Maderas. Ciencia y tecnología
ISSN: 0717-3644
ISSN: 0718-221X
Universidad del Bío-Bío

Urra-González, Cynthia; Ramos-Maldonado, Mario
Un enfoque de machine learning para la predicción de la calidad de tableros contrachapados
Maderas. Ciencia y tecnología, vol. 25, 36, 2023
Universidad del Bío-Bío

DOI: https://doi.org/10.4067/s0718-221x2023000100436

Disponible en: https://www.redalyc.org/articulo.oa?id=48575202038

ISSN impresa 0717-3644
ISSN online 0718-221X

Maderas. Ciencia y tecnología 2023 (25): 36, 1-22

DOI: 10.4067/s0718-221x2023000100436

A MACHINE LEARNING APPROACH FOR PLYWOOD QUALITY PREDICTION

UN ENFOQUE DE MACHINE LEARNING PARA LA PREDICCIÓN DE LA CALIDAD DE TABLEROS CONTRACHAPADOS

***Cynthia Urra-González*[1,♠]**

https://orcid.org/0000-0003-3899-5499

***Mario Ramos-Maldonado*[2]**

https://orcid.org/0000-0001-9498-6373

RESUMEN

Dado el impacto que tiene en la productividad y en la reducción de costos, la toma de decisiones es uno de los aspectos más requeridos en la industria. En la fabricación de tableros, la calidad del producto es función de múltiples variables, especialmente de la variabilidad de la madera. Esta calidad depende, entre otros factores, de la adherencia entre chapas o resistencia a la tracción perpendicular. El objetivo principal de este estudio fue evaluar un enfoque de Machine Learning, esto es modelos de aprendizaje automático, que permitan predecir la adherencia bajo condiciones de operación industrial, en la etapa de encolado y pre-prensado. Las principales variables de control que determinan esta adherencia son los tiempos operacionales, la cantidad de adhesivo, las condiciones ambientales y la temperatura en la chapa. Usando la metodología de analítica de datos Knowledge Discovery in Databases, se evaluaron algoritmos de Redes Neuronales Artificiales y Máquina de Soporte Vectorial. La función Sigmoid entregó resultados de acierto global (accuracy sobre el 66 %) y precisión en encontrar resultados correctos (casi 70 %). Al usar la función Relu se obtuvó un recall (sobre el 74 %), lo que muestra su aptitud para identificar la realidad. Estos resultados muestran la viabilidad de usar inteligencia artificial en la predicción de procesos complejos. Muchos espacios de mejora se abren con un adecuado pretratamiento de las variables de proceso a objeto de obtener mejores resultados. El aporte de este trabajo radica en la definición de una metodología para ser usada en plantas industriales, en particular en la fabricación de tableros contrachapados, y en mostrar la factibilidad del uso de datos industriales y Machine Learning en la predicción de la calidad del producto.

Palabras claves: Algoritmos, aprendizaje supervisado, industria de la madera, ingeniería de datos, modelos predictivos, Machine Learning, tableros contrachapados, Redes Neuronales Artificiales.

ABSTRACT

Because of the impact on productivity and cost reduction, decision making in industrial processes is one of the most required aspects in the industry. Specifically in the panel industries, product quality depends on multiple variables, especially wood variability. Among other factors, quality depends on the adhesion of veneers or perpendicular tensile strength. The main objective of this study was to evaluate a Machine Learning approach to predict the adhesion under industrial conditions in the gluing and pre-pressing stage. The control variables that determine this adhesion are mainly: operational times, amount of adhesive, environmental conditions, and veneer temperature. Using Knowledge Discovery in Databases data analytics methodology, Artificial Neural Networks and Support Vector Machine were evaluated. The sigmoid activation function was used

[1]Universidad del Bío-Bío. Facultad de Ingeniería. Departamento Ingeniería Industrial. Concepción, Chile.
[2]Universidad del Bío-Bío. Facultad de Ingeniería. Departamento Ingeniería en Maderas. Concepción, Chile.
♠Autor de correspondencia: cynthiaurragonzalez@gmail.com
Recibido: 24.06.2022 Aceptado: 25.04.2023

with 3 hidden layers and 245 neurons. In addition to the Adam optimizer, Multi-LayerPerceptron, Artificial Neural Networks delivered the best accuracy levels of over 66 %. Sigmoid showed an accuracy of over 66 %, precision fit good to find positive results (70 %). Relu function obtained the best recall (over 74 %) showing a good capacity to identify reality. Results show that it is not sufficient to generate a data set using the averages of each process variable, since it is difficult to obtain better results with the algorithms evaluated. This work contributes to defining a methodology to be used in plywood plants using industrial data to train and validate Machine Learning models.

Keywords: Algorithms, supervised learning, wood industry, data engineering, predictive models, Machine Learning, plywood, artificial neural networks.

INTRODUCCIÓN

El control y la optimización de procesos industriales requiere de la exhaustiva recopilación y análisis de datos. Los problemas de calidad de un producto pueden involucrar múltiples variables de entrada y salida que no son fáciles de modelar y optimizar (Köksal *et al*. 2011). Generalmente, se utilizan técnicas estadísticas para descubrir patrones a partir de los datos recopilados. Sin embargo, los grandes volúmenes de datos poseen varios inconvenientes: valores perdidos, ruido, tiempo extenso de procesamiento de cómputo. Las técnicas tradicionales de análisis de datos se han mostrado incapaces de modelar las relaciones complejas entre las variables y de predecir los valores de características desconocidas para una nueva muestra (Dogan y Birant 2021, Yan *et al*. 2020). Actualmente, las técnicas de Machine Learning (ML) o aprendizaje automatizado se han mostrado robustas en la modelación de procesos complejos multivariados y no lineales. Como ha sido señalado por Dogan y Birant (2021) el uso de ML es una rama de la inteligencia artificial que se utiliza para el descubrimiento de conocimientos "ocultos" de grandes volúmenes de datos, sean estos patrones, correlaciones, relaciones o anomalías. Estos modelos se validan a partir de datos históricos para predecir con precisión eventos futuros.

Las técnicas de ML han tenido éxito para resolver problemas en diversas aplicaciones de la industria manufacturera, tales como; sistemas inteligentes para la toma de decisiones (Cheng *et al*. 2018, Kujawińska *et al*. 2018), arreglos para mantenimiento de máquinas, predicciones de fallos, estimación de consumo energético de máquinas (Cupek *et al*. 2018, Gandhi *et al*. 2018, Nedelkoski y Stojanovski 2017, Pavlyshenko 2016, Zhang *et al*. 2017). Otros trabajos han abordado la calidad del producto y el control de las variables del proceso que involucran datos multivariados, donde las bases de datos pueden estar relacionadas a las características del producto, máquinas, parámetros de operación, nivel de experiencia del operador, tipo de turno, entre otros factores (Huang *et al*. 2018, Rostami *et al*. 2015, Wang 2013).

En la industria de la madera, la modelación de procesos usando inteligencia artificial ha estado circunscrita al uso de datos experimentales, sin embargo, un número cada vez mayor de trabajos están usando datos industriales. Esto es debido a la mayor incorporación de sensores, redes de datos y bases de datos de alta performance (Ramos y Aguilera 2021).

Específicamente en la industria de tableros se busca tener el control del proceso productivo y disminuir la variabilidad y la cantidad de productos rechazados. Este proceso de fabricación se basa en las siguientes etapas (Teihuel 2007):

- Recepción y almacenamiento: los rollizos de madera se almacenan en una cancha de acopio para ser sometidos a riego a través de aspersores, con el fin de conservar la calidad de la materia prima.
- Macerado: la madera proveniente de la etapa anterior ingresa a túneles de macerado, siendo sometidos a un baño de agua caliente, por un tiempo de 15h a 20 h. Esta etapa se realiza para disminuir la resistencia mecánica de la madera y así proporcionar las condiciones adecuadas para la próxima etapa.
- Debobinado: la madera ya ablandada pasa a través de tornos debobinadores, donde se obtienen laminas largas y aplanadas, denominadas chapas.
- Secado: las chapas son introducidas al interior de un secador continuo para reducir el contenido de humedad de la chapa. Las chapas se dejan en reposo por varias horas, para que disipen temperatura y humedad.
- Encolado: consiste en la aplicación del adhesivo en las chapas de madera. Este proceso es realizado

mediante encoladoras de rodillos, los cuales regulan la cantidad de adhesivo aplicada. Las principales variables que afectan a la calidad del tablero son: tipo y cantidad de adhesivo.

En la industria se utiliza generalmente urea-formaldehído o fenol-formaldehído, debido a la rápida reacción que posee en la etapa de prensado y curado, en donde se opera a elevadas temperaturas, relación de encolado y tiempo de encolado.

- Armado: formación del tablero con la superposición de las láminas que componen el tablero contrachapado, la cantidad de láminas va a depender del espesor final que se desee producir. Es por esto que la variable principal de esta etapa es el tiempo de armado, el cual podría influir en la adherencia del adhesivo.

- Prensado: los tableros encolados son ingresados a una preprensa en frío que hace presión sobre los tableros y los consolida. Posteriormente son ingresados a una prensa a temperatura elevada por un ciclo establecido de prensado. Es por esto que las variables más significativas en esta etapa del proceso son temperatura y tiempo de prensado, ya que de no ser las adecuadas el tablero puede presentar características que afecten a la calidad. Una de ellas es el "soplado" del tablero, es decir, la separación de las capas centrales que contienen el tablero, generalmente ocasionado por una acumulación excesiva de humedad en las capas internas del tablero, ya sea por fraguado del adhesivo o bien porque el tablero permaneció un tiempo inadecuado en reposo (Poblete y Peredo 1990).

