

Análisis de vivienda saludable con Inteligencia Artificial en zona periurbana Jayllihuaya Puno – 2023

Analysis of healthy housing with Artificial Intelligence in the periurban area of Jayllihuaya Puno – 2023

Magali G. Gonzales-Paco^A, Sandra Flores-Asencio^B, Pablo C. Tapia-Catacora^C y Darwin Pino-Cordero^D

ORCID: 0000-0001-9378-0787^A, ORCID: 0000-0002-5712-7163^B, ORCID: 0000-0002-8608-4203^C y ORCID: 0000-0002-8761-5184^D

(Recepción: 18/01/2024 y aceptación 26/04/2024)

Resumen— La tendencia actual en la construcción de viviendas en la zona periurbana de Jayllihuaya es creciente, se observó que el terreno de la zona es estable, carente de humedad y con escasa presencia de antenas de emisión electromagnética. El estudio buscó establecer vínculo entre la inteligencia artificial y las viviendas saludables utilizando modelos de clasificación supervisada. Las características analizadas de las viviendas estuvieron relacionadas con el acceso a los servicios básicos, iluminación, contaminación acústica, material predominante de la vivienda, combustible utilizado en la cocina y acceso a la tecnología. La información sobre las viviendas se recolectó mediante encuestas de preguntas cerradas, procesadas posteriormente con la finalidad de predecir mediante modelos de algoritmos de clasificación. Los resultados mostraron que el modelo de algoritmo máquinas de vectores de soporte (SVM) logró clasificar correctamente las viviendas con una precisión de 82%, los modelos de algoritmos de k-vecinos más cercanos (KNN) y los bosques aleatorios (RF) alcanzaron una precisión de 82% y 86.3% respectivamente, con la particularidad que los dos últimos, no lograron clasificar las viviendas consideradas saludables. Estas precisiones alcanzadas evidencian que para clasificar viviendas en saludables y no saludables es mejor utilizar modelos basados en SVM.

Palabra clave: Bosques aleatorios, k-vecinos más cercanos, máquinas de vectores de soporte, vivienda saludable

Abstract— The current trend in the construction of housing in the peri-urban area of Jayllihuaya is growing, it was observed that the terrain in the area is stable, devoid of humidity and with little presence of electromagnetic emission antennas. The study sought to establish a relationship between artificial intelligence and homes considered healthy, using supervised classification models. The analyzed characteristics of the homes were related to access to basic services, lighting, noise pollution, predominant material of the home, fuel used in cooking and access to technology. The information about the homes was collected through surveys with closed questions, subsequently processed with the purpose of predicting using classification algorithm models. The results showed that the support vector machines (SVM) algorithm model managed to correctly classify the homes with an accuracy of 82%, the k-nearest neighbors (KNN) algorithm models and the random forests (RF) achieved a accuracy of 82% and 86.3% respectively, with the particularity that the last two failed to classify the homes considered healthy. These precisions achieved show that to classify homes as healthy and unhealthy it is better to use models based on SVM.

Keyword: Healthy housing, k-nearest neighbors, random forests, support vector machines

A. Magali Gianina Gonzales Paco, Universidad Nacional del Altiplano Puno, mggonzales@unap.edu.pe

B. Sandra Flores Asencio, Universidad Nacional del Altiplano Puno, sflores@unap.edu.pe

C. Pablo Cesar Tapia Catacora, Universidad Nacional del Altiplano Puno, pctapia@unap.edu.pe

D. Darwin Pino Cordero, Universidad Nacional del Altiplano Puno, dpino@unap.edu.pe

1 INTRODUCCIÓN

La vivienda será cada vez más importante para la salud debido a los patrones demográficos y tecnológicos cambiantes, así mismo, el número de personas que pasan en casa una mayor proporción de su tiempo se duplicará para el 2050 [1]. El desempleo creciente [2], los patrones climáticos cambiantes, subrayan la importancia de que la vivienda brinde protección contra el frío, el calor y otros fenómenos meteorológicos extremos a fin de promover comunidades resilientes [3]. En el marco de la nueva agenda de desarrollo sostenible agenda 2030 de la ONU, el uso y adopción de las nuevas tecnologías de la información y la comunicación en el hogar como la

Domótica [4], [5], están empezando a inducir cambios en el uso y la función de la vivienda, lo cual plantea nuevos retos en la medida que constituye una de las instancias primarias de las relaciones sociales, de la interacción familiar, de la vida cotidiana y de la estructura de la ciudad [6].

