Tecnología en marcha
ISSN: 2215-3241
revistatm@tec.ac.cr
Instituto Tecnológico de Costa Rica
Costa Rica

Predicción de la calidad en revestimientos moldeados para puertas mediante el uso de minería de datos

Troncoso-Espinosa, Fredy; Castro-Albornoz, Karen

Predicción de la calidad en revestimientos moldeados para puertas mediante el uso de minería de datos
Tecnología en marcha, vol. 35, núm. 1, 2022
Instituto Tecnológico de Costa Rica, Costa Rica
Disponible en: https://www.redalyc.org/articulo.oa?id=699872860010
DOI: https://doi.org/10.18845/tm.v35i1.5395

Artículos científicos

Predicción de la calidad en revestimientos moldeados para puertas mediante el uso de minería de datos

Quality prediction in molded door skin using data mining

Fredy Troncoso-Espinosa
Universidad del Bío-Bío, Chile
ftroncos@ubiobio.cl
https://orcid.org/0000-0002-9972-3123

Karen Castro-Albornoz
Promasa S.A., Chile
kcastroalbornoz@gmail.com

DOI: https://doi.org/10.18845/tm.v35i1.5395
Redalyc: https://www.redalyc.org/articulo.oa?id=699872860010

Recepción: 01 Octubre 2020
Aprobación: 21 Enero 2021

Resumen:

Un revestimiento moldeado para puertas es un tablero de madera de alta densidad que es utilizado como el principal componente en la fabricación de puertas. Para asegurar su comercialización, se debe cumplir con exigentes normas de calidad, siendo la principal norma aquella que mide la fuerza necesaria para desprender el revestimiento de la estructura de una puerta. Los ensayos de calidad son realizados cada dos horas y sus resultados son obtenidos luego de aproximadamente cinco horas. Si los resultados muestran que los revestimientos están fuera del estándar de calidad exigido, se generan pérdidas económicas debido a este tiempo de espera. Esta investigación propone el uso de minería de datos mediante técnicas de machine learning para predecir en forma continua esta medida de calidad y reducir las pérdidas económicas asociadas a la espera de los resultados. Para la aplicación de minería de datos, se creó una base de datos en base al registro histórico de las variables del proceso productivo y de los ensayos de calidad. La metodología empleada es el descubrimiento de conocimiento en bases de datos KDD (Knowledge Discovery in Databases). La aplicación de esta metodología permitió identificar las principales variables que afectan la calidad de los revestimientos y entrenar cuatro algoritmos de machine learning para predecir su calidad. Los resultados muestran que el algoritmo que mejor predice la calidad es Neural Net y permiten demostrar que la implementación del algoritmo Neural Net reducirá las pérdidas económicas asociadas a la espera de los resultados de los ensayos de calidad.

Palabras clave: Minería de datos, machine learning, revestimiento para puertas, calidad, producción.

Abstract:

A door skin is a high-density wooden board and is the main component in the manufacture of doors. To ensure its commercialization, it must comply with demanding quality standards, the main one that measures the force necessary to detach the door skin from the structure of a door. Quality tests are carried out every two hours and the results are obtained after five hours. If the results show that the door skins are outside the required quality standard financial losses are generated during this waiting time. This research proposes the use of data mining using machine learning techniques to continuously predict this measure of door skin quality and reduce the economic losses associated with waiting for quality tests. For the use of data mining, a database was created using historical record of the variables of the production process and quality tests. The methodology used is the discovery of knowledge in KDD databases (Knowledge Discovery in Databases). The application of this methodology allowed identifying the main variables that affect the quality of the door skin and training four machine learning algorithms to predict the quality. The results show that the algorithm that best performance is Neural Net and allows to show that the implementation of the Neural Net algorithm will reduce the economic losses associated with waiting for the results of the quality tests.

Keywords: Data mining, machine learning, door skin, quality, manufacture.

Introducción

Con miras a potenciar la satisfacción del cliente, aumentar la confianza de inversionistas, consolidar una gestión y administración eficiente y eficaz, es que la calidad de un producto resulta de gran importancia y debe ser considerada en cada nivel dentro de las industrias de fabricación modernas. Dada la intensa competencia industrial y clientes cada vez más exigentes e insatisfechos, es que la calidad debe mejorarse continuamente y las empresas deben adoptar nuevos procesos de evaluación y predicción de la calidad con respecto a los productos y servicios que ofrecen [1, 2].