- Acondicionamiento: almacenamiento del tablero en un lugar adecuado para que el adhesivo termine su proceso de curado.

Distintos estudios realizados a lo largo del tiempo en la industria de contrachapados no han considerado el proceso productivo completo. Esto se debe principalmente a la variabilidad de la madera, por tratarse de un material de origen biológico lo cual juega un rol esencial. Asimismo, la cantidad de variables y parámetros a tener en cuenta durante el proceso confirman la complejidad a controlar. En la Tabla 1, se presenta una revisión sobre las variables de proceso que inciden en la calidad de un tablero contrachapado, que se han concentrado en determinar las condiciones óptimas de operación.

Tabla 1: Revisión de variables significativas en el proceso de fabricación de un tablero contrachapado.

Research	T y H de chapa	Tipo de adhesivo	Flujo adhesivo	Tipo de madera	Nudos en la madera	H y T de secado	Distribución del adhesivo	t en la prensa	T y P en la prensa
(Toksoy *et al*. 2006)				X					
(Vick 1999)					X				
(Aydin y Colakoglu 2005)	X				X	X			
(Demirkir *et al*. 2013)	X	X	X	X		X			
(Li *et al*. 2020)		X	X						
(Bekhta y Salca 2018)	X						X		
(Kamal *et al*. 2017)					X				
(Özşahin *et al*. 2019)	X					X			
(Bekhta *et al*. 2020)				X				X	X

*T: temperatura; t: tiempo; H: humedad; P: presión.

De acuerdo con lo expresado en la Tabla1, se observa la cantidad de variables a controlar, aspectos que coinciden con la práctica industrial, a la que agregamos las condiciones ambientales en plantas donde el gradiente de temperatura también es importante y no existen acondicionamiento de clima (Urra 2021).

Normalmente, el control de calidad se realiza mediante tomas de muestras desde la línea de producción. Se realizan ensayos de laboratorio que definen las condiciones de operación. Por muchos años, la ausencia de monitoreo en tiempo real de las variables de proceso ha llevado a esta industria a considerar la experiencia de los operadores al momento de tomar decisiones con los problemas de latencias en la oportunidad de la decisión. Por ejemplo, las propiedades de resistencia mecánica medidas en el laboratorio no son confiables para la predicción en tiempo real dado su desfase temporal (Young *et al*. 2013). Sin embargo, en la última década, la incorporación de sensores, redes de datos y bases de datos industriales está facilitando el control del proceso y un mejor ajuste de la calidad del producto final.

Por su parte, la inteligencia artificial, y más específicamente ML, está en el centro de la Industria 4.0 y presenta enormes oportunidades para la optimización de los procesos productivos. La estructura compleja de los datos puede ser descubierta con técnicas de ML. Aquí, la práctica y las pruebas son claves, así como las metodologías de ingeniería de datos. Este enfoque ha vuelto las técnicas de ML muy poderosa en problemas de naturaleza compleja o con alta variabilidad (Ramos y Aguilera 2021).

En la Tabla 2, se muestran algunas investigaciones usando ML en la industria de tableros. Se muestra por tipo de problema de tablero, método de solución o algoritmo que se utilizó.

Tabla 2: Investigaciones sobre el control de proceso y calidad de la industria maderera enfocada en data minning.

Research	Tipo de problema	Área	Método de solución	Tipo de algoritmo	Rendimiento	Precisión	Cantidad de datos	Naturaleza de los datos
(Carty 2011)	1-2	2	1	3-4 (BRT)	2	-	1	1-2
(Tiryaki y Aydın 2014)	2	5	1	1-4	2	-	1	2
(Melo y Miguel 2016)	1-2	4	1	1	2	-	1	2
(Pang *et al*. 2020)	1-2	3	2	-	-	-	-	-
(Hazir *et al*. 2020)	1-2	5	1	2	1	-	1	1-2
(Demirkir *et al*. 2013)	2-3	3	1	1	1	-	1	2
(García *et al.* 2012)	3	3	1	1-4	1	-	1	2
(Suárez y Amador 2009)	5	-	-	-	-	-	-	-
Investigación (Urra 2021)	1-2-3-4	3	1	1-2-3-5	1	2	1	1

Celdas con un guion (-) indica no aplica o investigación no contiene la información.
Tipo de problema: 1 (variabilidad del proceso); 2 (control de calidad); 3 (control de proceso); 4 (análisis calidad); 5 (fundamento teórico). Área: 1 (MDF); 2 (MDP); 3 (tablero contrachapado); 4 (paneles aglomerados); 5 (madera).
Método de solución: 1 (ML-minería de datos); 2 (inteligencia artificial).
Tipo de algoritmo: 1 (Redes neuronales); 2 (SVM); 3 (Árbol de decisión); 4 (Regresión); 5 (Naive Bayes).
Rendimiento: 1 (=<70 %); 2 (>70 %). Precisión: 1 (<60 %); 2 (60<x<70 %); 3 (>70 %). Cantidad de datos: 1 (< 7000)
Naturaleza de los datos: 1 (datos en línea del proceso); 2 (Datos experimentales).

Como puede observarse en la Tabla 2, las investigaciones señaladas se enfocan y analizan como máximo 2 algoritmos de ML por estudio. Estos buscan predecir los resultados de ensayos físicos-mecánicos a través de las variables seleccionadas del proceso de fabricación con datos provistos por experimentos de laboratorio. Sólo dos de ellos incluyen datos en línea del proceso productivo (Carty 2011, Hazir *et al*. 2020). La totalidad de estos estudios incluyen y seleccionan en sus modelos variables desde el inicio hasta el final del proceso productivo, sin analizar la influencia y la capacidad predictiva de las variables por subetapas del proceso lo que los vuelve incompletos.

La presente investigación tiene como propósito analizar las variables de proceso de la etapa de encolado y pre-prensado de la fabricación de tableros contrachapados, usando con datos industriales capturados en línea del proceso productivo. Se tiene como objetivo evaluar diferentes algoritmos de ML que permitan validar un modelo que pueda predecir con la mejor efectividad posible la calidad del tablero.

MATERIALES Y MÉTODOS

Recursos materiales y equipamiento

1. Servidor de pruebas: las características del servidor de prueba contaban con un procesador de Intel Core i7 1.50 GHz, un RAM de12 gb y un sistema de 64 bits.

2. Programas

Para la elaboración de los algoritmos analizados en esta investigación se utilizaron los lenguajes de programación recomendados para el análisis de datos y ML, Python y R (Manrique 2020), Adicionalmente se utilizó

el software Anaconda, puesto que cuenta con los editores de código Jupyter y Spyder. Las versiones de los programas y bibliotecas con sus versiones utilizadas se mencionan en la Tabla 3 y Tabla 4 respectivamente.

Tabla 3: Programas utilizados en el estudio.

Sotfware	R studio	Software Anaconda	Spyder	Jupyter	Python
Versión	4.1.0	5.2.0	3.2.8	5.5.0	3.5.5

Tabla 4: Bibliotecas utilizadas en Python.

Bibliotecas Python	Numpy	Pandas	Scikit-Learn	Tensor Flow	Keras	Matplotlib	Nltk
Versión	1.14.0	0.23.0	0.19.1	1.10	2.2.2	2.2.2	3.30

Descripción de metodología

Se empleó la metodología KDD (Knowledge Discovery in Databases), una de las más utilizadas en ML, comenzando con un pre procesamiento y limpieza de los datos, con el fin de obtener modelos que contengan las variables de entradas más significativas y que expliquen la variable de salida/respuesta. Posteriormente, se valid**ó** el modelo a través de conjuntos de datos de prueba. La calidad de los resultados depende de la calidad de dichos datos (Dogan and Birant 2021).

Selección de datos

Se realizó una revisión de la literatura y conocimiento del proceso en planta (verificando los datos disponibles), determinando las variables que podrían incidir en la calidad (Figura 1).

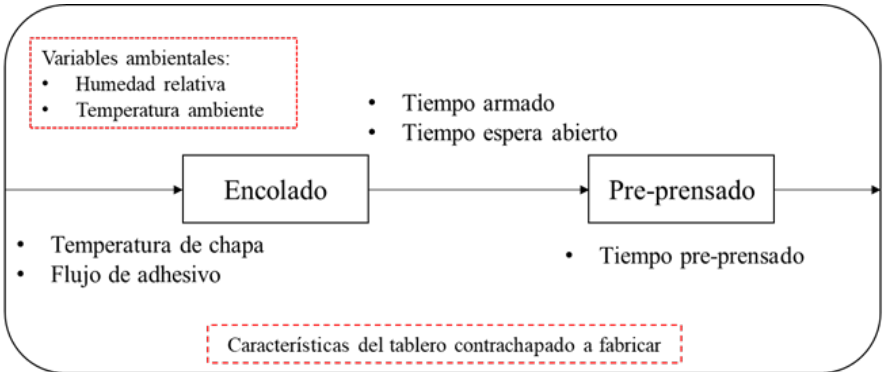

Figura 1: Representación de variables seleccionadas del proceso de encolado y pre-prensado de un tablero contrachapado.

