

MACHINE LEARNING PARA PREDECIR LA CALIDAD DEL SECADO DE CHAPAS EN LA INDUSTRIA DE TABLEROS CONTRACHAPADOS DE *Pinus radiata*

MACHINE LEARNING TO PREDICT VENEER DRYING QUALITY IN THE *Pinus radiata* PLYWOOD INDUSTRY

Mario Ramos-Maldonado^{1,*}

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

Thays Duarte-Sepúlveda¹

<https://orcid.org/0009-0004-7864-7020>

Francisco Gatica-Neira²

<https://orcid.org/0000-0002-1968-9384>

Diego Venegas-Vásquez¹

<https://orcid.org/0000-0002-7376-6272>

RESUMEN

El Aprendizaje automático o Machine learning es una herramienta que está siendo utilizada para optimizar procesos industriales de alta complejidad. En la industria de producción de paneles contrachapados, el secado de chapas es uno de los procesos más importantes ya que permite obtener productos con alta calidad. La naturaleza biológica y alta variabilidad estructural de la madera hace que su procesamiento industrial sea multivariado y difícil de controlar. La gran cantidad de variables presentes y la posibilidad cada vez más frecuente de medirlas en tiempo real están permitiendo la disponibilidad de una gran cantidad de datos. En la actualidad, el enfoque basado en datos y las técnicas de Inteligencia Artificial, específicamente Machine learning, pueden permitir enfoques robustos de predicción y de control. En la industria de procesos con altos niveles de automatización es posible viabilizar la toma de decisiones para predecir la calidad del producto, monitoreando las variables de control explicativas. El objetivo de este trabajo fue evaluar el comportamiento de tres algoritmos de Machine learning para predecir la calidad del proceso de secado de chapas a partir de un número considerable de variables de entrada capturadas de un proceso industrial real. Se utilizó la plataforma Weka y código Python. Se evaluaron los algoritmos: K-Nearest-Neighbor, eXtreme Gradient Boosting y Support Vector Machine. Se realizó reducción de variables y dimensionalidad de análisis de correlación y de componentes principales. Los resultados demostraron que eXtreme Gradient Boosting logró una precisión del 76 % en la predicción de resultados de calidad. Finalmente, se concluye que la metodología de ingeniería de datos y los algoritmos fueron eficientes para predecir los datos industriales.

Palabras claves: Algoritmos de aprendizaje supervisado, calidad de la madera, modelo predictivo, secado de chapas, optimización de procesos.

¹Universidad del Bío-Bío. Facultad de Ingeniería. Departamento de Ingeniería en Maderas. Grupo de Investigación en Sistemas Inteligentes. Concepción, Chile.

²Universidad del Bío-Bío. Facultad de Ciencias Empresariales. Departamento de Economía y Finanzas. Concepción, Chile. Grupo de Investigación Socioeconómica Medioambiental (ISMA) (GI2320140). Concepción, Chile

*Autor de correspondencia: mramos@ubiobio.cl

Recibido: 31.10.2023 Aceptado: 30.06.2024

ABSTRACT

Machine learning is a tool that is being used to optimize highly complex industrial processes. In the plywood panel production industry, veneer drying is one of the most important processes as it allows to obtain high quality products. The biological nature and high structural variability of wood turns its industrial processing multivariate and difficult to control. The large number of variables present and the increasing possibility to measure them in real time are enabling the availability of a large amount of data. Nowadays, data-driven approach and Intelligence Artificial techniques, specifically Machine Learning can enable robust prediction and control approaches. In the process industry, with high levels of automation, it is possible to enable decision making to predict product quality by monitoring explanatory control variables. The objective of this work was to evaluate a Machine Learning algorithm capable of predicting the quality of the veneer drying process from a considerable number of input variables captured from a real industrial process. The Weka platform and Python code were used. Three algorithms were evaluated: K-Nearest-Neighbor, eXtreme Gradient Boosting and Support Vector Machine. Variable and dimensionality reduction, correlation analysis and Principal Component Analysis were performed. The results showed that eXtreme Gradient Boosting achieved an accuracy of 76 % in predicting quality scores. Finally, it is concluded that both the data engineering methodology and the evaluated algorithms were efficient in predicting industrial data.

Keywords: Machine learning algorithms, wood quality, predictive model, veneer drying, process optimization.

INTRODUCCIÓN

En un contexto de industria del futuro, la ingeniería de datos resulta clave para la toma de decisiones. La disponibilidad actual de capacidad de cómputo y de almacenamiento viabilizan el uso de internet industrial (Internet of Things, IoT) y el tratamiento de datos en tiempo real en procesos industriales. Los algoritmos de análisis basados en Inteligencia Artificial (IA) están haciendo posible la toma de decisiones en línea y la predicción de fenómenos difíciles de modelar con técnicas convencionales. La disponibilidad de sensores y técnicas de máquinas de aprendizaje automático (Machine Learning, ML) permiten la captura, despliegue y análisis inteligente de datos capaz de asociarse a estrategias modernas de gestión como el mantenimiento prescriptivo (Wu *et al.* 2007).

Así, la problemática científica y tecnológica radica en encontrar los algoritmos apropiados para un análisis inteligente de los datos, esto es algoritmos capaces de integrar las variables del proceso, aprender en función de la experiencia de los expertos tomadores de decisión y seguir “aprendiendo” de los muchos datos capturados. En la industria, la gran disponibilidad de datos y el uso de Manufacturing Execution Systems (MES) están facilitando una mayor comprensión del proceso productivo (Mihel *et al.* 2024).

El constante avance tecnológico e informático y la introducción de nuevas herramientas para medición y control hacen que sea cada vez mayor la cantidad de datos recopilados durante los procesos industriales (Gatica-Neira y Ramos-Maldonado 2020). Esta cantidad de información genera varios inconvenientes tales como tiempo extenso de procesamiento de datos, valores perdidos durante el análisis y ruido excesivo por la variabilidad de estos (Dogan y Birant 2021). Por otra parte, los modelos analíticos basados en la fenomenología del proceso son difíciles, costosos y llevan mucho tiempo en desarrollar (Ellis *et al.* 2020). En contraste, el enfoque basado en datos y las técnicas de inteligencia artificial (IA) tienen la capacidad de desarrollar modelos sustitutos con muy buenas aproximaciones (Tabassum *et al.* 2023). Entre estas técnicas destacan los algoritmos de Machine Learning (ML), los cuales consisten en una rama de la IA que se utilizan para el descubrimiento de conocimientos “ocultos” de grandes volúmenes de datos, sean estos patrones, correlaciones, o anomalías. ML tiene la capacidad de aprender y adaptarse a nuevas situaciones (Shalev-Shwartz y Ben-David 2014). En un enfoque de ML, la calidad del resultado depende, entre otros, de la cantidad de ejemplos de entrenamiento. ML no se restringe al tipo y el comportamiento de las variables o a la relación entre la respuesta y variable de control (Lu *et al.* 2007). El enfoque ML también ofrece mayor precisión en la predicción de acciones futuras, ya que, a partir de datos históricos, conociendo experiencias positivas y negativas, se pueden modelar comportamientos futuros, y sacar conclusiones para situaciones posteriores (Frey *et al.* 2019).

En resumen, ML aporta las siguientes ventajas: considera el fenómeno como caja negra y se concentra en las entradas y salidas del sistema, permite una conexión directa con bases de datos para aplicaciones reales, logra un tratamiento de gran volumen de datos y permite una capacidad de aprender y adaptarse a nuevos contextos (Mobarak *et al.* 2023).

Los enfoques ML se clasifican en dos categorías principales: aprendizaje supervisado (Parvin *et al.* 2013) y aprendizaje no supervisado (Minaei-Bidgoli *et al.* 2014). La clasificación supervisada se refiere al uso de software de procesamiento que es guiado por el usuario para especificar categorías en las que se deben clasificar los datos. El agente observa pares de datos, de entrada y salida, a modo de ejemplo para aprender una función que modele la salida según la entrada; por lo tanto, en los datos utilizados para construir el modelo se encuentra la información que se desea predecir (Russell y Norvig 2021). La clasificación no supervisada es una metodología en la cual el usuario no proporciona clases de muestra para el análisis de resultados. Luego, el software determina qué datos están relacionados y los agrupa en clases (Pillay *et al.* 2021).

El aprendizaje supervisado puede ser abordado a partir de problemas de regresión y problemas de clasificación. La regresión resuelve relaciones entre variables dependientes e independientes, y se utiliza para hacer proyecciones tales como ingresos por ventas de un negocio (Urrea-González y Ramos-Maldonado 2023). La clasificación utiliza algoritmos para asignar datos de prueba en categorías específicas. Reconoce entidades dentro de un conjunto de datos e intenta sacar conclusiones sobre cómo esas entidades deben etiquetarse o definirse (Rokach 2010). Para abordar esta temática se utilizan técnicas de redes neuronales, máquinas de vectores de soporte y árboles de decisión, entre otros (Parvin *et al.* 2015).

