



## Importancia de la inteligencia artificial en la evaluación de la biomasa forestal en el Perú

### Importance of artificial intelligence on forest biomass evaluation in Perú

Álvaro Ormachea<sup>1,\*</sup>

---

**Recibido:** 13 noviembre 2023 | **Aceptado:** 18 junio 2024 | **Publicado en línea:** 25 julio 2024

**Citación:** Ormachea, A. 2024. Importancia de la inteligencia artificial en la evaluación de la biomasa forestal en el Perú. Revista Forestal del Perú 39(1): 11-21. DOI: <https://doi.org/10.21704/rfp.v39i1.2065>

---

#### Resumen

En este artículo se explican conceptos básicos de la Inteligencia Artificial (IA) ligados a la estimación de la biomasa forestal, y se revisan tres investigaciones realizadas en zonas tropicales que emplean modelos de IA conjuntamente con las tradicionales ecuaciones alométricas. Los resultados muestran que los métodos basados en IA tienen mayor precisión y capacidad para relacionar variables clave en el desarrollo de la biomasa forestal que las ecuaciones alométricas. Estos hechos resaltan la necesidad de que los ingenieros forestales peruanos deban desarrollar capacidades para el uso de la IA en la estimación de la biomasa forestal en el país. Este desarrollo de capacidades implicaría un currículo más exigente en matemática, estadística y ciencias informáticas para los ingenieros forestales; así como la instalación de una extensa red de parcelas permanentes para la creación de una sólida base de datos de las variables que intervienen en el desarrollo de la biomasa forestal. Se propone iniciar la discusión al respecto entre la comunidad forestal peruana con la finalidad de no perder oportunidades en el mercado de bonos de carbono, que según los casos revisados, requerirá mediciones más precisas que las actuales realizadas mediante el uso de las ecuaciones alométricas.

**Palabras clave:** inteligencia artificial, machine learning, ecuación alométrica, biomasa forestal

---

<sup>1</sup> Investigador independiente.

\* Autor de Correspondencia: [manuelormachea@yahoo.es](mailto:manuelormachea@yahoo.es)

### Abstract

This article explains basic concepts of Artificial Intelligence (AI) linked to forest biomass estimation, and reviews three research studies carried out in tropical areas using AI models with traditional allometric equations. The results show that AI-based methods have greater accuracy and ability to relate key variables in forest biomass development than allometric equations. These facts highlight the need for Peruvian forest engineers to develop capabilities for the use of AI in the estimation of forest biomass in the country. This capacity building would imply a more demanding curriculum in mathematics, statistics and computer science for forest engineers; as well as the installation of an extensive network of permanent plots for the creation of a solid database of the variables involved in the development of forest biomass. It is proposed to start a discussion on the subject among the Peruvian forestry community in order not to lose opportunities in the carbon credits market, which, according to the cases reviewed, will require more precise measurements than the current ones carried out through the use of allometric equations.

**Key words:** artificial intelligence, machine learning, allometric equation, forest biomass

### Introducción

A inicios del 2023 se reportó en medios periodísticos una noticia que daba cuenta de un estudio realizado por la Universidad de Cambridge afirmando que la mayoría de los programas de compensación de carbono sobreestiman significativamente los niveles de deforestación que pretenden prevenir o evitar (Greenfield, 2023). El estudio fue realizado por un equipo de científicos de la Universidad de Cambridge y fue publicado en la revista científica *Science*, donde se revisó 26 proyectos de Reducción de Emisiones Derivadas de la Deforestación (REDD) en cinco países tropicales, entre ellos el Perú. En base a este estudio, Greenfield (2023) explica que una investigación periodística realizada entre *The Guardian*, *Die Zeit* y *SourceMaterial* revisó tres estudios científicos de proyectos REDD (uno de ellos el de la U. de Cambridge), los cuales obtuvieron créditos de carbono otorgados por Verra<sup>1</sup> a través de los Verified Carbon Standard (VCS). En este reporte periodístico se señala que solo un puñado de los proyectos mostraron evidencias de reducción en la deforestación, indicando que el 94% de los créditos de carbono otorgados no tuvieron impacto en el clima. Además, según el estudio de la Universidad de Cambridge, la amenaza a los bosques habría sido sobrestimada en promedio unas cuatro veces.

La nota periodística mencionada afirma que un alto funcionario de Verra responsable de temas legales, política y mercados respondió argumentando que sus proyectos enfrentan condiciones locales únicas que los estándares no pueden medir y que trabajan con expertos para continuamente actualizar sus metodologías y asegurar un consenso científico, y que es absolutamente incorrecto decir que el 90% de los certificados REDD de Verra no tienen valor por interpolar un pequeño número de proyectos. Luego reporta declaraciones de David Coomes, uno de los autores del estudio de la Universidad de Cambridge y que trabajó en los inicios de la certificación de bonos de carbono, quien afirma que los grandes montos de dinero entregados por los proyectos de carbón fueron generados debido a altos estándares y metodologías que continuamente son reforzadas en cooperación con gobiernos, científicos y comunidades. En ese mismo sentido, Greenfield (2023) reporta que Julia Jones, otra coautora, sostiene que el sistema de medición de reducción de emisiones debe ser urgentemente corregido para que el mercado de carbón prospere, y que, si no se aprende de los errores, habrá un gran riesgo de que los inversionistas estén dispuestos a pagar para evitar la deforestación en los trópicos. Entonces entre los actores del mercado de carbono existe una preocupación por mejorar la forma en que se mide la biomasa.