Las viviendas de la zona periurbana de Jayllihuaya y el bienestar de las familias que lo habitan es una relación biunívoca donde se establecen determinadas condiciones para ser considerada vivienda saludable. Para la Organización Panamericana de la Salud (PAHO) [7] una vivienda saludable es un refugio que propicia un estado de completo bienestar físico, mental y social. Con un enfoque mundial, también hace referencia a su estructura física y a la medida en que favorece la salud física,

incluido por ser estructuralmente sólida [8].

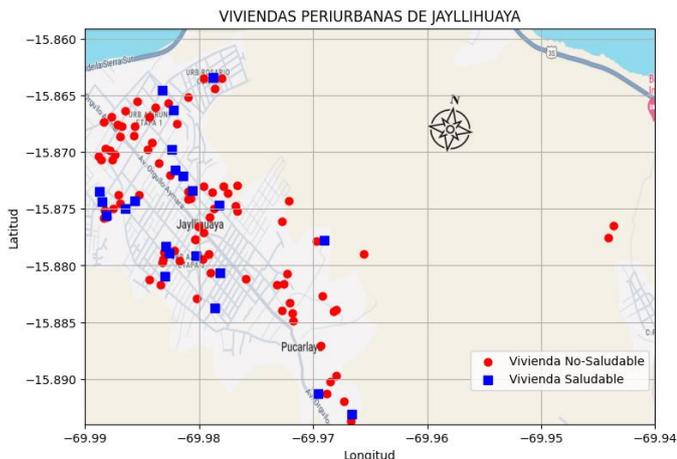


Fig. 1. Ubicación de las viviendas en base a sus coordenadas geográficas (Latitud, Longitud). Mapa referencial: OpenStreetMap

El objetivo de la investigación está enfocado al uso del aprendizaje automático supervisado para determinar si las viviendas encuestadas de la zona periurbana de Jayllihuaya, son consideradas saludables o no en base a la información recopilada de las viviendas en mención mediante entrevistas directas utilizando preguntas, filtros y categorías de la Sección 2 de la Encuesta Demográfica y de Salud Familiar ENDES - 2022, elaborada por el Instituto Nacional de Estadística e Informática (INEI) [9]. Posterior a ello, validar las viviendas clasificadas como saludables considerando los criterios básicos establecidos por la Organización Mundial de la Salud [7], todo ello en estricto cumplimiento del artículo 97 del Capítulo IV Secreto Estadístico y Confidencialidad de la Información, establecido en el Decreto Supremo 043-2001-PCM de la legislación peruana.

En un contexto donde la transversalidad de la inteligencia artificial [10] proporciona nuevas formas de analizar datos, se encuentra el aprendizaje automático que permite completar con mejor eficiencia las tareas de clasificación. Una buena alternativa para resolver el problema de clasificar viviendas en saludables o no, es utilizar algoritmos de clasificación de aprendizaje supervisado como: las Máquinas de Vectores de Soporte (SVM: support vector machine) [11] con sus parámetros ajustados a $kernel='rbf'$, $gamma='auto'$, $C=2$, K-Vecinos más cercanos (KNN: k-nearest neighbors) [12] con su parámetro $k=3$, y los Bosques Aleatorios (RF: random forest) [13], [14] con parámetros ajustados a $n_estimators=90$, $criterion='gini'$, $max_features='sqrt'$, $oob_score=True$, $bootstrap=True$, $max_samples=2/3$.