La literatura científica reporta soluciones basadas en datos para resolver diversos problemas industriales (Diez-Oliván *et al.* 2019). La mayoría de las aplicaciones están basadas en el uso de algoritmos de ML como máquinas de soporte vectorial (Support Vector Machines SVM), algoritmos de aprendizaje supervisado (Cortes y Vapnik 1995) y Random Forest. Si bien originalmente fueron diseñadas para resolver problemas de clasificación lineales y binarios, actualmente su uso se ha generalizado a problemas no lineales, tanto en predicción como en clasificación (Abe 2010, Mehra y Gupta 2013). Por ejemplo, Chen y Guestrin (2016) han utilizado SVMs para clasificar imágenes entre diferentes categorías. Mientras que Gosselin *et al.* (2014) no sólo distinguen entre categorías, sino también entre intracategorías utilizando Fisher Vectors y SVM. Turhan y Serdar (2013) diferencian entre distintas especies de árboles, mientras que Koskinen *et al.* (2013) clasifican automáticamente la calidad de la madera. Por otro lado, Boosting es un tipo de meta-algoritmo de aprendizaje automático que convierte un grupo de clasificadores débiles en un único clasificador fuerte (Freund y Schapire 1999). Es interesante mencionar que, aunque las SVMs son clasificadores fuertes (strong classifiers), también pueden ser utilizadas como clasificadores débiles en conjunto con técnicas de boosting (Alfaleh *et al.* 2023).

Los tableros contrachapados son materiales muy demandados en la industria de la construcción, ya que tienen peso ligero, brindan alta resistencia, estabilidad de forma, propiedades de aislamiento acústico y baja conductividad térmica (Gradov *et al.* 2022). Su proceso de fabricación cuenta con un diverso número de variables de entrada y salida en cada una de sus etapas, entre las que se pueden mencionar (Teihuel 2007): recepción y almacenamiento, macerado, debobinado, secado, encolado, armado, prensado y acondicionamiento. Las propiedades de la madera son sensibles frente a las variables de proceso, por lo que resulta complejo encontrar una función que explique las características del producto final. La madera por naturaleza es anisotrópica, es decir, exhibe distintas propiedades en diferentes direcciones dentro de una pieza. Además, es higroscópica, es decir que tiene la capacidad de absorber y liberar humedad del entorno, lo cual hace que pueda sufrir cambios dimensionales y deformaciones en función de la humedad y la temperatura ambiente (Duarte 2023). Adicionalmente, su composición química es variable, por ejemplo, en maderas blandas como el pino, el contenido de celulosa oscila entre el 40 % y el 44 %, hemicelulosa entre 25 % - 29 %, lignina entre 25 % - 31 %, y extractos entre 1 % - 5 %. Para maderas duras como por ejemplo, el eucalipto, el contenido de celulosa varía entre el 43 % y el 47 %, la hemicelulosa entre 25 % - 35 %, la lignina entre 16 % - 24 %, y el contenido de extractos entre 2 % - 8 % (Venegas-Vásconez *et al.* 2023).

La fabricación de tableros contrachapados requiere varias etapas, siendo el secado de chapas una etapa fundamental. Este proceso consiste en eliminar un determinado contenido de humedad de la madera para que sea compatible con el sistema adhesivo durante el encolado (Kehr 2007). Por otro lado, el contenido de humedad presente en las chapas afecta directamente las propiedades físico-mecánicas del tablero, tales como resistencia y estabilidad dimensional (Aydin 2014). El contenido de humedad óptimo varía según el tipo de madera, la composición del adhesivo, el tipo de panel y otros factores (Gradov *et al.* 2022). Demirkir *et al.* (2013) afirman que el contenido de humedad óptimo en tableros contrachapados debe ser inferior al 7 %. Aydin y Colakoglu (2005) afirman que el secado es responsable de hasta el 70 % del consumo total de energía en la producción de tableros contrachapados. Jia *et al.* (2019) estimaron que el secado por sí solo contribuye con más del 30 % al potencial de calentamiento global en la producción total de contrachapado en China, que en 2018 fue equivalente a 6,3 toneladas de CO₂ emitidas por metro cúbico de producto terminado. Por lo tanto, la optimización del proceso de secado es un tema de considerable interés tanto desde el punto de vista económico como ambiental.

El equipo de secado consiste en bandas transportadoras que llevan la chapa de madera a una determinada velocidad la cual es determinada por las condiciones iniciales de la materia prima (Moisan 2007). En este movimiento se transfiere calor a la chapa, cuidando que ésta se mantenga plana y sin deformaciones para lo cual se cuida la velocidad de transferencia de calor, y la difusión de humedad (Lutz 1978). Durante el proceso de secado, la humedad de la madera se elimina y la madera se contrae; esta contracción puede generar fuerzas o tensiones internas que actúan en diferentes direcciones. Las tensiones de secado pueden generar la formación de defectos como deformaciones, agrietamiento, entre otros (Rice 1988). Los factores que inciden en la calidad del secado se muestran en la Figura 1.

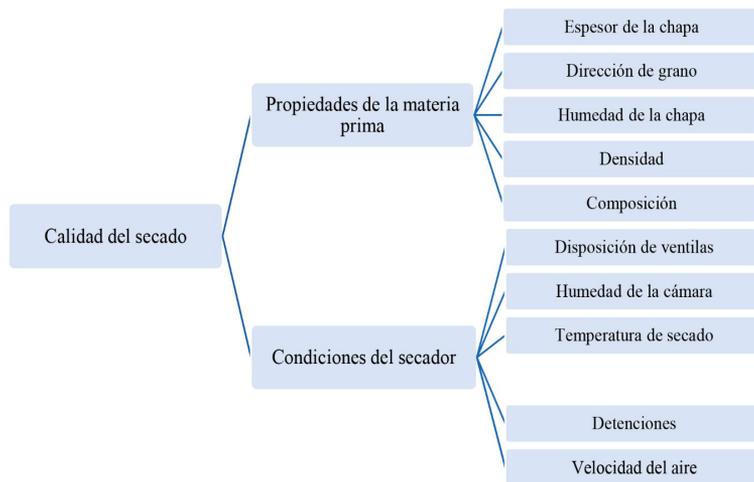


Figura 1: Factores que inciden en la calidad del secado. Adaptado de Navarrete (2020).

El secado de la madera es un proceso donde se controlan cuidadosamente la temperatura, la humedad y la circulación del aire. El propósito principal es reducir la humedad en la madera hasta alcanzar un nivel adecuado para su uso (Navarrete 2020).

El modelado y la predicción del secado de chapas usando datos industriales no ha sido mayormente tratado usando herramientas de IA. Esto se explicaría por la falta de disponibilidad de datos y su captura masiva, lo cual estaría ocurriendo actualmente por la entrada en operación de sistemas MES y la mayor capacidad de almacenamiento, generalmente en la “nube” Navarrete (2020), Ramos-Maldonado y Aguilera-Carrasco (2021) y Duarte (2023).

El uso de IA es cada vez más recurrente en estudios para mejoramiento de la calidad del producto final en la industria de la madera (Troncoso-Espinosa *et al.* 2021). Por ejemplo, Ramos-Maldonado y Aguilera-Carrasco (2021) han abordado las tendencias y oportunidades que brinda la Industria 4.0 en los procesos de fabricación de madera. Gradov *et al.* (2022) desarrollaron un modelo de secado continuo de chapas de madera basado en balances de masa y energía, optimizando la energía requerida en el proceso a través de ANOVA. Demir (2023) utilizó redes neuronales artificiales (RNA) para determinar el efecto en la resistencia de unión de paneles contrachapados de la resina de urea-formaldehído (UF) mezclada con taninos obtenidos de bellotas de roble turco (*Quercus aegilops* L.). Aquí, se determinaron modelos de predicción con el mejor rendimiento mediante comparaciones estadísticas y gráficas entre datos experimentales y valores de predicción obtenidos del análisis de RNA. Özşahin *et al.* (2019) determinaron con la ayuda de RNA, la temperatura óptima de secado de paneles contrachapados sin disminuir sus propiedades mecánicas. Encontraron que la temperatura óptima de secado era 165 °C, 162 °C y 161 °C para el contrachapado de pino escocés (*Pinus sylvestris* L.) para obtener las mejores propiedades de resistencias al corte, resistencias a la flexión y módulos de elasticidad, respectivamente. Urra-González y Ramos-Maldonado (2023) utilizaron ML para predecir la adherencia en condiciones de operación industrial durante la etapa de encolado y prensado de tableros de madera.