<sup>1</sup>Verra es una organización líder en el mundo que gestiona los mercados voluntarios de carbono.

sa forestal. Entre estos actores destacan quienes elaboran los estándares para la certificación de carbono y los científicos que están procurando siempre desarrollar métodos más exactos y precisos para medir su contenido en la biomasa forestal. Quienes deben de estar atentos a su trabajo son los ingenieros forestales dedicados a proyectos de secuestro de carbono, los cuales ya se desarrollan en el Perú.

Esta polémica sobre la medición de la biomasa forestal ha motivado que el presente estudio revise los métodos basados en la Inteligencia Artificial (IA) que se usan en la comunidad de profesionales forestales del mundo, además de analizar cómo está al respecto en el Perú. Tomando en cuenta que en los últimos años la IA ha aparecido entre la comunidad científica como una nueva y potente forma de procesar la información y generar conocimiento que favorezca al bienestar de la humanidad, este artículo pretende llamar la atención de la comunidad forestal peruana sobre la necesidad de comenzar a pensar en la aplicación de la IA en la medición de la biomasa forestal de los bosques peruanos. En ese sentido, el artículo tiene como objetivo exponer los conceptos básicos y algunos métodos de la IA que se están utilizando en la medición y estimación de la biomasa forestal en otros países de clima tropical, así como discutir las condiciones para que la IA sea empleada en los proyectos que requieran la medición exacta y estimación precisa de la biomasa forestal. Greenfield (2023) considera que es pertinente comenzar a discutir este tema, pues las ideas y acciones que puedan surgir permitirán que los proyectos REDD en el Perú, y otros que se fundamentan en la medición exacta y precisa de la biomasa, puedan tener la seriedad que los mercados exigen.

### Principios de la Inteligencia Artificial (IA)

En los últimos tiempos se ha comenzado a hablar mucho sobre la IA. Esta ha surgido repentinamente como la nueva y gran innovación tecnológica que, según múltiples especialistas

e instituciones de desarrollo, va a generar importantes cambios en prácticamente todas las actividades humanas. Para entenderla y tener una idea de su funcionamiento y posibles alcances, es necesario revisar brevemente de qué se trata. Theobald (2017) explica que dentro de lo que se entiende como ciencia computacional se encuentra lo que los informáticos llaman ciencia de datos (*data science*), que son los métodos y sistemas para obtener conocimiento y perspectivas de los datos mediante las computadoras y, que entre una de estas *data science* se halla la inteligencia artificial. Además, sostiene que la inteligencia artificial comprende la habilidad que tienen las computadoras de ejecutar tareas inteligentes o simular habilidades cognitivas. Rouahiainen (2018) define a la IA como la capacidad de las computadoras para analizar grandes volúmenes de información, usar algoritmos, aprender de los datos y utilizar lo aprendido en la toma de decisiones tal y como lo haría un ser humano, pero con una proporción de error significativamente menor. Asimismo, comenta que las capacidades de aprender y tomar decisiones son particularmente importantes, ya que estos procesos están creciendo exponencialmente con el tiempo.

Para poder explicar cómo es que las computadoras pueden ejecutar tareas inteligentes o simular habilidades cognitivas es necesario que se vea lo que es el *Machine Learning* (ML), el *Data Mining* (DM) y el proceso mediante el cual las computadoras llegan a “aprender”. Hurwitz y Kirsch (2018) afirman que ML es la forma en que la IA permite que un sistema logre aprender a partir de los datos más que a partir de la programación, lo cual no es un proceso simple. Para tener idea de cómo funciona este proceso es necesario revisar el concepto de DM. Theobald (2017) afirma que DM es una disciplina hermana del ML que se enfoca en descubrir diferentes patrones en las grandes bases de datos. Estas disciplinas se relacionan ya que ambas hacen uso de algoritmos, análisis de asociación y análisis de regresión para analizar los datos. Pero se diferencian en que el ML se enfoca en el proceso de auto-aprendizaje y modelamiento de datos para generar predic-

ciones para el futuro, mientras que el DM se concentra en la limpieza de grandes bases de datos para obtener lecciones del pasado. Theobald (2017) explica que, para el proceso de aprendizaje de un modelo, los datos se reparten en dos grupos: el primer grupo se usa para entrenar y desarrollar el modelo con la adecuada precisión, luego se puede probar el modelo con el grupo de datos remanente. Hurwitz y Kirsch (2018) afirman que el proceso de aprendizaje del ML usa una variedad de algoritmos que en forma iterativa aprenden a partir de los datos con la finalidad de mejorar, describir y predecir resultados.

