



Inteligencia Artificial aplicada a la ingeniería geotécnica: Comparación de LightGBM y XGBoost para la predicción de propiedades mecánicas del suelo

Artificial Intelligence applied to geotechnical engineering: Comparison of LightGBM and XGBoost for predicting soil mechanical properties

Inteligência Artificial aplicada à engenharia geotécnica: Comparação de LightGBM e XGBoost para previsão de propriedades mecânicas do solo

ARTÍCULO ORIGINAL

Jaime Yelsin Rosales Malpartida

jrosalesm@uni.pe

<https://orcid.org/0000-0003-4574-5172>

Facultad de Ingeniería Civil

Universidad Nacional de Ingeniería, Lima, Perú

Recibido 02 de Mayo 2025 | Arbitrado y aceptado 02 de Mayo 2025 | Publicado el 09 de Julio 2025

RESUMEN

Este estudio investiga las capacidades no lineales de los modelos de predicción LightGBM y XGBoost, técnicas de machine learning dentro del campo de la Inteligencia Artificial (IA), para estimar cuatro propiedades mecánicas clave del suelo: la máxima densidad seca (MDS), el contenido óptimo de humedad (OCH), la relación de soporte de California al 100% (CBR100) y al 95% (CBR95) de la MDS. Determinar estas propiedades es fundamental en el diseño estructural de pavimentos, pero los métodos tradicionales de laboratorio suelen ser costosos y requieren tiempo. En este trabajo se utilizaron 201 registros experimentales obtenidos del Laboratorio N°2 de Mecánica de Suelos de la Facultad de Ingeniería Civil de la UNI (FIC-UNI), incorporando como variables de entrada tres parámetros granulométricos (grava, arena y finos) y tres relacionados con los límites de consistencia (límite líquido, límite plástico e índice de plasticidad). Ambos modelos fueron entrenados utilizando optimización por grid search para ajustar sus hiperparámetros y maximizar el rendimiento predictivo. Los resultados, evaluados mediante métricas como RMSE, MAE y MAPE, indican que el modelo XGBoost supera en precisión al modelo LightGBM para la predicción de las cuatro propiedades mecánicas analizadas.

Palabras clave: Inteligencia artificial, Machine learning, LightGBM, XGBoost, Propiedades mecánicas del suelo, Modelos de predicción, Carreteras.

ABSTRACT

This study explores the nonlinear capabilities of the LightGBM and XGBoost prediction models—machine learning techniques within the field of Artificial Intelligence (AI)—to estimate four key mechanical properties of soil: maximum dry density (MDD), optimum moisture content (OMC), and the California Bearing Ratio at 100% (CBR100) and 95% (CBR95) of MDD. Determining these properties is essential for pavement structural design; however, traditional laboratory methods are often time-consuming and costly. In this research, 201 experimental records were used, obtained from Soil Mechanics Laboratory No. 2 of the Faculty of Civil Engineering at UNI (FIC-UNI). The input dataset included three granulometric parameters (gravel, sand, and fines) and three consistency limits (liquid limit, plastic limit, and plasticity index). Both models were trained using grid search optimization to fine-tune their hyperparameters and maximize predictive performance. The results, evaluated using metrics such as RMSE, MAE, and MAPE, show that the XGBoost model outperformed LightGBM in prediction accuracy for all four analyzed mechanical properties.

Keywords: Artificial intelligence, Machine learning, LightGBM, XGBoost, Soil mechanical properties, Prediction models, Roads.

RESUMO

Este estudo explora as capacidades não lineares do LightGBM e a determinação dessas propriedades é essencial para o projeto estrutural do pavimento; Entretanto, os métodos laboratoriais tradicionais costumam ser demorados e caros. Nesta pesquisa foram utilizados 201 registros experimentais, obtidos no Laboratório de Mecânica dos Solos n° 2 da Faculdade de Engenharia Civil da UNI (FIC-UNI). O conjunto de dados de entrada incluiu três parâmetros granulométricos (cascalho, areia e finos) e três limites de consistência (limite de liquidez, limite de plasticidade e índice de plasticidade). Ambos os modelos foram treinados usando otimização de busca em grade para ajustar seus hiperparâmetros e maximizar o desempenho preditivo. Os resultados, avaliados usando métricas como RMSE, MAE e MAPE, mostram que o modelo XGBoost superou o LightGBM em precisão de previsão para todas as quatro propriedades mecânicas analisadas.

Palavras-chave: Inteligência artificial, Aprendizado de máquina, LightGBM, XGBoost, Propriedades mecânicas do solo, Modelos de predição, Estradas.

1. Introducción

La infraestructura de transporte es esencial para promover el crecimiento y el desarrollo económico (Banerjee et al., 2020 y Ghani et al., 2016). Al mejorar y construir nuevos accesos de movilidad, la infraestructura de transporte reduce costos y mejora el acceso a los mercados y al trabajo, estimulando la integración económica, la competencia y las economías de aglomeración. Además, ayuda a aprovechar oportunidades de inversión en capital humano, con un mejor acceso a escuelas y atención médica. Otros beneficios intangibles incluyen facilitar la difusión de conocimientos sobre nuevas tecnologías y el desarrollo de patentes (Banerjee et al., 2012; Calderón y Servén, 2014; Agrawal et al., 2017; Gibbons et al., 2019). Por lo tanto, las naciones invierten en vías de comunicación para asegurar su desarrollo,

Por ejemplo, en Colombia durante el 2023 se atendió 4.087 km de la red vial nacional, lo cual representó empleos para 30408 personas y una inversión total por 2 billones de pesos (USD 116,6 millones). Y en Perú, Según el Organismo Supervisor de la Inversión en Infraestructura de Transporte de Uso Público (Ositrán) la inversión en infraestructuras de transporte cerró el 2023 con USD 723,4 millones, lo que representa un incremento del 33,2 % respecto al 2022. La inversión en obras viales es importante como la calidad del diseño del pavimento.

Según Das, B. M., & Sivakugan, N. (2018) se entiende por pavimento a toda la estructura de la vía, desde las capas inferiores hasta la capa de asfalto o concreto, sobre la que rueda el tránsito, por ello se deben tomar en cuenta asuntos como el volumen del tráfico, el uso que tendrá la vía, el tipo de vehículo que transitará y, sobre todo, tener un conocimiento profundo de los factores que rigen el comportamiento del suelo debido a que este no es uniforme. De acuerdo con Özyanki, A. (2019) el estudio del suelo es importante para garantizar la vida útil del pavimento, puesto que permite conocer las propiedades del suelo y determinar su comportamiento, en caso contrario se construirán carreteras que más tarde presentarán problemas como hundimientos y daños a la capa asfáltica.

De acuerdo con Carrillo, A. (2016) el estudio del suelo comprende el análisis de características físicas como: la granulometría, los límites de consistencia, contenido de humedad, propiedades mecánicas como el valor de la relación de soporte de California

(CBR) y aquellas que ayudan a mejorar las propiedades del suelo como la compactación.

La granulometría mide la distribución del tamaño de las partículas (en porcentaje), para ello, el ensayo de análisis granulométrico determina esta característica haciendo pasar una muestra de suelo por una sucesión de tamices que retienen una cantidad de suelo.

Los límites de consistencia o límites de Atterberg (propuestos a principios del siglo XX) definen el estado mecánico del suelo y su capacidad para soportar cargas externas, según Al-Adhadh, et al. (2020) y Zhou, B., & Lu, N. (2021). Para este propósito se requiere realizar varias mediciones. Los límites ayudan a conocer la respuesta del suelo en función de su contenido de humedad y se considera: el límite líquido (LL), que se realiza para evaluar la cantidad de humedad necesaria para que el suelo cambie de un estado semilíquido a uno plástico; el límite plástico (PL), que se realiza para hallar la cantidad de humedad necesaria para convertir el estado del suelo de plástico a semisólido; el límite de contracción (SL), que se requiere para hallar el contenido de humedad de manera que el suelo cambie de semisólido a sólido; y el índice de plasticidad (IP), que se calcula como la diferencia entre el LL y el PL.

Como afirma Katte, et al. (2019) el CBR se usa para calcular la resistencia característica de la subrasante, la misma que sirve para determinar los requisitos mínimos del pavimento y evaluar si el material a usar es el adecuado. El ensayo de CBR se usa para determinar el espesor de los pavimentos flexibles según Li, T., Kong, L., & Liu, B. (2020), evaluar la capacidad portante y la resistencia a la deformación de la subrasante, la subbase, los materiales de base, así como, los materiales reciclados usados en los pavimentos según X. Ji et al (2020). Como señala Mina, E. (2019) cuando un suelo carece de capacidad portante, el valor del CBR será bajo y viceversa. Considerando a Sandoval-Vallejo, E. A., & Rivera-Mena, W. A. (2019) y Bogado, G. O et al. (2017) comentan que, si bien es cierto que el ensayo de CBR es aceptado para el diseño de pavimentos, no siempre es posible elaborarlo ya que el tiempo que requiere su realización es aproximadamente 4 días y el costo es 135 dólares en promedio, y la poca disposición de tiempo y limitaciones económicas en etapas previas al proyecto. Frente a esta situación lo que normalmente se realiza en los proyectos es calcular un valor estimado representativo de CBR para diseñar el pavimento según el MTC.

De acuerdo con Özgan, E., Serin, S., & Vural, I. (2015) y Vafoeva, O. (2020) la compactación es una propiedad vinculada a la resistencia del suelo y se produce cuando la cantidad de aire en el suelo disminuye. Esta densificación del suelo mejora las características físicas y mecánicas del suelo, trayendo como consecuencia el aumento de su capacidad de carga, y para ello se requiere un ensayo de Proctor con el que se obtienen dos parámetros de compactación: la máxima densidad seca (MDS), que es el estado del suelo cuando alcanza su punto máximo de densificación y el óptimo contenido de humedad (OCH), que es la cantidad de agua necesaria para alcanzar la máxima densidad seca, según Chirinos Quispe, J. C. (2016) y Millan-Romero, E., & Millan-Paramo, C. (2020).

