diferentes pruebas, las que consideran distintos subconjuntos de atributos, priorizados según su

capacidad predictiva. Este desempeño fue medido mediante accuracy, precision y recall.

Tabla 5. Accuracy, precision y recall para Red Neuronal y distintos subconjuntos de variables.

Prueba	Nº variables	Accuracy	Precision (P)	Recall (R)
1	53	99,87% [+/-0,15%]	99,91%	99,82%
2	40	99,87% [+/-0,18%]	99,91%	99,82%
3	30	99,87% [+/-0,18%]	99,91%	99,82%
4	20	98,83% [+/-0,91%]	98,67%	98,98%
5	10	94,22% [+/-1,34%]	94,59%	93,79%
6	5	72,96% [+/-3,52%]	91,24%	50,80%

Dado a que el corte de la hoja de celulosa es una falla crítica, por su gran impacto económico, se decide priorizar el desempeño predictivo del algoritmo por sobre un ajuste más sencillo con una menor cantidad de variables. Por esta razón se elige la prueba 1, AutoMLP con 53 variables como el mejor algoritmo entrenado. La utilización de este algoritmo para la predicción del corte de la hoja de celulosa implica una reducción considerable del impacto económico al evitar la falla. Para evaluar el impacto de su implementación, se evaluó el caso hipotético en que el algoritmo hubiese estado implementado para predecir los 47 cortes de hoja de celulosa considerados en el periodo de estudio. Para esto, se consideró el valor del recall que representa el número de cortes de hoja correctamente predichos. Este valor fue de 99,82% considerando 53 atributos. Por lo que la reducción estimada de costos en los 47 cortes, dado este nivel de recall, es de $20,4 \text{ MM US\$} * 0,9982 = 20,3 \text{ MM US\$}$.

Para comprender de manera general el patrón de corte de la hoja de celulosa, en la figura 3 se muestra un

árbol construido con los resultados obtenidos mediante al algoritmo Decision Tree. Este árbol contribuye a un mejor control de las variables más importantes para impedir que ocurran nuevos cortes de hojas de celulosa. Su análisis muestra que la variable que más influye en el corte de la hoja de celulosa es Combi Press Top Felt Suction Box. Si esta variable se mantiene sobre -46.4 kPa se produce el corte de la hoja de celulosa. Si se mantiene por sobre este valor, el corte depende de la variable Velocidad Foudrinier. Si esta velocidad esta entre 63.3 y 126.7 metros/minuto se produce el corte. Si la velocidad es superior a 126.7 metros/minuto y la variable Temperatura Vapor a Secador está por debajo de $163.1 \text{ }^\circ\text{C}$ existe alta probabilidad de corte. En términos generales, basado en las tres variables más relevantes, si la variable Combi Press Top Felt Suction Box se mantiene bajo 46.4 kPa , la Velocidad de Foudrinier sobre 126.7 metros/minuto y la Temperatura Vapor a Secador sobre $163.1 \text{ }^\circ\text{C}$, debería evitarse a lo menos el 50% de los cortes de hoja de celulosa.

