

Predicción del comportamiento de ocupantes de edificios de oficina en base a variables de confort térmico mediante machine learning

Gastón Arias Aravena ¹ | Fredy Troncoso Espinosa ² | Jaime Soto-Muñoz ³ | Maureen Trebilcock Kelly ⁴

Recibido: 22-12-2022 | Versión final: 21-09-2023

Resumen

Los trabajadores de oficinas se encuentran la mayor parte del tiempo al interior de un edificio, y con ello, las variables físico-ambientales comienzan a presentar una importancia en la productividad y desempeño de ellos. Este estudio relaciona los modelos de *machine learning* con el comportamiento de los ocupantes y la productividad autoevaluada que presentan, mediante el uso de diferentes modelos. Estos modelos se implementaron para reconocer y comparar cuáles de ellos permiten estimar de mejor forma este comportamiento, en particular, la productividad autoevaluada que las personas sienten en su espacio de trabajo. Para ello, se recogieron las variables físico-ambientales y la percepción de los ocupantes de diversos edificios de oficina en la ciudad de Concepción. Este estudio logra comparar el desempeño de cuatro modelos de *machine learning* (árbol de decisiones, *K-Nearest Neighbor*, modelo de bayes y red neuronal), el desempeño de estos se midió mediante los indicadores denominados *Accuracy*, *Precision* y *Recall*. Estos modelos se implementaron tanto para una base de datos original como en una base de datos balanceada, para luego comparar los resultados obtenidos. Se puede establecer que existe una relación entre las variables físico-ambientales y la productividad autoevaluada de los trabajadores. Así mismo, se puede mencionar que la red neuronal es el modelo que mejor describe esta relación y, por ende, el que mejor desempeño logra. Este estudio permite un acercamiento a comprender el comportamiento de los ocupantes desde una perspectiva del *machine learning*.

Palabras clave: confort térmico; productividad; edificios de oficina; máquinas de aprendizaje

Citación

Arias Aravena, G. et al. (2023). Predicción del comportamiento de ocupantes de edificios de oficina en base a variables de confort térmico mediante machine learning. *ACE: Architecture, City and Environment*, 18(53), 11958. <https://doi.org/10.5821/ace.18.53.11958>

Predicting Occupant Behavior in Office Buildings Based on Thermal Comfort Variables Using Machine Learning

Abstract

Office workers spend most of their time inside a building, and as a result, physical-environmental variables begin to play a crucial role in their productivity and performance. This study establishes a connection between machine learning models and the behavior of occupants and the self-assessed productivity they exhibit, through the use of various models. These models were implemented to identify and compare which of them better estimate this behavior, particularly the self-assessed productivity that individuals experience in their workplace. To accomplish this, physical-environmental variables, and the perceptions of occupants from various office buildings in the city of Concepción were collected. This study successfully compares the performance of four machine learning models (decision tree, *K-Nearest Neighbor*, Bayesian model, and neural network). Their performance was measured using indicators known as *Accuracy*, *Precision*, and *Recall*. These models were applied to both an original database and a balanced database, followed by a comparison of the results obtained. It can be established that there is a relationship between physical-environmental variables and the self-assessed productivity of workers. Furthermore, it can be mentioned that the neural network is the model that best describes this relationship and, therefore, achieves the highest performance. This study provides an approach to understanding occupant behavior from a *machine learning* perspective.

Keywords: thermal comfort; productivity; office buildings; machine learning

¹ Magister en Ingeniería Industrial. (ORCID: [0000-0002-0488-6158](https://orcid.org/0000-0002-0488-6158)) Universidad del Bío-Bío. ² Doctor en Sistemas de Ingeniería. (ORCID: [0000-0002-9972-3123](https://orcid.org/0000-0002-9972-3123)) Universidad del Bío-Bío, ³ Doctor en Arquitectura. (ORCID: [0000-0002-5767-6189](https://orcid.org/0000-0002-5767-6189)) Universidad del Bío-Bío, ⁴ Doctora en Arquitectura. (ORCID: [0000-0002-1984-0259](https://orcid.org/0000-0002-1984-0259)) Universidad del Bío-Bío. Correo de contacto: garias@ubiobio.cl

1. Introducción

La calidad del ambiente interior tiene un efecto sobre el bienestar de las personas, y dentro de esto, la comodidad térmica es la variable que presenta mayor impacto en la comodidad de los ocupantes. La arquitectura, a través del diseño de los edificios, se enfoca en crear espacios adecuados para los ocupantes. Según la American Society of Heating, Refrigerating and Air-Conditioning Engineers-55, (2017) la sensación térmica de las personas es una expresión subjetiva de un individuo que evalúa si un entorno térmico específico puede considerarlo cómodo o no. En ello, la adaptación psicológica es importante en la sensación, aceptabilidad y preferencia térmica. La importancia antes mencionada, se basa en que la temperatura de confort se relaciona con los cambios en las expectativas térmicas debido a una combinación de niveles de control percibido y a una gran diversidad de experiencias térmicas en el edificio (Brager & Baker, 2009; Hellwig et al., 2006; Schweiker & Wagner, 2016). En otras palabras, la sensación de control sobre los equipos y mecanismos de acondicionamiento térmicos afecta a la temperatura de confort de los ocupantes. El uso del edificio y el comportamiento de los ocupantes son dinámicos y complejos, ya que la operación del edificio no siempre es la misma y tiene un efecto sobre los ocupantes. Autores han podido demostrar, por ejemplo, que el cruce de datos más significativo es el que se realizó entre la temperatura media efectiva exterior y la temperatura de neutralidad térmica interior (Mui & Chan, 2003; Trebilcock et al., 2020). Sin embargo, los profesionales dedicados al diseño y operación de edificios presentan dificultades para analizar la ocupación y comportamiento de las personas. Por lo tanto, se hace necesario buscar nuevas metodologías para la predicción del comportamiento de los ocupantes (Mora et al., 2021; Yan & Hong, 2018; Hong et al., 2016; García et al., 2013).

Además, si se consideran los efectos directos del confort térmico en trabajadores, por ejemplo, de edificios de oficinas, existen estudios que han demostrado una correlación entre la productividad y el entorno térmico en edificios de oficina (Haynes, 2008). Esta productividad en edificios de oficina es compleja de medir, dada la amplia gama de actividades que realizan los trabajadores (archivar, escribir, caminar por interiores, caminar por exteriores, conversar, etc.) y las diversas posiciones dentro de la organización. No obstante, la autoevaluación del trabajador es aceptada como una medida de productividad efectiva, en relación con el puesto de trabajo, ya que permite una comparación entre las diversas actividades laborales (Soto-Muñoz et al., 2022). En este sentido, aprovechando las herramientas de análisis complejo, la autoevaluación de productividad en oficinas se ha podido estudiar en relación con múltiples variables del espacio construido (Soto-Muñoz, et al., 2022). Hoy en día es posible recoger esta información, y con ello, generar bases de datos robustas que relacionen estas variables, sin embargo, las bases de datos por sí solas no generan un gran aporte, estas toman relevancia cuando permiten la extracción de información útil para describir y comprender el fenómeno originador de estos datos (Rodríguez & Amador, 2009).

Así entonces, el trabajador en la oficina presenta una interacción con muchas variables, lo que implica que pueden emerger grandes colecciones de datos. La comprensión de estos datos requiere de una serie de procedimientos y técnicas que se deben ir desarrollando de forma consecutiva con el objetivo de generar conocimiento útil (Romero et al., 2021). (Donkers et al., 2022) reconoce la importancia de las bases de datos que se originan en la información de los edificios, ya que permiten comprender el desempeño y predecir situaciones en el ámbito de la arquitectura, ingeniería y construcción. Así pues, este proceso se denomina KDD (Knowledge Discovery in Databases), el cual se puede definir como la búsqueda de patrones válidos, novedosos y útiles en bases de datos (Castro et al., 2018). Desde otro punto de vista, esta metodología transforma la información de bajo nivel a información significativa (Riquelme et al., 2006). El descubrimiento y la transformación se realiza mediante la aplicación de modelos de *machine learning* y minería de datos (Camana, 2016). En este sentido, los modelos de árbol de decisiones son significativos para describir y predecir la percepción de los ocupantes en su puesto de trabajo respecto del ambiente que los rodea.

La conducta de las personas dentro de una edificación presenta tres características principales; es estocástica, diversa y compleja (Yan et al., 2017). En general, los ocupantes y los edificios son sistemas dinámicos, y el comportamiento humano influye en el desempeño del edificio. Así, podemos seguir dos caminos; alterar el comportamiento de los ocupantes o bien mejorar la calidad de las predicciones (Darakdjian et al., 2019; Hong et al., 2017).