Kardani, Navid, et al (2021), Ekanayake, I. U, et al. (2022) y Raja, M. N. A., & Shukla, S. K. (2021) resaltan que anteriormente se propusieron numerosos métodos de computación suave (SC), como algoritmos de machine learning (ML) o métodos de la IA, para capturar el comportamiento complejo de las estimaciones. Hilloulin, B., & Tran, V. Q. (2022) y Adel, H. et al. (2022) afirman que los algoritmos ML se han aplicado eficazmente a numerosos problemas de ingeniería civil, por ejemplo, para estimar el CBR según Brito, J (2021) y los parámetros de compactación del suelo como señala Pacheco-Torgal, F. et al. (2014). Rajakumar, C. (2021) utilizó modelos de redes neuronales para predecir el CBR de suelos estabilizados con desechos industriales; Kurnaz, T. F., & Kaya, Y. (2019) para predecir el CBR de suelos compactados; Suthar, M., & Aggarwal, P. (2018) para predecir el CBR de suelos con cenizas estabilizadas con cal, además, su aplicación es posible en otras disciplinas de la ingeniería civil según Valderrama-Purizaca, F. J, et al. (2021).

En el estudio actual, se consideran las técnicas del modelo Máquina de impulso de gradiente de luz (Light GBM) y del impulso de gradiente extremo (XGBoost) para predecir las propiedades mecánicas del suelo como máxima densidad seca (MDS), óptimo contenido de humedad (OCH), la relación de soporte de California al 100% de la MDS (CBR100) y la relación de soporte de California al 95% de la MDS (CBR95). Según Tran, V. Q. (2022) el concepto de ambos algoritmos es "impulso", lo que sugiere que atribuye todos los valores de salida estimados de un conjunto de aprendizaje "débiles" para establecer un aprendizaje "fuerte" mediante la implementación de métodos de entrenamiento aditivos.

En primer lugar, el algoritmo Light GBM requiere menos tiempo de cálculo, muestra una alta eficiencia computacional y mejores resultados de predicción. Además, procesa datos masivos complejos, no lineales y de alta dimensión de manera eficiente. Según Liang, Bu, Tan, Pan, Yi, Kong y Fan, es un algoritmo de árbol de decisión que aumenta el gradiente (GBDT), pues se ha utilizado ampliamente en la predicción de la calidad del aire, la predicción de la teledetección oceánica, y la predicción de incumplimiento de préstamos de redes P2P, entre otros como afirma Hao, X, et al. (2022). Este método comprende dos algoritmos principales (es decir, muestreo unilateral basado en gradiente (GOSS) y agrupación de características exclusivas (EFB)) y emplea el fenómeno de análisis de componentes principales (PCA) para eliminar el ruido o la redundancia según Zhang, C., Lei, X., & Liu, L. (2021). Chun, P. J., Izumi, S., & Yamane, T. (2021) idearon una técnica automatizada de detección de grietas basada en el procesamiento de imágenes utilizando el algoritmo LightGBM que arrojó una exactitud del 99,7% y una precisión del 68,2%, lo que representa el mayor rendimiento de este método de IA.

En segundo lugar, desarrollado por Chen, T. et al. (2015), el algoritmo XGBoost es un tipo mejorado y sofisticado de modelo de conjunto basado en árboles, y el 'impulso de gradiente' es aplicable en varios casos, por ejemplo, regresión y problemas relacionados con la clasificación como menciona Liu, J. et al. (2022).

El Boosting (Impulsando) es un método conjunto que tiene como objetivo crear un modelo eficaz a partir de una colección de modelos deficientes. Esto se logra separando los datos de entrenamiento en cuatro partes, cada una de las cuales se utiliza para entrenar modelos distintos o un modelo único que tiene una variedad de configuraciones, y luego combinando los resultados con la ayuda de un voto mayoritario según Al Daoud, E. (2019).

Tenga en cuenta que, Nguyen, H. et al. (2019) utilizaron el algoritmo XGBoost formulando un modelo para la predicción de la velocidad máxima de las partículas (PPV) inducida por las voladuras en Vietnam y se reveló que el nivel de precisión de este método es superior al de la máquina de vectores de soporte (SVM), regresión de bosque aleatorio (RF) y redes neuronales tipo k (KNN).

Según Duan, J. et al. (2021) y Imran, H. et al. (2022) concluyeron que el modelo competitivo imperialista (ICA) de XGBoost produjo la mayor precisión y superó a los modelos de IA restantes en la predicción de la resistencia a la compresión del hormigón RA. Liu, J., Zhang, S., & Fan, H. (2022) encontraron que los resultados modelados de XGBoost son eficaces y tienen un mayor rendimiento en el preprocesamiento de datos en el caso de la evaluación del riesgo crediticio. Nguyen, N. H. et al (2022) implementaron el modelo XGBoost para la estimación de predicciones de resistencia a la compresión de concreto de ultra alto rendimiento (UPHC) empleando 931 puntos de datos de la colección de mezclas UHPC con 17 variables de entrada.

Este estudio proporciona una idea del potencial de los modelos de machine learning de XGBoost y Light GBM, basados en técnicas convencionales o de conjunto, para predecir los valores de la máxima densidad seca (MDS), óptimo contenido de humedad (OCH), la relación de soporte de California al 100% de la MDS (CBR100) y la relación de soporte de California al 95% de la MDS (CBR95) utilizando como variables de entrada propiedades como la granulometría y los límites de consistencia.

2. Metodología

2.1. Base de datos experimental

Se recopilaron en total 201 datos experimentales de muestras de suelos del Laboratorio de Mecánica de Suelos de la Facultad de Ingeniería Civil – UNI correspondientes a proyectos de carreteras peruanas desde 2011 hasta el 2020, de las cuales se utilizaron 80% para el escenario de entrenamiento y 20% para el escenario de prueba o validación de los modelos LightGBM y XGBoost. Para la implementación de los modelos se empleó el lenguaje de programación Python v.3.10.12.

Con respecto a la selección de variables (o características), se incluyeron aquellas muestras de suelo con resultados de los ensayos de granulometría y límites de consistencia, clasificación SUCS y AASHTO y CBR, excluyéndose a las muestras de suelo de proyectos que no correspondían a obras de carreteras.

Para la presente investigación se utilizaron en total 10 variables de los ensayos del laboratorio las cuales son: tres variables de la granulometría (grava, arena y finos), tres variables de los límites de consistencia (límite líquido, límite plástico e índice de

plasticidad), dos variables de la compactación (MDS y OCH) y dos variables del CBR (CBR al 95% y CBR al 100%). Las cuales fueron analizadas para ser usadas como variables de entradas de los modelos de machine learning y se descartaron las variables de la clasificación SUCS y AASTHO por ser poco significativas para la estimación. Con respecto a las salidas, los modelos propuestos de machine learning se desarrollaron para predecir: la MDS, el OCH, el CBR al 100% de la MDS y el CBR al 95% de la MDS. Como se puede observar, se propone estimar un único valor en cada modelo, similar a una regresión, cada predicción requiere un modelo específico de machine learning, y para ello se usaron 6 variables de entrada las cuales son: contenido de grava (G), contenido de arena (A), contenido de finos (F), límite líquido (LL), límite plástico (PL) y el índice de plasticidad (PI). Por otra parte, para predecir el CBR al 100% de la MDS y el CBR al 95% de la MDS, en una primera instancia se utilizaron las seis variables de entrada (G, A, F, LL, LP e IP), y en un segundo momento fueron agregados dos nuevas variables, la MDS y el OCH, calculando un total de 8 variables.

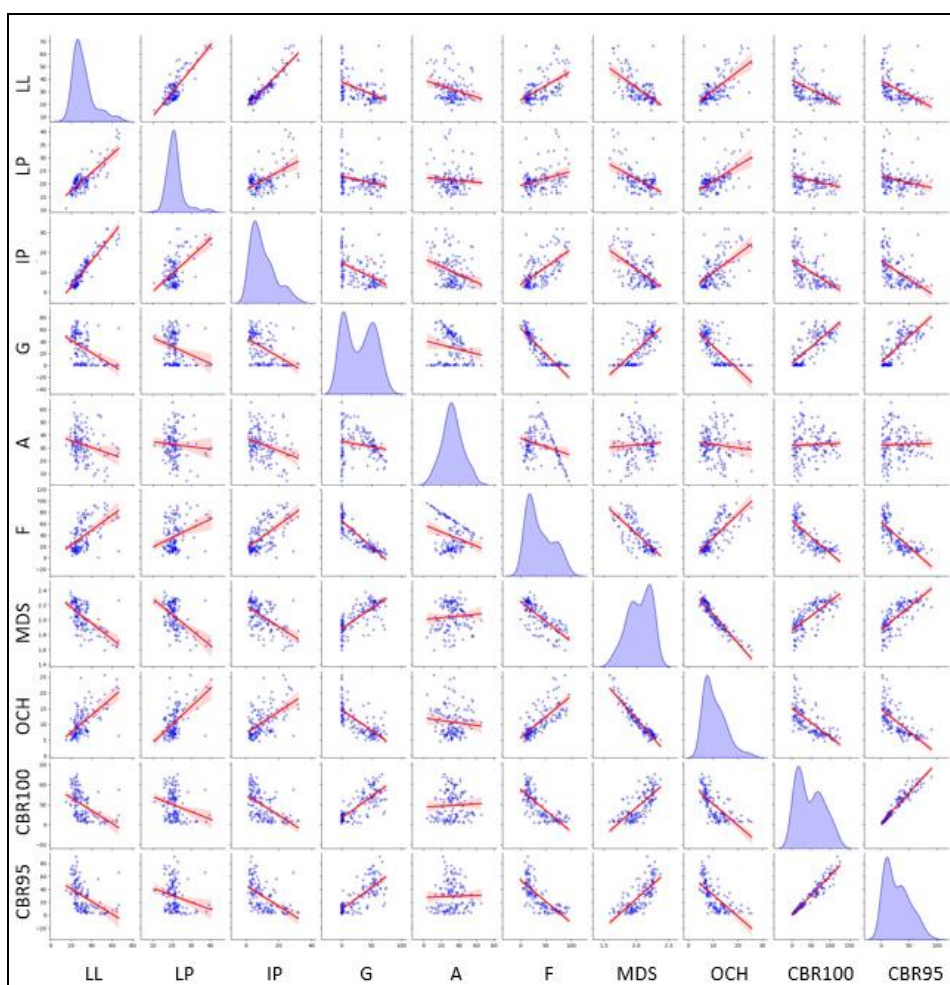


Fig. 1. Diagramas de pares que muestran variaciones de magnitudes de variables entre sí.

para el desarrollo de modelos cuyas estadísticas descriptivas se ilustran en la Tabla 1. La figura 1 muestra los detalles de la matriz de pares en forma de variación de la magnitud de las variables entre sí. seis parámetros de entrada las cuales son: Gravas, arena, finos, limite líquido, limite plástico e índice de plasticidad, para predecir MDS, OCH, el CBR al 100% de la MDS y el CBR al 95% de la MDS.