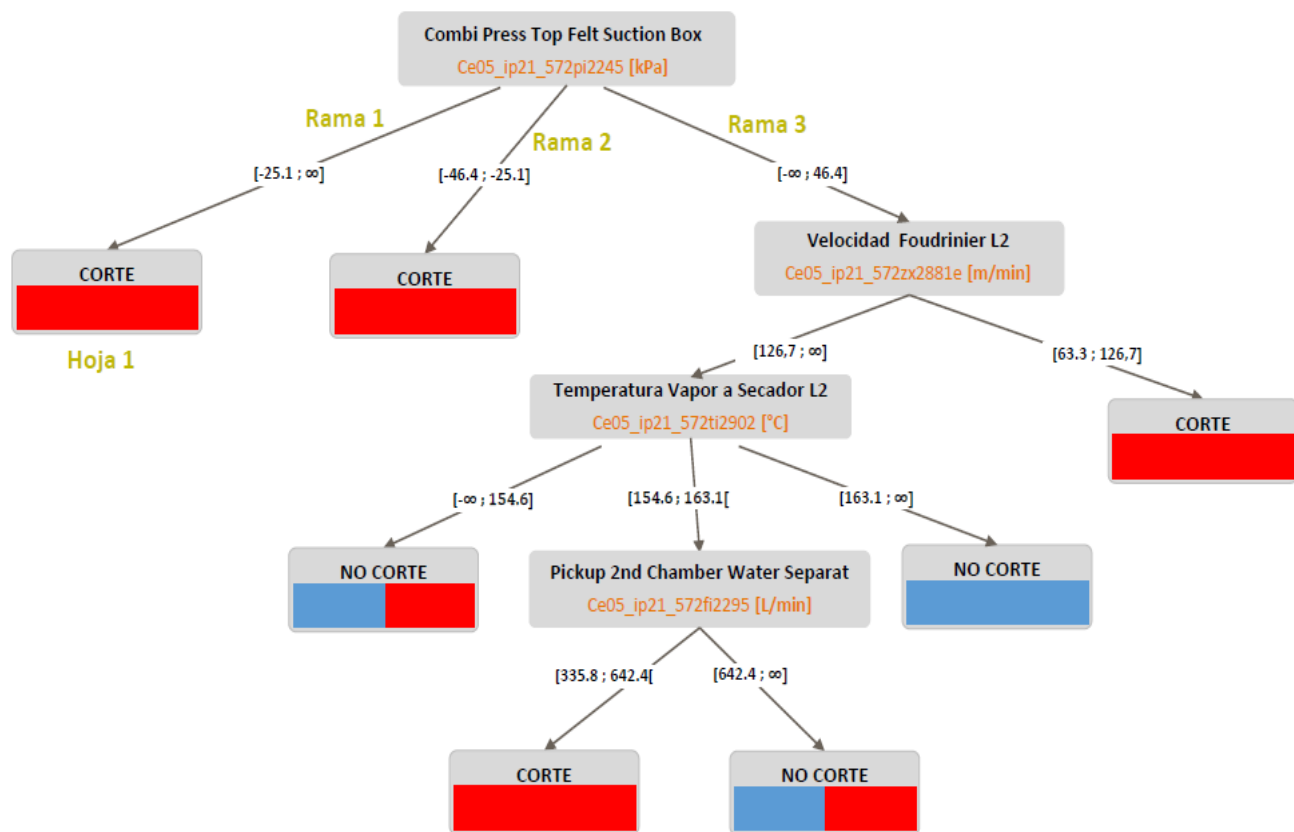


Fig. 3. Árbol de decisión asociado al corte de la hoja de celulosa.

V. CONCLUSIONES

No se encontró un método para la predicción de fallas en el corte de hoja de celulosa en la literatura por lo que se puede considerar esta investigación como inédita.

Los algoritmos de machine learning utilizados fueron capaces de identificar el patrón para la predicción del corte de la hoja de celulosa. El algoritmo de mayor desempeño predictivo fue AutoMLP. El algoritmo Decision Tree permitió analizar el patrón más general en base a las tres variables más importantes, lo que garantiza la reducción de a lo menos un 50% de los cortes de hoja.

Dado el alto valor de Recall del algoritmo AutoMLP entrenado con 53 variables, bajo el supuesto que este se hubiese implementado durante los 17 meses de estudio, la reducción del impacto económico de la falla hubiese bajado en 20.3 MM US\$.

Como la implementación de este algoritmo permitirá la predicción de esta falla con una ahora de anticipación, se podrán generar alertas preventivas en tiempo real y un plan de acción para prevenir los cortes.

El patrón general para prevenir los cortes, dada una alerta preventiva, será mantener Combi Press Top Felt Suction Box bajo -46.4 kPa, la Velocidad de Foudrinier sobre 126.7 metros/minuto y la Temperatura Vapor a Secador sobre 163.1 °C.