Masonite Chile S.A. [3] es una empresa dedicada a la fabricación de revestimientos moldeados para puertas. Los revestimientos moldeados para puertas son tableros de fibra de madera de alta densidad. Estos tableros son fabricados en un 94% de fibra de madera de pino insigne, un 5% de resina (fenol formaldehído) y un 1% de cera parafínica. Su espesor es de 3.2 mm, densidad 1.020 g/cm3, alto de 1800 a 2451 mm y ancho de 470 a 927 mm.

Los revestimientos moldeados para puertas y representan el principal producto de la empresa y son exportadas principalmente a Estados Unidos, con un muy buen posicionamiento en este mercado.

Para asegurar una buena comercialización, se debe cumplir con requisitos de normas estadounidenses de calidad. La principal medida de calidad es una medida de la fuerza que se debe ejercer para desprender el revestimiento de la estructura de una puerta. Esta medida de fuerza se expresa en Newton [4].

Los ensayos para esta medida de esta calidad son realizados cada dos horas y los resultados son obtenidos luego de cinco horas. Si el resultado de la prueba es inadecuado se ajustan las variables productivas para mejorar la calidad y se califica la producción de revestimientos de las últimas cinco horas como de calidad inadecuada. Esta producción de calidad inadecuada se destina a otros mercados a un menor precio de venta. En el caso de que los ensayos arrojen una calidad por sobre el estándar, se ajustan las variables productivas para ajustar la producción a la calidad estándar. Esta producción con una calidad sobre el estándar implica una producción a un costo mayor.

La principal variable que influye en esta medida de calidad es la cantidad de resina. La resina es un insumo productivo de alto costo, por lo que su utilización inadecuada es el factor más influyente en un sobre costo de producción.

La demora en los resultados de calidad conlleva al desconocimiento sobre los niveles adecuados de resina [5, 6] lo que acarrea perdidas económicas para la empresa por no ajuste al estándar.

Se ha observado el uso exitoso de Minería de Datos en la resolución de problemas relacionados con la calidad, posicionándose como una herramienta válida su predicción en las industrias manufactureras [1, 2, 7, 8]. Shi, Schillings y Boyd [9] construyeron y validaron modelos empíricos de redes neuronales artificiales para predecir la calidad de productos químicos y un proceso de mecanizado de placa de circuito impreso, cuyos resultados brindaron una mejor comprensión del proceso y una mejora en la calidad de este. Tseng [10] propone un enfoque integrado para abordar el problema del diseño económico de las soldaduras, aplicando una red neuronal de regresión general para predecir determinados parámetros de calidad de soldaduras. Mediante la aplicación del algoritmo de red neuronal de retropropagación y el método Taguchi, Chen, Lee, Deng y Liu [11] establecieron un predictor de calidad para analizar la relación entre la configuración de los parámetros del proceso de fabricación y la calidad del producto final en la deposición química de vapor mejorada con plasma de la fabricación de semiconductores. Agarwal y Shivpuri [12] aplicaron el clasificador bayesiano para predecir la calidad de la microestructura y propiedades mecánicas en el laminado en caliente de bandas en una empresa del acero. Austin y otros [13] emplearon arboles de decisión para clasificar a pacientes con insuficiencia cardíaca y predecir la calidad en términos de asistencia sanitaria en una muestra poblacional de pacientes de Ontario, Canadá. Ribeiro, Sanfins y Belo [8] implementaron métodos de Maquinas de Vectores de Soportes para predecir la calidad del procedimiento de tratamiento de aguas residuales en una planta ubicada en el norte de Portugal, detectando los parámetros claves en el manejo eficiente de dicho

procedimiento. Thiede y otros [2] emplearon diversas técnicas de minería de datos para predecir diferentes parámetros de calidad en la producción de baterías, cuyos resultados permitieron mejorar la planificación y el control de tal producto.

Masonite Chile S.A. lleva el registro histórico de las variables de su proceso productivo y de calidad. Dado los buenos resultados obtenidos en la predicción de la calidad en productos de empresas manufactureras, esta investigación plantea la utilización de minería de datos para entrenar técnicas de machine learning que permitan la estimación continua de la calidad y reducir las pérdidas económicas asociadas a la demora de los resultados de la prueba de calidad.

Metodología de resolución

La metodología utilizada es Knowledge Discovery in Databases (KDD). Fayyad, Piatetsky-Shapiro, Smyth y Uthurusamy [14] definen KDD como "el proceso no trivial de identificar patrones válidos, novedosos, potencialmente útiles y, en última instancia, comprensibles en los datos". Este proceso consta de los pasos mostrados en la figura 1 [15]. Es necesario mencionar que la Minería de Datos es un paso dentro del proceso KDD que consiste en la aplicación de técnicas de machine learning, estadísticas y técnicas de visualización para descubrir y presentar patrones y conocimientos comprensibles [16].