Con respecto a las variables a predecir, se puede observar que el valor de la MDS disminuye mientras aumenta el OCH, es decir tienen una correlación fuerte negativa de -0.9 la cual también se puede apreciar en la Fig. 2, también vemos que la MDS disminuye cuando aumenta la cantidad de los finos con una correlación de fuerte de -0.7. El CBR al 95% y al 100% de la MDS aumenta mientras aumenta la MDS o las gravas con una correlación positiva de 0.7

Además, se puede apreciar claramente que el OCH aumenta al aumentar los finos (con correlación positiva de 0.7) o al aumentar el LL, LP e IP (con correlación positiva de 0.6), además disminuye al aumentar las gravas. Notamos también que el CBR al 95% de la MDS y el CBR al 100% de la MDS tienen una correlación positiva de 1.0.

Fig. 2. Matriz de coeficientes de correlación.

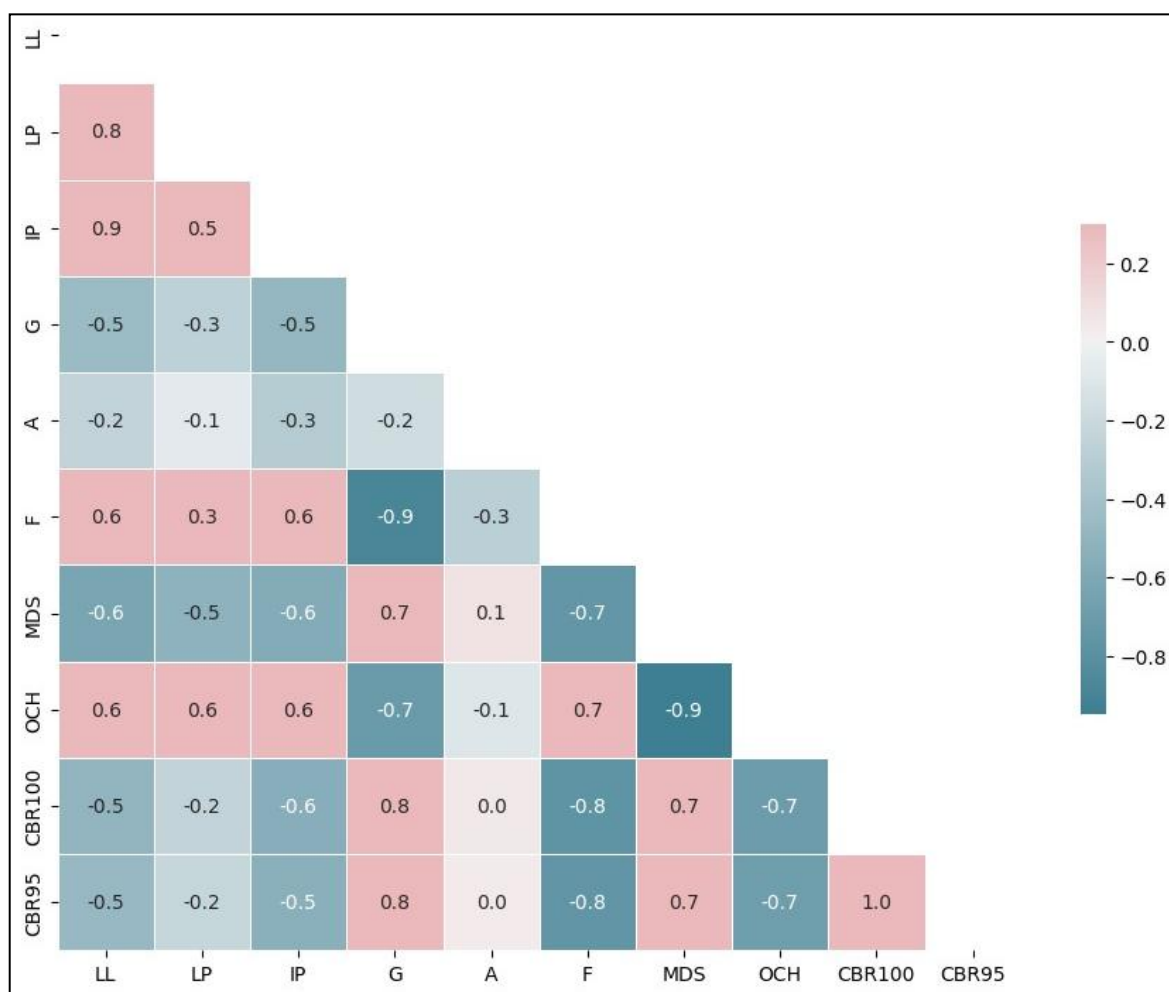


Tabla 1. Estadísticas descriptivas de los datos de entrada utilizados en el desarrollo de modelos.

Estadística descriptiva	LL	LP	IP	G	A	F	MDS	OCH	CBR100	CBR95
Media	31.24	20.26	10.98	26.83	31.86	41.30	2.05	10.45	40.51	24.62
Error estándar	0.66	0.31	0.46	1.59	0.81	1.70	0.01	0.28	2.25	1.43
Mediana	29.00	20.00	11.00	25.00	31.80	39.00	2.08	9.30	29.50	16.90
Moda	28.00	19.00	12.00	0.00	33.00	16.00	2.08	7.80	18.80	2.80
Desviación estándar	9.29	4.41	6.56	22.54	11.42	24.06	0.17	3.99	31.84	20.24
Varianza muestral	86.29	19.45	43.04	508.23	130.34	578.93	0.03	15.95	1014.09	409.67
Kurtosis	3.20	5.72	0.48	-1.25	0.03	-0.95	-0.46	1.70	-0.52	0.09
Asimetría	1.68	1.85	0.89	0.29	0.24	0.41	-0.42	1.23	0.76	0.97
Rango	51.20	30.10	30.34	74.00	61.20	95.80	0.85	21.00	123.50	88.40
Mínimo	15.20	10.60	1.55	0.00	4.20	0.00	1.59	4.60	1.70	1.60
Máximo	66.40	40.70	31.89	74.00	65.40	95.80	2.44	25.60	125.20	90.00
Suma	6278.61	4071.58	2207.03	5393.30	6404.50	8302.30	411.64	2100.80	8143.20	4949.20
Cuenta	201.00	201.00	201.00	201.00	201.00	201.00	201.00	201.00	201.00	201.00
Coefficiente de variación	0.30	0.22	0.60	0.84	0.36	0.58	0.08	0.38	0.79	0.82

2.2 Modelos de machine learning

2.2.1 Descripción general del Light GBM

Light GBM es una forma de árbol de decisión de aumento de gradiente que se establece sobre una técnica de árbol de decisión propuesta por Microsoft Research según Sun, Lei, et al. (2021). Light GBM es un enfoque poderoso para resolver problemas de regresión, clasificación y otros problemas de machine learning. Utiliza menos memoria y supera en precisión de predicción. Busca mejorar el rendimiento informático para abordar de forma más eficaz problemas de predicción de grandes datos según Liang, Weizhang, et al. (2020). El algoritmo Light GBM incluye dos técnicas innovadoras: muestreo unilateral basado en gradientes (GOSS), que se considera que gestiona un gran conjunto de datos, y agrupación de características exclusivas (EFB), que pretende gestionar numerosas características de datos sin problemas de sobreajuste. Basado en el método del histograma y el enfoque de crecimiento de las hojas del árbol, Light GBM mejora el entrenamiento y minimiza el uso de la memoria. Los puntos flotantes posteriores se dividen en pequeños contenedores. A continuación, el histograma con ancho se construye utilizando estos contenedores. Las estadísticas

requeridas (suma de gradiente y recuento de muestras en cada contenedor) se encuentran en el histograma hasta que se examinan los datos por primera vez. El mejor punto de segmentación se puede determinar utilizando el valor discreto del histograma Zeng, Hong, et al. (2021). El costo de almacenamiento y medición se puede reducir empleando este método. La fig. 3 muestra las técnicas de crecimiento nivelado y en hojas sabias. Según el enfoque de crecimiento nivelado, las hojas de la misma capa se dividen al mismo tiempo. Es mejor optimizar con numerosos subprocesos y mantener bajo control la complejidad del modelo. Además, las hojas de una misma capa se manipulan mucho, a pesar de que tienen una adquisición de información diferente. La ganancia de información según Kodaz, Halife, et al. (2009) representa la disminución proyectada de la entropía como resultado de separar los nodos en atributos que pueden derivarse como:

$$IG(C, A) = F_n(C) - \sum_{v \in \text{Values}(V)} \frac{|C_v|}{C} F_n(C_v)$$

$$F_n(C) = \sum_{b=1}^B -pb \log_2 pb$$

Donde $F_n(C)$ es la información de entropía de la colección C , pb es la proporción de C a clase b , B son las diversas clases, C es el valor del atributo V , y C_v es el subconjunto de C para el cual el atributo tiene valor v . Ciertas hojas con una ganancia mínima de información son redundantes, lo que genera un aumento significativo en el consumo de memoria y hace que este método sea ineficaz.

El enfoque de crecimiento en hojas, en particular, es más exitoso ya que solo divide la hoja con mayor ganancia de información en la misma capa (Fig. 4). Además, mediante esta técnica se pueden inducir árboles con una gran profundidad; Para evitar el sobreajuste, se impone un límite de profundidad máxima durante el crecimiento del árbol según Kodaz, Halife, et al. (2009).

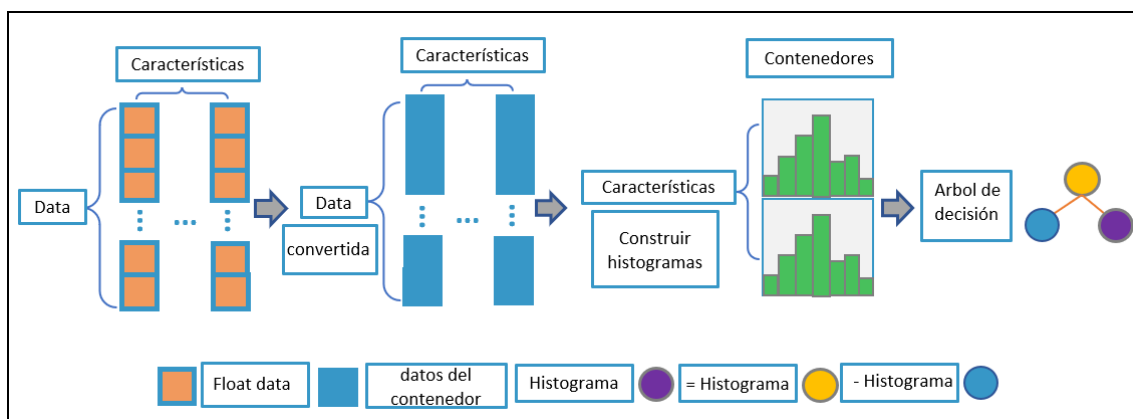


Fig. 3. Árbol de decisión basado en histograma.