A pesar de la creciente importancia que ha tenido la IA en los procesos industriales, no existe más información sobre su uso en los procesos de secado de tableros contrachapados. En este trabajo, se evalúan algoritmos de ML, tales como eXtreme Gradient Boosting (XGBoost), Support Vector Machine (SVM) y

K-Nearest-Neighbor (KNN), para predecir la calidad del secado de chapas en la industria de paneles contrachapados de pino radiata (*Pinus radiata* D. Don). De acuerdo con lo expuesto, esta investigación contribuye a los estudios existentes ya que aporta a la modelación del problema prediciendo la calidad del tablero utilizando datos industriales.

MATERIALES Y MÉTODOS

Materiales

Se utilizó un computador Asus equipado con un procesador Intel Core i9-13900 KF con 32 núcleos, 3.0 Ghz. de velocidad de procesamiento y 65 Gb. de memoria RAM. Para la elaboración de los algoritmos que analizaron los datos se utilizó Python y Visual Studio Code (Manrique-Rojas 2020). Los lenguajes de programación para el análisis de datos fueron ML, Python (Versión 3.11.4) (2023) y Weka (Versión 3.9.6) (2022) y las librerías Numpy, Scikit-Learn, TensorFlow, Keras, Pandas y Matplotlib.

Los datos fueron obtenidos de un proceso industrial de fabricación de tableros contrachapados de pino radiata (*Pinus radiata* D. Don). Se trata de una planta que produce aproximadamente 350 mil m³ de tableros al año. El origen de los trozos se cataloga como trozos de bosques podados o no podados. Los datos son obtenidos desde los PLC de la planta vinculados al secador continuo de chapas.

La especie utilizada en la producción de tableros contrachapados es pino radiata (*Pinus radiata* D. Don) proveniente de la zona centro-sur de Chile. En la planta industrial donde se obtuvieron los datos se dispone de secadores continuos de chapas. El ingreso de las chapas a los secadores se produce desde diferentes entradas a través de sistemas transportadores. En este proceso la distribución de la temperatura es uniforme en cada chapa, con lo que se garantiza un secado homogéneo y controlado.

Métodos

Etapas del proceso y recopilación de datos

En la Figura 2 se muestran las etapas del proceso y recopilación de datos industriales para el proceso de secado. Los datos son los elementos de mayor importancia en el aprendizaje automático e inteligencia artificial, pues sin estos es imposible entrenar un modelo y como consecuencia, la investigación y automatización a realizar no tendrían sentido. Con respecto a la función que cumplen en dicho modelo se tienen tres tipos de datos: de entrenamiento, de validación y de prueba (Navarrete 2020). Los datos provenientes del servidor OPC fueron recopilados y almacenados en un servidor PI System© a través del complemento PI DataLink de Excel. Los datos capturados desde diferentes sensores fueron vehiculados por PI System© a su repositorio.

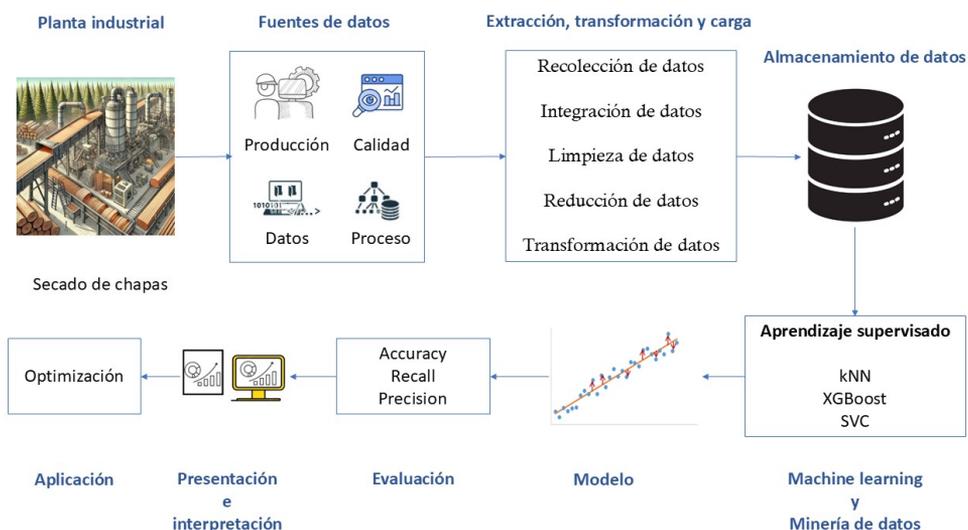


Figura 2: Implementación de la metodología KDD para el proceso de secado. Adaptado de (Dogan y Birant 2021).

Las variables de entrada fueron:

Origen de la troza (Or): se refiere al origen desde un bosque tratado o no, esto es si es un bosque tratado silviculturalmente (o podado) o no.

Espesor (E), en mm: se trata del espesor de las chapas producto del debobinado. Los espesores normales son 2,6 y 3,2 mm.

Secado: dividido en tres zonas (subíndices $_{1,2}$ y $_3$); en cada una de estas zonas se consideraron temperatura (T), contenido de humedad (CH, en %), porcentaje de apertura de las ventilas (%Ap) y sus respectivas desviaciones estándar (σT , σCH y $\sigma \%Ap$).

Temperatura y presión de vapor (T_v , P_v , respectivamente) y sus desviaciones estándar (σT_v y σP_v).

Un extracto de la base de datos mencionada se muestra en la Tabla 1. Cada línea de datos corresponde a un registro. El subíndice ($_{1,2,3}$) se refiere a cada una de las tres zonas del secador.

Tabla 1: Extracto de datos depurados.

Or	E	T_1	σT_1	CH_1	σCH_1	$\%Ap_1$	$\%Ap_1$	T_v	σT_v	P_v	σP_v	Crit $_1$	Crit $_2$
1	3,2	139,55	1,64	287,67	20,21	1,86	0,01	241,90	0,51	18,44	0,30	ExcesoCH	0
1	3,2	139,85	1,49	288,56	21,71	1,86	0,01	242,39	0,48	18,43	0,34	Optimo	1
1	3,2	139,69	1,63	284,67	24,93	1,86	0,01	242,53	0,40	18,54	0,37	Optimo	1

Las variables de salida o de respuesta que corresponden a las etiquetas de cada registro fueron consideradas al contenido de humedad de salida de la chapa de acuerdo a las siguientes categorías:

Exceso CH: corresponde al valor de humedad promedio de la chapa que supera el valor permitido de humedad.

Optimo: es el valor de humedad que se encuentra en el rango permitido.

Deficit CH: es el valor de humedad que se encuentra en un valor más bajo que el permitido.

Limpieza y reducción de datos

La limpieza de datos sirvió para eliminar valores atípicos como paradas de producción, valores perdidos y errores en la toma de datos de los sensores. Se completaron datos faltantes a través de técnicas de imputación de datos tales como valores medios, eliminación de columnas y realización de regresión. Esta depuración e imputación resultó en la eliminación de valores perdidos 'NaN'. En la Tabla 2 se muestra la cantidad de registros (también llamados instancias) en la depuración de datos del proceso de secado (número total de registros 7862). Se observa un cierto desequilibrio entre clases ya que la clase 0, Exceso CH, muestra una presencia de solo 5 % del total de registros. Esto es un inconveniente al momento de evaluar algoritmos robustos y es una limitación que se da al trabajar con datos industriales, debido a los procesos de innovación y aprendizaje acumulados en la empresa.

Tabla 2: Variables e instancias de la data set.

Variable	Clase	Instancias
ExcesoCH	0	395 (5 %)
Óptimo	1	4877 (62 %)
DeficitCH	2	2587 (32,9 %)
Total		7862 (100 %)

Los datos fueron divididos en conjuntos de entrenamiento (70 %) y prueba (30 %), utilizando la técnica `train_test_split` (Uniyal *et al.* 2022). Se realizó una normalización y estandarización de los datos utilizando `StandardScaler` para asegurar que todas las variables estén en una misma escala y tengan el mismo peso durante el análisis. Para la normalización de datos se utilizó la Ecuación 1 (Lagos-Biolley 2023):

$$X_{\text{norm}} = \frac{X - X_{\text{min}}}{X_{\text{max}} - X_{\text{min}}} \quad (1)$$

Donde: X_{norm} : dato normalizado entre 0 y 1, X : dato a normalizar, X_{min} : dato mínimo, X_{max} : dato máximo.

Finalmente, se utilizó la matriz de correlación para una posible reducción de variables y el Análisis de Componentes Principales (ACP) para la reducción de dimensionalidad.