El preprocesamiento de los datos procedentes de las

encuestas es un paso crucial en el campo del aprendizaje automático [15] porque ayuda a mejorar la calidad de los datos y así promover la extracción de información clasificadora a partir de ellos mediante los algoritmos de aprendizaje supervisado [16]. Esta etapa del preprocesamiento también se conoce también como la técnica de preparar (limpiar y organizar) los datos sin procesar para que sean los adecuados en la construcción y entrenamiento de los modelos de aprendizaje automático [17]. Así mismo, se considera como una técnica de minería de datos que transforma datos sin procesar en un formato comprensible y legible [18].

No obstante, todos los modelos de clasificación de aprendizaje supervisado para su entrenamiento y prueba [19] requieren las características preprocesadas de las viviendas, para este estudio se consideró 23 características y una con el resultado deseado, como describe la Tabla 1. La eficiencia de estos modelos se evaluó mediante métricas proporcionados por la biblioteca Scikit-learn [20] como son: la matriz de confusión [21] y el reporte de clasificación [22]. Con el objetivo de evaluar el nivel de precisión con que los modelos propuestos (SVM, KNN y RF) pueden clasificar correctamente si una vivienda es saludable o no es saludable.

Tabla 1.

Características clasificadoras de las viviendas evaluadas

Código	Característica de la vivienda
HV201	Fuente principal de abastecimiento de agua para tomar o beber
HV204	Tiempo que se demora en ir, recoger agua y volver
HV205	Tipo de servicio higiénico del hogar
HV206	Electricidad en el hogar
HV207	Tiene radio
HV208	Tiene televisor
HV213	Material predominante del piso de la vivienda
HV214	Material predominante de las paredes exteriores de la vivienda
HV215	Material predominante del techo de la vivienda
HV226	Combustible que utilizan más frecuentemente en su hogar para cocinar
HV242	Utilizan un cuarto especialmente para cocinar
HV243A	Tiene un teléfono móvil (celular)
SH42	El abastecimiento de agua para beber está disponible todo el día
SH61J	Tiene televisión por cable
SH61L	Tiene cocina de gas
SH61P	Tiene computadora
SH61Q	Tiene acceso a Internet en casa
SH61S	Tiene generador de electricidad
SH70	Qué tipo de alumbrado utiliza el hogar
SH76A	La vivienda tiene ventanas
SH76B	Ventanas con cristal
SH76C	Ventanas de madera
SH76E	Ventanas con cortinas/persianas
HV201	Fuente principal de abastecimiento de agua para tomar o beber

Fuente: Encuesta demográfica y de salud familiar ENDES-2022/Instituto Nacional de Estadística e Informática.

Tabla 2.

Datos de entrenamiento y prueba

Datos	Entrenamiento (train)	Prueba (test)
107 viviendas	85 (80 %)	22 (20 %)
Viviendas no saludables	66	19
Viviendas saludables	19	3

2 MATERIALES Y MÉTODOS

2.1 Ubicación del estudio

Las viviendas seleccionadas para el presente estudio se ubican en la zona periurbana de Jayllihuaya, distrito, provincia y departamento de Puno como se visualiza en la Fig. 1, los datos de dichas viviendas se organizan en dos vectores, el primero está conformado por las características propias de las viviendas descritas en la Tabla 1 y el segundo vector contiene la categorización de las viviendas en saludables y no saludables. Estos datos se separan en datos de entrenamiento y prueba utilizando la función *train_test_split* de la biblioteca scikit-learn, así muestra la Tabla 2.

2.2 Vivienda saludable

El enfoque de la Organización Mundial de la Salud (OMS) sobre las condiciones básicas para ser considerada una vivienda saludable [7].

- Acceso a la red eléctrica y servicios básicos.
- Ventilación adecuada y aislamiento térmico.
- Presencia mínima de contaminación acústica.
- Iluminación natural adecuada.
- Presencia mínima de señales electromagnéticas.
- Alejado de zonas húmedas.
- Construcción sobre terreno estable.
- No quemar combustible sólido en la cocina.