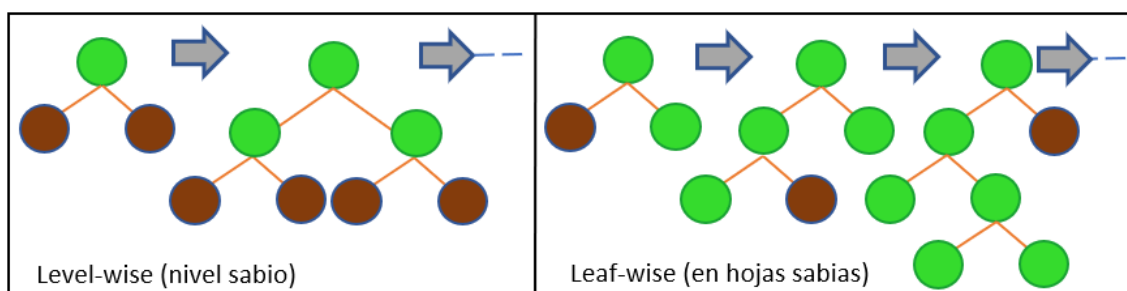


Fig.4. construcción del árbol de decisión.

2.2.2 Descripción general de XGBoost

XGboost es un algoritmo de machine learning recientemente establecido que ha encontrado un uso generalizado en un campo diferente. El algoritmo XGboost introducido por Chen y Guestrin es un enfoque novedoso implementado para máquinas de aumento de gradiente, particularmente para árboles de regresión de clasificación según Chen, T., & Guestrin, C. (2016, August). La técnica se basa en la idea de "impulso", que combina la predicción de aprendiz debil con enfoques de entrenamiento aditivo para construir un alumno poderoso. XGboost ayuda a evitar el sobreajuste, así como a optimizar las capacidades computacionales como mencionan Fan, Junliang, et al. (2018) y Ullah, Irfan, et al. (2022). Esto se logra minimizando las funciones objetivo, lo que permite combinar los términos de predicción y regularización mientras se mantiene la velocidad de procesamiento más rápida posible. Durante el proceso de

entrenamiento, la estimación similar de XGboost también se realiza automáticamente. Se describen los mecanismos de aprendizaje aditivo en XGboost. Para fijar los límites de un aprendiz débil, el primer aprendiz se ajusta al conjunto de datos de entrada completo y luego el segundo modelo se ajusta a los residuos de acuerdo con Ullah, Irfan, et al. (2022). La figura 5 muestra el esquema de XGboost. Antes de que se alcance la condición de parada, este enfoque aceptable se realiza varias veces. El total de las predicciones de cada alumno produce la predicción del modelo final. Las siguientes ecuaciones determinan la función genérica de predicción de fase.

$$f_i^p = \sum_{k=1}^l f_k(x_i) = f_i^{(p-1)} + f_i(x_i)$$

Donde $f_p(x_i)$ es el aprendiz en la fase p , $f_i^{(p)}$ y $f_i^{(p-1)}$ son la predicción de la fase p y $p-1$, y x_i son las variables de entrada.

Para evitar problemas de sobreajuste y al mismo tiempo perder el rendimiento computacional del modelo, XGBoost crea la fórmula analítica para evaluar la "bondad" del modelo basado en la funcionalidad original.

$$Obj^{(p)} = \sum_{k=1}^n l(\bar{y}_i, y_i) + \sum_{k=1}^p \sigma(f_i)$$

donde l denota la función de pérdida, n indica el número de observaciones y σ representa el término de regularización, el cual se describe a continuación:

$$\sigma(f) = \gamma T + \frac{1}{2} \lambda \|\omega\|^2$$

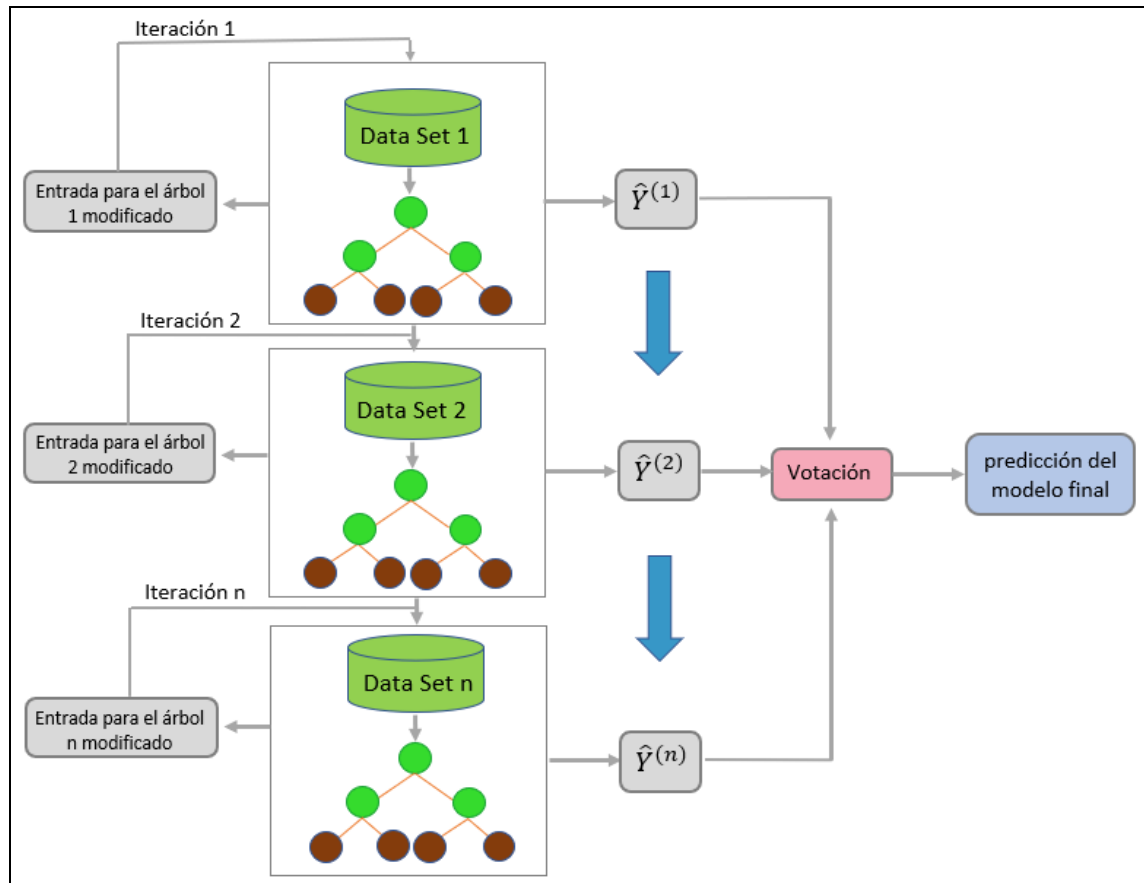


Fig. 5. Ilustración de la regresión XGBoost.

Dónde ω denota puntuaciones vectoriales en hojas, γ es la pérdida mínima necesaria para dividir aún más el [nodo de la hoja](#) y λ representan los [parámetros de regularización](#). Chen, T., & Guestrin, C. (2016) proporcionan una ilustración detallada de XGBoost.

2.3. Desarrollo de modelos

El modelo depende de una combinación de características que son esenciales para la eficacia del modelo. El método de grid search se utilizó como optimización de hiperparámetros para XGboost y Light GBM con el fin de identificar un rendimiento mejorado y una alta precisión. Este método estima el rendimiento para cada combinación de todos los hiperparámetros indicados y sus valores, y luego elige el valor

óptimo para los hiperparámetros. En la optimización de hiperparámetros de grid search, un subconjunto de puntos de datos se mantiene completamente enmascarado de los modelos y se utiliza como un "conjunto de prueba" para mejorar la precisión de los modelos de ML y eliminar cualquier posible sobreajuste. En este estudio, los parámetros optimizados del experimento revelaron un mayor rendimiento para la identificación y predicción de la máxima densidad seca, óptimo contenido de humedad, CBR95 y CBR100.

La optimización de los hiperparámetros mediante el grid search de los modelos XGBoost y Light GBM para la MDS, OCH, CBR100 y CBR95 con las 6 variables (G, A, F, LL, LP e IP) se muestra en las tablas 2, 3,5,6,8,9,11,12.

La optimización de los hiperparámetros mediante el grid search de los modelos XGBoost y Light GBM para el CBR100 y CBR95 con las 8 variables (G, A, F, LL, LP, IP, MDS y OCH) se muestra en las tablas 14,15,17 y 18.

2.4. Evaluación del modelo

Los modelos se evaluaron utilizando índices estadísticos (métricas de rendimiento) de uso común, a saber, error absoluto medio (MAE), error cuadrático medio (RMSE) y error porcentual absoluto medio (MAPE) de acuerdo con Amin, Muhammad N. et al. (2022); Bardhan, A. et al. (2021) y Bardhan, A. et al. (2021). Tenga en cuenta que los valores ideales de RMSE, MAE y MAPE es 0.

3. Resultados y discusión

3.1. Evaluación estadística

De las tablas 4,7,10,13, 16 y 19 se determinan los valores de las métricas de desempeño (RMSE, MAE y MAPE) para la MDS, OCH, CBR100 y CBR95, y luego para el CBR100 y CBR95 agregando la MDS y OCH (8 variables) con los enfoques LightGBM y XGBoost para el escenario de entrenamiento (training) y conjuntos de datos de prueba (testing).