El concepto de confort térmico se relaciona con la respuesta de la percepción de los ocupantes con respecto al ambiente térmico al que se ven expuestos (Lusett et al., 2020). Básicamente, permite evaluar qué tan conforme se siente un ocupante con su espacio desde el punto de vista de las variables ambientales (Pao et al., 2022). Este concepto se puede dimensionar mediante el método de Fanger, que busca establecer un indicador del grado de satisfacción que presentan los ocupantes (Fanger & Toftum, 2002). Sin embargo, este método ha dado paso a la generación de modelos de confort térmico adaptativos. Esto se deriva de comprender que el comportamiento de los ocupantes es dinámico y si ven alterado su equilibrio térmico, buscarán reestablecerlo mediante una estrategia de adaptación (Marincic et al., 2012; Yao et al., 2009). Así mismo, las preferencias térmicas juegan un papel fundamental en esta valoración. Dentro de los factores claves que afecta esta sensación de confort se encuentra la temperatura del aire y el lugar del cuerpo del ocupante donde impacta (Chludzińska & Bogdan, 2015; Alireza et al., 2019).

En general, medir la productividad en oficinas es una tarea compleja, esto por la naturaleza de la producción que se genera, entendiéndose que no son tangibles. Con ello, el indicador de productividad autoevaluada se considera una medida válida (Byrd & Rasheed, 2016; Lipczynska et al., 2018). Así mismo, esta productividad se puede ver afectada por el entorno físico de trabajo (Fassoulis & Alexopoulos, 2015). Con ello, se puede establecer que mejorar las condiciones y la calidad del ambiente interior trae consigo un potencial aumento en la productividad (Esfandiari et al., 2017) (Budaiova & Vilcekova, 2015). Roskams & Haynes (2019), plantean que los trabajadores tienden a distraerse más con un menor confort térmico derivando en una reducción en la productividad.

Ahora bien, se han planteado algunos estudios que relacionan la gestión y el confort térmico del espacio construido con modelos de *machine learning* y minería de datos (Gao et al., 2010; Kim et al., 2018). Así también, se estos modelos se han utilizado para determinar la importancia de las variables físico-ambientales (Du et al., 2019) y establecer patrones de interacción de los ocupantes con el espacio físico construido (D'Oca & Hong, 2014).

Las técnicas de minería de datos ofrecen una amplia gama de modelos de *machine learning* para implementar. No obstante, los modelos que permiten una mejor comprensión y relación de las variables físico-ambientales son: Árbol de Decisiones, Modelo K-NN, Modelo de Naive Bayes y la Red neuronal (Zhang et al., 2022). Estos modelos presentan distintos algoritmos para predecir y clasificar las variables (Berlanga et al., 2013; García & Gómez, 2006; Webb, 2010; Salas et al., 2020).

Este estudio tiene por objetivo proponer un modelo de *machine learning* para predecir la productividad autoevaluada de los trabajadores de oficina en la ciudad de Concepción, Chile. Permitiendo generar un acercamiento novedoso, desde la mirada de los modelos de predicción. Para ello, se confeccionó una base de datos, la cual posteriormente fue caracterizada desde el punto de vista de las variables térmicas para relacionar dichas variables con la productividad autoevaluada de los trabajadores. Así mismo, se seleccionó las variables que presentaron un mayor poder predictivo. Finalmente se implementaron cuatro tipos de modelos (árbol de decisiones, modelo K-NN, modelo de Bayes y red neuronal), con el objetivo de comparar dichos modelos en función de los indicadores: *Accuracy*, *Precision* y *Recall*.

2. Metodología de trabajo

2.1 Colección de datos

Este estudio utilizó una estrategia multicaso basada en un modelo no experimental, adoptando un diseño transversal y descriptivo. Se estudiaron 9 edificios de oficinas en la ciudad de Concepción, Chile. Se realizaron observaciones y se recopilaron datos de variables físico-ambientales, correspondientes a condiciones tales como; la temperatura del aire, la temperatura media radiante, la velocidad del aire, la humedad relativa. Además, este fenómeno fue examinado en el espacio de trabajo a través de la percepción de los ocupantes, tanto térmica como psicológica, mediante encuestas en el puesto de trabajo. Se tomaron precauciones para no intervenir en el curso natural de la jornada laboral de las oficinas. En el confort térmico, las características climáticas tienen un efecto directo en la comodidad de la persona, por lo que esta investigación abordó la evaluación en condiciones de invierno, primavera y verano. A ello, se suma múltiples consideraciones contextuales que, en esta investigación, implican una perspectiva cultural chilena representada por una selección de casos de organizaciones gubernamentales y privadas.

Para elaborar la base de datos se consideraron nueve edificios ubicados en diversas comunas de la ciudad de Concepción, Chile (Figura 1). La selección de los casos de estudio se generó a partir de cuatro criterios principales; Construcciones desde el año 1997 hacia adelante, edificaciones que no hayan sufrido modificaciones, cantidad de ocupantes mayor o igual a 20 personas y que el destino del edificio fuese oficinas.

Figura 1. Distribución de ubicación de los casos de estudio en Chile



Fuente: Elaboración propia a partir de planimetría de Google Maps y mapas de la región.

La ciudad de Concepción corresponde a un clima templado cálido con lluvias invernales y gran humedad atmosférica según la clasificación Köppen. Presenta una temperatura exterior media anual que se encuentra entre los 8,8°C y los 16,3°C y una oscilación media diaria de 7,3°C. Tiene una población de más de 200.000 habitantes.

La colección de datos se generó con las mediciones obtenidas de los distintos equipos e instrumentos que fueron posicionados en diferentes lugares del edificio (Figura 2). Estos equipos fueron previamente calibrados antes de la realización de este estudio. Estos equipos recogieron mediciones de variables tales como; la temperatura del aire, la velocidad del aire, la humedad relativa, la temperatura húmeda natural, la temperatura de globo y temperatura de bulbo húmedo.

Figura 2. Ubicación de equipos de medición en el edificio



Fuente: Elaboración propia.

Luego, estos datos fueron relacionados con variables de percepción de momento del ocupante, las cuales fueron recogidas a través de la encuesta de percepción de momento. Dicha encuesta logra recoger la percepción y preferencia de los ocupantes en variables y atributos como; la sensación y preferencia térmica, la sensación y preferencia en el movimiento, la calidad del aire, el uso o modificación de los equipos de acondicionamiento climático, la productividad autoevaluada, entre otras. Así mismo, se identificaron y caracterizaron cada uno de los trabajadores entrevistados. El estudio recopiló información durante las estaciones de invierno, primavera y verano. En resumen, se generó una colección de datos compuesta por 54 variables (Tabla nº1) para un total de 584 trabajadores, lo que permitió generar una base de datos con 2.604 observaciones.

Tabla 1. Variables medidas y encuestadas con su abreviatura presentes en la base de datos

Variable	Descripción	Variable	Descripción
Estación	Estación climática en la que se realiza el levantamiento.	A_VrI	El trabajador realizó modificaciones de control al ventilador individual.
Ciudad	Ciudad a la que pertenecen los datos.	A_VrC	El trabajador realizó modificaciones de control al ventilador compartido.
Caso	Identificación del edificio en estudio.	A_CrI	El trabajador realizó modificaciones de control al calefactor individual.
HT	El edificio cuenta con calefacción.	A_CrC	El trabajador realizó modificaciones de control al calefactor compartido.
HVAC	El edificio cuenta con ventilación y aire acondicionado.	A_T	El trabajador realizó modificaciones de control al termostato.
MM	El edificio cuenta con modo mixto.	Tw (C)	Temperatura húmeda natural.
Trimestre	Informa el trimestre donde se realiza la encuesta.	Tg (C)	Temperatura de Globo.
Hora	Jornada del día donde se realiza la encuesta	Ta (C)	Temperatura del Aire.
TSV	Voto de sensación térmica del ocupante.	RH (%)	Humedad Relativa.
TPV	Preferencia térmica de trabajador.	Va (m/s)	Velocidad del Aire.
AcAir	Aceptabilidad de la calidad del aire trabajador en su puesto de trabajo.	Tr (C)	Temperatura operativa.