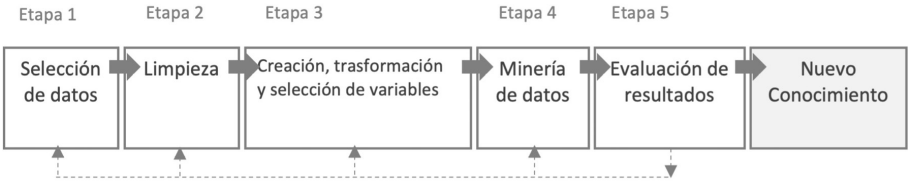

FIGURA 1
Procesos de datos dentro de la metodología Knowledge Discovery in Databases KDD

La primera etapa es la selección de datos, donde se localizan las fuentes de datos y el tipo de información a utilizar. Para tener un apropiado proceso de Minería de Datos, es que antes de integrar algún dato es muy importante definir la variable a predecir y conocer bien los atributos involucrados [17, 18].

La segunda busca mejorar la calidad de los datos y tener información más confiable, por lo que se procede a la limpieza de los datos para mejorar la precisión y eficiencia del proceso de Minería de Datos. Esta etapa incorpora el análisis de datos faltantes, detección y eliminación de datos redundantes, datos atípicos y análisis de datos fuera de rango [19, 20].

La tercera etapa es la creación, transformación y selección de variables, modificando la forma de los datos para generar nuevas variables cuyo formato sea el más apropiado para enriquecer la información con la que se entrenará el modelo, y, en consecuencia, se obtenga un mejor desempeño predictivo. Luego, se procede a identificar aquellos atributos que contribuyan mejor a predecir la variable de interés, permitiendo disponer de modelos más sencillos y que expliquen mejor el problema [21].

La cuarta etapa es la Minería de Datos. Esta etapa consiste en la aplicación de modelos y técnicas que permitirán explorar y extraer patrones y relaciones relevantes que puedan existir en los datos en análisis. En este caso se utilizarán técnicas de machine learning las cuales aprenden el patrón general oculto en los datos y luego lo utilizan para generar una nueva predicción. Esta predicción, consiste en asignar un registro u observación a una categoría o clase previamente definida [22, 23]. En esta investigación se utiliza algoritmos de machine learning que permitan predecir tres clases diferentes. Para que un algoritmos de machine learning pueda predecir se considera una etapa de entrenamiento y prueba mediante el método conocido como Hold Out [24]. Este método divide los datos aleatoriamente en dos conjuntos mutuamente exclusivos, denominados conjunto de entrenamiento, el cual representa el 70% del total de los datos y conjunto de

prueba, que representa el 30% del total de los datos. Mediante la etapa de prueba, es posible medir el desempeño predictivo del algoritmo.

Los resultados obtenidos en la etapa de prueba son representados mediante una matriz llamada Matriz de Confusión [25]. Para comprender de mejor forma esta matriz, se considera solo dos clases: la clase 1 que correspondería a una calidad adecuada y la clase 0 que correspondería a una calidad inadecuada (sobre o bajo el estándar) [26]. La Matriz de Confusión tiene la forma que se muestra en el cuadro 1. En la Matriz de Confusión mostrada en cuadro 1, Tp representa los elementos de la clase 1 correctamente predichos por el modelo y Fn representa los elementos de la clase 1 incorrectamente predichos por el modelo (Error Tipo I). Tn representa los elementos de la clase 0 correctamente predichos por el modelo y Fp representa los elementos de la clase 0 incorrectamente predichos por el modelo (Error Tipo II). La Matriz de Confusión, permite obtener las siguientes medidas de desempeño:

Accuracy: La ecuación (1) representa la proporción total de predicciones que fueron correctamente clasificadas.

$$Accuracy = (TP + TN)/(TP + FP + FN + TN) \quad (1)$$

Recall: La ecuación (2) representa el porcentaje de observaciones que pertenecen a clase 1 y que fueron clasificados correctamente por el modelo.

$$Recall = TP/(TP + FN) \quad (2)$$

Precision: La ecuación (3) representa el porcentaje de elementos clasificados correctamente como clase 1 del total de elementos clasificados como clase 1.

$$Precision = TP/(TP + FP) \quad (3)$$

CUADRO 1
Matriz de confusión para dos clases

		Valor Real	
Clases		1	0
Valor Predicho	1	TP	FP
	0	FN	TN

Aplicación del proceso KDD para la predicción de la calidad

En este capítulo se detalla la aplicación de cada una de las etapas del proceso KDD para la definición de un modelo predictivo de la calidad de los revestimientos moldeados para puertas.