### Biomasa Forestal e Inteligencia Artificial

Con la finalidad de conocer la forma en que se aplica la IA en el cálculo de la biomasa en los bosques tropicales se realizó una búsqueda electrónica de artículos científicos enfocada en este tema. Estos artículos fueron revisados desde la óptica de un ingeniero forestal formado en la escuela de la dasometría clásica que se enseña en las facultades de ciencias forestales peruanas, y con el apoyo de un especialista en informática capacitado por la University of London (Inglaterra). La revisión de estos artículos ha permitido conocer, o al menos tener una idea acerca de algunos métodos de IA que se suelen emplear para medir la biomasa en bosques tropicales y plantaciones comerciales. A partir de ello, se puede identificar qué capacidades o tareas se tendría que comenzar a desarrollar y realizar respectivamente para poder aplicar la IA en la medición de la biomasa de nuestros bosques. Uno de los artículos revisados fue realizado por la Universidad Federal de Paraná, Brasil (Montaño *et al.* 2017) enfocado en bosques naturales; el segundo fue escrito por otro equipo de científicos de Tailandia (Wongchai *et al.* 2022) enfocado en plantaciones de tres especies de rápido crecimiento; y el tercero, fue elaborado por un equipo de investigadores chinos (Liu *et al.* 2023) que usaron métodos de IA para la predicción a futuro de la biomasa forestal en una región determinada.

### Modelos de IA y medición de biomasa en bosques tropicales

La investigación de Montaño *et al.* (2017) comparó tres modelos de IA con el tradicional método de ecuaciones alométricas que se emplea para calcular la biomasa de los bosques en base a la regresión estadística. Para ello evaluaron datos del diámetro a la altura del pecho (DAP), altura total (Ht), biomasa seca aérea (b) y densidad de la madera ( $\rho$ ) de 4004 árboles ubicados en Australia, Sudáfrica, Brasil, Camboya, Camerún, República Centroafricana, Colombia, Costa Rica, Guayana Francesa, Gabón, Ghana, Guadalupe, India, Indonesia, México, Madagascar, Malasia, Mozambique, Nueva Guinea, Perú, Puerto Rico, Tanzania, Venezuela y Zambia. El modelo alométrico para estimar la biomasa que emplearon fue el de Schumacher-Hall, expresado en la siguiente fórmula:

$$\lg(b) = \beta_0 + \beta_1 \lg(\text{DAP}) + \beta_2 \lg(\text{Ht}) + \beta_3 \lg(\rho)$$

Dónde: b es la biomasa forestal, DAP es el diámetro a la altura del pecho, Ht es la altura total,  $\rho$  es la densidad básica de la madera y  $\beta_n$  son coeficientes asignados para las tres variables.

Los tres modelos de IA evaluados corresponden a *K-Nearest Neighbors* (KNN), *Artificial Neural Networks* (ANN) y *Support Vector Machine* (SVM). Una breve descripción de los modelos de IA evaluados se muestra en el Cuadro 1. Montaño *et al.* (2017) reportan que sometieron los datos de las variables consideradas en la ecuación alométrica evaluada a la prueba de Lilliefors y Shapiro-Wilk. Los resultados de esta prueba mostraron que, tanto en su forma original como en la transformación logarítmica de los datos, ninguna de las variables se ajustaba a la distribución normal. Además, al realizar el análisis gráfico del error residual encontraron que estos no presentaban homocedasticidad. A partir de estos dos resultados, los autores señalaron que no se cumplen las condiciones para realizar el ajuste del modelo alométrico, y que este método no sería muy adecuado para evaluar la biomasa. Así mismo, señalan en la discusión de los datos que la bio-

Grupo	Descripción
KNN	Es un modelo no paramétrico, usado en DM, que se basa en los valores de los $n$ vecinos más cercanos para realizar la estimación. Para la estimación de la biomasa Montaña <i>et al.</i> (2017) emplearon la ecuación de distancia Euclidiana con las mismas variables del modelo alométrico mencionado de los tres árboles más cercanos. Es decir, un valor de $n = 3$ .
ANN	Se trata de un sistema de ML en el que se conmutan una serie de funciones matemáticas simulando la sinapsis biológica de células dispuestas en estratos. Las células interconectadas son funciones matemáticas lineales o sigmoideas que procesan datos y cuyos resultados alimentan a la siguiente célula, o función matemática. A cada estrato se le asigna un coeficiente, que se van ajustando a medida que se van probando y logrando la estimación estadísticamente más exacta. Un proceso algo análogo al de calibración de un modelo alométrico.
SVM	Es una técnica, de ML, en la que la data es procesada para seleccionar los datos que presentan una mayor probabilidad de tener un patrón lineal, para lo cual se emplea la función Kernel. Esta función otorga a cada grupo de datos seleccionado los parámetros $C$ y $\gamma$ . El parámetro $C$ determina la relación entre la exactitud y la complejidad del modelo generado por cada grupo de datos, mientras que $\gamma$ mide el sesgo y la variancia de ellos. Montaña <i>et al.</i> (2017) evaluaron 150 combinaciones de estos parámetros para encontrar la combinación con mejor correlación.