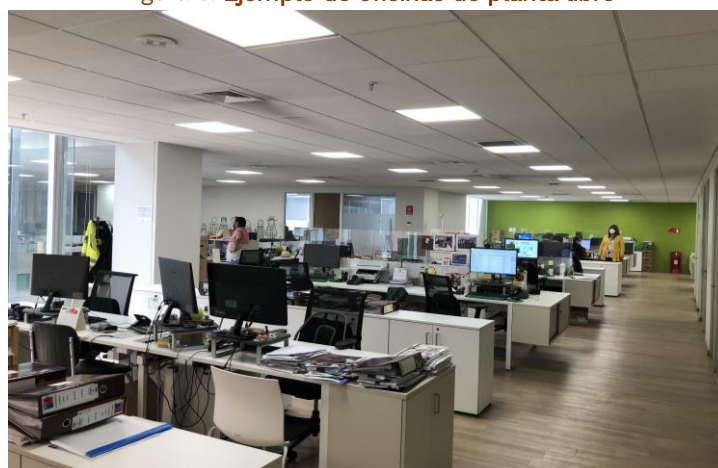
AcTer	Aceptabilidad de ambiente térmico del trabajador en su puesto de trabajo.	WBGT-i (C)	Temperatura de bulbo húmedo IN: En presencia o ausencia de irradiación solar.
AcLuz	Aceptabilidad lumínica que recibe el trabajador en su puesto de trabajo.	WBGT-o (C)	Temperatura de bulbo húmedo OUT: En presencia de irradiación solar.
AcRui	Aceptabilidad del nivel de ruido del trabajador en su puesto de trabajo.	PMV	Predicted Mean Vote (Voto Medio Previsto).
MA	Percepción del movimiento del aire por el trabajador.	PPD (%)	Porcentaje de Personas insatisfechas.
Prod	Productividad autopercebida por el trabajador.	EXT Temp (°C)	Temperatura Exterior.
CLO	Arropamiento que presenta el trabajador.	TipOf	Establece el tipo de oficina.
MET	Actividad que estuvo realizando el trabajador en los últimos 15 minutos.	Gen	Genero del trabajador.
AT_Vn	Identifica operación de ventanas o ventilación natural.	Edad	Edad del trabajador.
AT_SC	Identifica operación de Aire acondicionado.	Horing	Horario de ingreso.
AT_Vr	Identifica operación de ventilador.	HorSCol	Horario de salida a colación.
AT_CC	Identifica operación de losa radiante.	HorECol	Horario de ingreso a colación.
AT_Cr	Identifica operación del calefactor.	HorSal	Horario de salida.
A_V	El trabajador realizó modificaciones de control a las ventanas.	TiemEst	Tiempo que lleva trabajando en la estación de trabajo.
A_P	El trabajador realizó modificaciones al control a las persianas o cortinas.	%jorn	Porcentaje de la jornada que permanece en la estación de trabajo.
A_Lm	El trabajador realizó modificaciones al control de alguna lampara individual.	TConf	Temperatura de Confort.
A_L	El trabajador realizó modificaciones de control a la luminaria general.	MAP	Preferencia del movimiento del aire del trabajador.

Fuente: Elaboración propia.

En general, la mayoría de los edificios presentes en este estudio utilizan un sistema de acondicionamiento climático mixto (calefacción y aire acondicionados), llegado a alcanzar el 51,1% de los datos. Luego, el uso únicamente de aire acondicionado llega al 25,9%, y en último lugar, los edificios que solo contemplan calefacción alcanzan el 23,0% del total.

Por otro lado, se observa que existe una cantidad similar de mujeres y hombres entre los trabajadores en todas las oficinas, alcanzado una distribución total de 51,3% de hombres y un 48,7% de mujeres. La distribución de edades está encabezada por trabajadores que tienen entre los 36 a los 45 años, representando el 36,6% de las observaciones, seguido por el rango de 26 a 35 años, con un 24,8%. Estos trabajadores están distribuidos en su mayoría en plantas libres, lo que para este estudio representa el 75,6% del total de casos (Figura 3).

Figura 3. Ejemplo de oficinas de planta libre



Fuente: Elaboración propia.

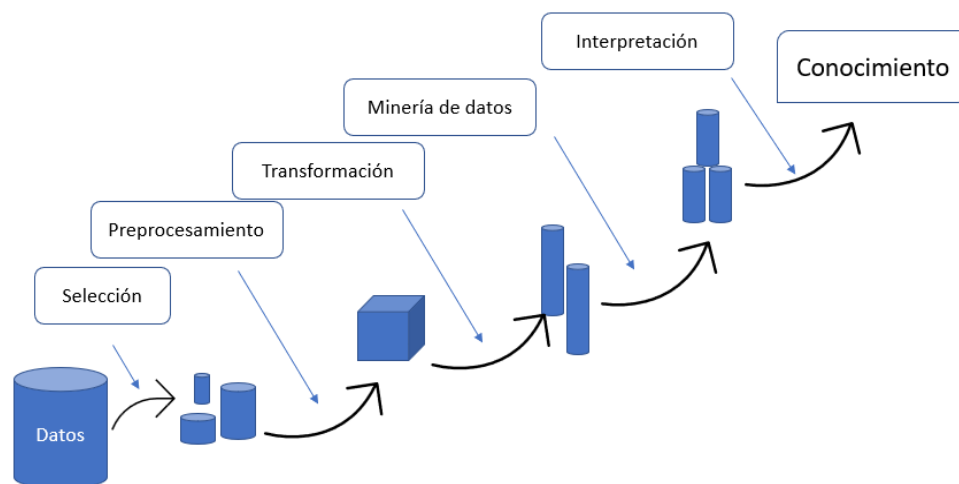
Luego, cuentan con oficina compartida el 17,5% de las observaciones. Finalmente, con oficina privada existe un 7,3% de los datos. Esto implica que, en su mayoría, los trabajadores comparten condiciones ambientales similares en su espacio de trabajo.

Los trabajadores de oficina perciben, en general, su productividad como normal, alcanzado el 81,2% de los datos. Se puede definir normal, como una condición habitual asumida de forma intrínseca por los ocupantes en el desempeño de sus funciones (Soto-Muñoz et al., 2022b). Aquellos que considera su productividad por encima de lo normal llegan tan sólo a un 8% y, por último, aquellos que tiene una percepción de productividad bajo de lo normal son un 10,8%. En los casos analizados, los trabajadores expresan una sensación térmica en general neutra y se inclinan a tener una leve tendencia hacia el cálido. Esto se ve reflejado en que el 62,8% de las observaciones se concentran entre una sensación ligeramente fresca, neutra y ligeramente cálida. Por otro lado, la preferencia térmica de los trabajadores también se aprecia en general como sin cambios, alcanzado el 50% de los datos. No obstante, prefieren un ambiente un poco más frío. Esto concuerda con otros estudios que establecen que los trabajadores prefieren una temperatura más fría en su ambiente de trabajo.

2.2 Predicción del comportamiento de ocupantes

Los modelos de *machine learning* son modelos de clasificación y/o predicción que aprenden el patrón oculto de los datos y, luego, generan la predicción (Espinosa, 2020). Para poder llegar a la implementación de estos modelos, se requiere de la aplicación de la metodología KDD (Knowledge Discovery in Databases). Esta metodología consiste en una serie de procedimientos que se aplican de forma secuencial a una base de datos para generar conocimiento útil (Figura 4) (Romero et al., 2021). Para llevar a cabo esta aplicación, es necesario primero estudiar y caracterizar la base de datos.

Figura 4. Proceso KDD



Fuente: Elaboración propia.

Del estudio y caracterización de la base de datos, para este estudio, se puede establecer que existen datos categóricos binominales, polinominales y numéricos. Estos últimos de naturaleza continua y entera. Es importante dejar en claro que este estudio busca predecir la productividad autoevaluada de los trabajadores, por lo tanto, a esta variable se le denomina *label*, *etiqueta* o *variable a predecir*. Desde esa perspectiva, se puede establecer que esta variable es de tipo categórica ordinal.

Se construyó a partir de la pregunta *¿Cómo considera su productividad actual en su puesto de trabajo?* Este atributo puede tomar hasta cinco valores, como se aprecia en la Tabla 2.

Tabla 2. Distribución de respuestas de productividad autoevaluada

Valor asociado	Respuesta entregada	Distribución de respuestas
-2	Mucho menos que lo normal	0,5%
-1	Un poco menos que lo normal	7,5%
0	Normal	81,2%
1	Un poco más que lo normal	8,9%
2	Mucho más que lo normal	1,8%

Fuente: Elaboración propia, en base a la escala de Likert.

Luego, se analiza la relación entre la variable de productividad autoevaluada y la estación climática. Para esto, se utiliza el método estadístico de Chi-Cuadrado, el cual, mediante una comparación entre el cálculo teórico y empírico, determina si una variable depende de otra (Lastre et al., 2018). Con un total de 8 grados de libertad (valor obtenido de la matriz de comparación de variables) y un nivel de significancia del 5%, y con el valor de $X_{cal}=24,4$ comparado con $X_{teo}=15,5$, se concluye que existe una dependencia. Sin embargo, al estudiar la asociación entre estas variables mediante el coeficiente V de Cramer, se obtiene un valor de 0,07, el cual resulta estadísticamente no significativo. Por lo anterior, se descarta una dependencia significativa entre las dos variables. Este proceso se realiza para determinar si es necesario la implementación de modelos independientes por estación, llegando a concluir que no es necesario esta consideración, ya que no existe una dependencia.

2.3 Preprocesamiento de los datos

Para la implementación de modelos de machine learning, es necesario llevar a cabo la limpieza de la base de datos con el objetivo de identificar valores fuera de rango o anómalos, lo que comúnmente se denomina ruido. Del mismo modo, reemplazar datos nulos o perdidos en la base de datos.