Análisis de componentes principales y matriz de correlación

Se realizó un Análisis de Componentes Principales (ACP), el cual es un método de compresión de datos no supervisado basado en la correlación entre variables, que transforma un gran número de variables potencialmente correlacionadas en un pequeño número de factores no correlacionados (ortogonales) llamados componentes principales (Tharwat 2016), mientras que la información presente en el conjunto de datos inicial se conserva tanto como sea posible. De hecho, los ACP se construyen como una combinación lineal simple de las variables originales y son completamente no correlacionados. El ACP también puede eliminar la información irrelevante (es decir, ruido) presente en el conjunto de datos (Sadrara y Khorrami 2023).

Se utilizó la matriz de correlación para una posible reducción de variables. La matriz de correlación debe proporcionar información sobre las relaciones lineales entre las variables del conjunto de datos. Los valores que tengan una correlación cercana a 1 o -1 pueden ser valores que tienen una alta correlación y, por lo tanto, pueden no considerarse en el análisis final, dos Santos Freitas *et al.* (2022), Zhao *et al.* (2020).

Minería de datos

Las técnicas de minería de datos proporcionan una herramienta eficaz para descubrir conocimiento previamente desconocido, no trivial, útil y necesario para tomar decisiones (Penkova 2017). Se utilizó la metodología Knowledge Discovery in Databases (KDD) para la extracción automatizada a partir de grandes volúmenes de datos (Duarte 2023). KDD es una metodología secuencial y aplica iteraciones hasta obtener la información deseada. Los algoritmos de aprendizaje supervisado ML utilizados para modelar y predecir la calidad parcial del proceso se muestran en la Tabla 3.

Tabla 3: Algoritmos evaluados y sus hiperparámetros.

Algoritmo	Simbología	Fundamento	Referencia
K-Nearest Neighbors	KNN	Utilizado para clasificación y regresión. Asigna etiquetas de clase a nuevos puntos de datos considerando la clase mayoritaria de sus k vecinos más cercanos. El valor de k determina la complejidad de los límites de decisión. Se adopta ampliamente debido a su facilidad de comprensión e implementación.	Cover y Hart, (1952), Uniyal <i>et al.</i> (2022)
Support Vector Machine	SVM	Utilizado para clasificación y regresión. El objetivo es identificar un hiperplano óptimo que maximice el margen entre puntos de datos que pertenecen a diferentes clases. Puede manejar relaciones lineales y no lineales en los datos utilizando funciones kernel.	Foody y Mathur (2004)
Extreme Gradient Boosting	XGBoost	Se puede construir modelos predictivos a partir de árboles de decisión, integrando datos débiles. Un enfoque de XGBoost por etapas, realizando árboles subsiguientes que rectifican los errores cometidos por los árboles anteriores. Durante el entrenamiento, XGBoost utiliza técnicas de optimización de descenso de gradiente para minimizar efectos específicos de pérdida.	Chen y Guestrin (2016), He <i>et al.</i> (2018)

Evaluación de modelos

Para evaluar el rendimiento de los modelos se utilizó una matriz de confusión (Luque *et al.* 2019) la cual compara las predicciones realizadas por el modelo con los valores reales de las etiquetas de clase (Tabla 4). Los verdaderos positivos (VP) son las instancias que se clasificaron correctamente como positivas, los falsos positivos (FP) son las instancias que se clasificaron incorrectamente como positivas, falsos negativos (FN) las instancias que se clasificaron incorrectamente como negativas y verdaderos negativos (VN) las instancias que se clasificaron correctamente como negativas (Nakamura 2023).

Tabla 4: Matriz de confusión para clasificación binaria. Adaptado de (Dütsch y Gediga 2019).

		Predicción	
		P	N
Real	P	VP	FN
	N	FP	VN

Encontrados los valores de la Tabla 4, se calcularon las métricas de evaluación (Lu *et al.* 2023). Se determinó la exactitud o “ACCURACY” que es igual a la proporción de predicciones que el modelo clasificó correctamente (Ecuación 2), la precisión o “PRECISION” que es lo cerca que está el resultado de una predicción del valor verdadero (Ecuación 3), y la exhaustividad o “RECALL” que representa la tasa de verdaderos positivos, o la proporción entre los casos positivos bien clasificados del modelo, respecto al total de positivos (Ecuación 4) (Bressan *et al.* 2020).

$$ACCURACY = \frac{VP + VN}{VP + FN + FP + VN} \quad (2)$$

$$PRECISION = \frac{VP}{VP + FN} \quad (3)$$

$$RECALL = \frac{VP}{VP + FP} \quad (4)$$

RESULTADOS Y DISCUSIÓN

Análisis de componentes principales y matriz de correlación

La Matriz de Correlación (Figura 3) muestra los coeficientes de correlación entre las variables resultantes del Análisis de Componentes Principales (ACP), y varían entre -1 (correlación negativa perfecta) y 1 (correlación positiva perfecta), mientras que 0 señala ausencia de correlación (Easaw *et al.* 2023). En la Figura 3 se aprecia la alta correlación entre T_1 vs T_3 y σT_1 vs σT_3 (0,88 y 0,83, respectivamente) con lo que es posible descartar una de estas variables correlacionadas. El criterio usado fue la correlación mayor a 0,8. Se pudo comprobar que hasta este valor de correlación (0,8), la reducción de variables no afecta la variabilidad del proceso para los indicadores de performance (ACCURACY, PRECISION y RECALL). Por ello, la alta correlación permite remover T_3 y σT_3 . Con respecto a las variables restantes, su correlación fue baja, lo que indica que son independientes entre sí, razón por la cual no son removidas de la data.

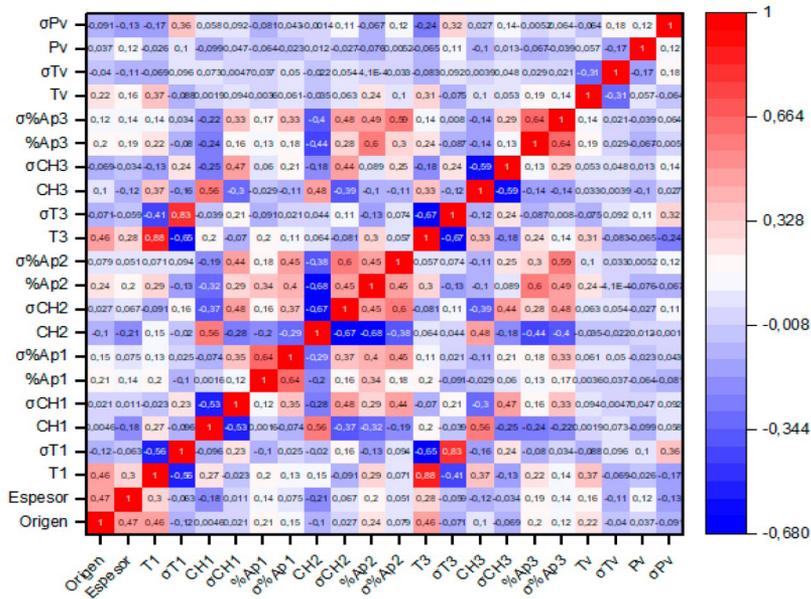


Figura 3: Matriz de correlación de variables generado por la plataforma Weka (Versión 3.9.6).

Los componentes principales fueron ordenados en función de su importancia, con el fin de explicar la variabilidad en los datos, a partir de examinar los coeficientes de carga de los atributos del informe ACP. Los resultados ACP fueron analizados con varianza acumulada y se escogió un valor de 90 % para este parámetro. Se obtuvo que el primer componente principal explicó el 64,6 % de la varianza acumulada, el segundo componente principal el 16,3 %, y el tercer componente principal el 10,4 %. Por tanto, con solo tres componentes se logra explicar el 90,3 % de la varianza acumulada en los datos.

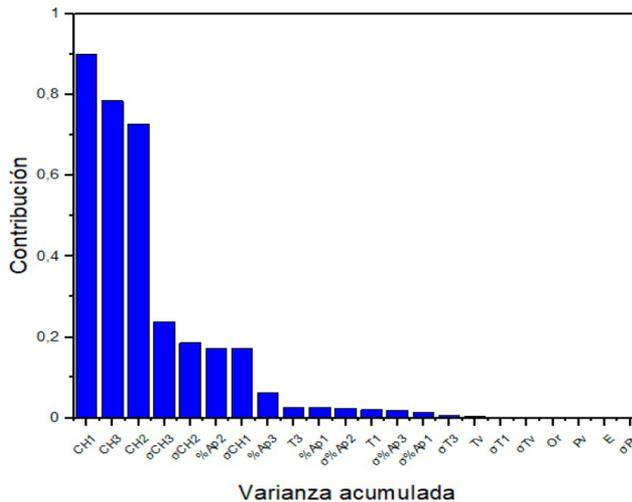


Figura 4: Varianza acumulada para el ACP.