**Cuadro 1.** Descripción de los modelos KNN, ANN y SVM. Fuente: Montaña *et al.* (2017).

masa correspondiente a los árboles grandes (DAP > 70 cm) muestra una mayor dispersión y que esta variable no se distribuye como una curva normal.

Luego, cuando se realiza el análisis estadístico sobre el sesgo, exactitud y precisión con los tres modelos de IA y el de regresión, los resultados mostraron que los modelos KNN, ANN y SVM tienen un alto poder de estimación de la biomasa de los árboles tropicales, comparable o mejor que el del modelo alométrico, en particular el modelo SVM mostró mejor precisión para los datos considerados. Concluyeron que los modelos de IA pueden ser considerados como una interesante alternativa frente a la técnica de regresión, especialmente cuando los datos no se distribuyen en forma normal y no existe homocedasticidad; que es el caso de la biomasa de los árboles de los bosques tropicales. Esta última parte de las conclusiones de estos autores es muy importante pues durante todos estos años en el Perú se ha estado midiendo la biomasa forestal empleando mayormente ecuaciones alométricas que, según al menos el trabajo de Montaña *et al.* (2017), indica que no son tan confiables como se pensaba. Considerando que los modelos de IA estarían presentando mayor fiabilidad estadística que las ecuaciones alométricas, es válido y necesario considerar hacer los esfuerzos necesarios para

poder aplicarlos en la medición de nuestros bosques; más aun tomando en cuenta la alta heterogeneidad de especies, las dimensiones y formas de los árboles en sus diferentes edades y la condición sociológica dentro de los ecosistemas en que se desarrollan.

### Modelos de IA en la medición de la biomasa en plantaciones forestales

Este estudio evalúa siete modelos ML para estimar la biomasa de tres especies de rápido crecimiento cultivadas a escala industrial como alternativa a las regresiones alométricas empleando variables dasométricas simples, como el DAP y la altura total o comercial de los árboles. Wongchai *et al.* (2022) afirman que la estimación de la biomasa de las plantaciones industriales de manera rápida y precisa se ha convertido en una necesidad, además resaltan que el uso de modelos ML aplicados en plantaciones de rápido crecimiento es relativamente nuevo, pues los pocos trabajos reportados se enfocan en pastos y árboles para climas relativamente fríos, ninguno para plantas de climas tropicales, donde los modelos de crecimiento de la biomasa solamente la estiman a través de ecuaciones alométricas de menor precisión. En ese sentido, su trabajo tuvo como objetivo usar modelos de ML para estimar la biomasa de plantaciones en condiciones tro-

picales de tres especies de *Eucalyptus camaldulensis*, *Acacia hybrid* y *Leucaena leucocephala* a partir de variables dasométricas simples y probar su precisión con las ecuaciones alométricas tradicionales.

Los modelos de ML probados fueron: *Decision Tree* (DT), *Random Forest* (RF), *Gradient Tree Boosting* (GTB), *Adaptive Boosting* (AdB), *Kernel Ridge Regression* (KRR), *Support Vector Machine* (VSM) y, *K-Nearest Neighbor* (KNN). La descripción realizada por Wongchai *et al.* (2022) de los primeros cinco modelos son difíciles de entender para quien no está familiarizado con el léxico técnico de la programación, los métodos de manejo de datos que emplean y las complejas matemáticas con las que procesan los datos. Por lo tanto, al autor se le hace complicado presentar una descripción de ellos que oriente a los lectores de este artículo. Los fundamentos de los dos últimos modelos son explicados en el Cuadro 1. En el Cuadro 2 se presentan las siete ecuaciones alométricas evaluadas por los autores, en las que las variables utilizadas son el diámetro a la altura del pecho (DAP) y altura total (H).

Con la finalidad de evaluar cuál de los modelos ML o la ecuación alométrica ofrece la predicción más precisa de la biomasa, los autores emplearon un método estadístico usado generalmente para estimar el desempeño de los modelos. Este método consistente en separar al azar la gran cantidad de datos y dividirlo en “n”

