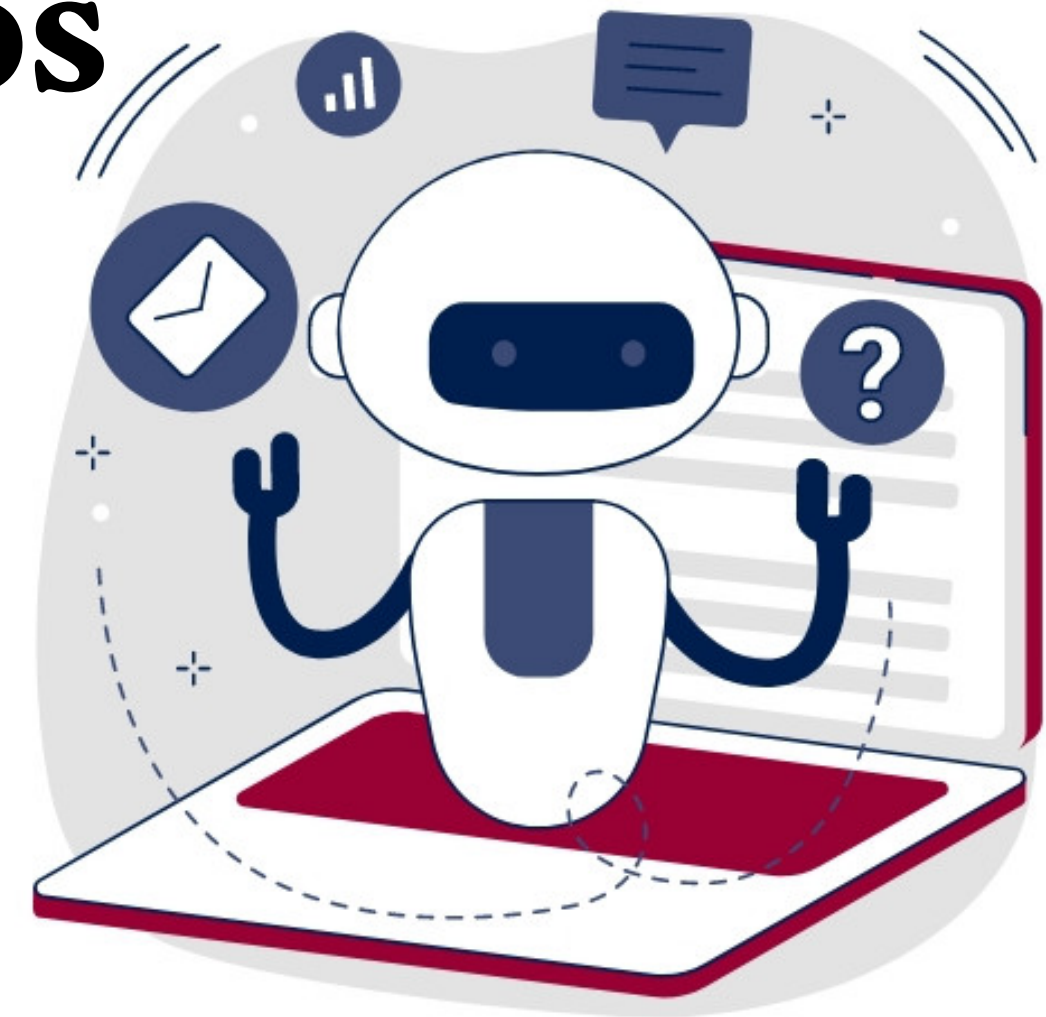


**INTELIGENCIA
ARTIFICIAL**

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

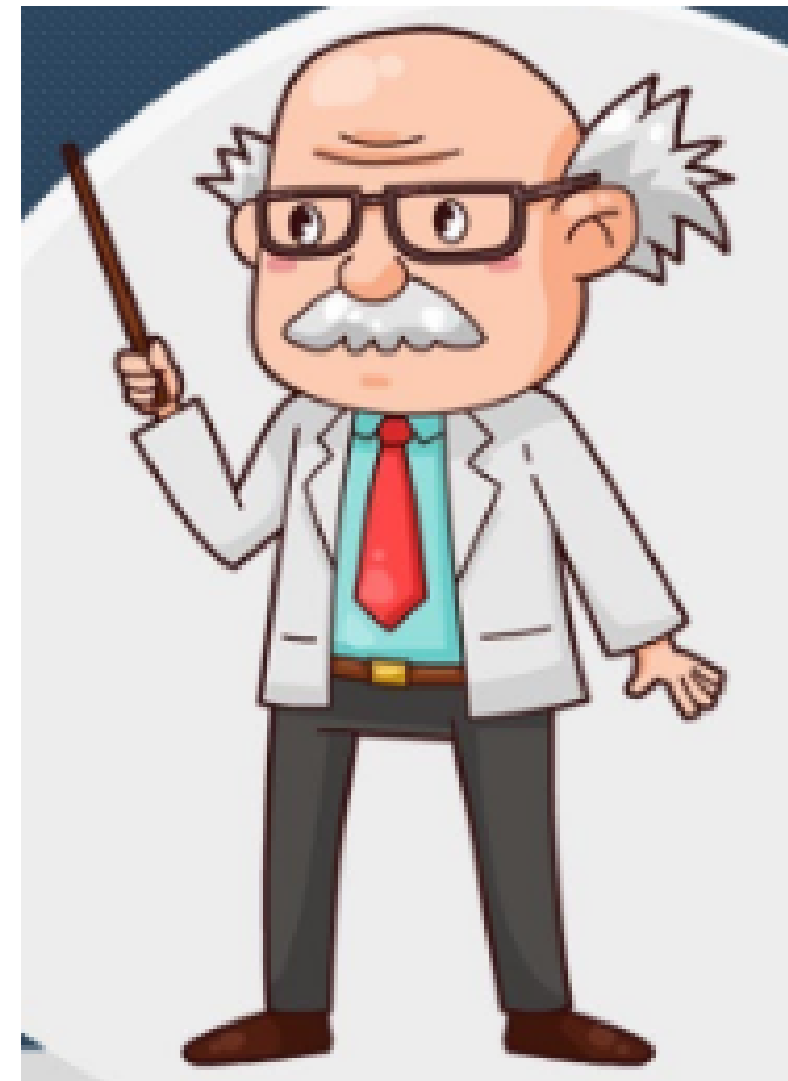


CHÓEZ GÓMEZ ANTHONY

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. 