2.3 Diseño de la investigación

Las encuestas proporcionan un conjunto de características que describen el estado actual de una vivienda, las mismas que se describen en la Tabla 1, lo que resultó en 107 viviendas encuestadas. Para aplicar algoritmos de inteligencia artificial que realice tareas de clasificación de aprendizaje supervisado, se requiere organizar dos vectores en función a las características de las viviendas, el primer vector constituye la entrada del algoritmo y el segundo vector la salida deseada, que es un valor numérico (0: vivienda no saludable, 1: vivienda saludable). Para analizar las características de las viviendas y clasificarlos en saludables y no saludables se utilizó tres algoritmos de clasificación supervisada: máquina de vectores de soporte (SVM - support vector machine), k-vecinos cercanos (KNN - k-nearest neighbors) y bosques aleatorios (RF - random forest).

2.4 Lenguaje de programación y bibliotecas

Python es un lenguaje de programación de alto nivel muy utilizado por la inteligencia artificial y la ciencia de datos, tiene estructuras de datos muy eficientes y utilizado crear las funciones clasificadoras.

Scikit-learn es una biblioteca de Python orientado al aprendizaje automático que estandariza el proceso de clasificación en cuatro procesos principales y son:

1. Selección de un modelo.
2. Establecer los valores iniciales del modelo.
3. Entrenamiento del modelo.
4. Probar la eficiencia de clasificación del modelo.

2.5 Análisis de las características de las viviendas

El análisis de los datos de las viviendas se realizó aplicando modelos de aprendizaje supervisado [23]: máquina de vectores de soporte (SVM), k-vecinos cercanos (KNN) y bosques aleatorios (RF), el hardware utilizado es una computadora con procesador Core i7 de 11va generación con sistema operativo Windows 10 pro v.22. las características de las viviendas analizadas están asociados a los servicios básicos, la ventilación, iluminación, contaminación acústica, señales electromagnéticas y ubicación de las viviendas.

La eficiencia de los modelos de clasificación de aprendizaje supervisado se evaluó en función a las clasificaciones alcanzadas de forma correcta e incorrecta, estas se resumen en dos métricas de clasificación:

- a) Matriz de confusión.

		Predicción	
		No saludable	Saludable
Real	No saludable	TN	FP
	Saludable	FN	TP

Fig. 2. Matriz de confusión

Donde:

- TN:** True Negatives, el modelo clasifica correctamente la vivienda no saludable.
- FP:** False Positive, el modelo clasifica incorrectamente la vivienda saludable.
- FN:** False Negative, el modelo clasifica incorrectamente la vivienda no saludable.
- TP:** True Positive, el modelo clasifica correctamente la vivienda saludable.

- b) Reporte de clasificación.

	Precision	Recall	Support
0	0.xx	0.xx	X1
1	0.xx	0.xx	X2
accuracy		0.xx	X1 + X2

Donde:

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

- X1:** Número de predicciones para viviendas no saludables.
- X2:** Número de predicciones para viviendas saludables.

3 RESULTADOS Y DISCUSIONES

3.1 Resultados alcanzados por los modelos

Tabla 3.

Matriz de confusión de los modelos de clasificación de aprendizaje supervisado (0: No saludable, 1: Saludable).

Modelo clasificador		Predicción	
Máquina de vectores de soporte (SVM)	0	17	2
	1	2	1
K-vecinos más cercanos (KNN)	0	18	1
	1	3	0
Bosques aleatorios (RF)	0	19	0
	1	3	0

Los resultados de la matriz de confusión Tabla 3 muestran que el modelo SVM tiene mejor rendimiento, porque clasifica correctamente una vivienda saludable y 17 viviendas no saludables, mientras que KNN solo clasifica correctamente 18 viviendas no saludables y RF solo clasifica correctamente 19 viviendas no saludables.

Tabla 4.

Reporte de clasificación alcanzada por los algoritmos de clasificación supervisados.