grupos, los que deben de contar con una cantidad de datos de tal manera que representen la amplitud de valores de todos los datos. Se separa uno de los grupos y con los datos de los n-1 grupos restantes, se realiza la calibración y ajuste de la precisión del modelo; se determinan los valores de los coeficientes del modelo (a, b y c) y los hiperparámetros óptimos de los modelos ML. Este proceso es lo que se conoce como aprendizaje del modelo (explicado brevemente líneas arriba). Luego, el grupo restante de datos se emplea para verificar la precisión de los modelos con las constantes e hiperparámetros calibrados. Wongchai *et al.* (2022) reportan que usaron el 90% de los datos para el proceso de aprendizaje del modelo y el 10% restante para la validación. Los valores de los coeficientes y las potencias que afectan las variables de las ecuaciones alométricas fueron determinados por el método del mínimo cuadrado, determinando que la segunda ecuación alométrica del Cuadro 2 es la que presentó el mayor coeficiente de regresión ( $R^2$ ), y cuyos valores de a y b calculados fueron 0.0938 y 2.4308 respectivamente. En el Cuadro 3 se muestran los valores de  $R^2$  y el cuadrado medio del error que obtuvieron para la ecuación alométrica seleccionada y para cada uno de los modelos ML evaluados. Se aprecia que el valor de  $R^2$  para la ecuación alométrica es de 0.9276 y el cuadrado medio del error es 4.2784, mientras que los valores correspondientes para los modelos ML tienen un valor de  $R^2$  superior a 0.9417 y que los cuadrados medio del error son bajos, no superan el valor de 3.5874. Tomando en cuenta estos resultados, Wongchai *et al.* (2022) afirman que los modelos ML podrían ofrecer mejor precisión que las ecuaciones alométricas debido al hecho que estos pueden analizar las interacciones entre los valores del DAP, la altura e inclusive la edad que contienen la inmensa cantidad de datos correspondiente a las plantaciones de las tres especies que evaluaron.

Wongchai *et al.* (2022) determinaron que el estimador de biomasa para estas tres especies que mejor funcionó fue el modelo RF, debido a que logró una precisión similar a los

Número	Ecuación Alométrica
1	Biomasa = a + b(DAP)
2	Biomasa = a(DAP + 1) <sup>b</sup>
3	Biomasa = a(H) <sup>b</sup>
4	Biomasa = a + b(H) + c(H) <sup>2</sup>
5	Biomasa = a(DAP x H) <sup>b</sup>
6	Biomasa = a(DAP <sup>2</sup> x H) <sup>b</sup>
7	Biomasa = a(DAP <sup>b</sup> x H) <sup>c</sup>

**Cuadro 2.** Ecuaciones alométricas evaluadas por Wongchai *et al.* (2022). Donde: DAP es el diámetro a la altura del pecho; H es la altura total o comercial de los árboles; a, b y c son coeficientes a los que se les asigna un valor numérico.

Ecuación/Modelo ML	R <sup>2</sup>	CME
Biom = 0.0938(DAP + 1) <sup>2.4308</sup>	0.9276	4.2784
DT	0.9417	3.5874
RF	0.9572	3.0923
GTB	0.9540	3.1997
AdB	0.9472	3.4687
KRR	0.9569	3.1291
SVM	0.9567	3.0420
K-NN	0.9546	3.1052

**Cuadro 1.** Valores de R<sup>2</sup> y el cuadrado medio del error para la ecuación alométrica seleccionada y modelos ML.

otros modelos ML y por requerir un ingreso de datos menos costoso. Cuando analizaron el modelo RF junto a la ecuación alométrica encontraron que ambos lograban una precisión similar para estimar una biomasa menor de 30 kg, pero por encima de este valor, el modelo RF presentó mejores resultados que los de la ecuación alométrica. Al analizar la variable edad en esta comparación observaron que la ecuación alométrica tiene una alta precisión para estimar la biomasa de los árboles de hasta los cuatro años en estas tres especies, y sugieren que debería desarrollarse otra ecuación alométrica para estimar la biomasa de árboles mayores a cuatro años de estas tres especies. En ese sentido, comentaron que el modelo RF puede tener la capacidad de reconocer los diferentes patrones de edad, y que el algoritmo del modelo RF puede ofrecer una predicción de la biomasa con mayor precisión.

Estos autores concluyen que el modelo RF estima exitosamente la biomasa de estas tres especies plantadas en el norte de Tailandia, empleando variables simples como el DAP, altura total y edad de los árboles; resaltando que la calibración de los parámetros del modelo fue crucial para lograrlo. Fundamentan esta conclusión afirmando que, si bien el DAP es la variable más influyente en la estimación de la biomasa, la inclusión de las variables altura total y edad puede mejorar la precisión pues existen interacciones entre ellas, y que la precisión de pierde significativamente si se emplean es-

tas variables en forma individual. Finalmente, recomiendan que, para mejorar el modelo, las futuras investigaciones deberían incluir otras características de los árboles, así como otras variables de tipo ambiental como la humedad, temperatura, precipitación, tipo de bosque, densidad de rodal y competencia dentro de las plantaciones.

### Estimación a futuro de la biomasa a escala regional

El trabajo realizado por Liu *et al.* (2023), es uno de los pocos que se han realizado en la predicción de la biomasa de una región. Para ello, se utilizaron dos modelos de ML denominados BP-ANN y SVM, mediante los cuales se espera mejorar el entendimiento de la productividad de los bosques y el almacenamiento de carbón en los ecosistemas forestales; así como evaluar la respuesta y adaptabilidad de los bosques al cambio climático en Beijing (China) considerando factores forestales y meteorológicos.