REFERENCIAS

- [1]B. Ranaganth y G. Viswanath, «Application of artificial neural network for optimizing cutting variables in laser cutting of 304 grade stainless steel,» *International Journal of Applied Engineering and Technology*, vol. 1, n° 1, pp. 106-112, 2011.
- [2]M. Durica, J. Frnda y L. Svabova, «Decision tree based model of business failure prediction for Polish companies,» *Oeconomia Copernicana*, vol. 10, n° 3, pp. 453-469, 2019.
- [3]G. Köksal, İ. Batmaz y M. C. Testik, «A review of data mining applications for quality improvement in manufacturing industry,» *Expert systems with Applications*, vol. 38, n° 10, pp. 13448-13467, 2011.
- [4]H. Poblete y R. Vargas, «Relacion entre densidad y propiedades de tableros HDF producidos por un proceso seco,» *Maderas. Ciencia y tecnología*, vol. 8, n° 3, pp. 169-182, 2006.
- [5]B. Kovalerchuk y E. Vityaev, «Data mining for financial applications,» *Data Mining and Knowledge Discovery Handbook*, pp. 1203-1224, 2005.
- [6]U. Fayyad, G. Piatetsky-Shapiro, P. Smyth y R. Uthurusamy, «Advances in knowledge discovery and data mining,» *American Association for Artificial Intelligence*, 1996.
- [7]A. K. Pandey y A. K. Dubey, «Neuro fuzzy modeling of laser beam cutting process,» *Applied Mechanics and Materials*, vol. 110, pp. 4109-4117, 2012.
- [8]M. Németh y G. Michalčonok, «Preparation and cluster analysis of data from the industrial production process for failure prediction,» *Research Papers Faculty of Materials Science and Technology Slovak University of Technology*, vol. 24, n° 39, pp. 111-116, 2016.
- [9]S. Ballı, «A data mining approach to the diagnosis of failure modes for two serial fastened sandwich composite plates,» *Journal of Composite Materials*, vol. 51, n° 20, pp. 2853-2862, 2017.
- [10]S. Dindarloo y E. Siami-Irdemoosa, «Data mining in mining engineering: results of classification and clustering of shovels failures data,» *International Journal of Mining, Reclamation and Environment*, vol. 31, n° 2, pp. 105-118, 2017.
- [11]E. e Oliveira, V. Miguéis, L. Guimarães y J. L. Borges, «Power Transformer Failure Prediction: Classification in Imbalanced Time Series,» *U. Porto Journal of Engineering*, vol. 3, n° 2, pp. 34-48, 2017.
- [12]A. Taghizadeh y N. Demirel, «Application of Machine Learning for Dragline Failure Prediction,» *E3S Web of Conferences*, vol. 15, p. 03002, 2017.
- [13]W. Chang, Z. Xu, M. You, S. Zhou, Y. Xiao y Y. Cheng, «A Bayesian Failure Prediction Network Based on Text Sequence Mining and Clustering,» *Entropy*, vol. 20, n° 12, p. 923, 2018.
- [14]K. Halteh, K. Kumar y A. Gepp, «Financial distress prediction of Islamic banks using tree-based stochastic techniques,» *Managerial Finance*, vol. 44, n° 6, pp. 759-773, 2018.
- [15]C.-H. Liu, C.-J. Lin, Y.-H. Hu y Z.-H. You, «Predicting the failure of dental implants using supervised learning techniques,» *Applied Sciences*, vol. 8, n° 5, p. 698, 2018.
- [16]B. Mohammed, I. Awan, H. Ugail y M. Younas, «Failure prediction using machine learning in a virtualised HPC system and application,» *Cluster Computing*, vol. 22, n° 2, pp. 471-485, 2019.
- [17]O. Sukhbaatar, T. Usagawa y L. Choimaa, «An artificial neural network based early prediction of failure-prone students in blended learning course,» *International Journal of Emerging Technologies in Learning (iJET)*, vol. 14, n° 19, pp. 77-92, 2019.
- [18]Z. Wang, W. Zhao y X. Hu, «Analysis of prediction model of failure depth of mine floor based on fuzzy neural network,» *Geotechnical and Geological Engineering*, vol. 37, n° 1, pp. 71-76, 2019.
- [19]V. S. Gujre y R. Anand, «Machine learning algorithms for failure prediction and yield improvement

during electric resistance welded tube manufacturing.» *Journal of Experimental & Theoretical Artificial Intelligence*, vol. 32, nº 4, pp. 601-622, 2020.

[20]P. du Jardin, «Forecasting corporate failure using ensemble of self-organizing neural networks,» *European Journal of Operational Research*, vol. 288, nº 3, pp. 869-885, 2021.

[21]R. Brachman y T. Anand, «The process of knowledge discovery in databases,» *Advances in knowledge discovery and data mining*, pp. 37-57, 1996.

[22]W. Frawley, G. Piatetsky-Shapiro y C. Matheus, «Knowledge discovery in databases: An overview,» *AI magazine*, vol. 13, nº 3, p. 57, 1992.

[23]F. H. Troncoso Espinosa y J. V. Ruiz Tapia, «Predicción de fuga de clientes en una empresa de distribución de gas natural mediante el uso de minería de datos,» *Universidad Ciencia y Tecnología*, vol. 24, nº 106, pp. 79-87, 2020.

[24]F. H. Troncoso, «Prediction of Recidivism in Thefts and Burglaries Using Machine Learning,» *Indian Journal of Science and Technology*, vol. 13, nº 6, pp. 696-711, March 2020.

[25]M. Kantardzic, *Data mining: concepts, models, methods, and algorithms*, John Wiley & Sons, 2011.

[26]F. H. Troncoso Espinosa, P. G. Fuentes Figueroa y I. R. Belmar Arriagada, «Predicción de fraudes en el consumo de agua potable mediante el uso de Minería de Datos,» *Universidad Ciencia y Tecnología*, vol. 24, nº 104, pp. 58-66, 2020.

[27]C. Romero y S. Ventura, «Data mining in education,» *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 3, nº 1, pp. 12-27, 2013.

[28]D. Larose y C. Larose, *Discovering knowledge in data: an introduction to data mining*, John Wiley & Sons, 2014.

RESUMEN CURRICULAR



Freddy Troncoso Espinosa, Doctor en Sistemas de Ingeniería, Universidad de Chile, Ingeniero Civil Industrial Universidad del Bío-Bío, Chile. Académico e Investigador Departamento de Ingeniería Industrial, Universidad del Bío-Bío. Concepción, Chile



Yamil Avello Betancur, Magíster en Ingeniería Industrial, Universidad del Bío-Bío, Ingeniero Constructor, Universidad del Bío-Bío. Jefe de proyectos, Departamento de Ingeniería, Planta de Celulosa Arauco Nueva Aldea, Concepción, Chile



Luis Martínez Flores, Ingeniero Civil Industrial, Magíster en Ingeniería Industrial, Universidad del Bío-Bío, Chile. Ayudante de investigación y docente tiempo parcial departamento Ingeniería Industrial, Universidad del Bío-Bío, Concepción, Chile.