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



OBEJETIVO

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.

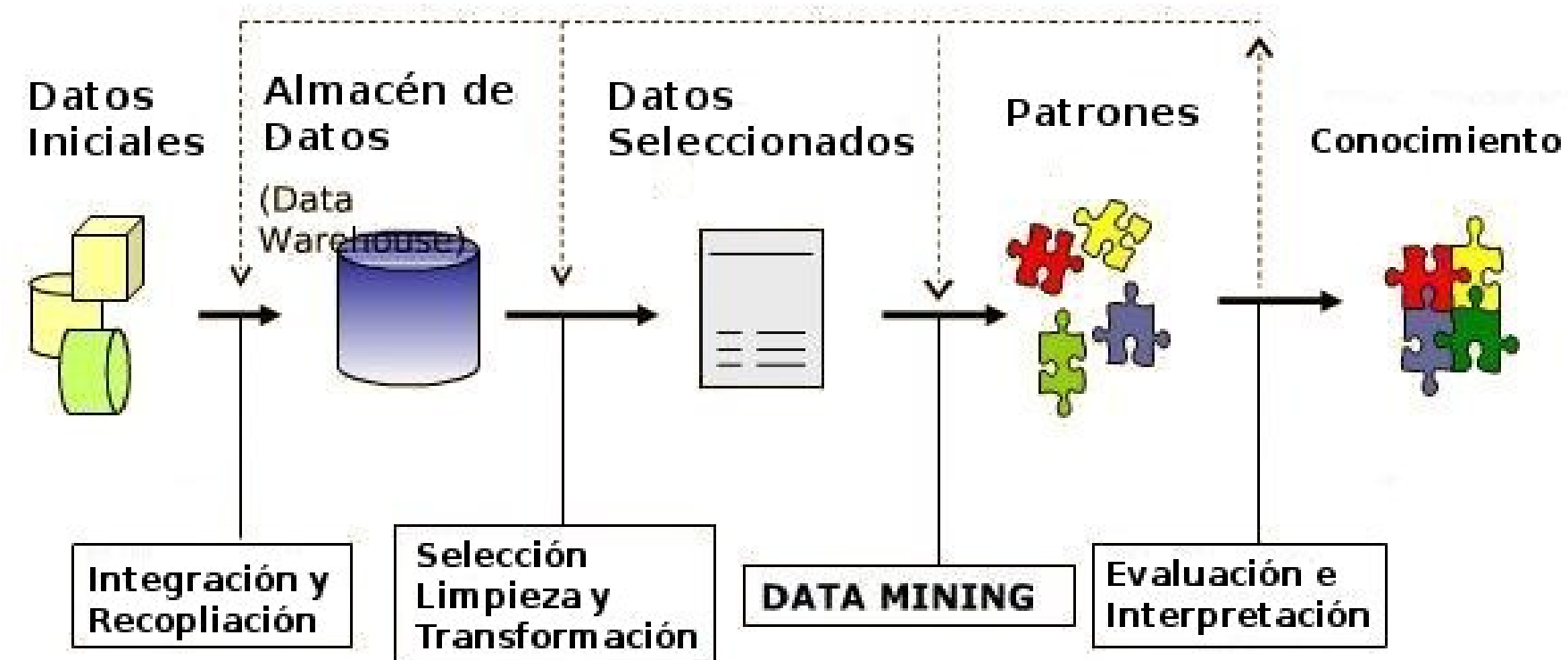


METODOLOGÍA

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.