Los datos forestales fueron obtenidos de tres períodos de inventarios forestales en Beijing, que incluye información de 214 parcelas permanentes y 428 parcelas de registro de datos. Las variables de las parcelas permanentes consideradas fueron el número de parcela, la pendiente, coordenadas GPS, cobertura vegetal, cobertura arbustiva, cobertura herbácea, edad promedio, densidad arbórea, densidad de copas, altura promedio de los árboles, DAP promedio, especies arbóreas dominantes, área basal y volumen por hectárea. Los datos meteorológicos fueron obtenidos de la *China Meteorological Science Data Sharing Service Network*, de donde se tomó información de 21 estaciones meteorológicas de Beijing desde el año 2004 hasta el 2018. Las variables meteorológicas consideradas fueron la temperatura promedio anual y precipitación media anual.

La biomasa a futuro en los puntos donde están las parcelas permanentes fue proyectada modelando el crecimiento forestal, para lo cual se tuvo que calcular la biomasa básica que estas contienen. La biomasa básica está compuesta por la biomasa de los árboles, arbustos y hierbas. La biomasa de los árboles fue estimada

a través de una ecuación de ajuste lineal que toma en cuenta la biomasa y el volumen de almacenaje, por tipo de bosque y edad propuesta por Xu *et al.* (2007), mientras que la biomasa de los arbustos y hierbas fue estimada por el método establecido por Wang *et al.* (2014). La data meteorológica a futuro fue obtenida utilizando el software Climate AP, que puede predecir datos climáticos a futuro basándose en el Quinto Reporte de Evaluación de Panel Internacional de Expertos en Cambio Climático (IPCC).

Liu *et al.* (2023) reportan que el modelo ML denominado BP-ANN (*Back-Propagation Artificial Neural Network*) tiene muchas ventajas para la predicción de la biomasa con mayor precisión que los modelos tradicionales de regresión lineal. Además, afirman que en el proceso de entrenamiento vía DM puede inferir datos perdidos o filtrar aquellos que identifica como extraños y a través del entrenamiento del modelo, este puede aprender las complejas relaciones no lineales entre las características de la biomasa. Por otro lado, afirman que el modelo SVM tiene como ventaja el poder manejar relaciones no lineales y grandes dimensiones de datos mediante la selección de ecuaciones Kernel para analizar y seleccionar los datos, y que esta selección le permite capturar con mayor precisión los factores clave que afectan a la biomasa. Otra ventaja identificada en este modelo es que puede conducir un entrenamiento iterativo y ajuste de parámetros que le permite mejorar su precisión. Como desventaja, señalan que este modelo puede tomar más tiempo y requerir mayores recursos computacionales.

Al evaluar la precisión de los modelos, Liu *et al.* (2023) encontraron que el modelo SVM tuvo un valor de  $R^2$  igual a 0.9 y el del modelo BP-ANN fue de 0.85. Además, concluyen que ambos modelos tienen una buena precisión de ajuste y que no hay una diferencia significativa en los resultados de predicción entre ellos, aunque el modelo BP-ANN puede presentar predicciones más volátiles. Sin embargo, comentan que para mejorar el modelamiento se debería ampliar el nivel de predicción, pasando

de los puntos fijos correspondientes a las parcelas permanentes a un nivel que cubra toda la región y, por otro lado, considerar la significancia de las actividades humanas como los cambios en la cobertura y patrones de paisaje. Con estas mejoras en las investigaciones consideran que se puede proveer a quienes manejan los recursos forestales la base científica que promueva el desarrollo forestal sostenible en Beijing y ayudar a establecer ecosistemas estables y saludables.

### **Condiciones para la Aplicación de la IA en la Medición de la Biomasa Forestal en el Perú**

De la información expuesta en los tres casos revisados se rescatan dos puntos importantes. El primero es que las ecuaciones alométricas que usualmente se usan para estimar la biomasa de los árboles en los bosques tropicales no estarían siendo muy precisas y confiables, y que la alternativa de los modelos de IA si lo estarían siendo, además de ser capaces de relacionar más variables de los árboles que lo que las ecuaciones alométricas permiten. Lo segundo es que, si es necesario aplicar los modelos de IA en el Perú, se necesitaría tres elementos que tal vez no se dispone plenamente. El primer elemento es el recurso humano capacitado en el desarrollo de algoritmos que se elaboren considerando las funciones matemáticas y estadísticas necesarias que interrelacionan las diversas variables que explican el crecimiento de la biomasa forestal en los complejos bosques tropicales. Para el diseño de los algoritmos es necesario además que estas personas dominen los lenguajes de programación y que puedan operar computadoras de gran potencia que sean capaces de procesar la gran cantidad de datos que se demandan los modelos de IA. El contar con este tipo de computadoras sería un segundo elemento necesario. Por último, y tal vez el más complicado de obtener sea el contar con las grandes bases de datos de las variables dasométricas y ambientales que combinadas entre ellas explican el desarrollo de la biomasa forestal.