Naturaleza de los datos

Los registros de las variables en línea fueron capturados por la plataforma PISystem®. La Figura 2 muestra la arquitectura de la captura y disponibilidad de datos. Se trata de una planta de tableros contrachapados ubicada en Chile, que produce un volumen superior a 300000 m^3/año de tableros de *Pinus Radiata.*

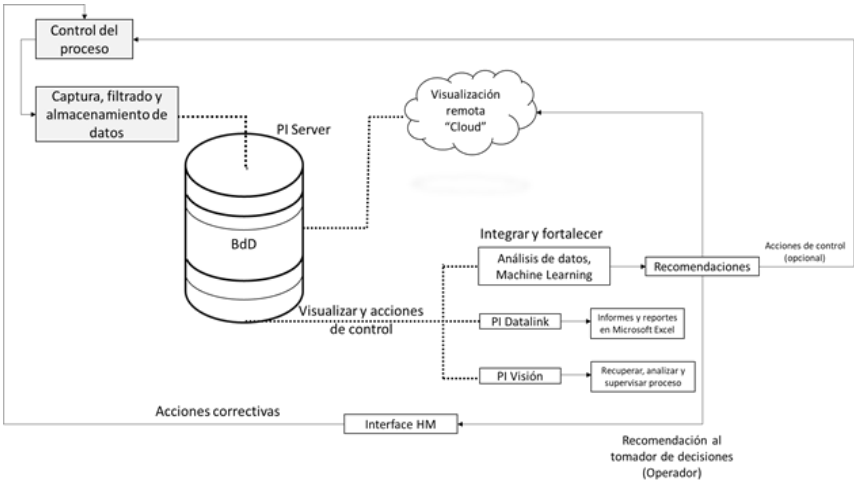

Figura 2: Captura y disponibilidad de datos mediante PISystem© (Osisoft 2020).

Las variables independientes del proceso productivo de tablero contrachapado que fueron consideradas son las mostradas en la Tabla 5.

Tabla 5: Descripción y naturaleza de las variables de proceso seleccionadas en el estudio.

Variable	Descripción	Data histórica 2020 (meses)	Time data App	Unidad
Flujo del adhesivo	Corresponde al flujo o gramaje de adhesivo en kg/min que es incorporado a las chapas en cada encoladora.	01-12	Cada 1 min	kg/min
Temperatura de la chapa	El control de temperatura de las chapas se realiza luego de la etapa de secado, puesto que se debe pasar por un periodo de estabilización para evitar problemas de adherencia.	01-12	Cada 1 s	°C
Tiempo de armado	Es el tiempo en que se completa el armado del tablero (desde la primera chapa a la última) luego de que cada chapa pase por las encoladoras.	01-12	Cada 10 min	min
Tiempo espera abierto	Es el tiempo que transcurre luego del armado del tablero hasta el ingreso de la pre-prensa.	09-12	Cada 10 min	min
Tiempo de pre-prensado	Tiempo que permanece el tablero en la pre-prensa. Consiste en aplicar presión al tablero encolado en frío.	01-12	Cada 1-5 min	min
Humedad relativa	Variables ambientales que influyen en la calidad de adherencia del tablero	01-12	Cada 30 min	%
Temperatura ambiente				°C

Tabla 6: Proporción de las variables de proceso capturadas.

	Flujo del adhesivo	Temperatura de la chapa	Tiempo de armado	Tiempo espera abierto	Tiempo de pre-prensado	Humedad relativa	Temperatura ambiente
Cantidad de registros	2500000	31000000	370000	17000	250000	17500	
Promedio	47,5 (kg/ min)	28,1 (°C)	9,32 (min)	4,34 (min)	6,22 (min)	44,2 %	24,7 (°C)
Desviación estándar	1,68	4,12	0,84	0,56	0,69	6,12	2,90

Como se observa en las Tabla 5 y Tabla 6, las variables de proceso fueron capturadas en distintos intervalos de tiempo, por lo que se tuvo una gran variabilidad en la cantidad de registros. Los únicos datos capturados fuera de línea corresponden a la resistencia adhesiva o tracción perpendicular. Estos registros de control de calidad se realizaban por medio de ensayos físico-mecánicos bajo la norma EN-314 (2007). La estructura de estos datos se caracteriza por el tipo de encoladora, turnos, número de la chapa del tablero donde fue tomada la muestra, espesor del tablero y tipo de uso (interior o exterior) según adhesivo aplicado.

Creación de conjunto de datos (data set)

Para cada variable e intervalo de tiempo de la Tabla 5, se calculó un valor promedio representativo que es registrado en el data set preliminar antes del pre procesamiento. Por su parte, los datos de ensayos de adherencia provenientes del laboratorio (datos capturados fuera de línea), que corresponden a la variable respuesta, fueron asignados al registro correspondiente por fecha de producción. Se realizó bajo los criterios de aceptación del valor de adherencia (criterio de aceptación: tablero utilizado para interiores ≥ 80 % y exteriores ≥ 85 %), se adicionó en la columna respuesta si el tablero fabricado en la fecha de producción fue aceptado o rechazado.

Por otro lado, se tuvieron las variables que indican las características del tablero fabricado, tales como espesor, calidad, uso (interno o externo) y también variables de operación industrial: número de la encoladora, turno al cual pertenece la producción (la jornada laboral es de 8 horas por lo que se tienen 3 turnos), jefe de turno (persona encargada de la producción); a esta variable se le realizó un label encoder (asignar un número random). Se observó en el análisis previo de los datos que las variables tiempo de armado y flujo del adhesivo puede variar drásticamente de un momento a otro, por lo que se calculó la varianza que presentan estas variables en cada turno como se expone en la Tabla 7.

Tabla 7: Ejemplo de data set preliminar (extracto de datos reales).

Producción	Jefe de turno	Espesor	Calidad	Uso	Turno	Encoladora	Línea	Temperatura ambiente (ºC)	Humedad relativa	Temperatura de la chapa (ºC)	Tiempo de armado (min)	Varianza de tiempo de armado	Tiempo de espera abierto (min)	Flujo de adhesivo (kg/min)	Varianza flujo de adhesivo	Respuesta
02-01-2020	1	18	1	Interior	2	1	6	27	53	26	10	7	4	55	2	Aceptado
04-01-2020	2	20	8	Exterior	3	2	5	30	40	40	9	1	2	40	1	Rechazado
10-06-2020	4	18	1	Interior	1	2	4	25	45	38	15	4	6	60	0,5	Aceptado

Preprocesamiento y transformación del data set

Para la creación del data set, se eliminaron valores atípicos, tales como; paradas de producción, valores perdidos y errores en la recolección de las variables. El criterio de calidad estuvo dado por la adherencia la que permitió establecer los criterios de aceptación o rechazo. Para datos faltantes o valores perdidos se realizaron técnicas de imputación de datos usadas en ciencia de datos.

Las variables cualitativas a cuantitativas fueron transformadas dependiendo del algoritmo evaluado y, cuando las variables no eran binarias, se procedió a una transformación a variables “dummy”.

Se crearon 2 data sets para realizar los experimentos.

a) Data set 1. Los datos faltantes de tiempo de espera abierto se imputaron a través de la técnica de regresión múltiple, obteniendo el data set 1 con un total de 1137 registros. En la Figura 3 se expone la frecuencia de estos registros en base al criterio de calidad.

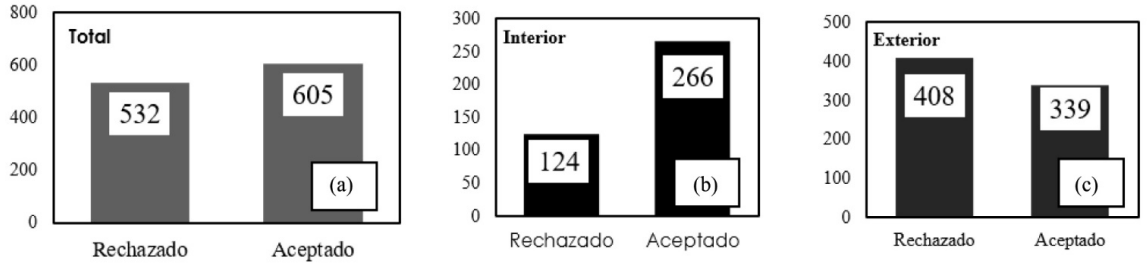

Figura 3: Frecuencia de tableros contrachapado; aceptados y rechazados bajo el criterio de calidad, adherencia. (a) Total datos, (b) interior, (c) exterior.

b) Data set 2. Al data set 1 se le adicionó la variable de tiempo de pre-prensado, la cual se encontraba disponible para tres encoladoras de las cuatro existentes. Para este data set se realizaron pruebas con los registros entregados disponibles (744 registros). Adicionalmente, se creó un subdataset, denominado data set 2.1, en el cual se realizaron pruebas con imputación de datos para la variable de tiempo de pre-prensado de la encoladora faltante, mediante técnicas de regresión múltiple y aleatoria (1137 registros).

Construcción de modelos predictivos

Los algoritmos evaluados y los parámetros que se modificaron se señalan en la Tabla 8. (Doshi *et al*. 2021).

Tabla 8: Algoritmos evaluados, con sus respectivos parámetros variables.

Algoritmo	Parámetros variables
Árbol de decisión	Numero de árbol,MTRY (número de predictores y/o variables)
Random Forest	Número de árbol, profundidad
K-NN	Número de vecinos (K)
Redes neuronales	Número de neuronas y capas ocultas, función de activación (Sigmoid, logistic, tanh y Relu), % de conjunto de datos de entrenamiento (70 % -80 %), epoch (ciclos).
SVM	Tipo de kernel, range (C)

Los valores de los parámetros mencionados en la Tabla 8 fueron optimizados según los resultados obtenidos, puesto que la metodología empleada en el desarrollo de algoritmos de ML es iterativa. El punto de partida en la iteración de los parámetros evaluados se realizó mediante un análisis de la búsqueda Grid, lo que permitió analizar el punto en donde el error del algoritmo va disminuyendo.