Selección de datos: Esta etapa implicó la identificación de las etapas del proceso productivo y la identificación de las principales variables de cada proceso. La figura 2 muestra un diagrama del proceso de fabricación de los revestimientos.

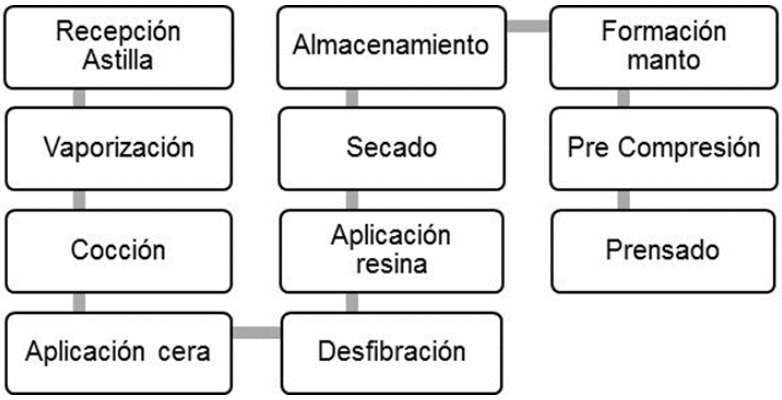

FIGURA 2
Procesos de proceso de fabricación de los revestimientos moldeados para puertas

La empresa no cuenta con un sistema unificado con el registro de las variables del proceso. La unificación de estas variables se realizó en forma manual, estableciendo como identificador de cada registro la fecha y hora en que se prensó cada revestimiento. Se identificó con la ayuda de expertos 22 variables que generan un impacto en la calidad de los revestimientos las cuales se muestran en el cuadro 2. Se consideró el mes como una variable adicional, debido a que características ambientales asociadas a cada mes, como temperatura y humedad, podían influir en las propiedades físicas de la astilla de madera, considerada el principal insumo productivo.

Cuadro 2. Variables que impactan en la calidad de los revestimientos moldeados para puertas.

CUADRO 2
Variables que impactan en la calidad de los revestimientos moldeados para puertas.

Proceso	Variable	Proceso	Variable
Vaporización	Temperatura digestor Flujo digestor Consumo vapor digestor Presión digestor Descarga digestor	Aplicación resina	Dosificación de resina flujo Dosificación de resina Presión del inyector
Cocción	Tiempo cocción astilla	Secado	Temperatura secado entrada
Aplicación cera	Flujo Dosificación de cera Temperatura cera	Formación Manto	Peso Manto Humedad manto Velocidad línea Energía específica Altura del Manto
Desfibrado	Separación de discos GAP Flujo astilla	Prensado	Temperatura prensado salida Tiempo prensado Espesor manto

Preprocesamiento: Se analizó cada variable en forma independiente. Se eliminó la variable temperatura de la cera por poseer una gran cantidad de datos faltantes. Se eliminó 58 registros con datos faltantes. Después se procedió a la detección de valores atípicos mediante la regla de tres sigmas [27]. Se identificaron 151 valores atípicos. Los valores faltantes y atípicos se reemplazaron mediante la estimación de un valor más apropiado. Para esto se buscó registros con calidades similares, utilizando como medida de similitud la distancia euclidiana [28]. Considerando todas las variables, se identificó aquellas con alta correlación entre sí mediante una matriz de correlación. Una alta correlación entre variables implica que cada una de ellas explica un fenómeno en forma similar. Por esta razón se decide dejar sólo una variable entre aquellas que presentan una correlación mayor a un 0.9 [24]. Las variables altamente correlacionadas fueron: Descarga Digestor, Flujo de Astillas y Espesor Manto. Se consideró la variable Descarga del Digestor.

Transformación y selección de variables: Para que los algoritmos de machine learning puedan ser entrenados y probados se categorizó la variable que mide la calidad. Esta transformación redefine la naturaleza de la variable, pasando de una variable numérica continua, que fluctúa en un rango de 289 y 625 Newton, a una variable categórica compuesta de cuatro categorías como se muestra en el cuadro 3.

CUADRO 3
Categorización de la calidad de los revestimientos moldeados para puertas

Notación	Rango
Calidad Inadecuada Baja	<425
Calidad Adecuada Baja	[425 - 460[
Calidad Adecuada Alta	[460 - 500[
Calidad Inadecuada Alta	[500 - ∞ [

Para la definición de estas categorías se fundamenta en el valor de aceptación de los estándares estadounidense. Con las clases definidas se procede a la identificación de las variables más importantes del proceso que influyen en la calidad, de manera de entrenar con estas los algoritmos de machine learning.