La desagregación de la incidencia individual de cada variable sobre la varianza del conjunto de datos muestra la naturaleza multivariada del problema (Figura 4). Este resultado es coherente con el fenómeno de secado, pues la humedad de cada una de las zonas, la apertura de ventilas y las temperaturas T_1 y T_3 son las variables de mayor importancia, quedando relegadas la presión de vapor, el origen de las trozas y el espesor de la chapa. La apertura de ventilas $\%Ap_2$ y $\%Ap_3$, tienen mayor influencia en los resultados que $\%Ap_1$. Esto se debe a que en esta zona se produce la evacuación del agua saturada de la chapa, generando alta humedad. La variación de $\%Ap_1$ (menor o mayor apertura) no es significativa en el conjunto total como $\%Ap_2$ y $\%Ap_3$. Por

otro lado, como en la zona 2 del secador se produce la evacuación del agua libre desde el interior de la chapa, por lo que la apertura %Ap₂ muestra mayor incidencia en los resultados. Esto puede explicarse por ser la zona donde una mayor cantidad de agua es evacuada desde los lúmenes de la madera (Duarte 2023).

Evaluación de algoritmos

Los algoritmos KNN, SVM y XGBoost se aplican sobre la base de datos después de análisis de depuración y componente principales.

KNN

El algoritmo KNN es un método de aprendizaje no paramétrico de clasificación y regresión, que asigna pesos a las contribuciones de los vecinos, de modo que los más cercanos contribuyen más a la media que los más distantes. Los vecinos se obtienen a partir de un conjunto de objetos utilizados en el entrenamiento del algoritmo (Wang *et al.* 2022). En la Figura 5 se muestra la proporción de predicciones que el modelo clasificó correctamente "ACCURACY" en función del parámetro K utilizando el algoritmo KNN. El mejor rendimiento se encontró cuando K fue igual a 1. Valores de K más pequeños pueden causar sobreajuste y sensibilidad a valores atípicos.

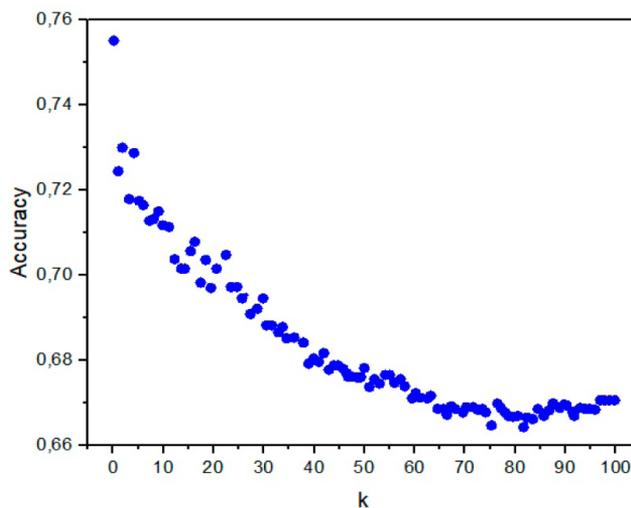


Figura 5: Mejor k del algoritmo KNN vs ACCURACY.

En la Tabla 5 se muestran los resultados del algoritmo KNN para el conjunto de datos. Con un valor de K igual a 1, se logró un ACCURACY = 69 %, RECALL= 56 %, precisión= 59 % y F1-Score 0,57. A medida que K aumenta, la precisión o "ACCURACY" disminuye, lo cual sugiere que el modelo tiene dificultades para recuperar las instancias positivas a medida que K se incrementa.

Tabla 5: Resultado del algoritmo KNN.

K	ACCURACY	RECALL	Precisión	F1-Score
1	0,69	0,56	0,59	0,57
10	0,67	0,48	0,63	0,51
40	0,66	0,44	0,64	0,46
70	0,66	0,43	0,64	0,45
90	0,66	0,41	0,62	0,41
100	0,66	0,41	0,65	0,40

La Figura 6 muestra la matriz de confusión del algoritmo KNN con $K=1$. El modelo obtuvo un rendimiento global del 69 %, con mejoras notables en las clases 0 y 2. La categoría 1 destacó con una precisión del 75 % y un RECALL del 79 %, logrando un equilibrado F1-score del 77 %. Aunque la categoría 0 tuvo una precisión del 39 % y un RECALL del 30 %, el F1-score fue del 34 %. La clase 2 mostró una precisión del 62 % y un RECALL del 58 %, resultando en un F1-score del 60 %. En resumen, el modelo mejoró su capacidad predictiva global, especialmente en la categoría 1, con un promedio macro F1 del 57 % y un F1-score ponderado del 69 %.

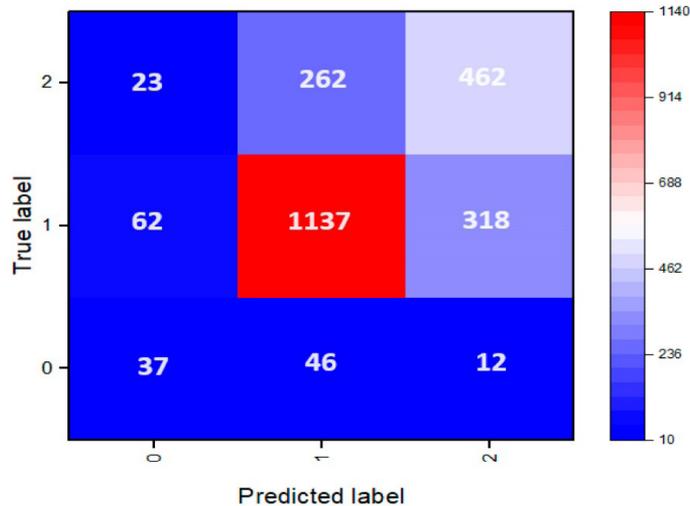


Figura 6: Matriz de Confusión del algoritmo KNN con $K=1$.

XGBoost

Se empleó el conjunto de datos con el algoritmo XGBoost, dividiéndolo en un 30 % para entrenamiento y configurando `Random_state=42` para resultados reproducibles. Se exploraron hiperparámetros como `'Learning_rate'` (0,01; 0,1; 0,2), `'n_estimators'` (3, 5, 7), y `'max_depth'` (100, 200, 300) mediante `grid_search`. Los valores óptimos encontrados fueron `'Learning_rate=0,1'`, `'Max_depth=7'`, y `'N_estimators=300'`. Se realizaron múltiples pruebas ajustando estos hiperparámetros para mejorar el modelo, cuyos resultados se muestran en la Tabla 6. Después de ajustar los hiperparámetros, se evaluaron las métricas para medir el rendimiento del modelo. Estas métricas proporcionan información sobre el desempeño del modelo en términos de aciertos, capacidad de identificación de instancias positivas y el equilibrio entre PRECISION y RECALL.

Tabla 6: Pruebas de hiperparámetros del algoritmo XGBoost.

Pruebas	Learning_rate	Max_depth	N_estimators	Accuracy	Precision	RECALL	F1-Score
1	0,1	7	300	0,76	0,72	0,58	0,61
2	0,01	3	100	0,66	0,59	0,44	0,46
3	0,01	5	100	0,69	0,63	0,48	0,50
4	0,1	5	100	0,72	0,69	0,54	0,57
5	0,1	7	100	0,74	0,69	0,55	0,58
6	0,1	7	200	0,75	0,69	0,56	0,59
7	0,2	5	200	0,74	0,69	0,57	0,60
8	0,2	5	300	0,75	0,59	0,57	0,60

En la Figura 7, la matriz de confusión corresponde a la mejor prueba realizada, donde se llevaron a cabo ajustes de los hiperparámetros Learning_rate=0,1, Max_depth=7, N_estimators=300 que arrojó un ACCURACY de 0,76. En resumen, el algoritmo XGBoost demostró un rendimiento aceptable en la clasificación de las clases, con un enfoque particularmente fuerte en la clase 1. Aunque algunas clases presentaron desafíos en términos de RECALL, en general, el modelo fue capaz de proporcionar predicciones sólidas en la mayoría de las instancias.

Al comparar las matrices de confusión el XGBoost trata mejor la superposición de las clases 1 y 2, ajustando la predicción de la clase 1.

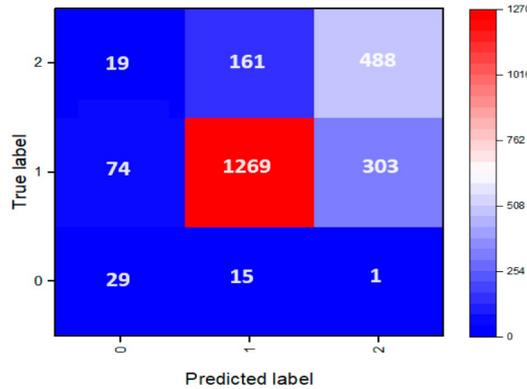


Figura 7: Matriz de Confusión del algoritmo XGBoost con mejor ACCURACY.