Las Redes Neuronales se implementaron en Python, variando los parámetros que se mencionan en la Tabla 8. Se probaron distintas combinaciones en un mismo modelo. La optimización se realizó bajo el algoritmo Adam.

Evaluación y validación de modelos predictivos

Por tratarse de un de un problema de clasificación, en este caso la clasificación de aceptación/rechazo del tablero, en la evaluación de los modelos se utilizó como métrica la matriz de confusión. Esta es una herramienta que permite calificar el desempeño de un algoritmo de aprendizaje supervisado (Düntsch y Gediga 2019, Luque *et al*. 2019).

		Predicho	
		P	N
Real	P	VP	FN
	N	FP	VN

$$Accuracy = \frac{VP+VN}{VP+FN+FP+VN}; \quad Recall = \frac{VP}{VP+FN};$$

$$Precisión = \frac{VP}{VP+FP}$$

Figura 4: Métricas de eficiencia en algoritmos de ML.

Donde, VP: Verdaderos positivos, FN: falsos negativos, FP: falsos positivos, VN: verdaderos negativos, accuracy: proporción de predicciones correctas en relación con el total, recall: proporción de casos que el algoritmo identifico como valores positivos, precisión: proporción de casos que el modelo predijo correctamente los valores positivos.

La última etapa del proceso KDD llamada validación, es donde se evaluó el desempeño de los diferentes algoritmos de ML explicados anteriormente. En ocasiones, esto se realiza dividiendo el conjunto de datos en tres partes: entrenamiento, validación y prueba. Estos últimos para visualizar el desempeño predictivo del algoritmo. En este estudio en particular se determinó utilizar datos para entrenamiento y validación, debido a que no se contaba con una gran cantidad de registros en el conjunto de datos (1137). Es por esto que se realizó la validación a través de validación cruzada, que es una metodología confiable para evaluar los algoritmos cuando se tienen registros limitados (Ochoa 2019).

Se dividió el conjunto de datos en entrenamiento y validación (70 % - 30 % y 80 % - 20 %, respectivamente). Esta proporción aparece con más frecuencia en la literatura. Con esto se obtuvieron las métricas del rendimiento del modelo (entrenamiento) y la validación con el 20 % - 30 % de los datos, es decir, se realizó la predicción con el conjunto de datos de validación y luego se evaluaron las predicciones acertadas a través de las métricas informadas en la Figura 4.

RESULTADOS Y DISCUSIÓN

Análisis de correlación

De acuerdo con el análisis de correlación de Pearson de las variables de operación, se determinó el nivel de interdependencia de las variables a través del análisis de componentes principales, se obtuvo la matriz representada en la Tabla 9.

Del análisis se obtuvo una mayor correlación entre la temperatura ambiente y la temperatura de la chapa con un valor de 0,49, seguido del tiempo de armado con el espesor del tablero contrachapado, con un valor de 0,39, lo que es razonable y acorde a lo expuesto por Demirkir *et al*. (2013) y Zavala y Valdivia (2004). Sin embargo, las correlaciones no son lo suficientemente altas (mayores a 0,8) como para disminuir variables del data set.

Tabla 9: Resultado de correlación: variables de operación seleccionadas en la etapa de encolado y pre-prensado.

	Espesor	T ambiente	HR	T de la chapa	t de armado	σ^2 turno t de armado	t espera abierto	σ^2 turno del flujo de adhesivo	Flujo adhesivo
Espesor	1								
Temperatura ambiente	0,07	1							
Humedad relativa	-0,020	-0,28	1						
Temperatura de la chapa	0,011	0,49	-0,07	1					
Tiempo de armado	0,39	0,25	-0,16	0,075	1				
Varianza turno tiempo de armado	0,18	0,18	-0,025	-0,13	0,36	1			
Tiempo espera abierto	-0,1	-0,37	0,35	-0,20	-0,20	-0,084	1		
Varianza turno del flujo de adhesivo	0,05	-0,07	0,05	-0,09	0,025	0,08	0,16	1	
Flujo adhesivo	0,19	0,31	-0,023	0,38	0,27	0,06	-0,17	-0,11	1

*Las celdas que se encuentran coloreadas muestran el grado de correlación, colores cálidos mayor correlación.

Modelos predictivos utilizados

Árbol de decisión

Mediante validación cruzada se realizó un barrido por la cantidad de árboles incluidos en el modelo. El parámetro MTRY es el número de predictores y/o variables que se seleccionan en cada división del árbol, con el fin de encontrar los parámetros óptimos, tanto para el ajuste del modelo como para la capacidad de predicción del algoritmo, obteniendo que a partir de una cantidad de 100 árboles y un MTRY igual a 3 el error se mantuvo constante. Como el análisis individual de los parámetros del modelo puede ignorar combinaciones óptimas, se realizó la búsqueda grid search para los parámetros mencionados en la Tabla 10.

Tabla 10: Valores de los hiperparámetros evaluados mediante grid search.

Hiperparámetro	Rango
Número de árboles (num_trees)	c (50, 100, 500, 1000, 5000)
MTRY	c (1, 5, 7)
Profundidad máxima (max_depth)	c (1, 3, 10, 20)

Mediante la optimización de los hiperparámetros se obtuvieron los resultados expuestos en la Tabla 11, dismnuyendo el error MSE del modelo en un 6 % y en la validación un 3 %.

Tabla 11: Resultados de árbol de decisión de tipo regresión, data set 1.

		Entrenamiento				Validación
	Paquete	N° de árbol	MTRY	Nro de variables independientes	Error MSE	Error MSE
Validación cruzada	Ranger	10	3	15	0,28	0,52
	Ranger	101	7	15	0,24	0,50
	Ranger	101	1	15	0,237	0,49
	Ranger	500	3	15	0,235	0,49
	Ranger	500	1	15	0,234	0,49

Para efectos de analisis posteriores, se determinó la importancia de las variables seleccionadas en la predicción de la adherencia del tablero, es decir, cuales de ellas tiene una mayor influencia en la calidad del producto. Dicho análisis, se relizó a través de la pureza de los nodos y empleando la técnica de permutación (Figura 5), coincidiendo con Demirkir *et al*. 2013 y Li *et al*. 2020 en que las variables más influyentes en la calidad del tablero (adherencia) fueron: el flujo de adhesivo que se adiciona en la etapa de encolado y la tem-

peratura ambiente.

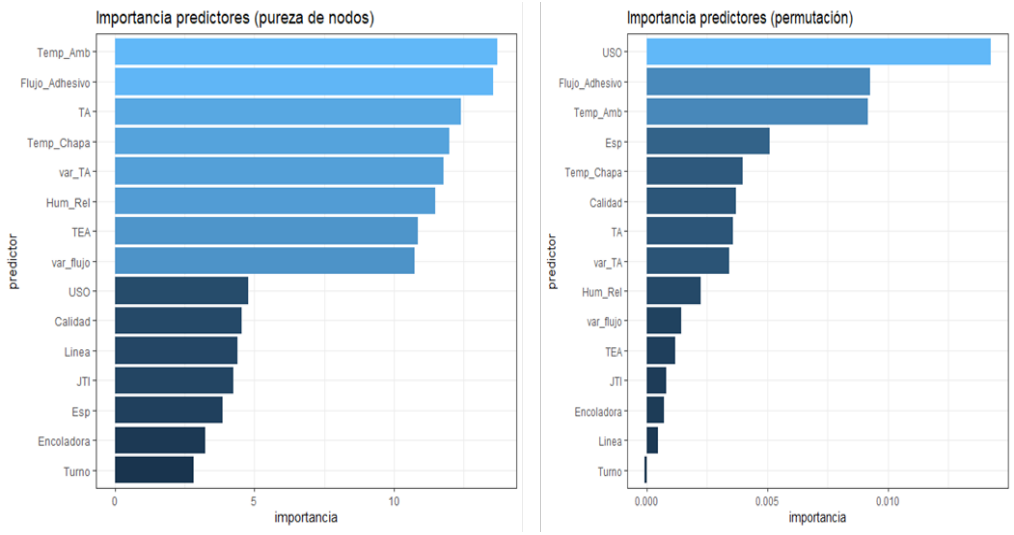

Figura 5: Importancia de los predictores (variables de entrada) para árbol de decisión-regresión.

De igual manera que con los árboles de decisión-regresión, se realizó una optimización de los hiperparámetros, Tabla 12.

Tabla 12: Resultados de árbol de decisión-clasificación, data set 1.

N°	Paquete	Accuracy	Recall	Precisión	Optimización
1	Ranger	57,29	58,58 %	36,40 %	-
2	Tree	55,36 %	60,46 %	19,12 %	-
3	Tree	53,25 %	51,14 %	49,63 %	Hiperparámetro de N°2
4	DecisionTreeClassifier	52,63 %	53,75 %	49,43 %	-
5	DecisionTreeClassifier	57,60 %	53,12 %	54,84 %	Hiperparámetro de N°4

Se obtuvo una mejora en la predicción de un 5 % al realizar la optimización de los parámetros del árbol de decisión-clasificación. Sin embargo, un accuracy de 57,6 % en la validación del modelo, no garantiza que la predicción de eventos futuros sea la adecuada. Por esto, para este estudio se realizaron experimentos para determinar la influencia de las variables de proceso más influentes en la calidad del tablero.