En cuanto a los valores fuera de rango, se refiere a aquellos datos que están más lejanos al resto de datos. Estos valores pueden originarse debido a errores en las mediciones o a una concentración inusual de datos muy alejados de la tendencia central. Para este estudio, se aplicó la técnica de la media y la desviación estándar. Esto implica eliminar los datos que se encuentren más alejados de un cierto límite, tanto por encima como por debajo de la media, basándose en un factor amplitud aplicado a la desviación estándar de los datos.

En el proceso de manejo de datos, se abordan los valores faltantes, valores atípicos y las inconsistentes presentes en la base de datos, siguiendo criterios estadísticos de tratamiento de variables (Vallejo Ballesteros et al., 2018). En la Tabla 3 se aprecian los datos faltantes, (luego de la limpieza) en la colección de datos de este estudio.

Tabla 3. Total de valores nulos

Variable	Valores Nulos	Variable	Valores nulos
TSV	17	PMV	292
TPV	17	PPD (%)	308
CLO	130	EXT Temp (°C)	332
Tw (C)	234	Gen	61
Tg (C)	234	Edad	70
Ta (C)	172	HorIng	62
RH (%)	298	HorSCol	90
Va (m/s)	342	HorSal	70
Tr (C)	234	TiemEst	61
WBGT-i (C)	308	%jorn	62
WBGT-o (C)	308	TConf	234

Fuente: Elaboración propia.

Posteriormente, estos valores faltantes se relacionaron con el resto de las variables y observaciones disponibles en la base de datos utilizando la técnica *Hot Deck*. Este procedimiento implica identificar la observación más similar y reemplazar el dato faltante con el valor presente en la observación seleccionada (Useche et al., 2022). Como resultado, se obtuvo una base de datos completa, sin registros faltantes; fundamental para su procesamiento en diversas herramientas o softwares.

2.4 Transformación de las variables

Para relacionar algunas variables presentes en la base de datos, se realizó la transformación o generación de nuevos atributos con el objeto de enriquecer la información contenida en la misma. A modo de ejemplo, el atributo fecha que se encontraba en la base de datos original se transformó al atributo de mes y año (Tabla 4). Se optó por esta transformación porque trabajar con fechas dificulta el procesamiento de datos.

Tabla 4. Transformación variable fecha

Original	Transformación	
Fecha	Mes	Año
27-06-2017	Junio	2017

Fuente: Elaboración propia.

Otro ejemplo de transformación de atributos fue el realizado a la variable original denominado hora, el cual también fue transformado para permitir un mejor análisis (Tabla 5). Aquí se establecieron rangos horarios para la transformación.

Tabla 5. Transformación variable hora

Desde	Hasta	Categoría
08:30	11:59	Mañana
12:00	15:59	Medio día
16:00	18:00	Tarde

Fuente: Elaboración propia.

En general, ambos atributos mantuvieron su naturaleza categórica durante el proceso de transformación de variables. Además de lo anterior, esta adaptación es necesaria para la implementación de diversos modelos, dado que algunos operan únicamente con datos categóricos o numéricos. Este proceso también implica la recategorización de los atributos y las variables presentes en la base de datos. Así, para llevar a cabo un análisis de las variables térmico-ambientales, así como las preferencias y percepciones de los ocupantes, utilizando las técnicas propias de la minería de datos, es necesario en una primera instancia convertir todas las variables de la base de datos a una forma categórica.

En el caso particular del atributo a predecir, se realiza una transformación de variable, agrupando las observaciones en solo dos categorías; productividad *inferior* para aquellas respuestas que son bajo productividad normal y productividad *normal* para aquellas respuestas que hayan sido normal o superior (Tabla 6). Esta reducción de categorías se realiza para implementar los diversos modelos de *machine learning*, y se basa en que una persona productiva se entiende como aquella que cumple los objetivos, por lo tanto, es posible reducir a 2 categorías: aquellas que cumplen (*normal*) y las que no (*inferior*).

Tabla 6. Recategorización de atributo

Categorías	Mucho menos de lo normal	Un poco menos de lo normal	Normal	Un poco más de lo normal	Mucho más de lo normal
Valor	-2	-1	0	1	2
Re-categorización	Inferior		Normal		

Fuente: Elaboración propia en base a escala de Likert.

2.5 Selección de variables

En este apartado, se busca obtener el subconjunto de variables que tenga la mayor relación con la variable a predecir. Para la selección de las variables, se utilizan indicadores estadísticos, siendo el primero chi-cuadrado, el cual se describe anteriormente. El indicador de ganancia de información mide la entropía de un variable. Por lo tanto, cuanto mayor sea la ganancia de información, más importante será la variable para la clasificación. Por último, se encuentra el índice de Gini que, si bien su campo de aplicación principal está relacionado con los ingresos, dentro de un país, también permite establecer una medida de desigualdad entre dos variables (Prasad et al., 2013). Badulescu (2020) estudió algunas variaciones del índice de Gini para la implementación en un modelo de árbol de decisiones. Con todo lo anterior, en la Tabla 7 se registran las 3 variables que tienen una mayor importancia a la hora de predecir la productividad autoevaluada.

Tabla 7. Atributos con mayor poder predictivo para la productividad

Nº	Atributo	Chi-Cuadrado	Ganancia de información	Índice de Gini
1	TPV	98,803	0,005	0,022
2	TSV	95,818	0,005	0,018
3	AcTer	91,173	0,005	0,017

Fuente: Elaboración propia.

2.6 Implementación de modelos y análisis de patrones

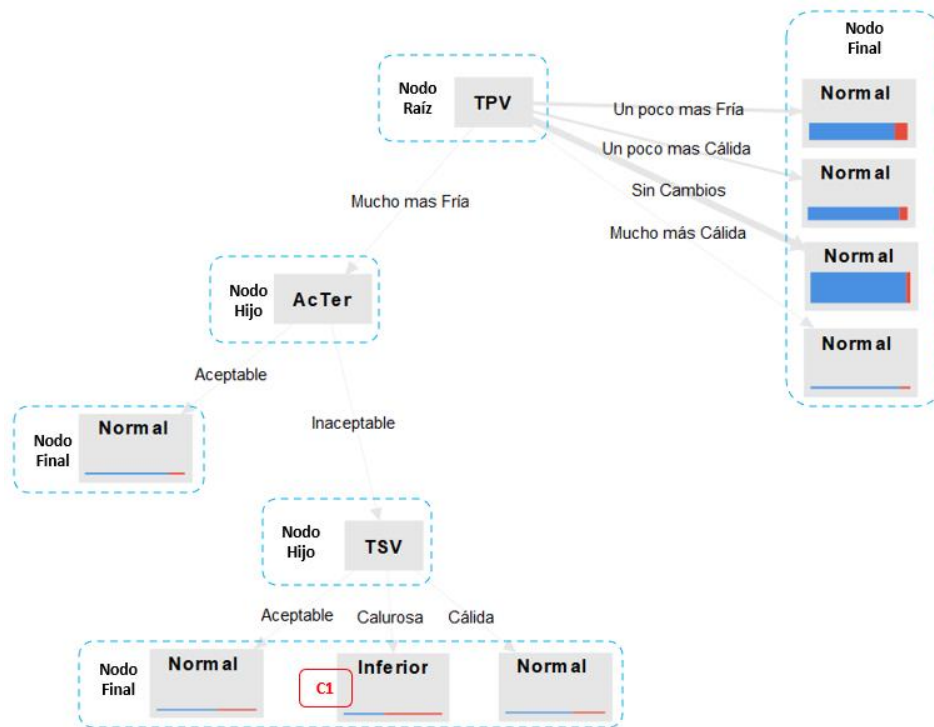
El primer modelo implementado es el Árbol de decisiones. Este es un modelo de clasificación que se basa en reglas y se estructura en forma de esquemas. Comienza con un nodo raíz y luego se generan nodos hijos con ramas que representan los valores de la variable. Para terminar, se alcanzan los nodos finales. Estos nodos finales se crean si el valor del atributo entrega la predicción del *label*, de lo contrario, continúa con el siguiente atributo. Esta ramificación se produce hasta la clasificación de todos los *label* (Barrientos et al., 2009; Perez-Pareja et al., 2020; Charbuty & Abdulazeez, 2021). La implementación del Árbol de Decisiones como primero modelo nos proporciona una visión general de lo que está describiendo la base de datos (Figura 5).

El patrón general de la base de datos sugiere que la productividad autoevaluada por los trabajadores es *normal en casi todos los casos*. Sin embargo, cuando los trabajadores experimentan una sensación térmica *calurosa*, su productividad disminuye y se vuelve *inferior* (C1 – Figura 6). Este fenómeno ha sido discutido previamente en la literatura, asociándose a un factor de estrés ambiental causado por una baja en el rendimiento del trabajador (Lamb & Kwok, 2016). Así mismo, podría deberse a la fatiga que podría sentir el trabajador debido a la temperatura calurosa. Además, de acuerdo con el modelo de confort térmico adaptativo, aquellos trabajadores que tienen una sensación *calurosa* tienden a preferir un ambiente térmico *mucho más frío*. Con base en el análisis del patrón general, se establece además como la clase de interés, aquellas observaciones que quedan clasificadas como *inferior*.