Modelo clasificador	Clasificación	Precision	Recall	Accurac y
SVM	0: No	0.89	0.89	0.82 (82%)
	1: Saludable	0.33	0.33	
KNN	0: No	0.86	0.95	0.82 (82%)
	1: Saludable	0.00	0.00	
RF	0: No	0.86	1.00	0.86 (86%)
	1: Saludable	0.00	0.00	

Respecto al reporte de clasificación Tabla 4 y comparando la exactitud (Accuracy) de los modelos aplicados se observa que el modelo SVM alcanza una exactitud de 82% al clasificar correctamente una vivienda, este algoritmo alcanza una precisión de 89% las viviendas no saludables y con un 33% las viviendas saludables. El modelo RF alcanza una exactitud de 86%, superior a los algoritmos de SVM y KNN lo cual es muy bueno. Considerando la precisión del modelo RF, clasifica muy bien las viviendas no saludables dejando de lado las viviendas saludables. Por tanto, se descarta su aplicación para

clasificar las viviendas de la zona periurbana de Jayllihuaya.

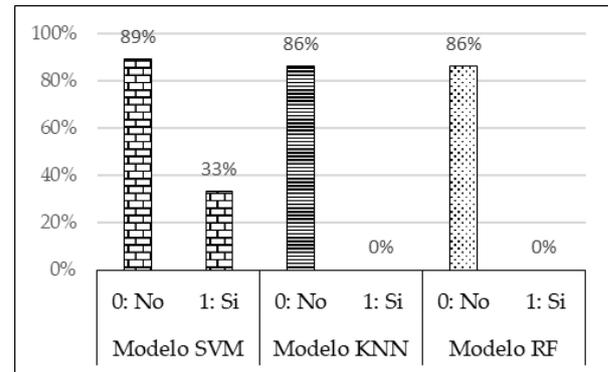


Fig. 3. Métrica **precisión**, alcanzada por los algoritmos de clasificación supervisado (0: vivienda no saludable, 1: vivienda saludable).

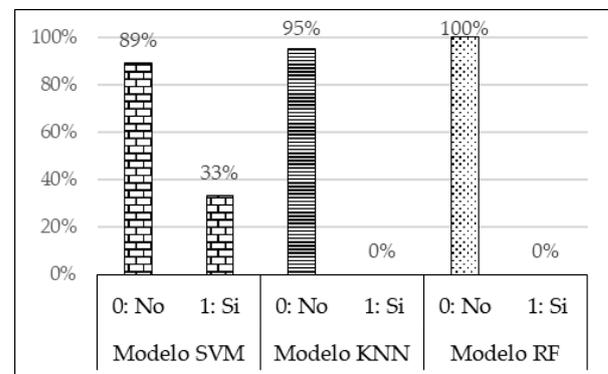


Fig. 4. Métrica **recall**, alcanzada por los algoritmos de clasificación supervisado (0: vivienda no saludable, 1: vivienda saludable).

Los resultados de las métricas *precision* y *recall* que se muestran en la Fig. 3 y Fig. 4 respectivamente, el modelo SVM tiene mejor precisión que los modelos KNN y RF. Además, de tres viviendas saludables proporcionadas en los datos de prueba (*test*) que se describen en la Tabla 2, el modelo SVM logra clasificar correctamente la vivienda saludable con un 33% de precisión y el 67% de las viviendas saludables lo clasifica en el grupo de *Falso Negativo*, es decir las viviendas son saludables pero el modelo lo clasifica como no saludable. En cambio, los modelos KNN y RF tienen una precisión del 0% al clasificar correctamente las viviendas saludables, pero si los clasifica el 100% de las viviendas saludables en el grupo de *Falso Negativo*, es decir que las viviendas son saludables pero los modelos los clasifican como no saludables.