Para identificar las variables más importantes, se mide el poder predictivo de cada uno de ellos, sobre la variable categorizada de calidad. Se selecciona aquellas variables con mayor poder predictivo para el entrenamiento de los modelos [29]. Para medir el poder predictivo de cada variable, se utilizó el método de la Ganancia de Información. La ganancia de información mide la parte de la información contenida en una variable que explica la variable a predecir. A mayor ganancia de información más importante es la variable para la predicción. Para medir la ganancia de información de cada variable fue necesario categorizar las variables numéricas del proceso. Las categorías se definieron en tres rangos de igual amplitud, que representan tres magnitudes de las variables: alta, media y baja. Luego se seleccionó entre todas las variables aquellas con un valor de ganancia de información mayor o igual al promedio 0.2320 [30]. El cuadro 4 muestra las variables seleccionadas y el valor normalizado de la ganancia de información de cada variable en orden descendente.

CUADRO 4
Medición del desempeño predictivo de cada variable
de producción mediante la ganancia de información

Variable	Valor normalizado de la Ganancia de información
Mes	1.0000
Velocidad línea	0.6128
Presión inyector	0.5977
Porcentaje dosificación resina	0.5916
Flujo dosificación resina	0.4819
Consumo Vapor	0.4796
Separación discos GAP	0.3750
Flujo Dosificación cera	0.2653

Minería de datos: Para entrenar los algoritmos de machine learning, es necesario el balance de los datos, de manera que el número de registros clasificados como calidad inadecuada, sean igual al número de registros clasificados como calidad adecuada. Esto permite que algoritmos de machine learning no presenten una tendencia de clasificación hacia la clase mayoritaria. Para mitigar este problema se utiliza la técnica de balance SMOTE (Synthetic Minority Oversampling Method) que genera nuevas instancias de la clase minoritaria para equilibrar la base de datos en base a la técnica del vecino más cercano [31].

La base de datos para el entrenamiento y prueba de los modelos quedó compuesta por 3845 registros. Se entrenó y probó cuatro modelos mencionados en el capítulo cuyos algoritmos se encuentran incorporados en

la librería Scikit-Learn del lenguaje de programación Python llamados: Decision Tree, Naive Bayes, Neuronal Net y KNN. De manera de optimizar el desempeño predictivo de cada algoritmo, se iteró en los distintos algoritmos y se ajustó los respectivos parámetros de cada algoritmo siguiendo el procedimiento que se muestra en el pseudocódigo de la figura 3 [15].

La predicción generada por un algoritmo de machine learning para el conjunto de prueba es un valor entre cero y uno para cada una de las cuatro categorías de la variable calidad. Este valor es conocido como confidence. Un confidence más cercano a 1 representa una mayor probabilidad de que un registro pertenezca a una determinada calidad, por lo que la calidad asignada a un registro será aquella que presente un mayor valor. En base a esto se obtuvo la matriz de confusión para cada algoritmo de machine learning entrenado. El cuadro 5 muestra los resultados de la predicción de la calidad del algoritmo Neuronal Net.