El siguiente modelo se denomina K-NN (K-Nearest Neighbor), es un clasificador basado en aprendizaje por analogía, que evalúa una observación y luego mediante distancia euclidiana identifica las observaciones que más se asemejan a la seleccionada, para luego realizar la clasificación (Madariga et al., 2022).

Para el modelo K-NN implementado en este estudio, se hizo variar el parámetro que establece la cantidad de observaciones cercanas a considerar para la clasificación, desde 1 hasta 10. Esto permitió establecer que el número óptimo de observaciones cercanas a considerar eran nueve. En otras palabras, para que modelo pudiera clasificar la clase de interés necesitó considerar las nueve observaciones cercanas.

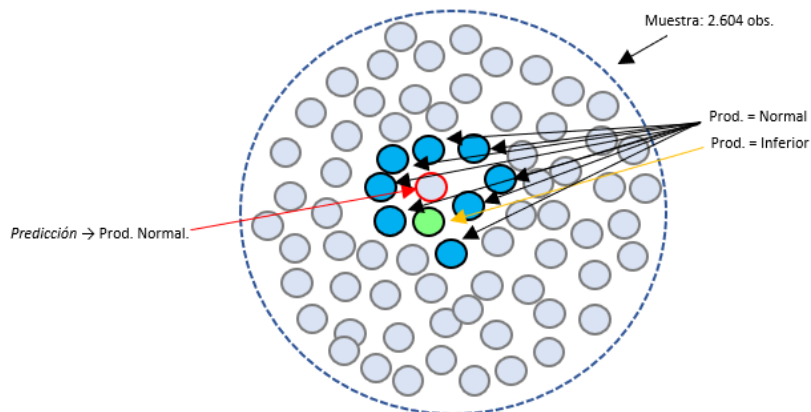
Figura 5. Árbol de decisiones para la base de datos en estudio



Fuente: Elaboración propia. En base a herramienta de análisis de datos RapidMiner.

En la Figura 6 se realiza una representación de la forma de clasificar del modelo K-NN. Allí, de los nueve vecinos cercanos a la predicción, ocho presentan una productividad *normal*, por lo tanto, la predicción será *normal*.

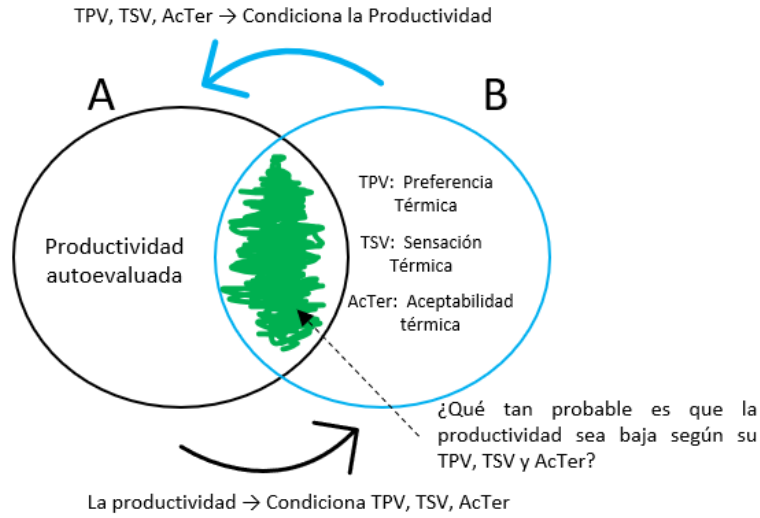
Figura 6. Esquema modelo K-NN para la base de datos estudiada



Fuente: Elaboración propia.

Por otro lado, el modelo Naive Bayes, es un modelo de clasificación probabilístico. Básicamente, analiza la probabilidad de que una observación pertenezca o no a una clase y se basa en el teorema de Bayes (Hernández et al., 2022) . Luego este modelo, es implementado en la herramienta de análisis de datos. En la Figura 7, se puede representar lo que está ocurriendo dentro del modelo. En otras palabras, lo que el modelo realiza es calcula la probabilidad. ¿Teniendo cierta preferencia, sensación y aceptabilidad térmica que tan probable es que la productividad autoevaluada sea inferior?

Figura 7. Esquema modelo de Naive Bayes para la base de datos estudiada

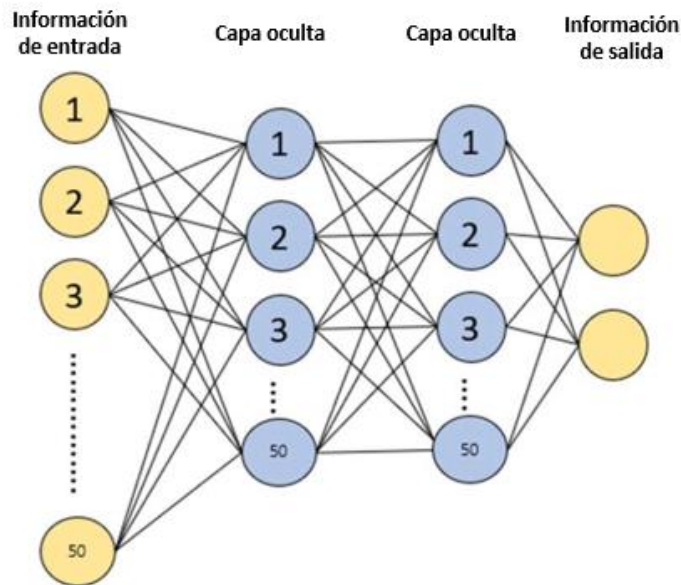


Fuente: Elaboración propia.

Por último, la red neuronal es un tipo de modelo por algoritmos que tratan de replicar el funcionamiento de una red neuronal biológica. Presenta una entrada de red la cual recibe la información y esta se va propagando hacia el resto de las neuronas. Dentro de los parámetros de red neuronal, al aumentar el número de capas ocultas mejoran la capacidad para de clasificación para problemas no lineales. Luego, esta se va ajustando con un algoritmo de retropropagación mediante ciclos de entrenamiento y fijando una tasa de aprendizaje (Salgado et al., 2018; Rady & Anwar, 2019; Sun et al., 2018) . Así, es capaz de predecir y clasificar una clase.

Para la base de datos, se realizan diversas iteraciones en los parámetros de la red neuronal como: cantidad de capas ocultas, tasas de aprendizaje y ciclos de entrenamiento. De estas diversas iteraciones se logró establecer que la red neuronal que generaba un mejor desempeño constaba de: una entrada de 50 nodos, dos capas ocultas, una tasa de 0,1 y 500 ciclos (Figura 8).

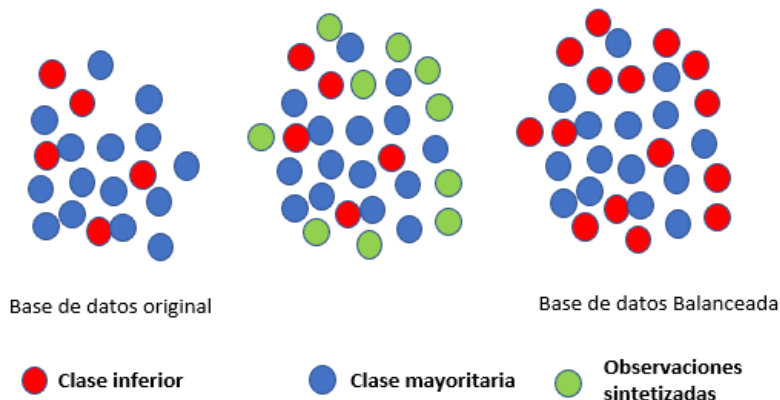
Figura 8. Representación Red Neuronal para la base de datos estudiada



Fuente: Elaboración propia.

Para un mejor análisis de los modelos, se implementó en la herramienta, de análisis de datos, un balance de las observaciones mediante la aplicación de la técnica SMOTE (Synthetic Minority Over-Sampling Technique – Técnica de Sobremuestreo de Minorías Sintéticas). Esta técnica de sobremuestreo genera observaciones sintéticas o artificiales con el objetivo de equilibrar la base de datos para la clase negativa como para la positiva (Spositto et al., 2020) (Figura 9).

Figura 9. Representación balance de colección de datos



Fuente: Elaboración propia.

2.7 Indicadores de comparación para los modelos

Los indicadores que se utilizaron para comparar los modelos son: *Accuracy*, *Precision* y *Recall*. El *Accuracy* representa las predicciones o clasificaciones correctamente realizadas $(VP+VN) / (VP+FP+FN+VN)$ (Tabla nº5). El *Precision* es un indicador que refleja una relación entre los elementos clasificados de forma correcta como positivos y el total de elementos clasificados como positivos $(VP / (VP+FN))$ (Tabla 5). El *Recall*, por su parte, corresponde a la relación entre las observaciones de la clase positiva que fueron correctamente clasificadas $(VP/(VP+FP))$ (Tabla 8) (Troncoso-Espinosa & Castro-Albornoz, 2021). Estas relaciones se generan a través de una matriz de confusión (Tabla 8).