3.2 El modelo máquina de vectores de soporte

Es un algoritmo de aprendizaje automático supervisado utilizado para clasificar de forma discriminativa las viviendas de la zona periurbana de Jayllihuaya, alcanzan un rendimiento óptimo con datos relativamente pequeños y con múltiples características como del presente estudio. Además, es un modelo clasificador con mejor desempeño, alcanzando una exactitud del 82%, para ello, se establecieron los valores

optimizados de sus parámetros de acuerdo con la siguiente tabla:

Tabla 5.

Parámetros del algoritmo clasificador SVM y sus valores

Parámetro	Valor establecido
<i>kernel</i>	(RBF) Función de base radial o gaussiana
<i>gamma</i>	auto
<i>coste (C)</i>	2

Con los parámetros de la Tabla 5, el modelo de algoritmo clasificador SVM alcanzó un score de clasificación inicial de 82.3% con los datos de entrenamiento y un score de 81.8% con los datos de prueba, con una variación de 0.05 puntos porcentuales, esto ratifica que el algoritmo logra una buena precisión clasificando viviendas.

3.3 Discusiones

Los algoritmos de clasificación de aprendizaje supervisado mostraron diferentes precisiones al momento de clasificar las viviendas saludables y no saludables, SVM (82.3%), KNN (82%) y RF (86.3%), los algoritmos KNN y RF su precisión de cálculo es inferior a la precisión alcanzada por el algoritmo SVM, además solo clasifica de forma correcta las viviendas no saludables. Sin embargo, [24] afirma que SVM logra mejores precisiones porque utiliza *límites de decisión* entre las clases, esta distancia de separación entre los puntos maximiza que la clasificación sea correcta. [11] complementa que la clasificación lineal SVM obtiene mejores resultados que una radial.

Una de las iniciativas sobre viviendas saludables en Perú, nace de las condiciones actuales de una vivienda y sus repercusiones en la salud de la población [7], vinculando la corresponsabilidad a los gobiernos locales y su ciudadanía. Esta iniciativa es el resultado de entrevistas a profesionales en temas de salud y vivienda, información bibliográfica y periodística con apreciación subjetiva [25]. En cambio, esta investigación se basa en resultados de aprendizaje automático supervisado cuyos datos proceden de viviendas encuestadas en la zona periurbana de Jayllihuaya - Puno.

Las viviendas ecológicas son una propuesta alternativa en términos de diseño y construcción de vivienda, están enfocada al bienestar de las personas que en ella habitan y reducir el impacto del deterioro del medio ambiente [26]. Es una investigación centrada en el proceso de construcción, pero las características de dichas viviendas pueden mejorar la precisión de los modelos utilizados en esta investigación.

La tecnología de simulación tridimensional permite emular el comportamiento de una vivienda saludable que garantiza la calidad de vida de sus residentes [27]. Este enfoque puede incorporar todas las características consideradas en esta investigación, lo cual es muy bueno.

En las encuestas de la investigación, no se consideró los factores externos a la vivienda como: los espacios verdes, el transporte (activo y público), la comunidad social y sus interacciones. Estas características dejamos a las futuras investigaciones para que puedan reforzar esta investigación.

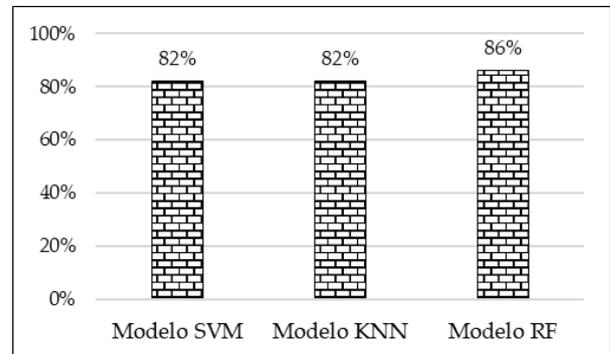


Fig. 5. Métrica de **exactitud** (accuracy) alcanzada por los algoritmos de clasificación supervisado.

Respecto a los resultados de la Fig. 5, aparentemente el modelo RF tiene el mayor porcentaje en clasificar las viviendas, ahora de acuerdo con la Fig. 3, este modelo no logra clasificar las viviendas saludables, en cambio el Modelo SVM si lo hace. Por tanto, nos quedamos con el modelo SVM para clasificar viviendas, aunque su exactitud de 82% sea inferior al modelo RF.