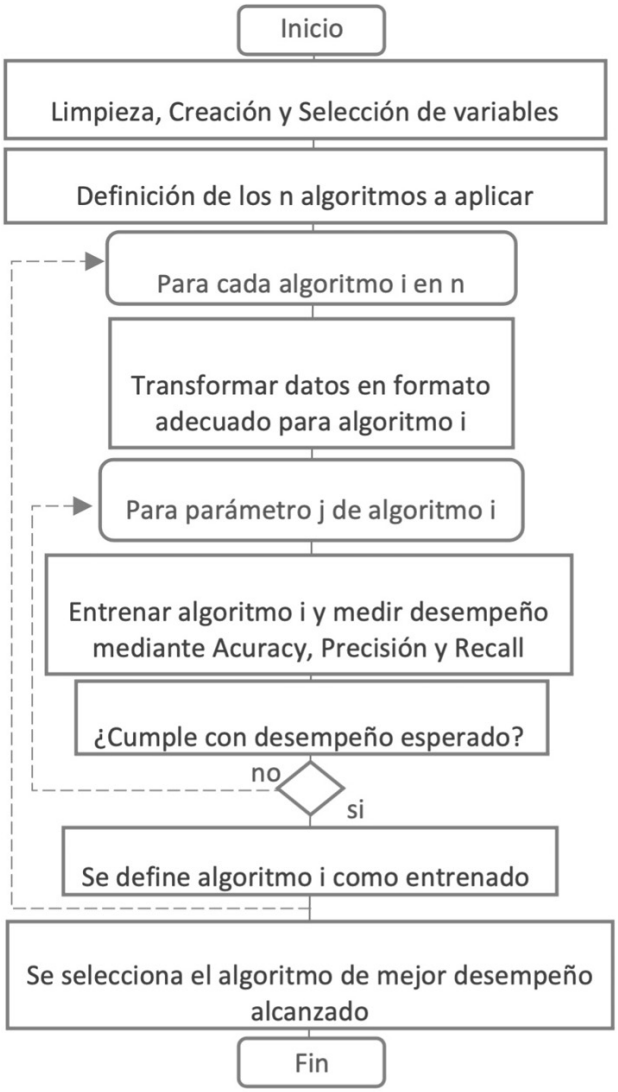

FIGURA 3
Pseudocódigo para la evaluación y selección de los algoritmos de machine learning

CUADRO 5
Desempeño predictivo de la red neuronal para cada calidad en los revestimientos moldeados para puertas.

Accuracy: 85.85%	Calidad Inadecuada Baja	Calidades Reales				Precision
		Calidad Adecuada Baja	Calidad Adecuada Alta	Calidad Inadecuada Alta		
Calidades Predichas	Calidad Inadecuada Baja	187	0	0	55	77.27%
	Calidad Adecuada Baja	7	278	4	0	96.19%
	Calidad Adecuada Alta	0	3	277	3	97.88%
	Calidad Inadecuada Alta	87	0	0	223	71.94%
Recall		66.55%	98.93%	98.58%	79.36%	1124

En el cuadro 6 se muestran los valores de Accuracy, Precision y Recall, de cada uno de los algoritmos entrenados. Se observa que el algoritmo Neural Net presenta los valores más altos en cada una de las medidas de desempeño.

CUADRO 6
Desempeño predictivo general de los modelos evaluados para predecir la calidad

	Modelo de clasificación			
Desempeño predictivo	Decision Tree	KNN	Naive Bayes	Neuronal Net
Accuracy	72,86%	68,24%	78,74%	85,85%
Recall Promedio	72,87%	68,24%	78,74%	85,86%
Precision Promedio	78,57%	67,84%	79,30%	85,82%

Estimación del impacto económico en la empresa y análisis de resultados

El objetivo de contar con un modelo predictivo es permitir el ajuste adecuado y oportuno de los niveles de resina. Esto se traduce en la reducción de los costos de producción derivados de una calidad inadecuada alta y en la reducción de pérdidas por ventas dada una calidad inadecuada baja. A la suma de ambas reducciones la llamaremos Beneficio Esperado por la Implementación del Algoritmo $_{B_e}$

La utilización de un algoritmo con mayor desempeño que otro, implica un mayor beneficio esperado, por lo que se decide utilizar el algoritmo Neural Net como el modelo predictivo de la calidad de los revestimientos moldeados para puertas.

Para estimar el beneficio esperado, se propone la ecuación (4).

$$B_e = R_p\, O_p + R_c\, O_c \tag{4}$$

Dónde es la ocurrencia promedio mensual de revestimientos bajo los estándares y
es la reducción esperada de pérdidas por ventas dada la utilización del algoritmo.
El valor representa la ocurrencia promedio mensual de revestimientos sobre los estándares y
la reducción de los cotos de producción, dada la utilización del algoritmo.

La reducción esperada de pérdidas por ventas dada la utilización del algoritmo, se obtiene mediante la multiplicación de la proporción de registros de calidad inadecuada baja predichos correctamente (Recall

Calidad Inadecuada Baja) β y las pérdidas por ventas. Las pérdidas por ventas se obtienen mediante la ecuación (5):

$$P_v = S\ h\ (V_e - V_p) \tag{5}$$

Dónde S representa las puertas producidos en una hora, h las horas de producción mientras se esperan los resultados de calidad, el precio del revestimiento exportado y el valor del revestimiento no exportado por una calidad inadecuada baja.

La reducción esperada en los costos de producción, dada la utilización del algoritmo, se obtiene mediante la multiplicación de la proporción de registros de calidad inadecuada alta predichos correctamente (Recall Calidad Inadecuada Alta) α y el costo del ajuste en el uso de resina. El costo del ajuste en el uso de resina se obtiene mediante la ecuación (6):

$$C_R = M\ q\ \ S\ h \tag{6}$$

En dónde M son los gramos de resina ajustados, q el precio del gramo de resina, S corresponde a los revestimientos producidos en una hora y h las horas de producción mientras se esperan los resultados de calidad.