Respecto a contar con recursos humanos capacitados cabe mencionar que la formación en matemática y estadística de los ingenieros forestales en el Perú no es muy exigente. Tradicionalmente esta alcanza hasta el análisis matemático de funciones cuadráticas, logarítmicas de ecuaciones de dos a tres variables, y del análisis estadístico de la distribución normal y chi cuadrado ( $X^2$ ) del DAP y altura de los árboles, así como el diseño y análisis de muestreo estadístico para realizar inventarios forestales de un conjunto de datos que puede procesarse con programas de hojas de cálculo en computadoras de escritorio, o inclusive laptops. Se ha visto que los algoritmos que se emplean en los modelos de DM y el ML se basan en funciones matemáticas más complejas que no son estudiadas en los cursos de ciencias básicas y de medición forestal con los que se forma un ingeniero forestal en el Perú, pero que en manos de un profesional de la *data science* que las conoce y maneja, pueden ser utilizadas para representar las múltiples relaciones de las variables dasométricas, físicas de la madera y hasta ambientales que juegan un rol en la formación de la biomasa forestal.

Esta situación conduce a pensar que, si el país necesita contar con ingenieros forestales capaces de aplicar modelos de IA, ML y DM, se debe comenzar a considerar una formación matemática y estadística más exigente. Y esto no solo se aplicaría para los ingenieros forestales que se orienten a la medición del recurso forestal, sino también a los que se enfoquen en el procesamiento industrial del mismo. Por ejemplo, sería ideal que un ingeniero forestal pueda desarrollar un algoritmo que determine la opción más eficiente de cortes de aserrío de las trozas de especies tropicales, que como se sabe, al tener diversas formas de sección transversal, ahusamiento y densidad de madera, se desperdicia madera al aserrarla; más aún, tomando en cuenta lo costoso que representa trasladar esa troza desde el bosque hasta el carro de aserrío. Esto supone que además este ingeniero forestal debería manejar lenguajes de programación para elaborar los algoritmos necesarios para ello. Tal vez por ahora no existan los medios para formar ingenieros

forestales con esas capacidades, así que tendría que recurrirse a formar equipos con ingenieros de sistemas, o ingenieros en IA que ya empiezan a formarse en la Universidad Nacional de Ingeniería (UNI) desde el año 2024 (Saavedra, 2023). Si bien estos ingenieros de las ciencias informáticas deben ser muy capaces en matemáticas, estadística, programación y otras capacidades técnicas, no se puede decir lo mismo sobre sus conocimientos de las ciencias ambientales que les permitirían entender cómo funciona un ecosistema forestal natural o creado por el hombre. Por ello, mientras se trabajaría en equipo con ingenieros informáticos, se debería mejorar la formación en matemática y estadística de los ingenieros forestales como preparación para realizar estudios superiores o complementarios en IA. Este tema queda en manos de las facultades de ciencias o ingeniería forestal del Perú.

Respecto a contar con computadoras de gran potencia, capaces de procesar en forma rápida y sin mucho costo la gran cantidad de datos que demandan los modelos de IA, no se percibe como gran problema que un presupuesto bien elaborado y justificado no pueda resolver. Este tema quedaría en manos de las instituciones encargadas de realizar investigación. Por último, un tema tan o más importante que el contar con profesionales capacitados en IA es el contar con la suficiente cantidad de datos que permitan elaborar y desarrollar los modelos de IA para estimar con precisión y confianza la biomasa forestal del Perú. Posiblemente en algún momento se podrá contar con ingenieros con altas capacidades para la elaboración y gestión de modelos de IA, o las más rápidas y poderosas computadoras capaces de procesar grandes cantidades de datos, pero si no se tienen los datos a procesar, no se tendrá nada.

En los tres casos revisados se ha visto que los datos son la materia prima para producir estimaciones precisas y lo más certeras posibles de la biomasa. Se sabe que la compleja geografía y climas del Perú han generado una cobertura vegetal de lo más diversa, de la que todavía se conoce poco por tener escasos datos sobre

ella. Por lo tanto, sería necesario contar con una red de puntos donde se instalen parcelas permanentes de medición de las más diversas variables físico-ambientales y dasométricas de los ecosistemas forestales, ello nos permitirá contar con suficientes datos para desarrollar modelos de IA para estimar con precisión la biomasa forestal. Esta red de puntos debe de ser lo suficientemente extensa para que cumpla con los criterios de representatividad, no solo biológica, sino sobre todo estadística. La accesibilidad a sitios remotos seguramente hará costosa la instalación, mantenimiento y mediciones periódicas frecuentes en esta red de parcelas. Considerando que el mercado de carbono genera suficientes ingresos económicos para financiar programas de desarrollo de las poblaciones donde se ubican los ecosistemas forestales, cabe pensar que parte de esos ingresos podrían cubrir los costos de operación de esta red de parcelas permanentes. Posiblemente, al comienzo podrían ser cubiertos por alguna institución privada o pública interesada en fomentar el mercado y los beneficios de los bonos de carbono.