Support vector machine

Se implementó SVM (Support Vector Machine) con 20 variables relevantes para reducir la dimensionalidad. Se dividió el conjunto en entrenamiento (70 %) y prueba (30 %) con validación cruzada de 5 pliegues para una evaluación sólida. Usando grid_search en Python, se encontraron los hiperparámetros óptimos: C=10 y Kernel 'Rbf'. Se reafirmó esta elección con 15 iteraciones de pruebas para consolidar la eficacia del modelo. La Tabla 7 muestra los resultados correspondientes al conjunto de pruebas de hiperparámetros realizadas en el algoritmo SVM variando los parámetros “C” y “Kernel”, además de los resultados correspondientes a estas pruebas en términos de métricas de rendimiento, incluyendo precisión, RECALL, F1-score y ACCURACY. El kernel “rbf” con “C=10” obtuvo la mejor precisión, pero esta elección depende del contexto y las características de los datos. Algunos kernels, como el polinómico y el RBF, mejoraron con valores más altos de “C” debido a la penalización de errores. Otros, como “Kernel=Linear” y “Kernel=Sigmoid”, mantuvieron alta precisión con valores bajos de “C”. Valores perfectos en métricas pueden indicar sobreajuste. La elección del kernel y “C” debe considerar las características del conjunto de datos y equilibrar el rendimiento en entrenamiento y generalización. “Kernel=Sigmoid” con “C=1” puede ser respaldado por las particularidades del conjunto de datos, pero se necesita una evaluación detallada.

Tabla 7: Pruebas de hiperparámetros del algoritmo SVM.

Pruebas	Parámetro C	Kernel	Acurracy	Precision	RECALL	F1-Score
1	0,1	Sigmoid	0,62	0,35	0,35	0,66
2	1	Sigmoid	0,54	0,37	0,36	0,36
3	10	Sigmoid	0,51	0,36	0,36	0,36
4	0,1	Poly	0,64	0,66	0,41	0,40
5	1	Poly	0,68	0,64	0,46	0,49
6	10	Poly	0,71	0,61	0,53	0,55
7	0,1	Rbf	0,64	0,80	0,36	0,31
8	1	Rbf	0,70	0,67	0,48	0,51
9	10	Rbf	0,73	0,68	0,55	0,59
10	0,1	Linear	0,65	0,43	0,38	0,39
11	1	Linear	0,65	0,46	0,39	0,59
12	10	Linear	0,65	0,42	0,39	0,70

En la Figura 8 se muestra la matriz de confusión para el algoritmo SVM. Se encontraron 352 verdaderos positivos, que se clasificaron correctamente como clase 0. Se tuvieron 24 falsos negativos, es decir instancias que se clasificaron incorrectamente como clase 0 en lugar de clase 1. Se hallaron 400 verdaderos negativos, que se clasificaron correctamente como clase 1, y se tuvieron 19 falsos positivos, los que se clasificaron incorrectamente como clase 1 en lugar de 0.

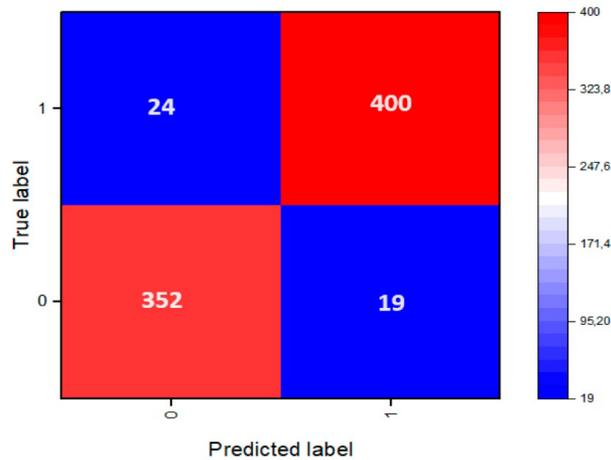


Figura 8: Matriz de Confusión del algoritmo SVM.

En resumen, la evaluación de los algoritmos muestra los mayores valores alcanzados de ACCURACY. (Tabla 8)

Tabla 8: Resumen de resultados de la evaluación de algoritmos.

Algoritmo	ACCURACY
XGBoost	76 %
SVM	73 %
KNN	69 %

Los resultados muestran que los algoritmos de Machine learning estudiados se ajustan razonablemente a buenas predicciones de calidad del secado, especialmente XGBoost.

Se usaron 7862 registros etiquetados según su calidad de secado. La clase ‘Exceso CH’ contaba solo con 395 instancias, este desbalance provocó que el modelo no lograra clasificar correctamente ninguna instancia de esta categoría. Se subraya la importancia de considerar el balance de los registros, ya que una gran brecha entre estas puede provocar errores en las predicciones, incluyendo el riesgo de sobreajuste o un entrenamiento deficiente del algoritmo.

La implementación de estos algoritmos (XGBoost, SVM, KNN) pueden ser compatibles con los análisis de RNA implementados por Özsahin *et al.* (2019) para la determinación de una temperatura óptima de secado y por Demir (2023) para determinar el efecto de resistencia de la unión de paneles contrachapados. También van en la línea de lo realizado por Urra-González y Ramos-Maldonado (2023) en la implementación de algoritmos de ML para predecir la adherencia en la etapa de encolado. Sin embargo, el uso de datos industriales lleva una mayor complejidad debido a que las condiciones de operación no están controladas como en un laboratorio. Los procesos industriales son también producto de una acumulación de conocimientos e innovaciones pasadas, haciendo más complejo su análisis. Un complemento para abordar de mejor manera esta problemática es la implementación de tecnología 3D, como lo recomiendan Acevedo *et al.* (2015).

El impacto de estas nuevas tecnologías, será mayor cuando estas actúen de manera sinérgica, formando un “ecosistema tecnológico”, donde se interrelacionen y refuercen dentro la empresa forestal, mejorando la productividad y la rentabilidad empresarial Gatica-Neira y Ramos-Maldonado (2022).

CONCLUSIONES

A partir de datos industriales, en este trabajo se evalúan tres algoritmos de ML para predecir la calidad del secado de chapas de pino radiata (*Pinus radiata* D. Don) en plantas de procesamiento de tableros contrachapados. La humedad de la cámara, la apertura de ventilas y las temperaturas son importantes en el resultado global y fundamentales para conseguir elevar la calidad del producto final.

A través de la evaluación de los tres algoritmos recomendados en la literatura, XGBoost, SVM y KNN, se observó que, si bien los tres mostraron resultados predictivos favorables, también se detectaron indicios de sobreajuste en algunas pruebas. El sobreajuste puede afectar la capacidad de generalización de los modelos a nuevos datos, lo que destaca la importancia de la selección cuidadosa del algoritmo.

Los mejores resultados se obtuvieron con el algoritmo XGBoost con un ACCURACY del 76 %. Por su parte SVM obtuvo un 73 % y KNN un 69 %. Se prueba que, con algoritmos de ML, y datos de origen industrial, los resultados son prometedores.

De acuerdo con los resultados obtenidos, esta investigación es un aporte a los estudios existentes, ya que aporta una modelación del problema, prediciendo la calidad del secado de chapas. Por ser un problema multivariado, el uso de ML es prometedor, ya que puede permitir al sector industrial disponer en el futuro de herramientas robustas de predicción y de análisis de condición operacional que lleven al ajuste de variables de control cuando las predicciones estén fuera de los estándares o metas.

Declaración de autorías

M. R-M.: Conceptualización, curación de datos, análisis formal, investigación, metodología, software, validación, visualización, redacción, revisión & edición, adquisición de fondos, recursos, administración del proyecto. T. D-S.: Software, curación de datos, investigación, redacción, revisión & edición. F. G-N.: Análisis formal, revisión & edición. D. V-V.: Curación de datos, investigación, redacción, revisión & edición.

Agradecimientos

Los autores agradecen el soporte del Proyecto Fondef ID22i10123: “Análisis Prescriptivo basado en Machine Learning, para la Operación de Plantas de Tableros Contrachapados de alta producción de pino radiata 4.0” de la Agencia Nacional de Investigación y Desarrollo, ANID del Ministerio de Cien, Tecnología, Conocimiento e Innovación de Chile.

Además, este trabajo contó con el soporte del Proyecto 2060360 IF/R: “Machine learning, emisión acústica y energía de corte para el monitoreo en tiempo real y predicción de calidad en madera aserrada en un contexto de industria 4.0” de la Vicerrectoría de Investigación y Postgrado, Universidad del Bío-Bío.