Finalmente, el Beneficio Esperado por la Implementación del Algoritmo queda expresado mediante la ecuación (7):

$$B_e = [\ S\ h\ (V_e - V_p\ )\ \beta]\ O_p + [(M\ q\ S\ h)\alpha]O_c \tag{7}$$

Para obtener Beneficio Esperado por la Implementación del Algoritmo se consideran datos reales de producción y el desempeño predictivo para las calidades inadecuada baja e inadecuada alta, representadas por el respectivo valor de Recall en el cuadro 5.

Los revestimientos producidos por la empresa en una hora (S) son aproximadamente 1200 unidad. Las horas de producción mientras se esperan los resultados de calidad (h) son 5 horas.

El precio del revestimiento exportado () es de 3,5 dólares por unidad y el precio del revestimiento no exportado () es de 3,1 dólares por unidad. La ocurrencia promedio mensual de revestimientos bajo los estándares es 2 veces al mes. La proporción de registros de calidad inadecuada baja predichos correctamente (Recall Calidad Inadecuada Baja) β es 66.55%.

La ocurrencia promedio mensual de revestimientos sobre los estándares es de 20 veces al mes. La proporción de registros de calidad inadecuada alta predichos correctamente (Recall Calidad Inadecuada Alta) α es 79.36%. Los ajustes en la cantidad de resina se realizan en forma discreta en 1% correspondiente a 4 gramos. Este ajuste se realiza hasta alcanzar el valor límite de ajuste de un 4%. El precio de cada gramo de resina es de 0,0018 dólares. Lo más frecuente es un ajuste en un 1% por lo que se tomará este valor modal como referencia.

Reemplazando los valores en ecuación (4), se obtiene:

$$B_e = [\ 1200 * 5 * (3{,}5 - 3{,}1) * 0{,}6655] * 2 + [(4 * 1200 * 5 * 0{,}0018) * 0{,}7936] * 20$$

$$= 3.880\ dólares/mes \tag{8}$$

Discusión

La pérdida económica asociada a la espera de los resultados de los ensayos de calidad se estima en US$67.968 al año, lo que se obtiene de las perdidas esperadas por calidad inadecuada baja más las pérdidas por calidad inadecuada alta.

La reducción de los costos esperados mensuales por la utilización del algoritmo Neural Net es de 3.880 dólares, lo que implica un beneficio anual esperado de 46.560 dólares al año.

Esto se traduce en una reducción de las perdidas asociado a la espera por los ensayos de calidad de un 68.5% al año.

Esta reducción en las perdidas en un antecedente fundamental para evaluar la implementación técnica del algoritmo en el proceso productivo.

Conclusiones

En la fabricación de revestimientos moldeadas para puertas, las ocho principales variables que afectan la principal medida de calidad, definida como aquella que mide la fuerza necesaria para desprender el revestimiento de la estructura de una puerta, son: mes, velocidad de la línea, presión del inyector, porcentaje dosificación de resina, flujo dosificación resina, consumo vapor, separación discos gap y flujo dosificación de la cera.

La influencia de estas variables era desconocida para la empresa, por lo que su conocimiento permitirá tener un mayor control sobre estas, aumentando los beneficios esperados.

La variable mes resulta de especial interés pues, el clima afecta las características de las astillas de madera utilizada para la obtención de la fibra de madera. Por esto, para estabilizar la calidad se recomienda almacenar la astilla de madera en un ambiente más controlado.

La identificación de estas variables permitió ajustar diferentes algoritmos de machine learning, incorporados en la librería Scikit-Learn del lenguaje de programación Python, e identificar el de mejor desempeño. En base a las medidas de desempeño accuracy, precisión y recall se determinó que el algoritmo de mejor desempeño fue Neural Net.

Mediante la utilización de valores reales de producción se demuestra que la implementación del algoritmo Neural Net permitiría reducir las pérdidas económicas asociadas a la espera de los resultados de los ensayos de calidad en un 68.5%.

Referencias

[1] H. Rostami, J.-Y. Dantan y L. Homri, «Review of data mining applications for quality assessment in manufacturing industry: support vector machines,» International Journal of Metrology and Quality Engineering, p. 401, 2015.

[2] S. Thiede, A. Turetskyy, A. Kwade, S. Kara y C. Herrmann, «Data mining in battery production chains towards multi-criterial quality prediction,» CIRP Annals, pp. 463-466, 2019.

[3] Masonite Chile S.A., «www.masonite.cl,» [En línea].

[4] C. Salgado, «Procediminetos para ensayos físico-mecánicos,» Masonite Chile, Cabrero, VIII región, Chile, 2012.