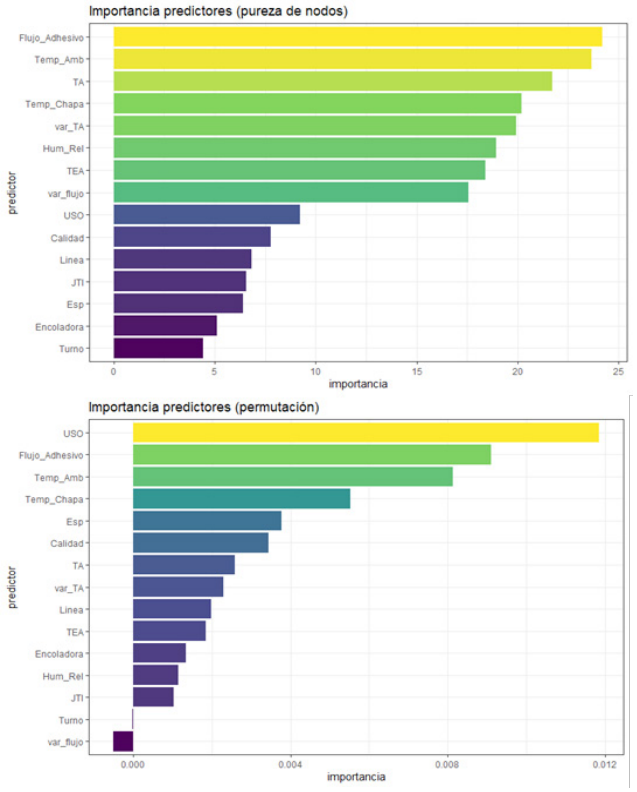

Figura 6: Importancia de las variables de entrada en la calidad (árbol de decisión-clasificación).

De los árboles de decisión evaluados preliminarmente se obtuvieron resultados similares en la proporción de las variables más importantes para modelos predictivos tanto para regresión como para clasificación (Figura 5 y Figura 6), las cuales están representadas por flujo de adhesivo, temperatura ambiente, tiempo de armado, temperatura de la chapa y el uso del tablero (para interiores o exteriores). Con estas variables y la aplicación del algoritmo Random Forest (RF), se obtuvieron los resultados presentados en la Tabla 13.

Tabla 13: Resultados de RandomForest variando el número de variables de entrada al modelo.

Paquete	Accuracy	Accuracy validación cruzada	Recall	Precisión	Variables
RandomForest	59,91 %	60,10 %	66,94 %	61,36 %	15
RandomForest	61,23 %	61,44 %	66,94 %	62,79 %	9

En base a la leve mejora en la predicción (aumento de un 1,3%), utilizando las variables seleccionadas en el modelo Random Forest, se realizaron experimentos con la utilización de estas variables y con la totalidad de las variables en los modelos predictivos Support Vector Machine (SVM), k-NN y Redes Neuronales, estableciendo distintos modelos en base a la selección de variables (Figura 7):

- Modelo 1: se seleccionaron todas las variables de estudio tales como: turno, jefe de turno, calidad, línea de producción, tipo de encoladora, espesor, uso, humedad relativa, tiempo de armado, varianza por turno de tiempo de armado, tiempo de espera abierto, flujo de adhesivo, temperatura ambiente y de la chapa.
- Modelo 2: se eliminaron las variables correspondientes al jefe de turno y línea de producción, manteniendo las variables espesor, uso, humedad relativa, tiempo de armado, varianza por turno de tiempo de armado, tiempo de espera abierto, flujo de adhesivo, temperatura ambiente y de la chapa.

- Modelo 3: se mantuvieron únicamente las variables de operación en la fabricación de tablero tales como humedad relativa, tiempo de armado, varianza por turno de tiempo de armado, tiempo de espera abierto, flujo de adhesivo, temperatura ambiente y de la chapa.

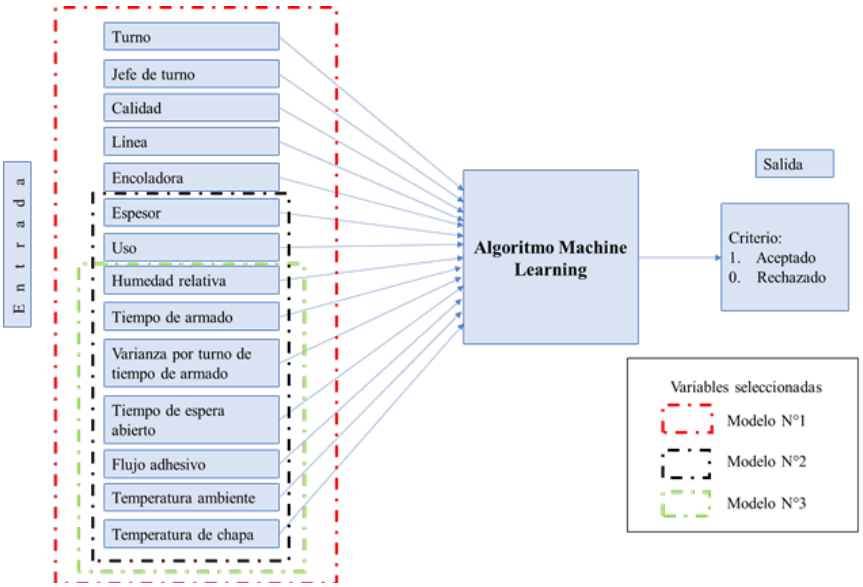

Figura 7: Variables seleccionadas en los algoritmos de ML evaluados, Data set 1.

SVM y K-NN

Para el algoritmo SVM de clasificación se modificó el tipo de kernel y el hiperparámetro C. Este último se optimizó a través de grid search (Figura 8). El kernel lineal entregó el mejor valor de C, el cual es igual a 0,1 (Figura 8a), ya que el error en la validación del algoritmo se vuelve contante, basta con utilizar ese valor para evitar costos computacionales innecesarios. Para el kernel radial y polinomial, se visualizó una mayor sensibilidad al cambio del valor de C (Figura 8b, Figura 8c). Dentro de los rangos evaluados se encontró que los mejores valores para el parámetro C en kernel radial es 10 y para polinomial es 5. Se realizaron réplicas del modelo variando estos parámetros para confirmar lo expuesto por la búsqueda grid search (Tabla 14).

Las mejores predicciones de SVM ocurrieron cuando se utilizó el hiperparámetro encontrado en el kernel lineal. Aquí se obtuvo una leve variación en las métricas de validación al aumentar la cantidad de variables de entrada del modelo. La diferencia radica principalmente en la capacidad del modelo en predecir de mejor forma los verdaderos negativos (en este caso los tableros que son aceptados bajo el criterio de calidad de adherencia), cuando se seleccionan todas las variables de estudio.

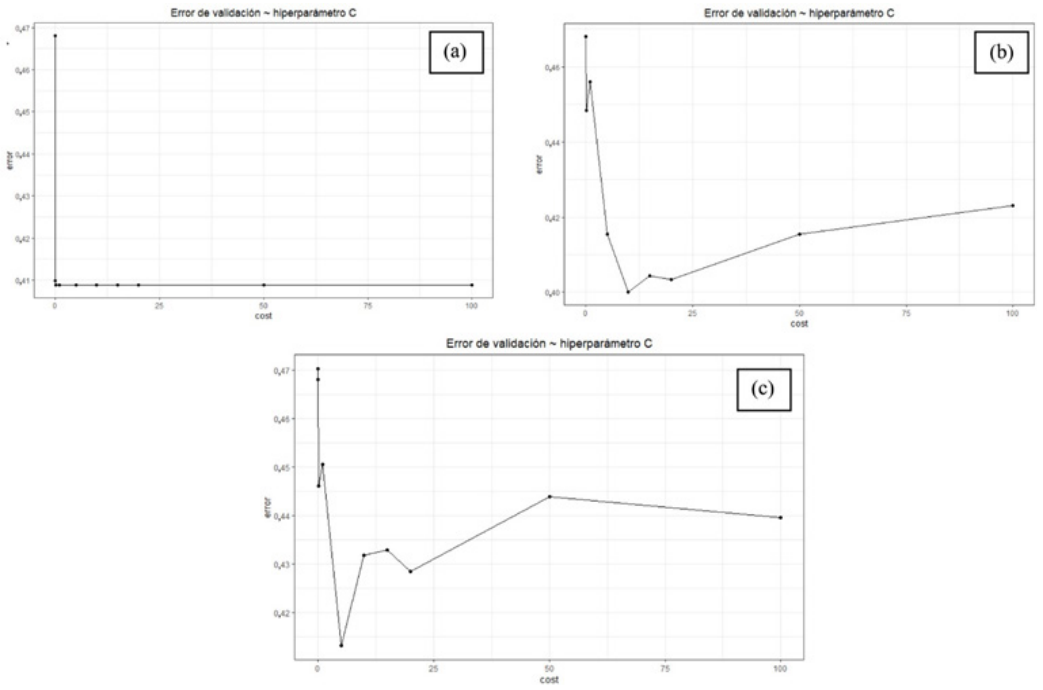

Figura 8: Error del modelo SVM según hiperparámetro C. Tipo de kernel: (a) lineal; (b) radial; (c) Polinomial.

Tabla 14: Resultados de SVM, Data set 1.