La investigación no considera aspectos como la cultura de paz, el ambiente familiar en la vivienda, el buen trato o los posibles riesgos psicosociales que puedan coexistir con las familias. Son temas para futuras investigaciones que puedan reforzar los resultados aquí alcanzados. Además, contrastar con resultados de viviendas ubicadas en zonas geográficas periurbanas distintas a las de Jayllihuaya.

4 CONCLUSIONES

Una vivienda saludable es un espacio físico adecuado que propicia las condiciones adecuadas para sus ocupantes reduciendo al máximo los factores de riesgo. El modelo de algoritmo que mejor clasifica una vivienda en función a sus características es la máquina de vectores de soporte (SVM), alcanza una precisión del 82% para una muestra de 107 viviendas encuestadas. Respecto a los demás algoritmos, la investigación demostró que para mejorar su precisión los modelos de bosques aleatorios requieren mayor cantidad de datos de entrenamiento y prueba, el algoritmo de k-vecinos más cercanos depende de su parámetro k y una métrica de distancia, su parámetro k es difícil de sincronizar cuando clasifica datos categóricos como son las de una vivienda.

5 REFERENCIAS

- [1] Organización Mundial de la Salud (OMS), "Informe mundial de la OMS sobre el envejecimiento y la salud.", 2015.
- [2] P. Baker, M., Keall, M., Au, E. L., & Howden-Chapman, "Home is where the heart is--most of the time.," *New Zeal. Med. journal*, 120(1264), U2769., 2007.