REFERENCIAS

Abe, S. 2010. Variants of Support Vector Machines. In Springer (pp. 163-226). https://doi.org/10.1007/978-1-84996-098-4_4

Acevedo, C.; Ramos-Maldonado, M.; Aguilera, C.; Monsalve-Lozano, D. 2015. Optimización 3D de patrones de corte para trozas de Pino radiata con cilindro central defectuoso. *Maderas. Ciencia y Tecnología* 17(2): 421-434. <https://doi.org/10.4067/s0718-221x2015005000039>

Alfaleh, A.; Ben, K.N.; Eldin, S.M.; Alturki, M.; Elbadawi, I.; Kumar, R. 2023. Predicting thermal conductivity and dynamic viscosity of nanofluid by employment of Support Vector Machines: A review. *Energy Reports* 10: 1259-1267. <https://doi.org/10.1016/j.egy.2023.08.001>

- Aydin, I. 2014.** Effects of veneer drying at high temperature and chemical treatments on equilibrium moisture content of plywood. *Maderas. Ciencia y Tecnología* 16(4): 445-452. <https://doi.org/10.4067/S0718-221X2014005000036>
- 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>
- Bressan, T.S.; Kehl de Souza, M.; Girelli, T.J.; Junior, F.C. 2020.** Evaluation of machine learning methods for lithology classification using geophysical data. *Computers and Geosciences* 139: e104475. <https://doi.org/10.1016/j.cageo.2020.104475>
- Chen, T.; Guestrin, C. 2016.** XGBoost: A scalable tree boosting system. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 13-17-August-2016*, 785-794. <https://doi.org/10.1145/2939672.2939785>
- Cortes, C.; Vapnik, V. 1995.** Support-Vector Networks. *Machine Learning* 20: 273-297. <http://dx.doi.org/10.1007/bf00994018>
- Cover, T.M.; Hart, P.E. 1952.** Approximate formulas for the information transmitted by a discrete communication channel. In *IEEE Transactions on Information Theory* 24:1. <https://doi.org/10.1109/TIT.1967.1053945>
- Demir, A. 2023.** Determination of the effect of valonia tannin when used as a filler on the formaldehyde emission and adhesion properties of plywood with artificial neural network analysis. *International Journal of Adhesion and Adhesives* 123: e103346 <https://doi.org/10.1016/j.ijadhadh.2023.103346>
- Demirkir, C.; Colak, S.; Aydin, I. 2013.** Some technological properties of wood-styrofoam composite panels. *Composites Part B: Engineering* 55: 513-517. <https://doi.org/10.1016/j.compositesb.2013.07.024>
- Diez-Olivan, A.; Del Ser, J.; Galar, D.; Sierra, B. 2019.** Data fusion and machine learning for industrial prognosis: Trends and perspectives towards Industry 4.0. *Information Fusion* 50: 92-111. <https://doi.org/10.1016/j.inffus.2018.10.005>
- Dogan, A. Birant, D. 2021.** Machine learning and data mining in manufacturing. In *Expert Systems with Applications* 166: e114060. <https://doi.org/10.1016/j.eswa.2020.114060>
- dos Santos-Freitas, M. M.; Barbosa, J. R.; dos Santos-Martins, E. M.; da Silva-Martins, L. H.; de Souza-Farias, F.; de Fátima-Henriques, L.; da Silva, N. 2022.** KNN algorithm and multivariate analysis to select and classify starch films. *Food Packaging and Shelf Life* 34: e 100976. <https://doi.org/10.1016/j.fpsl.2022.100976>
- Duarte, T. 2023.** *Uso de técnicas de Machine Learning para predecir la calidad de tableros contrachapados*. Habilitación Profesional de Ingeniería Civil Química, Universidad del Bío-Bío, Departamento Ingeniería en Madera. Concepción, Chile
- Dütsch, I.; Gediga, G. 2019.** Confusion Matrices and Rough Set Data Analysis. *Journal of Physics: Conference Series* 1229(1): e012055. <https://doi.org/10.1088/1742-6596/1229/1/012055>
- Easaw, N.; Lee, W.S.; Lohiya, P.S.; Jalan, S.; Pradhan, P. 2023.** Estimation of correlation matrices from limited time series data using machine learning. *Journal of Computational Science* 71: e102053. <https://doi.org/10.1016/j.jocs.2023.102053>
- Ellis, J.L.; Alaiz-Moretón, H.; Navarro-Villa, A.; McGeough, E.J.; Purcell, P.; Powell, C. D.; O’Kieily, P.; France, J.; López, S. 2020.** Application of meta-analysis and machine learning methods to the prediction of methane production from in vitro mixed ruminal micro-organism fermentation. *Animals* 10(4): e720. <https://doi.org/10.3390/ani10040720>
- Foody, G.M.; Mathur, A. 2004.** A relative evaluation of multiclass image classification by support vector machines. *IEEE Transactions on Geoscience and Remote Sensing* 42(6): 1335-1343. <https://doi.org/10.1109/TGRS.2004.827257>

- Freund, Y.; Schapire, R.E. 1999.** A Short Introduction to Boosting. *Journal of Japanese Society for Artificial Intelligence* 14(5): 771-780. <http://www.yorku.ca/gisweb/eats4400/boost.pdf>
- Frey, U.J.; Klein, M.; Deissenroth, M. 2019.** Modelling complex investment decisions in Germany for renewables with different machine learning algorithms. *Environmental Modelling and Software* 118: 61-75. <https://doi.org/10.1016/j.envsoft.2019.03.006>
- Gatica-Neira, F.; Ramos-Maldonado, M. 2020.** Public policies and networks for the development of 4.0 technologies in Chile. *PAAKAT: Revista de Tecnología y Sociedad* 19(10): 1-28. <https://doi.org/10.32870/Pk.a10n19.475>
- Gatica-Neira, F.; Ramos-Maldonado, M. 2022.** Limits to the Productivity in Biobased Territorial SMEs. *SAGE Open* 12(2): 1-15. <https://doi.org/10.1177/21582440221099294>
- Gosselin, P.H.; Murray, N.; Jégou, H.; Perronnin, F. 2014.** Revisiting the Fisher vector for fine-grained classification. *Pattern Recognition Letters* 49: 92-98. <https://doi.org/10.1016/j.patrec.2014.06.011>
- Gradov, D.V.; Yusuf, Y.O.; Ohjainen, J.; Suuronen, J.; Eskola, R.; Roininen, L.; Koironen, T. 2022.** Modelling of a continuous veneer drying unit of industrial scale and model-based ANOVA of the energy efficiency. *Energy* 244: e122673. <https://doi.org/10.1016/j.energy.2021.122673>
- He, H.; Zhang, W.; Zhang, S. 2018.** A novel ensemble method for credit scoring: Adaption of different imbalance ratios. *Expert Systems with Applications* 98: 105-117. <https://doi.org/10.1016/j.eswa.2018.01.012>
- Jia, L.; Chu, J.; Ma, L.; Qi, X.; Kumar, A. 2019.** Life cycle assessment of plywood manufacturing process in China. *International Journal of Environmental Research and Public Health* 16(11): e2037. <https://doi.org/10.3390/ijerph16112037>
- Kehr, R. 2007.** *Evaluación de programas de secado continuo en Chapas de Pinus radiata D. Don*. Tesis. Universidad Austral de Chile. Facultad de Ciencias Forestales. Valdivia, Chile
- Koskinen, J.; Vaarala, T.; Alatalo, J.; Heikkilä, T. 2013.** Automated Quality Classification of Wooden Parts for Flexible Manufacturing. *Journal of Engineering Technology* 2(1): 239-243. http://dx.doi.org/10.5176/2251-3701_2.1.60
- Lagos-Biolley, L. 2023.** *Uso de Machine Learning para la optimización de cámaras de secado de chapas de Pino radiata*. Habilitación Profesional de Ingeniería Civil en Industrias de la madera, Universidad del Bío-Bío. Departamento Ingeniería en Madera. Concepción, Chile.
- Lu, G.; Zeng, L.; Dong, S.; Huang, L.; Liu, G.; Ostadhassan, M.; He, W.; Du, X.; Bao, C. 2023.** Lithology identification using graph neural network in continental shale oil reservoirs: A case study in Mahu Sag, Junggar Basin, Western China. *Marine and Petroleum Geology* 150: e106168. <https://doi.org/10.1016/j.marpetgeo.2023.106168>
- Lu, Z.; Lin, F.; Ying, H. 2007.** Design of decision tree via kernelized hierarchical clustering for multiclass support vector machines. *Cybernetics and Systems* 38(2): 187-202. <https://doi.org/10.1080/01969720601139058>
- Luque, A.; Carrasco, A.; Martín, 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>
- Lutz, J. 1978.** *Wood Veneer: Log Selection, Cutting, and Drying*. Forest Service U.S Department of Agriculture. Technical Bulletin 1577. <https://ageconsearch.umn.edu/record/157864/files/tb1577.pdf>
- Manrique-Rojas, E. 2020.** Machine Learning: análisis de lenguajes de programación y herramientas para desarrollo. *Revista Ibérica de Sistemas e Tecnologías de Informação* Edición no regular E28 Abril: 586-599.
- Mehra, N.; Gupta, S. 2013.** Survey on Multiclass Classification Methods. *International Journal of Computer Science and Information Technologies* 4(4): 572-576. <https://www.ijcsit.com/docs/Volume%204/Vol4Issue4/ijcsit2013040408.pdf>