Librería/paquete	Programa	Modelo	Kernel	C	Acorace	Recall	Precisión
E1071	R studio	1	linear	0,01	55,51 %	63,89 %	38,01 %
SVC	Python	2	linear	0,01	53,95 %	45,05 %	53,19 %
E1071	R studio	1	linear	0,1	59,91 %	69,74 %	43,30 %
SVC	Python	2	linear	0,1	57,02 %	70,27 %	54,55 %
E1071	R studio	1	linear	100	59,91 %	69,74 %	43,80 %
SVC	Python	2	linear	100	53,95 %	54,95 %	52,59 %
E1071	R studio	1	radial	10	59,03 %	61,11 %	63,64 %
SVC	Python	2	rdf	10	54,39 %	50,45 %	53,33 %
E1071	R studio	1	radial	100	54,62 %	57,14 %	59,50 %
SVC	Python	2	rbf	100	57,90 %	56,76 %	56,76 %
E1071	R studio	1	polynomial	5	56,39 %	58,73 %	61,16 %

*La fila destacada informa la corrida con mejores métricas del algoritmo.

En la Tabla 14 se observa que la optimización del hiperparámetro es adecuado para cada kernel. Además, que independientemente del número de variables seleccionadas y el tipo de kernel, se obtienen resultados similares con eficiencias aproximadas a 59 %.

Se observa en la Tabla 15 las corridas del modelo N °2 en un amplio rango del valor de K desde 1 a 120, obteniendo las mejores métricas para accuracy y recall con un K igual a 50. Adicionalmente, no se encontró una diferencia significativa en la capacidad predictora en comparación con los demás algoritmos evaluados anteriormente, ya que al igual que SVM se obtiene un accuracy cercano al 60%. Sin, embargo, se obtuvieron mejores resultados para la métrica recall, es decir, se identificoo con mayor exactitud los tableros aceptados.

Tabla 15: Variación de k en el modelo N°2, K-NN.

Valor de K	Accuracy %	Recall %	Validación cruzada "accuracy" %
1	54,63	52,89	-
5	57,27	63,64	55,16
8	55,95	62,81	56,49
10	56,39	62,81	57,38
30	57,71	68,60	57,59
40	59,91	71,07	57,48
50	61,23	72,72	58,47
100	60,35	73,55	57,80
120	59,47	75,21	59,65

*La fila destacada informa la corrida con mejores métricas del algoritmo.

Redes neuronales artificiales (RNA)

Para los modelos de RNA se varió la selección de variables de entrada según los modelos mostrados en la Figura 7.

Redes neuronales simples

Las RNAs simples se evaluaron en el programa R studio, variando la cantidad de variables de entrada, el número de neuronas y la función de activación (logistic y tangencial). Los análisis se realizaron con ayuda de la librería neuralnet. Para los modelos1 y 2, se presentan estos resultados en la Tabla 16 y Tabla 17 respectivamente. Se realizaron 10 réplicas.

La distribución de redes neuronales se define como c(x,y), donde se expone el número de neuronas por capas. Por ejemplo: el vector c(6,7) indica que la primera capa de entrada cuenta con 6 neuronas y la segunda con 7, siendo esta la capa oculta. Si el vector es por ejemplo c(4,5,6), tendrá dos capas ocultas de 5 y 6 neuronas.

Tabla 16: Resultados RNA simples, modelo 1, data set 1.

Entrenamiento (75 % del data set 1)				Validación (25 % del data set 1)		
Distribución de neuronas	Función de activación	Steps	Error	Accuracy	Recall	Precisión
c(6,6)	"logistic"	40233	63,30 %	55,99 %	48,48 %	53,04 %
c (9,2)	"logistic"	61943	50,96 %	57,96 %	50,83 %	55,67 %
c (10,1)	"logistic"	33468	47,53 %	60,56 %	54,55 %	58,06 %
c (12)	"logistic"	13232	38,56 %	53,17 %	49,24 %	49,62 %
c(6,6)	"tanh"	15369	72,37 %	54,25 %	32,58 %	51,19 %
c (9,2)	"tanh"	12269	65,92 %	56,69 %	51,52 %	53,54 %
c (10,1)	"tanh"	1115	58,54 %	54,58 %	37,88 %	51,55 %
c (12)	"tanh"	22696	55,14 %	56,69 %	40,15 %	54,64 %

*La fila destacada informa la corrida con mejores métricas del algoritmo.

Tabla 17: Resultados RNA simples, modelo 2, data set 1.

Entrenamiento (75 % del data set 1)				Validación (25 % del data set 1)		
Distribución de neuronas	Función de activación	Steps	Error	Accuracy	Recall	Precisión
c(6,6)	"logistic"	48884	66,84 %	57,71 %	50,38 %	54,94 %
c (9,2)	"logistic"	25957	74,29 %	57,68 %	41,06 %	56,26 %
c (10,1)	"logistic"	97528	68,37 %	59,86 %	48,48 %	58,18 %
c (12)	"logistic"	19081	63,79 %	62,60 %	52,27 %	61,61 %
c(6,6)	"tanh"	82428	81,74 %	56,34 %	38,64 %	54,26 %
c (9,2)	"tanh"	21301	85,24 %	58,10 %	65,90 %	54,04 %
c (10,1)	"tanh"	97224	85,22 %	57,75 %	40,91 %	56,25 %
c (12)	"tanh"	32225	71,71 %	54,58 %	43,18 %	51,35 %

*La fila destacada informa la corrida con mejores métricas del algoritmo.

De las RNA evaluadas se obtuvo un mejor rendimiento con la función de activación logistic tanto para los modelos 1 y 2. Al disminuir las variables de entrada (modelo 2) se obtuvo un accuracy de 63 % en la validación. Sin embargo, existe un error en el entrenamiento de aproximadamente 64 % a diferencia de utilizar el modelo 1 (todas las variables seleccionadas), donde el error de entrenamiento disminuye a un 47 % con métricas similares a la validación.

Redes neuronales, Multi-Layer Perceptrón (MLP)

Al obtener mejores resultados de redes neuronales simples con la selección de variables mencionadas en el modelo N °2, se siguió esta línea en RNAs de tipo MLP con la función de activación Relu, variando el número de neuronas (dentro de los rangos evaluados anteriormente, de 6 a 30 neuronas) y la cantidad de ciclos en el entrenamiento. Los principales resultados se exponen en la Tabla 18 y Tabla 19.

Tabla 18: Experimentos de MLP con función de activación Relu, modelo N°2, dataset 1.

Entrenamiento 80% de data set					Validación cruzada		
Capa oculta	Distribución de neuronas	Epoch (ciclos)	Size	Accuracy	Accuracy	Recall	Precisión
1	c(12,6)	100	10	61,06 %	65,79 %	74,26 %	59,06 %
1	c(12,6)	200	10	60,95 %	64,47 %	73,27 %	57,81 %
1	c(12,6)	300	10	60,95 %	64,47 %	73,27 %	57,81 %
1	c(12,6)	500	10	60,29 %	64,47 %	71,29 %	58,06 %
1	c(12,6)	1000	10	60,84 %	64,47 %	74,26 %	57,69 %
1	c(20,10)	100	10	60,29 %	64,91 %	67,33 %	59,13 %
1	c(20,10)	200	10	61,61 %	63,60 %	61,39 %	58,49 %

*La fila destacada informa la corrida con mejores métricas del algoritmo.

En la Tabla 19 se realizaron corridas con las mismas instancias expuestas en la Tabla 18 (número de neuronas y epochs (ciclos), pero con distinta proporción del data set para entrenamiento y validación, 70 % y 30 % respectivamente. Se obtuvo una predicción correcta de 59 % a 63 %, siendo similar a lo obtenido por la partición del data set en 80 % de entrenamiento y 20% de validación (Tabla 18), obteniendo predicciones correctas de aproximadamente 65 %. Con relación al accuracy de la etapa de entrenamiento no se presentaron diferencias al disminuir la cantidad de registros, puesto que se mantuvo en 59 % a 60 % para ambos casos.

En la Tabla 20 se recogen los valores correspondientes a la evaluación del modelo de RNA de tipo MLP con una mayor cantidad de neuronas distribuidas en diferentes capas ocultas, variando a su vez la función de activación. El primer valor de la distribución de neuronas hace referencia a la capa de entrada y los demás va-

lores constituyen las neuronas de cada capa oculta, la capa de salida es binaria (tablero aceptado o rechazado). Por ejemplo, en la primera fila de la Tabla 20 se tiene 60 neuronas en la capa de entrada y 3 capas ocultas con 90, 60 y 35 neuronas.

Tabla 19: MLP con función de activación Relu, modelo N°2, entrenamiento 70 %, dataset1.

Entrenamiento 70% de data set					Validación cruzada		
Capa oculta	Distribución de neuronas	Epoch (ciclos)	Size	Accuracy	Accuracy	Recall	Precisión
1	c(12,6)	1000	10	52,08 %	55,85 %	0	0
2	c(12,12,6)	1000	10	87,92 %	59,94 %	54,97 %	54,61 %
1	c(20,10)	100	10	59,25 %	61,40 %	74,83 %	54,59 %
1	c(20,10)	200	10	60,38 %	61,99 %	68,21 %	55,68 %
1	c(20,10)	1000	10	59,75 %	63,16 %	75,50 %	56,16 %

*La fila destacada informa la corrida con mejores métricas del algoritmo.

Tabla 20: Experimentos de MLP, modelo N°1, dataset 1.