- Para el caso de la MDS se puede apreciar que en el escenario de entrenamiento y prueba el modelo XGBoost supera al modelo LightGBM en todas las métricas de desempeño en cuanto al RMSE, MAE y MAPE.
- Para el caso de la OCH se puede apreciar que en el escenario de entrenamiento y prueba el modelo XGBoost supera al modelo LightGBM en todas las métricas de desempeño en cuanto al RMSE, MAE y MAPE.
- Para el caso de la CBR100 se puede apreciar que en el escenario de entrenamiento y prueba el modelo XGBoost supera al modelo LightGBM en todas las métricas de desempeño en cuanto al RMSE, MAE y MAPE.
- Para el caso de la CBR95 se puede apreciar que en el escenario de entrenamiento y prueba el modelo XGBoost supera al modelo LightGBM en todas las métricas de desempeño en cuanto al RMSE, MAE y MAPE.
- Para el caso de la CBR100 con las 8 variables (G, A, F, LL, LP, IP, MDS, OCH) se puede apreciar que en el escenario de entrenamiento y prueba el modelo XGBoost supera al modelo LightGBM en todas las métricas de desempeño en cuanto al RMSE, MAE y MAPE.
- Para el caso de la CBR95 con las 8 variables (G, A, F, LL, LP, IP, MDS, OCH) se puede apreciar que en el escenario de entrenamiento y prueba el modelo XGBoost supera al modelo LightGBM en todas las métricas de desempeño en cuanto al RMSE, MAE y MAPE.

Dado que los valores de RMSE, MAE y MAPE en el caso de XGBoost son menores en comparación a LightGBM para todo el conjunto de datos (entrenamiento y prueba), por lo que el modelo XGBoost representa el mayor rendimiento con respecto al modelo LightGBM obteniendo menores errores en las predicciones para la MDS, OCH, CBR100, CBR95 y (CBR100 y CBR95 con 8 variables).

3.1.1. Hiperparámetros y métricas de rendimiento considerando las 6 variables (G, A, F, LL, LP e IP) para predecir el MDS, OCH, CBR100 Y CBR95.

3.1.1.1. Predicción de la máxima densidad seca (MDS).

Tabla 2. Hiperparámetros ajustados u optimizados de Light GBM para predecir la MDS.

Parámetros	Descripción	Valor	Rango
leargning_rate	Para reducir el paso de gradiente	0.01	0.01-0.2
max_depth	Profundidad del árbol	1	1 - 8
n_estimators	Construir el máximo número de árboles posible	600	100-600
colsample_bytree	Proporción de submuestra de columnas al construir cada árbol	0.7	0.1-0.70
num_leaves	Número máximo de hojas	7	1 - 8

Tabla 3. Hiperparámetros ajustados u optimizados de XGboost para predecir la MDS.

Parámetros	Descripción	Valor	Rango
leargning_rate	Para reducir el paso de gradiente	0.01	0.01-0.2
max_depth	Profundidad del árbol	1	1 - 8
n_estimators	Construir el máximo número de árboles posible	600	100-600
min_child_weight	Número mínimo de instancias necesarias para cada nodo	5	1-5
colsample_bytree	Proporción de submuestra de columnas al construir cada árbol	0.7	0.1-0.7

Tabla 4. Comparación entre el rendimiento del modelo LightGBM y XGBoost para predecir la MDS.

Escenario	Métricas de rendimiento	Light GBM	XGBoost
Training	RMSE	0.09354	0.09179
	MAE	0.07274	0.07055
	MAPE	3.65086	3.53706
Testing	RMSE	0.10251	0.10094
	MAE	0.07871	0.07793
	MAPE	3.88486	3.84061

3.1.1.2. Predicción del Óptimo Contenido de Humedad (OCH).

Tabla 5. Hiperparámetros ajustados u optimizados de Light GBM para predecir el OCH.

Parámetros	Descripción	Valor	Rango
leargning_rate	Para reducir el paso de gradiente	0.01	0.01-0.2
max_depth	Profundidad del árbol	1	1 - 8
n_estimators	Construir el máximo número de árboles posible	600	100-600
colsample_bytree	Proporción de submuestra de columnas al construir cada árbol	0.7	0.1-0.70
num_leaves	Número máximo de hojas	8	1 - 8

Tabla 6. Hiperparámetros ajustados u optimizados de XGboost para predecir el OCH.

Parámetros	Descripción	Valor	Rango
leargning_rate	Para reducir el paso de gradiente	0.01	0.01-0.2
max_depth	Profundidad del árbol	1	1 - 8
n_estimators	Construir el máximo número de árboles posible	600	100-600
min_child_weight	Número mínimo de instancias necesarias para cada nodo	5	1-5
colsample_bytree	Proporción de submuestra de columnas al construir cada árbol	0.7	0.1-0.7

Tabla 7. Comparación entre el rendimiento del modelo LightGBM y XGBoost para predecir el OCH.

Escenario	Métricas de rendimiento	Light GBM	XGBoost
Training	RMSE	2.30796	2.19852
	MAE	1.6748	1.6027
	MAPE	16.1972	15.8269
Testing	RMSE	2.4461	2.3066
	MAE	1.7764	1.6928
	MAPE	18.5069	17.9821

3.1.1.3. Predicción del CBR al 100% de la MDS (CBR100).

Tabla 8. Hiperparámetros ajustados u optimizados de Light GBM para predecir el CBR100.

Parámetros	Descripción	Valor	Rango
leargning_rate	Para reducir el paso de gradiente	0.01	0.01-0.2
max_depth	Profundidad del árbol	5	1 - 8
n_estimators	Construir el máximo número de árboles posible	600	100-600
colsample_bytree	Proporción de submuestra de columnas al construir cada árbol	0.1	0.1-0.70
num_leaves	Número máximo de hojas	6	1 - 8

Tabla 9. Hiperparámetros ajustados u optimizados de XGboost para predecir el CBR100.

Parámetros	Descripción	Valor	Rango
leargning_rate	Para reducir el paso de gradiente	0.01	0.01-0.2
max_depth	Profundidad del árbol	2	1 - 8
n_estimators	Construir el máximo número de árboles posible	400	100-600
min_child_weight	Número mínimo de instancias necesarias para cada nodo	5	1 - 5
colsample_bytree	Proporción de submuestra de columnas al construir cada árbol	0.7	0.1-0.70

Tabla 10. Comparación entre el rendimiento del modelo LightGBM y XGBoost para predecir el CBR100.

Escenario	Métricas de rendimiento	Light GBM	XGBoost
Training	RMSE	14.91129	13.65742
	MAE	11.54861	10.58901
	MAPE	56.96979	52.49169
Testing	RMSE	17.12007	16.75096
	MAE	12.67346	12.51784
	MAPE	58.42492	52.85221

3.1.1.4. Predicción del CBR al 95% de la MDS (CBR95).

Tabla 11. Hiperparámetros ajustados u optimizados de Light GBM para predecir el CBR95.

Parámetros	Descripción	Valor	Rango
leargning_rate	Para reducir el paso de gradiente	0.01	0.01-0.2
max_depth	Profundidad del árbol	3	1 - 8
n_estimators	Construir el máximo número de árboles posible	600	100-600
colsample_bytree	Proporción de submuestra de columnas al construir cada árbol	0.1	0.1-0.70
num_leaves	Número máximo de hojas	6	1 - 8

Tabla 12. Hiperparámetros ajustados u optimizados de XGBoost para predecir el CBR95.

Parámetros	Descripción	Valor	Rango
leargning_rate	Para reducir el paso de gradiente	0.01	0.01-0.2
max_depth	Profundidad del árbol	2	1 - 8
n_estimators	Construir el máximo número de árboles posible	600	100-600
min_child_weight	Número mínimo de instancias necesarias para cada nodo	5	1 - 5
colsample_bytree	Proporción de submuestra de columnas al construir cada árbol	0.7	0.1-0.70

Tabla 13. Comparación entre el rendimiento del modelo LightGBM y XGBoost para predecir el CBR95.

Escenario	Métricas de rendimiento	Light GBM	XGBoost
Training	RMSE	9.76944	8.017
	MAE	7.24873	5.98203
	MAPE	55.99586	45.12968
Testing	RMSE	12.39321	12.33617
	MAE	9.15759	8.89962
	MAPE	59.09576	49.22769

3.1.2. Hiperparámetros y métricas de rendimiento considerando las 8 variables (G, A, F, LL, LP, IP, MDS y OCH) para predecir el CBR100 Y CBR95.

3.1.2.1. Predicción del CBR al 100% de la MDS (CBR100) utilizando las 8 variables (G, A, F, LL, LP, IP, MDS y OCH).

Tabla 14. Hiperparámetros ajustados u optimizados de Light GBM para predecir el CBR100 utilizando las 8 variables (G, A, F, LL, LP, IP, MDS y OCH).

Parámetros	Descripción	Valor	Rango
leargning_rate	Para reducir el paso de gradiente	0.01	0.01-0.2
max_depth	Profundidad del árbol	3	1 - 8
n_estimators	Construir el máximo número de árboles posible	600	100-600
colsample_bytree	Proporción de submuestra de columnas al construir cada árbol	0.1	0.1-0.70
num_leaves	Número máximo de hojas	6	1 - 8

Tabla 15. Hiperparámetros ajustados u optimizados de XGBoost para predecir el CBR100 utilizando las 8 variables (G, A, F, LL, LP, IP, MDS y OCH).

Parámetros	Descripción	Valor	Rango
leargning_rate	Para reducir el paso de gradiente	0.1	0.01-0.2
max_depth	Profundidad del árbol	1	1 - 8
n_estimators	Construir el máximo número de árboles posible	100	100-600
min_child_weight	Número mínimo de instancias necesarias para cada nodo	5	1 - 5
colsample_bytree	Proporción de submuestra de columnas al construir cada árbol	0.1	0.1-0.70

Tabla 16. Comparación entre el rendimiento del modelo LightGBM y XGBoost para predecir el CBR100 utilizando las 8 variables (G, A, F, LL, LP, IP, MDS y OCH).

Escenario	Métricas de rendimiento	Light GBM	XGBoost
Training	RMSE	14.53778	13.31218
	MAE	11.53853	10.44234
	MAPE	59.14108	53.97036
Testing	RMSE	16.07289	15.44851
	MAE	12.74025	12.21478
	MAPE	61.23694	56.40786

3.1.2.2. Predicción del CBR al 95% de la MDS (CBR95) utilizando las 8 variables (G, A, F, LL, LP, IP, MDS y OCH).

Tabla 17. Hiperparámetros ajustados u optimizados de Light GBM para predecir el CBR95 utilizando las 8 variables (G, A, F, LL, LP, IP, MDS y OCH).

Parámetros	Descripción	Valor	Rango
learning_rate	Para reducir el paso de gradiente	0.01	0.01-0.2
max_depth	Profundidad del árbol	3	1 - 8
n_estimators	Construir el máximo número de árboles posible	600	100-600
colsample_bytree	Proporción de submuestra de columnas al construir cada árbol	0.1	0.1-0.70
num_leaves	Número máximo de hojas	6	1 - 8

Tabla 18. Hiperparámetros ajustados u optimizados de XGBoost para predecir el CBR95 utilizando las 8 variables (G, A, F, LL, LP, IP, MDS y OCH).