Mihel, A.M.; Lerga, J.; Krvavica, N. 2024. Estimating water levels and discharges in tidal rivers and estuaries: Review of machine learning approaches. *Environmental Modelling and Software* 176: e106033. <https://doi.org/10.1016/j.envsoft.2024.106033>

Minaei-Bidgoli, B.; Parvin, H.; Alinejad-Rokny, H.; Alizadeh, H.; Punch, W.F. 2014. Effects of resampling method and adaptation on clustering ensemble efficacy. *Artificial Intelligence Review* 41(1): 27-48. <https://doi.org/10.1007/s10462-011-9295-x>

Mobarak, M.H.; Mimona, M.A.; Islam, M.A.; Hossain, N.; Zohura, F.T.; Imtiaz, I.; Rimon, M. I. H. 2023. Scope of machine learning in materials research-A review. *Applied Surface Science Advances* 18: e100523. <https://doi.org/10.1016/j.apsadv.2023.100523>

Moisan, R. 2007. Modelo de determinación de rendimiento para el proceso de elaboración de paneles en planta Nueva Aldea. <http://repobib.ubiobio.cl/jspui/handle/123456789/2353>

Nakamura, K. 2023. A practical approach for discriminating tectonic settings of basaltic rocks using machine learning. *Applied Computing and Geosciences* 19: e100132. <https://doi.org/10.1016/j.acags.2023.100132>

Navarrete, C. 2020. Evaluación de método predictivo para variables de secado de chapas en planta de paneles Arauco Nueva Aldea. Habilitación Profesional de Ingeniería Civil Química. Universidad del Bío-Bío. Departamento Ingeniería en Madera. Concepción, Chile.

Özsahin, S.; Demir, A.; Aydin, 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>

Parvin, H.; Alinejad-Rokny, H.; Minaei-Bidgoli, B.; Parvin, S. 2013. A new classifier ensemble methodology based on subspace learning. *Journal of Experimental and Theoretical Artificial Intelligence* 25(2): 227-250. <https://doi.org/10.1080/0952813X.2012.715683>

Parvin, H.; Mirnabibaboli, M.; Alinejad-Rokny, H. 2015. Proposing a classifier ensemble framework based on classifier selection and decision tree. *Engineering Applications of Artificial Intelligence* 37: 34-42. <https://doi.org/10.1016/j.engappai.2014.08.005>

Penkova, T.G. 2017. Principal component analysis and cluster analysis for evaluating the natural and anthropogenic territory safety. *Procedia Computer Science* 112: 99-108. <https://doi.org/10.1016/j.procs.2017.08.179>

Pillay, T.; Cawthra, H.C.; Lombard, A.T. 2021. Integration of machine learning using hydroacoustic techniques and sediment sampling to refine substrate description in the Western Cape, South Africa. *Marine Geology* 440: e106599. <https://doi.org/10.1016/j.margeo.2021.106599>

Ramos-Maldonado, M.; Aguilera-Carrasco, C. 2021. Trends and Opportunities of Industry 4.0 in Wood Manufacturing Processes. In *Engineered Wood Products for Construction*. Intech Open. <https://doi.org/10.5772/intechopen.99581>

Rice, R. 1988. Mass transfer, creep and stress development during the drying of red oak. <https://vtechworks.lib.vt.edu/bitstreams/68c1415d-0bf7-4743-a6af-ffed0ed4ea9b/download>

Rokach, L. 2010. Ensemble-based classifiers. *Artificial Intelligence Review* 33(1-2): 1-39. <https://doi.org/10.1007/s10462-009-9124-7>

Russell, S.J.; Norvig, P. 2021. Artificial intelligence: a modern approach. Pearson Educational Inc.. Fourth Edition. <https://thuvienso.hoasen.edu.vn/bitstream/handle/123456789/8967/Contents.pdf?sequence=3>

Sadrara, M.; Khorrani, M. K. 2023. Principal component analysis–multivariate adaptive regression splines (PCA-MARS) and back propagation-artificial neural network (BP-ANN) methods for predicting the efficiency of oxidative desulfurization systems using ATR-FTIR spectroscopy. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* 300: e 122944. <https://doi.org/10.1016/j.saa.2023.122944>

Shalev-Shwartz, S.; Ben-David, S. 2014. *Understanding Machine Learning: From Theory to Algorithms*. Cambridge University Press. <http://www.cs.huji.ac.il/~shais/UnderstandingMachineLearning>

Tabassum-Abbasi, T.; Abbasi, T.; Patnaik, P.; Ali, B.M.J.; Abbasi, S.A. 2023. Artificial intelligence-based optimization of a novel process for generating volatile fatty acid energy precursors from date palm waste. *Bio-mass Conversion and Biorefinery* <https://doi.org/10.1007/s13399-023-04812-7>

Teihuel, J. 2007. Propuesta de alternativas de solución para el transporte de residuos de madera sólida en la industria de tableros contrachapados. Tesis, Universidad Austral de Chile, Facultad de Ciencias Forestales, Valdivia, Chile <http://cybertesis.uach.cl/tesis/uach/2007/fift263p/doc/fift263p.pdf>

Tharwat, A. 2016. Principal Component Analysis (PCA): An Overview. *International Journal of Applied Pattern Recognition* 3: 197-240. <https://doi.org/10.1504/IJAPR.2016.079733>

Troncoso-Espinosa, F.H.; Avello-Betancur, Y.G.; Martínez-Flores, L.A. 2021. Prediction of cellulose sheet cutting using Machine Learning. *Universidad Ciencia y Tecnología* 25(110): 109-118. <https://doi.org/10.47460/uct.v25i110.481>

Turhan, K.; Serdar, B. 2013. Support vector machines in wood identification: The case of three Salix species from Turkey. *Turkish Journal of Agriculture and Forestry* 37(2): 249-256. <https://doi.org/10.3906/tar-1205-47>

Uniyal, S.; Purohit, S.; Chaurasia, K.; Rao, S.S.; Amminedu, E. 2022. Quantification of carbon sequestration by urban forest using Landsat 8 OLI and machine learning algorithms in Jodhpur, India. *Urban Forestry and Urban Greening* 67: e127445. <https://doi.org/10.1016/j.ufug.2021.127445>

Urta-González, C.; Ramos-Maldonado, M. 2023. A machine learning approach for plywood quality prediction. *Maderas. Ciencia y Tecnología* 25: e36. <https://doi.org/10.4067/S0718-221X2023000100436>

Venegas-Vásquez, D.; Arteaga-Pérez, L.E.; Aguayo, M.G.; Romero-Carrillo, R.; Guerrero, V.H.; Tipanluisa-Sarchi, L.; Alejandro-Martín, S. 2023. Analytical Pyrolysis of *Pinus radiata* and *Eucalyptus globulus*: Effects of Microwave Pretreatment on Pyrolytic Vapours Composition. *Polymers* 15(18): e3790. <https://doi.org/10.3390/polym15183790>

Wang, H.; Xu, P.; Zhao, J. 2022. Improved KNN algorithms of spherical regions based on clustering and region division. *Alexandria Engineering Journal* 61(5): 3571-3585. <https://doi.org/10.1016/j.aej.2021.09.004>

Wu, S.J.; Gebrael, N.; Lawley, M.A.; Yih, Y. 2007. A neural network integrated decision support system for condition-based optimal predictive maintenance policy. *IEEE Transactions on Systems, Man, and Cybernetics Part A: Systems and Humans* 37(2): 226-236. <https://doi.org/10.1109/TSMCA.2006.886368>

Zhao, S.; Chen, C.; Luo, Y. 2020. Probabilistic Principal Component Analysis Assisted New Optimal Scale Morphological Top-Hat Filter for the Fault Diagnosis of Rolling Bearing. *IEEE Access* 8: 156774-156791. <https://doi.org/10.1109/ACCESS.2020.3019638>