Tabla 8. Matriz de confusión para una variable con dos categorías

Clases		Valor Real	
		Positivo	Negativo
Valor Predicho	Positivo	Verdadero Positivo (VP)	Falso positivo (FP)
	Negativo	Falso Negativo (FN)	Verdadero negativo (VN)

Fuente: En base a una matriz de confusión.

Los indicadores de clasificación y/o predicción se realizaron mediante la aplicación de la técnica *Hold-Out*. Esto divide la base de datos en un conjunto de prueba y un conjunto de entrenamiento. El primero nos permitirá identificar que tan bien logra predecir la variable en cuestión, y el segundo conjunto de datos permite el aprendizaje y entrenamiento del modelo (Maillard et al., 2021). En esta base de datos se utiliza la clase de interés la productividad definida como *inferior*. Para los modelos implementados para la base de datos original (sin ser balanceada), se generan las siguientes matrices de confusión:

Tabla 9. Matriz de confusión para el árbol de decisiones

Clases		Valor Real	
		Verdadero normal	Verdadero inferior
Valor Predicho	Pred. Normal	714	60
	Pred. Inferior	4	3

Fuente: Elaboración propia en base a RapidMiner.

De la Tabla 9, se puede determinar los indicadores correspondientes para el árbol de decisiones, el *Accuracy* corresponde a 91,98% (717/781), el *Precision* para la clase de interés alcanza el 42,8% (3 / (4+3)) y el *Recall* alcanza el 4,76% (3 / (3+60)).

Tabla 10. Matriz de confusión para el modelo K-NN

Clases		Valor Real	
		Verdadero normal	Verdadero Inferior
Valor Predicho	Pred. Normal	695	50
	Pred. Inferior	23	13

Fuente: Elaboración propia en base a RapidMiner.

Para el modelo K-NN las predicciones quedan registradas en la Tabla 10. Presenta los siguientes valores de indicadores; el *Accuracy* corresponde a 90,65% (708/781), el *Precision* alcanza el 36,11% (13 / (23+13)) y el *Recall* alcanza el 20,53% (13 / (13+50)).

Tabla 11. Matriz de confusión para el modelo de Bayes

Clases		Valor Real	
		Verdadero normal	Verdadero Inferior
Valor Predicho	Pred. Normal	699	49
	Pred. Inferior	19	14

Fuente: Elaboración propia en base a RapidMiner.

El modelo de Bayes entrega las predicciones plasmadas en la Tabla 11. Presenta los siguientes valores de indicadores; el *Accuracy* corresponde a 91,29% (713/781), el *Precision* alcanza el 42,42% (14 / (14+19)) y el *Recall* alcanza el 22,22% (14 / (14+49)).

Tabla 12. Matriz de confusión para el modelo de Red Neuronal

Clases		Valor Real	
		Verdadero normal	Verdadero Inferior
Valor Predicho	Pred. Normal	715	59
	Pred. Inferior	3	4

Fuente: Elaboración propia en base a RapidMiner.

Para la red neuronal, logra predecir de acuerdo con los registros de la Tabla 12. Presentando los siguientes valores para cada indicador; el *Accuracy* corresponde a 92,06% (719/781), el *Precision* alcanza el 57,14% (4 / (4+3)) y el *Recall* alcanza el 6,35% (4 / (4+59)).

3. Discusión de resultados y comparación de modelos

La comparación de los modelos se realizó mediante indicadores estadísticos de predicción y/o clasificación, que son: *Accuracy*, *Precision* y *Recall*. Este análisis se llevó a cabo tanto para la base de datos original como para la base de datos balanceada (Tabla 13).

Tabla 13. Desempeño predictivo para diversos modelos

Modelo implementado	B.D. Original			B.D. Balanceada		
	Accuracy	Precision	Recall	Accuracy	Precision	Recall
Árbol de decisiones	91,98%	42,86%	4,76%	67,97%	66,93%	71,03%
K-NN	90,65%	36,11%	20,63%	63,16%	71,24%	44,15
Naive Bayes	91,29	42,42%	22,22%	65,25%	74,50%	46,38%
Red Neuronal	92,06%	57,14%	6,35%	91,99%	90,58%	93,73%

Fuente: Elaboración propia.

Como era de esperar, la base de datos original tiende a predecir la productividad como *normal*. Esto se debe a que casi el 90% de los datos corresponde a esa categoría (estimación posterior a recategorización). La clase de interés *inferior* es más difícil de predecir, ya que se puede apreciar un bajo valor en los indicadores de *Accuracy*, *Precision* y *Recall*. El *Accuracy* promedia el 92%, en la base de datos original; en general, se aprecia alto debido a la naturaleza de los datos, ya que clasifica a la mayoría de las observaciones como *normal*. Por lo tanto, la mayoría de estas observaciones corresponden a esa categoría. Este valor disminuye casi un 20% en la base de datos balanceada, alcanzando el 72% en promedio, pero permite generar un incremento en los demás indicadores de predicción, que en ese caso se enfocan en la clase de interés.

El *Precision*, por su parte, tiene un promedio cercano al 44%, un valor bajo de desempeño para la base de datos original. En la base de datos balanceada, mejora a casi un 76% en promedio. Por último, el *Recall*, en la base de datos balanceada, es el indicador más bajo. Esto se debe a que no puede clasificar de buena forma la clase de interés debido a la baja cantidad de observaciones que se presentan. Los modelos tienen en promedio un 13,49% en dicho indicador, el cual mejora notoriamente al sintetizar más observaciones, alcanzando casi el 63%.

En el caso de modelo K-NN, dada la naturaleza del algoritmo que clasifica por analogía, teniendo en cuenta la baja cantidad de observaciones de la clase de interés, el modelo no puede predecir de buena forma dicha clase. Sin embargo, para la base de datos balanceada, su desempeño mejora en gran medida debido a que se presentan más observaciones de forma sintética. Esto es muy similar a lo que ocurre con el modelo de Bayes, ya que las probabilidades de que una observación pertenezcan a la clase de interés son muy bajas, y tiende a clasificar las observaciones como *normales*. Así también, ve mejorado su desempeño para la base de datos balanceada.

Por último, la red neuronal presenta el indicador de *Precision* más bajo de todos los modelos, pero su *Recall* presenta un gran desempeño, ambos para la base de datos original. Esto implica que el modelo en general puede predecir de mejor forma la clase de interés, a pesar de la baja cantidad de observaciones de dicha clase. Para la base de datos balanceada, este es el modelo que alcanza el mejor desempeño. En ese sentido, se podría decir que, en general, la red neuronal es un modelo que, a diferencia de los otros, puede encontrar patrones que son un poco más difíciles de ver a simple vista. Esto se puede deber a que opera de manera similar a neuronas humana, lo que le permite ampliar la gama de agrupaciones y/o clasificaciones. Además de lo anterior, aplica el algoritmo de retropropagación, que se va ejecutando de forma iterativa para mejorar las predicciones.

La importancia de este estudio radica en generar un aporte a la comprensión del comportamiento de los ocupantes, sus preferencias y cómo su entorno térmico afecta su productividad. Esto se logra mediante una mirada desde la perspectiva de la minería de datos y los modelos de *machine learning*. Poder predecir la productividad de los trabajadores de oficina puede permitir generar indicadores de desempeño de la compañía, así como anticipar los problemas que podrían estar afectado a la productividad de los trabajadores. Dicho sea de paso, estos problemas puede que no sean tan claros debido la naturaleza de las tareas que realizan los trabajadores de oficinas, como se menciona en capítulos anteriores. Con todo lo anterior, este estudio brinda un acercamiento a la comprensión de los patrones de comportamiento de los ocupantes de edificios.

4. Conclusiones

Se puede apreciar que las variables de sensación, preferencia y el ambiente térmico, intervienen en la productividad autoevaluada del trabajador. Esto se debe a que las tres primeras variables permiten predecir la productividad autoevaluada del trabajador. La implementación de estos métodos y técnicas podrían traer consigo la generación de nuevos criterios para el diseño, bienestar térmico y ergonómicos de los ocupantes, con un enfoque en la productividad de los trabajadores. En su mayoría,

los trabajadores de oficina perciben el aire quieto, pero desean un mayor movimiento del aire. Esto podría deberse a un factor psicológico, en el cual el trabajador busca mantenerse alerta en el trabajo o bien reducir esa sensación de fatiga que podría ser causada por el calor.

En el proceso de implementación de los modelos de *machine learning*, el árbol de decisiones es un modelo que permite establecer un panorama completo de la base de datos en estudio. Concediendo, una comprensión de los patrones que estos datos presentan y, con ello, optimizar el proceso para la posterior implementación de otros modelos.