Parámetros	Descripción	Valor	Rango
learning_rate	Para reducir el paso de gradiente	0.2	0.01-0.2
max_depth	Profundidad del árbol	1	1 - 8
n_estimators	Construir el máximo número de árboles posible	100	100-600
min_child_weight	Número mínimo de instancias necesarias para cada nodo	1	1 - 5
colsample_bytree	Proporción de submuestra de columnas al construir cada árbol	0.4	0.1-0.70

Tabla 19. Comparación entre el rendimiento del modelo LightGBM y XGBoost para predecir el CBR95 utilizando las 8 variables (G, A, F, LL, LP, IP, MDS y OCH).

Escenario	Métricas de rendimiento	Light GBM	XGBoost
Training	RMSE	8.61229	8.06015
	MAE	6.44289	5.97108
	MAPE	50.99823	47.62836
Testing	RMSE	11.67941	11.6123
	MAE	8.84248	8.77718
	MAPE	54.61247	49.75896

3.2. Comparación de pendientes de regresión

Se investigó el rendimiento de los modelos desarrollados utilizando la pendiente de regresión de la tendencia de la línea al trazar los resultados experimentales en el eje x y los resultados predichos en el eje y (Figuras 6 al 11). Este tipo de investigación para evaluar el modelo de IA ha sido practicado previamente por numerosos investigadores como Khan, M. Imran, et al. (2022); Iqbal, Mudassir, et al. (2021) y qbal, Mudassir, et al. (2021). También se ha informado que un valor de la pendiente superior a 0.80 interpreta una buena concordancia entre los resultados experimentales y los previstos según Jalal, Fazal E., et al. (2021).

Los resultados predichos de los modelos se compararon con los resultados experimentales y se muestran en la Fig. 6 para el caso de la MDS, y se pudo ver que los dos modelos se correlacionan de forma moderada con los resultados experimentales, es decir, con pendientes de 0.6199 y 0.6205 de la línea de regresión para el conjunto de entrenamiento de LightGBM y XGBoost respectivamente. Asimismo, los conjuntos de datos de prueba también mostraron una correlación moderada de las pendientes de la línea de regresión en comparación con la pendiente de ajuste ideal (pendiente = 1) como se muestra en la Fig.6. Los datos de prueba arrojaron pendientes positivos de 0.6236 y 0.618 de la línea de regresión para LightGBM y XGBoost respectivamente. Análogamente se pudo apreciar que, para el OCH, CBR100 y CBR95 (figuras del 7 al 11) dado que los valores de las pendientes de regresión resultaron ser menores a 0.8 se deduce que ambos modelos interpretan una moderada correlación entre los resultados experimentales y predichos.

Se puede ver que en el caso de la MDS y OCH (figuras 6 y 7) las pendientes para los datos de entrenamiento el modelo XGBoost son mayores que LightGBM, sin embargo, las pendientes para los datos de prueba son bajos en el modelo XGBoost en comparación de LightGBM. Por otra parte, de las figuras 8 al 11 para el caso del CBR100 y CBR95 (considerando 6 y 8 variables de entrada) se aprecia que las pendientes para los datos de entrenamiento y prueba el modelo XGBoost son mayores que LightGBM mostrando más robustez.

3.2.1. Pendientes de regresión considerando las 6 variables (G, A, F, LL, LP e IP) para predecir la MDS, OCH, CBR100 y CBR95.

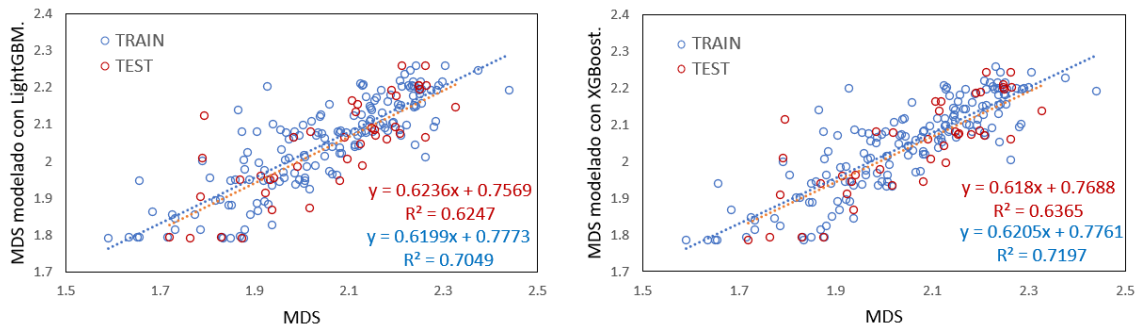


Fig. 6. Comparación de pendientes de regresión de los modelos desarrollados para la máxima densidad seca (MDS). En la izquierda se aprecia la pendiente de regresión para el modelo LightGBM, y a la derecha la pendiente de regresión para el modelo XGBoost.

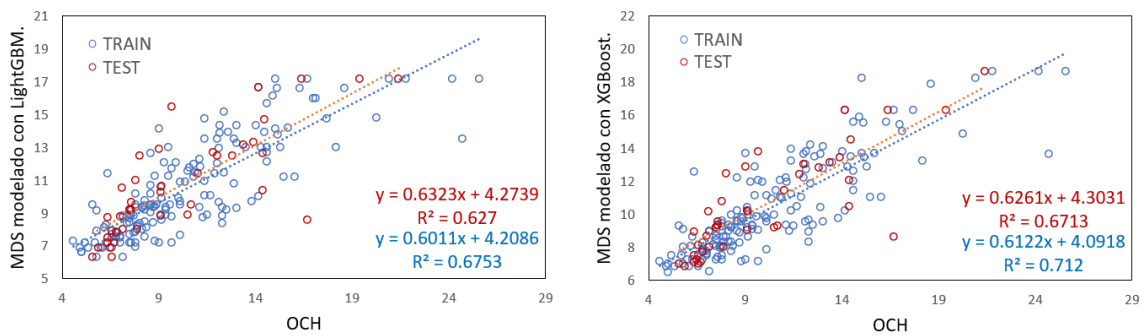


Fig. 7. Comparación de pendientes de regresión de los modelos desarrollados para el óptimo contenido de humedad (OCH). En la izquierda se aprecia la pendiente de regresión para el modelo LightGBM, y a la derecha la pendiente de regresión para el modelo XGBoost.

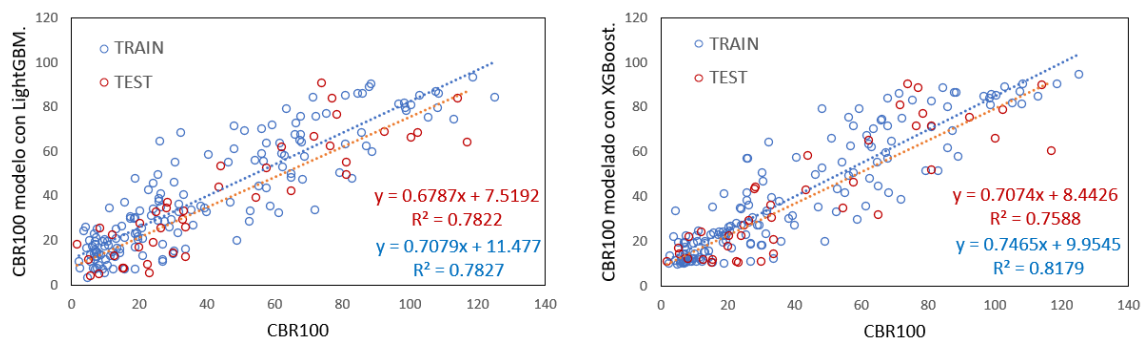


Fig. 8. Comparación de pendientes de regresión de los modelos desarrollados para la relación de soporte de California al 100% de la MDS (CBR100). En la izquierda se aprecia la pendiente de regresión para el modelo LightGBM, y a la derecha la pendiente de regresión para el modelo XGBoost.

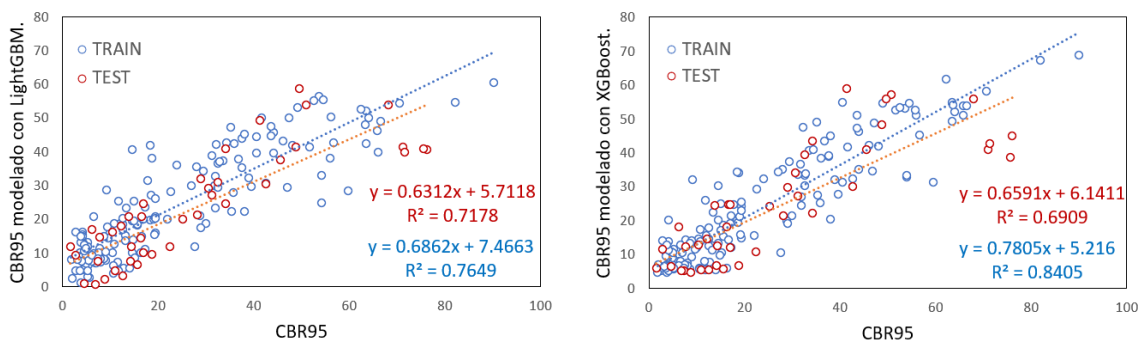


Fig. 9. Comparación de pendientes de regresión de los modelos desarrollados para la relación de soporte de California al 95% de la MDS (CBR95). En la izquierda se aprecia la pendiente de regresión para el modelo LightGBM, y a la derecha la pendiente de regresión para el modelo XGBoost.

3.2.2. Pendientes de regresión considerando las 8 variables (G, A, F, LL, LP, IP, MDS y OCH) para predecir el CBR100 y CBR95.