Entrenamiento 70% de data set						Validación cruzada		
Función activación	Capa oculta	Distribución De neuronas	Epoch (ciclos)	Size	Accuracy	Accuracy	Recall	Precisión
Sigmoid	3	c(60,90,60,35)	100	10	67,67 %	66,08 %	67,72 %	69,95 %
Sigmoid	3	c(60,90,60,35)	500	10	70,06 %	64,33 %	64,05 %	59,39 %
Sigmoid	3	c(60,90,60,35)	1000	10	69,94 %	64,91 %	65,36 %	59,88 %
Sigmoid	3	c(65,40,20,15)	100	10	68,68 %	65,20 %	68,25 %	68,62 %
Sigmoid	3	c(65,40,20,15)	100	10	67,92 %	65,20 %	61,44 %	61,04 %
Tanh	3	c(65,40,20,15)	100	10	68,18 %	64,04 %	52,29 %	61,54 %
Relu	3	c(65,40,20,15)	100	10	99,50 %	56,72 %	54,25 %	51,55 %
Relu, Sigmoid	3	c(60,50,90,15)	100	10	78,87 %	61,99 %	46,41 %	59,66 %
Sigmoid	3	c(80,50,50,15)	100	10	68,93 %	63,74 %	60,13 %	59,35 %
Sigmoid	1	c(65,15)	100	10	65,79 %	65,20 %	49,02 %	64,66 %

*La fila destacada informa la corrida con mejores métricas del algoritmo.

De las mejores instancias realizadas que se presentaron en la Tabla 20, se evaluó la capacidad predictiva bajo los mismos parámetros para el modelo N°2 y modelo N°3 ya que estos modelos contienen una menor selección de atributos (Tabla 21).

De acuerdo a lo observado en la Tabla 22, los análisis realizados con el data set 2 (se incluye el tiempo de pre-prensado) no presentaron eficiencias superiores al data set 1.

Tabla 21: RNA MLP para modelos 2 y 3, dataset 1.

Entrenamiento					Validación cruzada		
Función activación	Capa oculta	Distribución De neuronas	Modelo	Accuracy	Accuracy	Recall	Precisión
Sigmoid	3	c(60,90,60,35)	2	61,76 %	60,53 %	67,97 %	54,74 %
Sigmoid	3	c(60,90,60,35)	3	59,37 %	59,06 %	37,91 %	56,31 %
Relu	3	c(65,40,20,15)	2	84,03 %	56,73 %	62,75 %	51,34 %
Relu	3	c(65,40,20,15)	3	86,67 %	57,31 %	64,71 %	51,83 %
Sigmoid	3	c(80,50,50,15)	2	61,26 %	61,70 %	72,55 %	55,50 %
Sigmoid	3	c(80,50,50,15)	3	59,12 %	58,77 %	27,45 %	58,33 %

Tabla 22: Resumen de las mejores instancias de RNAs MLP para data set 2.

Conjunto de datos	Accuracy	Recall	Precisión	RNAs MLP 245 neuronas con 3 capas ocultas y función de activación Sigmoid
Data set 2	59,82%	42,45%	60,81%	
Data set 2.1	61,11%	49,01%	56,92%	

Comparativa de los mejores resultados de cada algoritmo

En la Tabla 23 se resumen los resultados de los algoritmos evaluados.

Tabla 23: Resumen de las mejores instancias de los algoritmos predictivos evaluados.

Algoritmo	Descripción	Modelo	Accuracy	Recall	Precisión
Árbol de decisión	Decisión Tree Classifier	1	57,6 %	53,12 %	54,84 %
Árbol de decisión	Random Forest	2	61,44 %	66,94 %	62,79 %
SVM	E1071, radial	2	59,03 %	61,11 %	63,64 %
RNA	Simple, logistic, c(10,1)	2	62,8 %	52,27 %	61,61 %
RNA	MLP, Relu, c(12,6)	2	65,9 %	74,26 %	59,06 %
RNA	MLP, Sigmoid, 245 neuronas con 3 capas ocultas	1	66,08 %	67,72 %	69,95 %

Se ha concluido en diversos estudios que las redes neuronales tienen mejores rendimientos predictivos cuando se tienen múltiples variables que describen la variable respuesta (Curteanu y Cartwright 2011, Demirkir *et al*. 2013, Miguel *et al*. 2018). Lo anterior se debe a que presentan parámetros que se pueden variar para obtener mejores métricas de rendimientos, como los que se evaluaron en este estudio: función de activación, cantidad de neuronas, capas ocultas, entre otros. Ello concuerda con lo observado en la Tabla 23 donde se obtuvieron los mejores resultados para RNAs de tipo MLP con una predicción correcta de un 66,08 %.

Este tipo de estudios de ML dependen en gran parte de la calidad de los datos que se poseen, de la variabilidad y el desfase de los tiempos de producción, entre otros (Dogan y Birant 2021). Por tanto, es primordial tener una base de datos de calidad que pueda brindar información correcta y verídica a los algoritmos a fin de que permita explicar la variabilidad del proceso y por ende predecir correctamente la adherencia del tablero.

Para obtener mejores rendimientos en la capacidad predictiva se requiere de una mayor cantidad de registros tanto para el entrenamiento como para la validación del algoritmo. Las variables independientes deben ser representativas de la variable respuesta, adherencia.

CONCLUSIONES

A partir de este estudio, se concluye que las variables independientes del proceso productivo de la etapa de encolado y pre-prensado tienen una gran influencia en la calidad de un tablero contrachapado.

La mayor correlación se obtuvo entre la temperatura de la chapa y la temperatura ambiente. Por ello fue necesario incluir todas estas variables en los modelos estudiados. De acuerdo con los resultados obtenidos, mediante el uso de árboles de decisión se determinó que, en las etapas de encolado y pre-prensado, las variables más influyentes sobre la adherencia son el flujo del adhesivo en las encoladoras, la temperatura ambiente y la temperatura de la chapa.

De los 4 modelos evaluados las RNAs entregaron mejores eficiencias en la predicción de calidad adhesiva. Para el entrenamiento fue necesario aumentar la cantidad de ciclos (epoch). Se obtuvieron buenos resultados al usar las funciones Relu y Sigmoid. La función Sigmoid entregó mejores resultados de acierto global (accuracy

sobre 66 %) y precisión en encontrar resultados correctos (casi 70 %). Al usarla función Relu se obtuvo un mejor recall (sobre el 74 %), lo que muestra su buena aptitud para identificar la realidad.

Los resultados demuestran que los conjuntos de datos de origen industrial son válidos para obtener buenos resultados de predicción del proceso, aun teniendo ciertas variables tomadas fuera de línea tal como la resistencia a la tracción perpendicular del tablero en este estudio. Este trabajo ha permitido mostrar que la aplicación de la metodología de analítica de datos al ámbito industrial tiene plena validez y que el enfoque de ML como herramienta de modelación de las etapas de encolado y pre-prensado es altamente factible. En un enfoque de Industria 4.0, la existencia cada vez mayor de monitoreo en tiempo real de las variables en la industria de tableros está permitiendo la evaluación de variadas técnicas de Inteligencia Artificial. Muchos trabajos están aún por realizarse. Los autores, en conjunto con la industria, están llevando adelante varios avances al respecto, especialmente en pos de disponer sistemas de recomendación en tiempo real o gemelos digitales robustos.

DECLARACIÓN DE AUTORÍAS

C. U. G.:Curación de datos, análisis formal, investigación, metodología, recursos, software, validación. M. R. M.:Conceptualización, adquisición de fondos, investigación, metodología, administración del proyecto, recursos, supervisión.

REFERENCIAS

Aydin, I.; Colakoglu, G. 2005. Formaldehyde emission, surface roughness, and some properties of plywood as function of veneer drying temperature. *Drying Technology* 23(5): 1107-1117. https://doi.org/10.1081/DRT-200059142

Bekhta, P.; Salca, E.A. 2018. Influence of veneer densification on the shear strength and temperature behavior inside the plywood during hot press. *Construction and Building Materials* 162: 20-26. https://doi.org/10.1016/j.conbuildmat.2017.11.161

Bekhta, P.; Sedliačik, J.; Bekhta, N. 2020. Effects of Selected Parameters on the Bonding Quality and Temperature Evolution Inside Plywood During Pressing. *Polymers* 12(5): 1035.https://doi.org/10.3390/polym12051035

Carty, D.M. 2011. An analysis of boosted regression trees to predict the strength properties of wood composites. Master's thesis, University of Tennessee, Knoxville, USA. https://trace.tennessee.edu/utk_gradthes/954

Cheng, Y.J.; Chen, M.H.; Cheng, F.C.; Cheng, Y.C.; Lin, Y.S.; Yang, C.J. 2018. Developing a decision support system (DSS) for a dental manufacturing production line based on data mining. *Applied System Innovation* 1(2): e17. https://doi.org/10.3390/asi1020017

Cupek, R.; Ziebinski, A.; Zonenberg, D.; Drewniak, M. 2018. Determination of the machine energy consumption profiles in the mass-customised manufacturing. *International Journal of Computer Integrated Manufacturing* 31(6): 537-561. https://doi.org/10.1080/0951192X.2017.1339914

Curteanu, S.; Cartwright, H. 2011. Neural networks applied in chemistry. I. Determination of the optimal topology of multilayer perceptron neural networks. *Journal of Chemometrics* 25(10): 527-549. https://doi.org/10.1002/cem.1401

Demirkir, C.; Özsahin, Ş.; Aydin, I.; Colakoglu, G. 2013. Optimization of some panel manufacturing parameters for the best bonding strength of plywood. *International Journal of Adhesion and Adhesives* 46:14-20. https://doi.org/10.1016/j.ijadhadh.2013.05.007

Dogan, A.; Birant, D. 2021. Machine learning and data mining in manufacturing. *Expert Systems with Applications* 166: e114060. https://doi.org/10.1016/j.eswa.2020.114060

Doshi, R.; **Kant Hiran, K.**; **Kumar Jain, R.**; **Lakhwani, K. 2021.** Machine Learning: Master Supervised and Unsupervised Learning Algorithms with Real Examples. BPB Publications: India. 294 pp ISBN: 978-93-91392-352.