- [3] Organización Internacional del Trabajo (OIT), "Perspectivas sociales y del empleo en el mundo: tendencias 2016. Ginebra," 2016.
- [4] L. F. Herrera Quintero, "Viviendas inteligentes (Domótica)," *Rev. Ing. e Investig.*, vol. 25, no. 2, 2005. <https://doi.org/10.15446/ing.investig.v25n2.14639>
- [5] A. Gabriel, "Qué es Domótica - CEDOM | Asociación Española de Domótica e Inmótica," *CEDOM*. 2020.
- [6] J. Chaparro Mendivelso, "Domótica: la mutación de la vivienda," *Scripta Nova: Revista electrónica de geografía y ciencias sociales*. 2003.
- [7] W. (DC) Organización Panamericana de la Salud, "Directrices de la OMS sobre vivienda y salud [Internet]," *Internet*, 2022. <https://www.ncbi.nlm.nih.gov/books/NBK583405/> (accessed Dec. 13, 2023).
- [8] S. Iddi, K. Muindi, H. Gitau, and B. Mberu, "Characterization of Healthy Housing in Africa: Method, Profiles, and Determinants," *J. Urban Heal.*, vol. 99, no. 1, 2022, <https://doi.org/10.1007/s11524-021-00603-5>
- [9] INEI 2018, "Dirección Nacional de Censos y Encuestas - ENDES," 2022. <https://proyectos.inei.gob.pe/endes/queesendes.asp> (accessed Mar. 01, 2024).
- [10] J. E. (Hans. Korteling, G. C. van de Boer-Visschedijk, R. A. M. Blankendaal, R. C. Boonekamp, and A. R. Eikelboom, "Human- versus Artificial Intelligence," *Front. Artif. Intell.*, vol. 4, 2021, <https://doi.org/10.3389/frai.2021.622364>
- [11] L. K. Ramasamy, S. Kadry, Y. Nam, and M. N. Meqdad, "Performance analysis of sentiments in Twitter dataset using SVM models," *Int. J. Electr. Comput. Eng.*, vol. 11, no. 3, 2021, <https://doi.org/10.11591/ijece.v11i3.pp2275-2284>
- [12] S. Zhang, X. Li, M. Zong, X. Zhu, and D. Cheng, "Learning k for kNN Classification," *ACM Trans. Intell. Syst. Technol.*, vol. 8, no. 3, 2017, <https://doi.org/10.1145/2990508>
- [13] J. J. Espinosa Zúñiga, "Aplicación de algoritmos Random Forest y XGBoost en una base de solicitudes de tarjetas de crédito," *Ing. Investig. y Tecnol.*, vol. 21, no. 3, pp. 1-16, Jul. 2020, <https://doi.org/10.22201/fi.25940732e.2020.21.3.022>
- [14] J. L. Speiser, M. E. Miller, J. Tooze, and E. Ip, "A comparison of random forest variable selection methods for classification prediction modeling," *Expert Systems with Applications*, vol. 134, 2019. <https://doi.org/10.1016/j.eswa.2019.05.028>
- [15] C. Janiesch, P. Zschech, and K. Heinrich, "Machine learning and deep learning," *Electron. Mark.*, vol. 31, no. 3, 2021, <https://doi.org/10.1007/s12525-021-00475-2>
- [16] K. Keykhosravi, A. Hamednia, H. Rastegarfar, and E. Agrell, "Data preprocessing for machine-learning-based adaptive data center transmission," *ICT Express*, vol. 8, no. 1, 2022, <https://doi.org/10.1016/j.icte.2022.02.002>
- [17] A. A. Akimov, D. R. Valitov, and A. I. Kubryak, "DATA PREPROCESSING FOR MACHINE LEARNING," *Научное обозрение. Технические науки (Scientific Rev. Tech. Sci.*, no. №2 2022, 2022, <https://doi.org/10.17513/srts.1391>
- [18] K. Goyal, "Data Preprocessing in Machine Learning: 7 Easy Steps To Follow," *upGrad*, 2021.
- [19] K. Dhasaradhan and R. Jaichandran, "Performance analysis of machine learning algorithms in heart disease prediction," *Concurr. Eng. Res. Appl.*, vol. 30, no. 4, 2022, <https://doi.org/10.1177/1063293X221125231>
- [20] W. De Vazelles, C. J. Carey, Y. Tang, N. Vauquier, and A. Bellet, "Metric-learn: Metric Learning Algorithms in Python," *J. Mach. Learn. Res.*, vol. 21, 2020.
- [21] Z. Khanam, B. N. Alwasel, H. Sirafi, and M. Rashid, "Fake News Detection Using Machine Learning Approaches," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1099, no. 1, 2021, <https://doi.org/10.1088/1757-899X/1099/1/012040>
- [22] S. De Gheselle *et al.*, "Machine learning for prediction of euploidy in human embryos: in search of the best-performing model and predictive features," *Fertil. Steril.*, vol. 117, no. 4, 2022, <https://doi.org/10.1016/j.fertnstert.2021.11.029>
- [23] S. Uddin, A. Khan, M. E. Hossain, and M. A. Moni, "Comparing different supervised machine learning algorithms for disease prediction," *BMC Med. Inform. Decis. Mak.*, vol. 19, no. 1, 2019, doi: <https://doi.org/10.1186/s12911-019-1004-8>
- [24] Bharadwaj, K. B. Prakash, and G. R. Kanagachidambaresan, "Pattern Recognition and Machine Learning," in *EAI/Springer Innovations in Communication and Computing*, 2021. doi: https://doi.org/10.1007/978-3-030-57077-4_11
- [25] R. Santa María and P. Lima, "LA INICIATIVA DE VIVIENDA SALUDABLE EN EL PERU," *Rev. Peru. Med. Exp. y salud PÚblica*, vol. 25, no. 4, 2008.
- [26] M. A. Polo and J. Sanchez, "Estudio De Mercado Para La Comercialización De Viviendas Ecológicas En El Distrito De Chiclayo - Año 2016," 2016.
- [27] A. E. Barragán and P. E. Ochoa, *Diseño de viviendas ambientales de bajo costo.*, vol. 5, no. 1. 2014. <https://doi.org/10.18537/mskn.05.01.06>