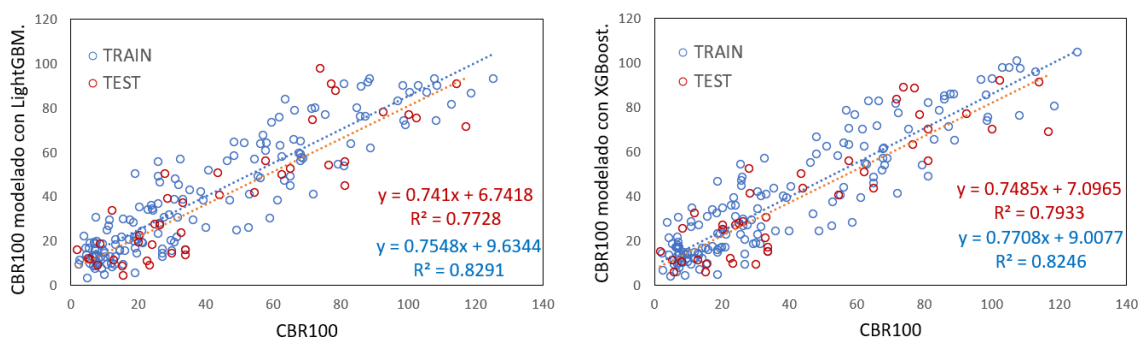


Fig. 10. Comparación de pendientes de regresión de los modelos desarrollados para la relación de soporte de California al 100% de la MDS (CBR100) agregando las variables de la MDS y OCH. En la izquierda se aprecia la pendiente de regresión para el modelo LightGBM, y a la derecha la pendiente de regresión para el modelo XGBoost.

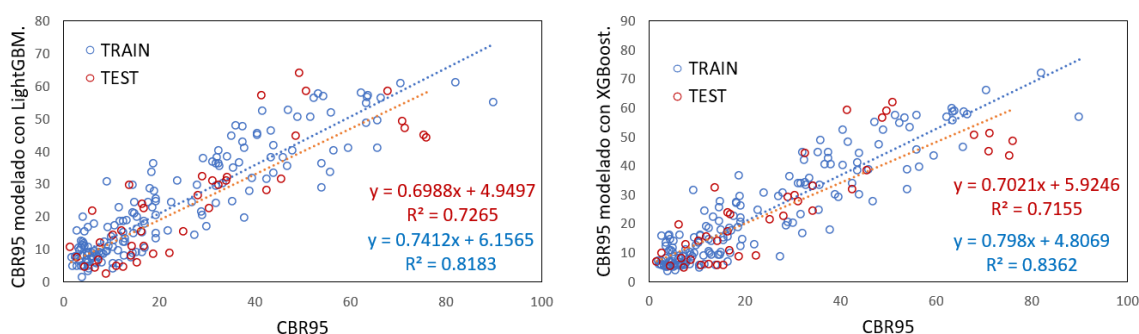


Fig. 11. Comparación de pendientes de regresión de los modelos desarrollados para la relación de soporte de California al 95% de la MDS (CBR95) agregando las variables de la MDS y OCH. En la izquierda se aprecia la pendiente de regresión para el modelo LightGBM, y a la derecha la pendiente de regresión para el modelo XGBoost.

4. Conclusiones

- El análisis de correlación mediante el diagrama de pares mostró que la máxima densidad seca (MDS) tiende a disminuir conforme aumentan el contenido óptimo de humedad (OCH), los límites de consistencia y el porcentaje de finos, lo cual es consistente con el comportamiento típico de suelos más plásticos o con mayor proporción de partículas finas, que requieren más agua para compactarse y presentan menor compactabilidad. Asimismo, los valores del CBR al 95% y 100% de la MDS aumentan cuando incrementa la MDS o el contenido de gravas, lo cual refleja una mayor resistencia estructural del suelo al esfuerzo cortante, atribuida a una mejor compactación y a la presencia de partículas gruesas que mejoran el soporte.

- Para la predicción de MDS y OCH, el modelo XGBoost alcanzó su rendimiento óptimo con una tasa de aprendizaje de 0.01, una profundidad máxima de 1, 600 árboles, mínimo de 5 instancias por nodo, y una proporción de submuestreo de columnas de 0.7.

- En el caso del CBR100 y CBR95, considerando seis variables de entrada (grava, arena, finos, límite líquido, límite plástico e índice de plasticidad), el mejor rendimiento del modelo XGBoost se logró con:

- CBR100 → tasa de aprendizaje 0.01, profundidad 2, 400 árboles, 5 instancias, submuestreo de 0.7.

- CBR95 → tasa de aprendizaje 0.01, profundidad 2, 600 árboles, 5 instancias, submuestreo de 0.7.
- Al considerar ocho variables de entrada (agregando MDS y OCH), el rendimiento óptimo del modelo XGBoost para la predicción del CBR fue:
 - CBR100 → tasa de aprendizaje 0.1, profundidad 1, 100 árboles, 5 instancias, submuestreo de 0.1.
 - CBR95 → tasa de aprendizaje 0.2, profundidad 1, 100 árboles, 1 instancia, submuestreo de 0.4.
- En términos de precisión, el modelo XGBoost superó al modelo LightGBM en todas las propiedades analizadas (MDS, OCH, CBR100 y CBR95), mostrando menores valores de error (RMSE, MAE y MAPE) y mayor ajuste entre valores predichos y experimentales.
- Las pendientes de regresión inferiores a 0.8 reflejan una correlación moderada entre los resultados experimentales y los predichos. En la predicción de MDS y OCH, XGBoost mostró mejor ajuste en los datos de entrenamiento en comparación con LightGBM, aunque su rendimiento en datos de prueba fue menor. En contraste, para el CBR100 y CBR95, XGBoost mostró mayor robustez tanto en entrenamiento como en prueba, superando a LightGBM especialmente al incorporar las ocho variables de entrada.

Referencias

- Banerjee, A., Duflo, E., Qian, N., 2020. On the road: access to transportation infrastructure and economic growth in China. *J. Dev. Econ.* 145 <https://doi.org/10.1016/j.jdeveco.2020.102442>.
- Banerjee, A., Duflo, E., Qian, N., 2012. On the road: access to transportation infrastructure and economic growth in China. *Work. Pap.* 50.
- Ghani, E., Goswami, A.G., Kerr, W.R., 2016. Highway to success: the impact of the Golden quadrilateral project for the location and performance of Indian manufacturing. *Econ. J.* 126 (591), 317–357. <https://doi.org/10.1111/econj.12207>.

- Calderón, C., Servén, L., 2014. Infrastructure, Growth, and Inequality: an Overview, Policy Research Working Papers. <https://doi.org/10.1596/1813-9450-7034>.
- Agrawal, A., Galasso, A., Oettl, A., 2017. Roads and innovation. Rev. Econ. Stat. 99 (3), 417–434. https://doi.org/10.1162/REST_a_00619.
- Gibbons, S., Lyytikäinen, T., Sanchis-Guarner, R., 2019. New road infrastructure: the effects on firms. J. Urban Econ. 110, 35–50.
- [B M. Das and N. Sivakugan, Principles of Foundation Engineering, 9th ed. USA: Cengage Learning, 2019.
- [A. Özyankı, “Soil classification by using artificial neural networks,” Master’s thesis, Near East University, Nicosi, 2019.
- A. Raad Al-Adhadh, H. Kadhem Sakban, and Z. Tawfiq Naeem, “Effect of Method of Soil Drying On Atterberg Limits and Soil Classification,” IOP Conf. Ser. Mater. Sci. Eng., vol. 739, no. 1, Feb. 2020, : 10.1088/1757-899X/739/1/012044.
- B. Zhou and N. Lu, “Correlation between Atterberg Limits and Soil Adsorptive Water,” J. Geotech. Geoenvironmental Eng., vol. 147, no. 2, Feb. 2021, doi: 10.1061/(ASCE)GT.1943- 5606.0002463.
- V. Y. Katte, S. M. Mfoyet, B. Manefouet, A. S. L. Wouatong, and L. A. Bezeng, “Correlation of California Bearing Ratio (CBR) Value with Soil Properties of Road Subgrade Soil,” Geotech. Geol. Eng., vol. 37, no. 1, pp. 217–234, Jan. 2019, doi: 10.1007/s10706-018-0604-x.
- T. Li, L. Kong, and B. Liu, “The California Bearing Ratio and Pore Structure Characteristics of Weakly Expansive Soil in Frozen Areas,” Appl. Sci., vol. 10, no. 21, pp. 1–22, Oct. 2020, doi: 10.3390/app10217576.
- X. Ji et al., “A Prediction Method for the California Bearing Ratio of Soil-Rock Mixture Based on the Discrete Element Method and CT Scanning,” Adv. Civ. Eng., vol. 2020, pp. 1–12, Sep. 2020, doi: 10.1155/2020/9794756.
- E. Mina, R. I. Kusuma, and N. Ulfah, “Utilization of steel slag and fly ash in soil stabilization and their effect to californi bearing ratio (CBR) value. (Case study:

- Kp. Kadusentar road Medong village Mekarjaya Subdistrict Pandeglang District),” in IOP Conference Series: Materials Science and Engineering, Dec. 2019, vol. 673, no. 1, doi: 10.1088/1757-899X/673/1/012034.
- E. A. Sandoval Vallejo y W. A. Rivera Mena, “Correlación del CBR con la resistencia a la compresión inconfnada”, Cienc. e Ing. Neogranadina, vol. 29, no. 1, pp. 135–152, Aug. 2019, doi: 10.18359/rcin.3478.
- Universidad Nacional de Ingeniería, “Tarifa de ensayos de laboratorio”, Lima, 2019. [En línea]. Disponible en: <http://www.lms.uni.edu.pe/labsuelos/tarifa/tarifa-general-2019.pdf>.
- Bogado, G. O., Pintos, N. A., Reinert, H. O., & Bressan, D. A. (2017). Correlación entre parámetros de compactación y propiedades geotécnicas en suelos tropicales misioneros.
- Ministerio de Transportes y Comunicaciones, “Manual de carreteras–Suelos, Geología, Geotécnia y Pavimentos. Sección: Suelos y Pavimentos”. Perú, 2014.
- E. Özgan, S. Serin, and İ. Vural, “Multi-faceted investigation and modeling of compaction parameters for road construction,” J. Terramechanics, vol. 60, pp. 33–42, Aug. 2015, doi: 10.1016/j.jterra. 2015.02.005.
- Vafoeva, O. (2020, July). Hydromechanical Method of Soil Compaction. In IOP Conference Series: Materials Science and Engineering (Vol. 883, No. 1, p. 012061). IOP Publishing.
- J. C. Chirinos Quispe, “Efecto de la energía de compactación en la densidad seca máxima y contenido óptimo de humedad del suelo granular de la cantera El Gavilán, 2015”, Tesis de licenciatura, Universidad Privada del Norte, 2016.
- E. Millan-Romero and C. Millan-Paramo, “Relationship of water parameters with the optimum moisture content in clay soils,” ARPN J. Eng. Appl. Sci., vol. 15, no. 15, pp. 1666–1671, 2020.
- I. Ekanayake, D. Meddage, U. Rathnayake, A novel approach to explain the black-box nature of machine learning in compressive strength predictions of concrete using