Desde otra perspectiva, la implementación de estos modelos podría ser una herramienta para mejorar la predicción del comportamiento de los ocupantes y, con ello, lograr optimizar el rendimiento de los edificios. En particular, la red neuronal presentó el mejor desempeño en base a los indicadores de predicción. Esto podría deberse a la naturaleza de este modelo, el cual en su algoritmo y construcción trata de replicar el funcionamiento de una neurona. Con ello, puede describir patrones complejos como los que presentan los seres humanos y relacionar estos patrones con variables físico-ambientales de manera más sencilla que el resto de los modelos.

Los modelos de *machine learning* se sustentan en la estadística. Esto implica que la respuesta podría deberse a una sucesión de iteraciones en el proceso. Así, se plantea que se podrían realizar más iteraciones de análisis sobre esta base de datos, lo que podría llevar hacia una mejora y, con ello, aumentar la certeza de estos resultados.

Agradecimientos

Este artículo forma parte del proyecto de investigación Fondecyt Regular N° 1201456 “Criterios innovadores de bienestar para el diseño y evaluación de edificios sustentables desde la perspectiva de los ocupantes” financiado por la Comisión Nacional de Investigación en Ciencia y Tecnología de Chile. Los autores agradecen al grupo de investigación “ARQUITECTURA Y CONSTRUCCIÓN INDUSTRIALIZADA - ACI” de la Universidad del Bío-Bío, por el apoyo brindado.

Autoría

Los autores suscriben de manera conjunta la totalidad del trabajo.

Conflicto de intereses: Los autores declaran que no hay conflicto de intereses.

Bibliografía

Alireza Zolfaghari, S., Teymoori, S., Mohammad Hooshmand, S., Izadi, M., & Afzalian, M. (2019). Experimental investigation of the effects of air temperature on the air movement acceptability in an office with UFAD system.

American Society of Heating, Refrigerating and Air-Conditioning Engineers - Standard 55. (2017). <https://hogiaphat.vn/upload/docs/ASHRAE55-version2017.pdf>

Badulescu, L. (2020). Experiments for a better Gini index splitting criterion for Data Mining Decision Trees algorithms. En Proceedings 24th International Conference on System Theory, Control and Computing (pp. 208–212). Sinaia, Romania: IEE. <https://doi.org/10.1109/ICSTCC50638.2020.9259691>

Barrientos Martínez, R. E., Cruz Ramírez, N., Acosta Mesa, H. G., Rabatte Suárez, I., Gogeochea trejo, M. del C., & Blázquez Morales, S. L. (2009). Árboles de decisión como herramienta en el diagnóstico médico. www.uv.mx/rm

Berlanga Silvente, V., Rubio Hurtado, M. J., & Vilà Baños, R. (2013). Cómo aplicar árboles de decisión en SPSS. Universitat de Barcelona. Institut de Ciències de l'Educació.

Brager, G., & Baker, L. (2009). Occupant satisfaction in mixed-mode buildings. *Building Research & Information*, 37(4), 369–380.

Budaiova, Z., & Vilcekova, S. (2015). Assessing the effect of indoor environmental quality on productivity at office work. *Selected Scientific Papers - Journal of Civil Engineering*, 10(1), 37–46. <https://doi.org/10.1515/sspjce-2015-0004>

Byrd, H., & Rasheed, E. O. (2016). The productivity paradox in green buildings. *Sustainability (Switzerland)*, 8(4). <https://doi.org/10.3390/su8040347>

Camana, R. (2016). Potenciales Aplicaciones de la Minería de Datos en Ecuador. *Revista Tecnológica ESPOL-RTE*, 29(1). <http://www.rte.espol.edu.ec/index.php/tecnologica/article/view/464/338>

Castro, L. F., Espitia, E., & Montilla, A. F. (2018). Applying CRISP-DM in a KDD process for the analysis of student attrition. *Communications in Computer and Information Science*, 885, 386–401. https://doi.org/10.1007/978-3-319-98998-3_30

Charbuty, B., & Abdulazeez, A. (2021). Classification Based on Decision Tree Algorithm for Machine Learning. *Journal of Applied Science and Technology Trends*, 2(01), 20–28. <https://doi.org/10.38094/jastt20165>

Chludzińska, M., & Bogdan, A. (2015). The effect of temperature and direction of airflow from the personalised ventilation on occupants' thermal sensations in office areas. *Building and Environment*, 85, 277–286. <https://doi.org/10.1016/j.buildenv.2014.11.023>

Daradjian, Q., Billé, S., & Inard, C. (2019). Data mining of building performance simulations comprising occupant behaviour modelling. *Advances in Building Energy Research*, 13(2), 157–173. <https://doi.org/10.1080/17512549.2017.1421099>

D'Oca, S., & Hong, T. (2014). A data-mining approach to discover patterns of window opening and closing behavior in offices. *Building and Environment*, 82, 726–739. <https://doi.org/10.1016/j.buildenv.2014.10.021>

Donkers, A., Yang, D., de Vries, B., & Baken, N. (2022). Semantic Web Technologies for Indoor Environmental Quality: A Review and Ontology Design. *Buildings*, 12(10). <https://doi.org/10.3390/buildings12101522>

Du, C., Li, B., Liu, H., Ji, Y., Yao, R., & Yu, W. (2019). Quantification of personal thermal comfort with localized airflow system based on sensitivity analysis and classification tree model. *Energy and Buildings*, (194), 1–11. <https://doi.org/10.1016/j.enbuild.2019.04.010>

Esfandiari, M., Zaid, S. M., Ismail, M. A., & Aflaki, A. (2017). Influence of indoor environmental quality on work productivity in green office buildings: A review. *Chemical Engineering Transactions*, 56, 385–390. <https://doi.org/10.3303/CET1756065>

Espinosa, F. H. T. (2020). Prediction of Recidivism in Thefts and Burglaries Using Machine Learning. *Indian Journal of Science and Technology*, 13(06), 696–711. <https://doi.org/10.17485/ijst/2020/v13i06/149853>

Fanger, P. O., & Toftum, J. (2002). Extension of the PMV model to non-air-conditioned buildings in warm climates.

Fassoulis, K., & Alexopoulos, N. (2015). The workplace as a factor of job satisfaction and productivity: A case study of administrative personnel at the University of Athens. *Journal of Facilities Management*, 13(4), 332–349. <https://doi.org/10.1108/JFM-06-2014-0018>

- Gao, Y., Tumwesigye, E., Cahill, B., & Menzel, K. (2010). Using Data Mining in Optimisation of Building Energy Consumption and Thermal Comfort Management.
- García Alvarado, R., Herrera Ojeda, R., & Soto Muñoz, J. (2013). Dispersión del comportamiento energético residencial; revisión de viviendas en el centro-sur de Chile. *ACE: Architecture, City and Environment*, 8(23), 165–192. <https://doi.org/10.5821/ace.8.23.2601>
- García Cambronero, C., & Gómez Moreno, I. (2006). ALGORITMOS DE APRENDIZAJE: KNN & KMEANS. *Inteligencia En Redes de Telecomunicación*, 23.
- Haynes, B.P. (2008). Office productivity: A Self-Assessed Approach to office evaluation. <https://www.researchgate.net/publication/254195836>
- Hellwig, R. T., Brasche, S., & Bischof, W. (2006). Thermal Comfort in Offices – Natural Ventilation vs. Air Conditioning. Conference Comfort and Energy Use in Buildings. Getting Them Right.
- Hernández, C. A., Martínez, F. H., & Giral, D. A. (2022). Evaluación del desempeño de modelos de decisión espectral predictivos. *Información Tecnológica*, 33(3), 149–158. <https://doi.org/10.4067/s0718-07642022000300149>
- Hong, T., Taylor-Lange, S. C., D'Oca, S., Yan, D., & Corgnati, S. P. (2016). Advances in research and applications of energy-related occupant behavior in buildings. *Energy and Buildings* (116), 694–702. <https://doi.org/10.1016/j.enbuild.2015.11.052>
- Hong, T., Yan, D., D'Oca, S., & Chen, C. fei. (2017). Ten questions concerning occupant behavior in buildings: The big picture. *Building and Environment*, 114, 518–530. <https://doi.org/10.1016/j.buildenv.2016.12.006>
- Kim, J., Zhou, Y., Schiavon, S., Raftery, P., & Brager, G. (2018). Personal comfort models: Predicting individuals' thermal preference using occupant heating and cooling behavior and machine learning. *Building and Environment*, 129, 96–106. <https://doi.org/10.1016/j.buildenv.2017.12.011>
- Lamb, S., & Kwok, K. C. S. (2016). A longitudinal investigation of work environment stressors on the performance and wellbeing of office workers. *Applied Ergonomics*, 52, 104–111. <https://doi.org/10.1016/j.apergo.2015.07.010>
- Lastre Baquerizo, D., Santana, M. P., & Tumbaco, O. L. (2018). Análisis estadístico de tablas de contingencia y chi-cuadrado para medir el flujo migratorio de origen y destino en el Ecuador año 2018. *ECUADORIAN SCIENCE JOURNAL*, 3(1), 23–30. <https://doi.org/10.26911/issn.2602-8077vol3iss1.2019pp23-30p>
- Lipczynska, A., Schiavon, S., & Graham, L. T. (2018). Thermal comfort and self-reported productivity in an office with ceiling fans in the tropics. *Building and Environment*, 135, 202–212. <https://doi.org/10.1016/j.buildenv.2018.03.013>
- Lusett, M. R., Calzada, J. R., & Roura, H. C. (2020). Buildings in warm humid weather. Tampico, tamaulipas case. Mexico. *ACE: Architecture, City and Environment*, 15(44), 1–24. <https://doi.org/10.5821/ace.15.44.5667>
- Madariga Fernández, C. J., Lao León, Y. A., Curra Sosa, D. A., & Lorenzo Martín, R. (2022). Empleo de algoritmos KNN en metodología multicriterio para la clasificación de clientes, como sustento de la planeación agregada. *Retos de La Dirección*, 16, 178–198.
- Maillard, G., Arlot, S., & Lerasle, M. (2021). Aggregated Hold-Out. In *Journal of Machine Learning Research* (Vol. 22). <http://jmlr.org/papers/v22/19-624.html>.
- Marincic, I., Ochoa, M., Antonio, J., & Río, D. (2012). Confort térmico adaptativo dependiente de la temperatura y la humedad. *ACE: Architecture, City and Environment*, 7(20), 27–46. <https://doi.org/10.5821/ace.v7i20.2572>