KDD Knowledge Discovery from Databases

Proceso de KDD



RECURSOS

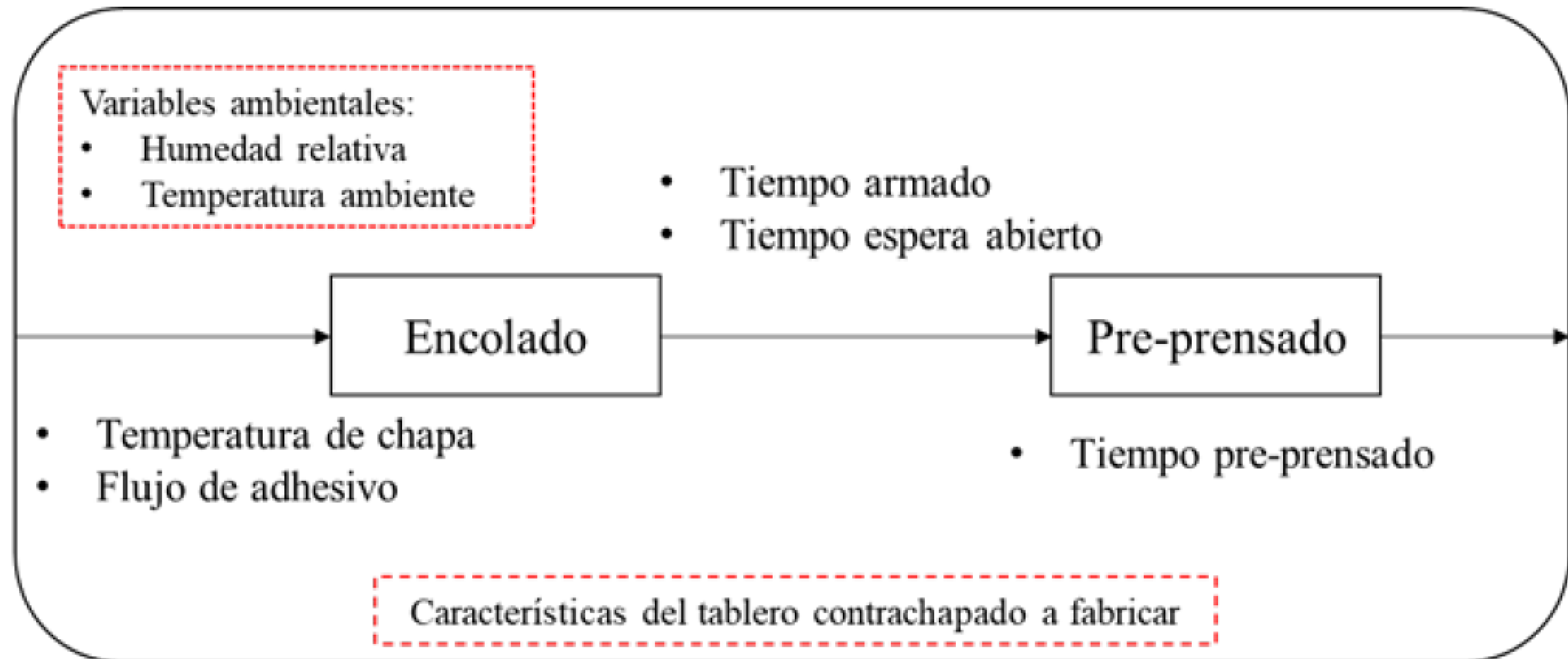
Servidor de pruebas: contaban con un procesador de Intel Core i7 1.50 GHz, RAM de 12 gb y un sistema de 64 bits



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

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

SELECCIÓN DE DATOS



CREACIÓN DE CONJUNTO DE DATOS (DATA SET)

Se llevó a cabo un proceso de recopilación y preprocesamiento de datos en un estudio relacionado con la adherencia de tableros de madera. Esto involucró calcular valores promedio para cada variable y periodo de tiempo registrarlos en un conjunto de datos preliminares.

Producción	Jefe de turno	Espesor	Calidad	Uso	Turno	Encoladora	Línea	Temperatura ambiente (°C)	Humedad relativa	Temperatura de la chapa (°C)	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

- Se eliminaron valores atípicos
- Se dio un criterio de calidad
- Datos faltantes
- Las variables cualitativas a cuantitativas fueron transformadas

Se crearon 2 data sets para realizar los experimentos.

- **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.
- **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).

CONSTRUCCIÓN DE MODELOS PREDICTIVOS

Los valores de los parámetros mencionados 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.

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)

EVALUACIÓN Y VALIDACIÓN DE MODELOS PREDICTIVOS

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.

		Predicho	
		P	N
Real	P	VP	FN
	N	FP	VN

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

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 identificó como valores positivos, precisión: proporción de casos que el modelo predijo correctamente los valores positivos.

$$Precisión = \frac{VP}{VP+FP}$$

RESULTADOS Y DISCUSIÓN

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.

	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

MODELOS PREDICTIVOS UTILIZADOS

Árbol de decisión

Se realizó un barrido por la cantidad de árboles incluidos en el modelo

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)

		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

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.



Conclusiones