- Shapley additive explanations (SHAP), *Case Studies, Constr. Mater.* (2022) e01059.
- N. Kardani, A. Bardhan, D. Kim, P. Samui, A. Zhou, Modelling the energy performance of residential buildings using advanced computational frameworks based on RVM GMDH, ANFIS-BBO and ANFIS-IPSO, *J. Build. Eng.* 35 (2021), 102105.
- M.N.A. Raja, S.K. Shukla, Predicting the settlement of geosynthetic-reinforced soil foundations using evolutionary artificial intelligence technique, *Geotextiles and Geomembranes* 49 (5) (2021) 1280–1293.
- B. Hilloulin, V.Q. Tran, Using machine learning techniques for predicting autogenous shrinkage of concrete incorporating superabsorbent polymers and supplementary cementitious materials, *J. Build. Eng.* (2022), 104086.
- H. Adel, M.I. Ghazaan, A.H. Korayem, Machine learning applications for developing sustainable construction materials, *Artificial Intelligence and Data Science in Environmental Sensing*, Elsevier (2022) 179–210.
- J. de Brito, C. Thomas, C. Medina, F. Agrela, *Waste and byproducts in cement-based materials: innovative sustainable materials for a circular economy*, Woodhead Publishing, 2021.
- F. Pacheco-Torgal, L.F. Cabeza, J. Labrincha, A.G. De Magalhaes, *Eco-efficient construction and building materials: life cycle assessment (LCA), eco-labelling and case studies*, woodhead Publishing, 2014.
- C. Rajakumar and G. Reddy Babu, “Experimental study and neural network modelling of expansive sub grade stabilized with industrial waste byproducts and geogrid,” *Mater. Today Proc.*, Aug. 2020, doi: 10.1016/j.matpr.2020.06.578
- T. Fikret Kurnaz and Y. Kaya, “Prediction of the California bearing ratio (CBR) of compacted soils by using GMDH-type neural network,” *Eur. Phys. J. Plus*, vol. 134, no. 7, Jul. 2019, doi: 10.1140/epjp/i2019-12692-0.

- M. Suthar and P. Aggarwal, "Predicting CBR Value of Stabilized Pond Ash with Lime and Lime Sludge Using ANN and MR Models," *Int. J. Geosynth. Gr. Eng.*, vol. 4, no. 1, p. 6, Mar. 2018, doi: 10.1007/s40891-017-0125-3.
- F. J. Valderrama Purizaca, D. A. Chávez Barturen, S. P. Muñoz Pérez, V. A. Tuesta-Monteza, and H. I. Mejía-Cabrera, "Importance of artificial neural networks in civil engineering: a systematic review of the literature," *ITECKNE*, vol. 18, no. 1, 2020, doi: 10.15332/iteckne.v18i1.2542.
- V.Q. Tran, Machine learning approach for investigating chloride diffusion coefficient of concrete containing supplementary cementitious materials, *Constr. Build. Mater.* 328 (2022), 127103.
- J. Liang, Y. Bu, K. Tan, J. Pan, Z. Yi, X. Kong, Z. Fan, Estimation of stellar atmospheric parameters with light gradient boosting machine algorithm and principal component analysis, *The Astronomical Journal* 163 (4) (2022) 153.
- X. Hao, Z. Zhang, Q. Xu, G. Huang, K. Wang, Prediction of f-CaO content in cement clinker: A novel prediction method based on LightGBM and Bayesian optimization, *Chemometrics and Intelligent Laboratory Systems* 220 (2022), 104461.
- C. Zhang, X. Lei, L. Liu, Predicting metabolite-disease associations based on LightGBM model, *Frontiers in Genetics* 12 (2021) 446.
- P.j. Chun, S. Izumi, T. Yamane, Automatic detection method of cracks from concrete surface imagery using two-step light gradient boosting machine, *Computer-Aided Civil and Infrastructure Engineering* 36(1) (2021) 61-72.
- T. Chen, T. He, M. Benesty, V. Khotilovich, Y. Tang, H. Cho, Xgboost: extreme gradient boosting, *R package version 0.4-2* 1 (4) (2015) 1–4.
- J. Liu, S. Zhang, H. Fan, A two-stage hybrid credit risk prediction model based on XGBoost and graph-based deep neural network, *Expert Systems with Applications* 195 (2022), 116624.

- E. Al Daoud, Comparison between XGBoost, LightGBM and CatBoost using a home credit dataset, *Int. J. Computer Inform. Eng.* 13 (1) (2019) 6–10.
- H. Nguyen, X.-N. Bui, H.-B. Bui, D.T. Cuong, Developing an XGBoost model to predict blast-induced peak particle velocity in an open-pit mine: a case study, *Acta Geophysica* 67 (2) (2019) 477–490.
- J. Duan, P.G. Asteris, H. Nguyen, X.-N. Bui, H. Moayedi, A novel artificial intelligence technique to predict compressive strength of recycled aggregate concrete using ICA-XGBoost model, *Engineering with Computers* 37 (4) (2021) 3329–3346.
- H. Imran, M. Ibrahim, S. Al-Shoukry, F. Rustam, I. Ashraf, Latest concrete materials dataset and ensemble prediction model for concrete compressive strength containing RCA and GGBFS materials, *Constr. Build. Mater.* 325 (2022), 126525
- J. Liu, S. Zhang, H. Fan, A two-stage hybrid credit risk prediction model based on XGBoost and graph-based deep neural network, *Expert Systems with Applications* 195 (2022), 116624.
- N.-H. Nguyen, J. Abell'an-García, S. Lee, E. Garcia-Castano, T.P. Vo, Efficient estimating compressive strength of ultra-high performance concrete using XGBoost model, *J. Build. Eng.* (2022), 104302.
- M.N. Amin, M. Iqbal, K. Khan, M.G. Qadir, F.I. Shalabi, A. Jamal, Ensemble treebased approach towards flexural strength prediction of FRP reinforced concrete beams, *Polymers* 14 (7) (2022) 1303.
- A. Bardhan, P. Samui, K. Ghosh, A.H. Gandomi, S. Bhattacharyya, ELM-based adaptive neuro swarm intelligence techniques for predicting the California bearing ratio of soils in soaked conditions, *Applied Soft Computing* 110 (2021), 107595.
- A. Bardhan, C. Gokceoglu, A. Burman, P. Samui, P.G. Asteris, Efficient computational techniques for predicting the California bearing ratio of soil in soaked conditions, *Engineering Geology* 291 (2021), 106239.

- L. Sun, M. Koopialipoor, D.J. Armaghani, R. Tarinejad, M. Tahir, Applying a metaheuristic algorithm to predict and optimize compressive strength of concrete samples, *Engineering with Computers* (2019) 1–13.
- W. Liang, S. Luo, G. Zhao, H. Wu, Predicting hard rock pillar stability using GBDT, XGBoost, and LightGBM algorithms, *Mathematics* 8 (5) (2020) 765.
- H. Zeng, C. Yang, H. Zhang, Z. Wu, J. Zhang, G. Dai, F. Babiloni, W. Kong, A lightGBM-based EEG analysis method for driver mental states classification, *Computational Intelligence and Neuroscience* 2019 (2019).
- H. Kodaz, S. Özs, en, A. Arslan, S. Günes,, Medical application of information gain based artificial immune recognition system (AIRS): Diagnosis of thyroid disease, *Expert Systems with Applications* 36 (2) (2009) 3086–3092.
- T. Chen, C. Guestrin, Xgboost: A scalable tree boosting system, in: *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, pp. 785–794.
- J. Fan, X. Wang, L. Wu, H. Zhou, F. Zhang, X. Yu, X. Lu, Y. Xiang, Comparison of Support Vector Machine and Extreme Gradient Boosting for predicting daily global solar radiation using temperature and precipitation in humid subtropical climates: A case study in China, *Energy Convers. Manag.* 164 (2018) 102–111.
- I. Ullah K. Liu T. Yamamoto R.E. Al Mamlook A. Jamal A comparative performance of machine learning algorithm to predict electric vehicles energy consumption: A path towards sustainability *Energy & Environment* 2021 0958305X211044998.
- M.I. Khan, M.H. Sutanto, K. Khan, M. Iqbal, M.B. Napiah, S.E. Zoorob, J.J. Klemes, A. Bokhari, W. Rafiq, Effective use of recycled waste PET in cementitious grouts for
- M. Iqbal, Q. Zhao, D. Zhang, F.E. Jalal, A. Jamal, Evaluation of tensile strength degradation of GFRP rebars in harsh alkaline conditions using non-linear genetic-based models, *Materials and Structures* 54 (5) (2021) 1–20.

- M. Iqbal, D. Zhang, F.E. Jalal, M., Faisal Javed, Computational AI prediction models for residual tensile strength of GFRP bars aged in the alkaline concrete environment, *Ocean Engineering* 232 (2021), 109134. [70] M.I. Khan, M.H. Sutanto, K. Khan, M. Iqbal, M.B. Napiiah
- F.E. Jalal, Y. Xu, M. Iqbal, B. Jamhiri, M.F. Javed, Predicting the compaction characteristics of expansive soils using two genetic programming-based algorithms, *Transportation Geotechnics* 30 (2021), 100608.
- Kodaz, H., Özşen, S., Arslan, A., & Güneş, S. (2009). Medical application of information gain based artificial immune recognition system (AIRS): Diagnosis of thyroid disease. *Expert Systems with Applications*, 36(2), 3086-3092.
- Carrillo, A. (2016). Comportamiento del suelo tropical peruano. Universidad Ricardo Palma. Lima-Perú. Pág, 1-35.

Financiamiento de la investigación

Con recursos propios.

Declaración de intereses

Declaro no tener ningún conflicto de intereses, que puedan haber influido en los resultados obtenidos o las interpretaciones propuestas.

Declaración de consentimiento informado

El estudio se realizó respetando el Código de ética y buenas prácticas editoriales de publicación.

Derechos de uso

Copyright© 2025 por **Jaime Yelsin Rosales Malpartida**



[Este texto está protegido por la Licencia Creative Commons Atribución 4.0 Internacional.](https://creativecommons.org/licenses/by/4.0/)

Usted es libre para compartir, copiar y redistribuir el material en cualquier medio o formato y adaptar el documento, remezclar, transformar y crear a partir del material para cualquier propósito, incluso comercialmente, siempre que cumpla la condición de atribución: usted debe reconocer el crédito de una obra de manera adecuada, proporcionar un enlace a la licencia, e indicar si se han realizado cambios. Puede hacerlo en cualquier forma razonable, pero no de forma tal que sugiera que tiene el apoyo del licenciante o lo recibe por el uso que hace.