- Mora, D., De Simone, M., & Chen Austin, M. (2021). Tecnologías para la detección de ocupación en edificios. *Revista Prisma Tecnológico*, 11(1), 17–22. <https://doi.org/10.33412/pri.v11.1.2530>
- Mui, K. W. H., & Chan, W. T. D. (2003). Adaptive comfort temperature model of air-conditioned building in Hong Kong. *Build. and Environment*, 38(6), 837–852. [https://doi.org/10.1016/S0360-1323\(03\)00020-9](https://doi.org/10.1016/S0360-1323(03)00020-9)
- Pao, S. L., Wu, S. Y., Liang, J. M., Huang, I. J., Guo, L. Y., Wu, W. L., Liu, Y. G., & Nian, S. H. (2022). A Physiological-Signal-Based Thermal Sensation Model for Indoor Environment Thermal Comfort Evaluation. *International Journal of Environmental Research and Public Health*, 19(12). <https://doi.org/10.3390/ijerph19127292>
- Perez-Pareja, F. J., García-Pazo, P., Jimenez, R., Escalas, T., & Gervilla, E. (2020). Dejar de Fumar, Terapia Cognitivo-conductual y Perfiles Diferenciales con Árboles de Decisión. *Clínica y Salud*, 31(3), 137–145. <https://doi.org/10.5093/CLYSA2020A12>
- Prasad, N., Kumar, P., & Naidu, M.M. (2013). An approach to prediction of precipitation using Gini Index in SLIQ decision tree. En Proceedings International Conference on Intelligent Systems, Modelling and Simulation (pp. 56–60). Bangkok, Thailand: IEEE <https://doi.org/10.1109/ISMS.2013.27>
- Rady, E. H. A., & Anwar, A. S. (2019). Prediction of kidney disease stages using data mining algorithms. *Informatics in Medicine Unlocked (Vol. 15)*. Elsevier Ltd. <https://doi.org/10.1016/j.imu.2019.100178>
- Riquelme, J. C., Ruiz, R., & Gilbert, K. (2006). Minería de Datos: Conceptos y Tendencias. *Revista Iberoamericana de Inteligencia Artificial*, 10, 11–18. <http://www.redalyc.org/articulo.oa?id=92502902>
- Rodríguez Suárez, Y., & Amador Díaz, A. (2009). Herramientas de Minería de Datos Data Mining Tools (Vol. 3, Issue 3).
- Romero, G. v, Toranzo Calderón, J. S., Jaremczuk, S. E., Gómez, J. C., & Verrastro, C. (2021). Predictor de deserción universitaria. www.frba.edu.ar/investigacion/proyecciones
- Roskams, M., & Haynes, B. (2019). Predictive analytics in facilities management: A pilot study for predicting environmental comfort using wireless sensors. *Journal of Facilities Management*, 17(4), 356–370. <https://doi.org/10.1108/JFM-03-2019-0008>
- Salas Ruiz, R. E., Rodríguez Rodríguez, J. E., & Hernández García, C. L. (2020). Algoritmo híbrido de redes neuronales artificiales con recocido simulado para predicción en minería de datos. <https://revistas.udistrital.edu.co/index.php/vinculos/>
- Salgado Reyes, N., Beltrán Morales, J., Guaña Moya, J., Escobar Teran, C., Nicolalde Rodríguez, D., & Chafra Altamirano, G. (2018). Modelo para predecir el rendimiento académico basado en redes neuronales y analítica de aprendizaje. <https://uvirtual.uce.edu.ec/>
- Schweiker, M., & Wagner, A. (2016). The effect of occupancy on perceived control, neutral temperature, and behavioral patterns. *Energy and Buildings*, 117, 246–259. <https://doi.org/10.1016/j.enbuild.2015.10.051>
- Soto Muñoz, J., Trebilcock Kelly, M., Flores-Alés, V., & Caamaño-Carrillo, C. (2022). Recognizing the effect of the thermal environment on self-perceived productivity in offices: A structural equation modeling perspective. *Building and Environment*, 210. <https://doi.org/10.1016/j.buildenv.2021.108696>
- Soto Muñoz, J., Trebilcock Kelly, M., Flores-Alés, V., & Ramírez-Vielma, R. (2022a). Understanding the perceived productivity of office occupants in relation to workspace thermal environment. *Building Research and Information*, 50(1–2), 152–170. <https://doi.org/10.1080/09613218.2021.1897501>
- Soto Muñoz, J., Trebilcock Kelly, M., Flores-Alés, V., & Ramírez-Vielma, R. (2022b). Understanding the perceived productivity of office occupants in relation to workspace thermal environment. *Building Research and Information*, 50(1–2), 152–170. <https://doi.org/10.1080/09613218.2021.1897501>

- Spositto, O., Blanco, G., Matteo, L., Levi, M., & Bossero, J. (2020). SMOTE, Algoritmo para balanceo de clases en estudio aplicado a la ganadería. www.cs.waikato.ac.nz/~ml/weka/
- Sun, Y., Huang, X., Kroening, D., Sharp, J., Hill, M., & Ashmore, R. (2018). Testing Deep Neural Networks.
- Trebilcock, M., Soto-Muñoz, J., & Piggot-Navarrete, J. (2020). Evaluation of thermal comfort standards in office buildings of Chile: Thermal sensation and preference assessment. *Building and Environment*, 183, 107158. <https://doi.org/10.1016/j.buildenv.2020.107158>
- Troncoso-Espinosa, F. H., & Castro-Albornoz, K. (2021). Predicción de la calidad en revestimientos moldeados para puertas mediante el uso de minería de datos. *Revista Tecnología En Marcha*, 35(1), 115–127. <https://doi.org/10.18845/tm.v35i1.5395>
- Useche Castro, L. M., Barrera Yayas, E. W., Pérez Parra, J. C., & Márquez Pérez, V. (2022). Árboles de regresión para optimizar la imputación Hot Deck.
- Vallejo Ballesteros, H. F., Guevara Iñiguez, E., & Medina Velasco, S. R. (2018). Minería de Datos. *RECIMUNDO*, 2(Esp), 339–349. <https://doi.org/10.26820/recimundo/2.esp.2018.339-349>
- Webb, G. I. (2010). Naïve Bayes. *Encyclopedia of Machine Learning*, 15, 713–714.
- Yan, D., & Hong, T. (2018). Definition and Simulation of Occupant Behavior in Buildings Annex 66 Final Report Operating Agents of Annex 66. www.iea-ebc.org
- Yan, D., Hong, T., Dong, B., Mahdavi, A., D'Oca, S., Gaetani, I., & Feng, X. (2017). IEA EBC Annex 66: Definition and simulation of occupant behavior in buildings. *Energy and Buildings*, 156, 258–270. <https://doi.org/10.1016/j.enbuild.2017.09.084>
- Yao, R., Li, B., & Liu, J. (2009). A theoretical adaptive model of thermal comfort - Adaptive Predicted Mean Vote (aPMV). *Building and Environment*, 44(10), 2089–2096. <https://doi.org/10.1016/j.buildenv.2009.02.014>
- Zhang, W., Wu, Y., & Calautit, J. K. (2022). A review on occupancy prediction through machine learning for enhancing energy efficiency, air quality and thermal comfort in the built environment. *Renewable and Sustainable Energy Reviews*, 167, 112704. <https://doi.org/10.1016/j.rser.2022.112704>