Düntsch, I.; Gediga, G. 2019. Confusion Matrices and Rough Set Data Analysis. *Journal of Physics Conference series,* 1229(1), 012055. https://doi.org/10.1088/1742-6596/1229/1/012055

EN. 2007. Plywood - Bonding quality - Part 1: Test Methods. EN 314-1.

Gandhi, K.; Schmidt, B.; Ng, A.H.C. 2018. Towards data mining based decision support in manufacturing maintenance. *Procedia CIRP* 72: 261-265. https://doi.org/10.1016/j.procir.2018.03.076

García-Fernández, F.; de Palacios, P.; Esteban, L.G.; Garcia-Iruela, A.; González-Rodrigo, B.; Menasalvas, E. 2012. Prediction of MOR and MOE of structural plywood board using an artificial neural network and comparison with a multivariate regression model. *Composites Part B: Engineering* 43(8): 3528-3533. https://doi.org/10.1016/j.compositesb.2011.11.054

Hazir, E.; Özcan, T.; Koç, K.H. 2020. Prediction of adhesion strength using extreme learning machine and support vector regression optimized with genetic algorithm. *Arabian Journal for Science and Engineering* 45: 6985-7004. https://doi.org/10.1007/s13369-020-04625-0

Huang, Y.; Pan, C.; Lin, S.; Guo, M. 2018. Machine-Learning Approach in Detection and Classification for Defects in TSV-Based 3-D IC. *IEEE Transactions on Components, Packaging and Manufacturing Technology* 8(4): 699-706. https://doi.org/10.1109/TCPMT.2017.2788896

Kamal, K.; Qayyum, R.; Mathavan, S.; Zafar, T. 2017. Wood defects classification using laws texture energy measures and supervised learning approach. *Advanced Engineering Informatics* 34: 125-135. https://doi.org/10.1016/j.aei.2017.09.007

Köksal, G.; Batmaz, İ.; Testik, M.C. 2011. A review of data mining applications for quality improvement in manufacturing industry. *Expert Systems with Applications* 38(10): 13448-13467. https://doi.org/10.1016/j.eswa.2011.04.063

Kujawińska, A.; Rogalewicz, M.; Muchowski, M.; Stańkowska, M. 2018. Application of cluster analysis in making decision about purchase of additional materials for welding process. In: *Smart Technology. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering.* Torres Guerrero, F.; Lozoya-Santos, J.; Gonzalez Mendivil, E.; Neira-Tovar, L.; Ramírez Flores, P.; Martin-Gutierrez, J. (eds.). 213: 10-20. Springer. https://doi.org/10.1007/978-3-319-73323-4_2

Li, W.; Zhang, Z.; Zhou, G.; Leng, W.; Mei, C. 2020. Understanding the interaction between bonding strength and strain distribution of plywood. *International Journal of Adhesion and Adhesives* 98: 102506. https://doi.org/10.1016/j.ijadhadh.2019.102506

Luque, A.; Carrasco, A.; de las Heras, A. 2019. The impact of class imbalance in classification performance metrics based on the binary confusion matrix. *Pattern Recognition* 91: 216-231. https://doi.org/10.1016/j.patcog.2019.02.023

Manrique, E. 2020. Machine Learning: análisis de lenguajes de programación y herramientas para desarrollo. *Revista Ibérica de Sistemas e Tecnologias de Informação* e28: 586-599. http://www.risti.xyz/issues/ristie28.pdf

Melo, R.R.; Miguel, E.P. 2016. Empregabilidade de redes neurais artificiais (RNA) na predição da quialidade de painéis aglomerados. *Revista Árvore* 40(5): 949-958. https://doi.org/10.1590/0100-67622016000500019

Miguel, E.P.; Melo, R.R.; Serenini Junior, L.; Menezzi, C.H.S.D. 2018. Using artificial neural networks in estimating wood resistance. *Maderas. Ciencia y Tecnología* 20(4): 531-543. https://dx.doi.org/10.4067/S0718-221X2018005004101

Nedelkoski, S.; Stojanovski, G. 2017. Machine learning for large scale manufacturing data with limited information. In: 13th IEEE International Conference on Control & Automation (ICCA). 70-75. https://doi.org/10.1109/ICCA.2017.8003037

Ochoa, L. 2019. Evaluación de Algoritmos de Clasificación utilizando Validación Cruzada. In: 17th LACCEI International Multi-Conference for Engineering, Education, and Technology. 471p. http://dx.doi.org/10.18687/LACCEI2019.1.1.471

OSISOFT. 2020. Visualizar datos de PI System. http://cdn.osisoft.com/learningcontent/pdfs/VisualizingPISystemDataWorkbook_Spanish.pdf

Özşahin, Ş.; Demir, A.; Aydın, İ. 2019. Optimization of Veneer Drying Temperature for the Best Mechanical Properties of Plywood via Artificial Neural Network. *Journal of Anatolian Environmental and Animal Sciences* 4(4): 589-597. https://doi.org/10.35229/jaes.635302

Pang, W.Y.; Qing, J.J.; Liu, Q.L.; Nong, G.Z. 2020. Developing an Artificial Intelligence (AI) system to patch plywood defects in manufacture. *Procedia Computer Science* 166: 139-143.https://doi.org/10.1016/j.procs.2020.02.036

Pavlyshenko, B. 2016. Machine learning, linear and bayesian models for logistic regression in failure detection problems. In: IEEE International Conference on Big Data (Big Data). 2046-2050. https://doi.org/10.1109/BigData.2016.7840828

Poblete, H.; Peredo, M. 1990. Tableros de desechos del debobinado de especies chilenas. *Bosque* 11(2): 45-58. http://revistas.uach.cl/pdf/bosque/v11n2/art05.pdf

Ramos-Maldonado, M.; Aguilera-Carrasco, C. 2021. Trends and Opportunities of Industry 4.0 in Wood Manufacturing Processes. In: *Engineered Wood Products for Construction*. Gong, M. (Ed.). IntechOpen: London, UK. https://doi.org/10.5772/intechopen.99581

Rostami, H.; Dantan, J.Y.; Homri, L. 2015. Review of data mining applications for quality assessment in manufacturing industry: support vector machines. *International Journal of Metrology and Quality Engineering* 6(4): e401. https://doi.org/10.1051/ijmqe/2015023

Suárez, Y.R.; Amador, A.D. 2009. Herramientas de minería de datos. *Revista Cubana de Ciencias Informáticas* 3(3-4): 73-80. https://www.redalyc.org/articulo.oa?id=378343637009

Teihuel, J. 2007. Propuesta de alternativas de solución para el transporte de residuos de madera sólida en la industria de tableros contrachapados. Tesis de Pregrado, Ingeniería en Maderas, Universidad Austral de Chile. http://cybertesis.uach.cl/tesis/uach/2007/fift263p/doc/fift263p.pdf

Tiryaki, S.; Aydın, A. 2014. An artificial neural network model for predicting compression strength of heat treated woods and comparison with a multiple linear regression model. *Construction and Building Materials* 62: 102-108.https://doi.org/10.1016/j.conbuildmat.2014.03.041

Toksoy, D.; Çolakoğlu, G.; Aydin, I.; Çolak, S.; Demirkir, C. 2006. Technological and economic comparison of the usage of beech and alder wood in plywood and laminated veneer lumber manufacturing. *Building and Environment* 41(7): 872-876. https://doi.org/10.1016/j.buildenv.2005.04.012

Urra, C. 2021. Desarrollo de modelo de predicción de calidad que permita la toma de decisiones en la etapa de encolado y pre prensado de un tablero contrachapado. Tesis Magíster Ingeniería Industrial, Universidad del Bío-Bío. Chile.

Vick, C.B. 1999. Adhesive bonding of wood materials. In: *Wood Handbook: Wood as an Engineering Material.* USDA Forest Service, Forest Products Laboratory: Madison, WI, USA. General Technical Report FPL; Chapter 9. GTR-113: 9.1-9.24. https://www.fs.usda.gov/research/treesearch/7139

Wang, K.S. 2013. Towards zero-defect manufacturing (ZDM)-a data mining approach. *Advances in Manufacturing* 1(1): 62-74. https://doi.org/10.1007/s40436-013-0010-9

Yan, H.; Yang, N.; Peng, Y.; Ren, Y. 2020. Data mining in the construction industry: Present status, opportunities, and future trends. *Automation in Construction* 119: 103331. https://doi.org/10.1016/j.autcon.2020.103331

Young, T.M.; Barbu, M.C.; Petutschnigg, A. 2013. The evolution of knowledge in forest products manufacturing. *Pro Ligno* 9(4): 22-27. http://www.proligno.ro/ro/articles/2013/4/Young_keynote_final.pdf

Zavala, D.; Valdivia, R. 2004. Transferencia de calor y su efecto en el proceso de prensado de tableros contrachapados. *Revista Chapingo. Serie Ciencias Forestales y del Ambiente* 10(1): 43-49. https://www.redalyc.org/articulo.oa?id=62910107

Zhang, Y.; Ren, S.; Liu, Y.; Si, S. 2017.A big data analytics architecture for cleaner manufacturing and maintenance processes of complex products. *Journal of Cleaner Production* 142: 626-641. https://doi.org/10.1016/j.jclepro.2016.07.123