[5] J. R. Eleotério, M. T. Filho y G. B. Júnior, «Propriedades físicas e mecânicas de painéis MDF de diferentes massas específicas e teores de resina,» Ciência Florestal, pp. 75-90, 2000.

[6] H. Poblete y R. Vargas, «Relacion entre densidad y propiedades de tableros HDF producidos por un proceso seco,» Maderas. Ciencia y tecnología, pp. 169-182, 2006.

[7] G. Köksal, İ. Batmaz y M. C. Testik, «A review of data mining applications for quality improvement in manufacturing industry,» Expert systems with Applications, pp. 13448-13467, 2011.

[8] D. Ribeiro, A. Sanfins y O. Belo, «Wastewater treatment plant performance prediction with support vector machines,» Industrial Conference on Data Mining, pp. 99-111, 2013.

[9] X. Shi, P. Schillings y D. Boyd, «Applying artificial neural networks and virtual experimental design to quality improvement of two industrial processes,» International Journal of Production Research, pp. 101-118, 2004.

[10] H.-Y. Tseng, «Welding parameters optimization for economic design using neural approximation and genetic algorithm,» The International Journal of Advanced Manufacturing Technology, pp. 897-901, 2006.

[11] W.-C. Chen, A. H. Lee, W.-J. Deng y K.-Y. Liu, «The implementation of neural network for semiconductor PECVD process,» Expert Systems with Applications, pp. 1148-1153, 2007.

[12] K. Agarwal y R. Shivpuri, «An on-line hierarchical decomposition based Bayesian model for quality prediction during hot strip rolling,» ISIJ international, pp. 1862-1871, 2012.

[13] P. Austin, J. Tu, J. Ho, D. Levy y D. Lee, «Using methods from the data-mining and machine-learning literature for disease classification and prediction: a case study examining classification of heart failure subtypes,» Journal of clinical epidemiology, pp. 398-407, 2013.

[14] U. Fayyad, G. Piatetsky-Shapiro, P. Smyth y R. Uthurusamy, Advances in knowledge discovery and data mining, American Association for Artificial Intelligence, 1996.

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

[16] U. Fayyad, G. Piatetsky-Shapiro y P. Smyth, «The KDD process for extracting useful knowledge from volumes of data,» Communications of the ACM, pp. 27-34, 1996.

[17] W. Frawley, G. Piatetsky-Shapiro y C. Matheus, «Knowledge discovery in databases: An overview,» AI magazine, pp. 57-57, 1992.

[18] L. A. Calvo Valverde y J. A. Mena Arias, «Evaluación de distintas técnicas de representación de texto y medidas de distancia de texto usando KNN para clasificación de documentos.,» Tecnología en Marcha, vol. 33, nº 1, pp. 64-79, 2020.

[19] L. A. Calvo Valverde y D. E. Alfaro Barboza, «Descubrimiento de reglas significativas mediante el uso de DTW basado en Interpolación Spline Cúbico.,» Tecnología en Marcha, vol. 33, nº 2, pp. 137-149, 2020.

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

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

[22] C. R. Morales, S. V. Soto y C. H. Martínez, «Estado actual de la aplicación de la minería de datos a los sistemas de enseñanza basada en web,» Actas del III Taller Nacional de Minería de Datos y Aprendizaje, TAMIDA2005, pp. 49-56, 2005.

[23] C. Romero y S. Ventura, «Data mining in education,» Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, pp. 12-27, 2013.

[24] J. Han, J. Pei y M. Kamber, Data mining: concepts and techniques, Elsevier, 2011.

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

[26] I.-C. Yeh y C.-h. Lien, «The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients,» Expert Systems with Applications, pp. 2473-2480, 2009.

[27] K. Lakshminarayan, S. Harp, R. Goldman y T. Samad, «Imputation of Missing Data Using Machine Learning Techniques,» KDD-96 Proceedings, pp. 140-145, 1996.

[28] B. Nguyen Cong, J. L. Rivero Pérez y C. Morell, «Aprendizaje supervisado de funciones de distancia: estado del arte,» Revista Cubana de Ciencias Informáticas, pp. 14-28, 2015.

[29] I. Guyon y A. Elisseeff, «An introduction to variable and feature selection,» Journal of machine learning research, pp. 1157-1182, 2003.

[30] K. Polat y S. Güneş, «A new feature selection method on classification of medical datasets: Kernel F-score feature selection,» Expert Systems with Applications, pp. 10367-10373, 2009.

[31] N. V. Chawla, K. W. Bowyer, L. O. Hall y W. Kegelmeyer, «SMOTE: Synthetic Minority Over-sampling Technique,» Journal of Artificial Inteligence Research 16, pp. 321-357, 2002.