La noticia con la que se inició este artículo, sobre que las mediciones de la deforestación en los bosques tropicales que dan pie al pago de los bonos de carbono estarían sobredimensionadas, y que tanto los científicos como quienes establecen los estándares al respecto deberían ponerse de acuerdo en mejorar los métodos de medir la biomasa forestal, conduce a pensar que los compradores de estos bonos podrían comenzar a perder interés en los bonos de carbono originados en los bosques tropicales si ese tema no se aclara. A partir de las investigaciones expuestas, queda claro que la medición de la biomasa forestal tiende al uso de modelos de IA, que ofrecen resultados más precisos y confiables que los obtenidos a través de las ecuaciones alométricas. Por ello, se concluye que, si el Perú quiere seguir participando en el mercado de carbono almacenado en su inmensa biomasa y patrimonio forestal, debería comenzar a pensar en formar ingenieros forestales capaces de conocer, entender y, si es posible, desarrollar algoritmos de IA para

la medición de la biomasa de sus bosques y plantaciones. Por ello, se espera que este artículo despierte el interés de la comunidad forestal peruana en la IA y se planteen iniciativas para contar pronto con los recursos humanos, equipos y bases de datos necesarias que permitirán, además de continuar participando en el mercado de carbono, contar con información precisa para tomar decisiones sobre el uso de la tierra o de la cobertura vegetal del Perú.

### **Agradecimientos**

Quiero agradecer a mi buen amigo Íbico Rojas, quien en base a sus conocimientos en informática e interés por la forestería me animó a escribir este artículo y por haberme orientado a entender los principios de la IA, su importancia y proyección en el mundo moderno, y además por haberme facilitado la información sobre los tres casos revisados.

### **Contribución de los autores**

AO: conceptualización, redacción-revisión y edición.

### **Conflicto de intereses**

Los autores no incurrir en conflictos de intereses.

### **Fuentes de financiamiento**

Esta investigación no recibió ninguna subvención específica de ninguna agencia de financiación, sector gubernamental ni comercial o sin fines de lucro.

### **Aspectos éticos / legales**

Los autores declaran no haber incurrido en aspectos antiéticos ni haber omitido normas legales.

### **ID ORCID**

Álvaro Ormachea

<https://orcid.org/0009-0005-3902-3765>

## Referencias

- Greenfield, P. 2023. Revealed: more than 90% of rainforest carbon offsets by biggest certifier are worthless, analysis shows (en línea). The Guardian, Reino Unido. Consultado 23 ago. 2023. Disponible en <https://www.theguardian.com/environment/2023/jan/18/revealed-for-est-carbon-offsets-biggest-provider-worthless-verra-aoe>.
- Hurwitz, J; Kirsch, D. 2018. Machine Learning for Dummies. (en línea) Nueva Jersey, Estados Unidos. IBM Limited Edition. 75 p. Visitado el 3 de noviembre 2023. Disponible en <https://www.ibm.com/downloads/cas/GB8ZMQZ3>.
- Liu, J; Yue, C; Pei, C; Li, X; Zhang, Q. 2023. Prediction of Regional Forest Biomass Using Machine Learning: A Case Study of Beijing, China. Forests 14(59):1008. DOI: <https://doi.org/10.3390/f14051008>.
- Montaño, RA; Sanquetta, C R; Wojciechowski, J; Mattar, E.; Dalla, AP. 2017. Artificial Intelligence Models to Estimate Biomass of Tropical Forest. (en línea) POLIBITS 56:29-37. DOI: <https://doi.org/10.17562/PB-56-4>.
- Rouahianen, L. 2018. Inteligencia Artificial, 101 cosas que debes saber hoy sobre nuestro futuro. Editorial Planeta. 22 p. Consultado 22 may. 2024. Disponible en [https://proassetspdl.com/cdnstatics2.com/usuarios/libros\\_contenido/arxius/40/39307\\_Inteligencia\\_artificial.pdf](https://proassetspdl.com/cdnstatics2.com/usuarios/libros_contenido/arxius/40/39307_Inteligencia_artificial.pdf).
- Saavedra, N. 2023. UNI enseñará Ingeniería de Inteligencia Artificial desde 2024: ¿qué se sabe sobre esta nueva carrera? (en línea). La República, Perú, 4 oct. Consultado 04 oct. 2023. Disponible en <https://larepublica.pe/sociedad/2023/10/03/uni-ensenara-ingenieria-de-inteligencia-artificial-desde-2024-que-se-sabe-sobre-esta-nueva-carre-ra.177540>.
- Theobald, O. 2017. Machine Learning for Absolute Beginners. (en línea) Second Edition. 128 p. Consultado 05 nov. 2023. Disponible en <https://bmansoori.ir/book/Machine%20Learning%20For%20Absolute%20Beginners.pdf>.
- Wang, H; Niu, SK; Shao, X; Zhang, C. 2014. Study on biomass estimation methods of understory shrubs and herbs in forest ecosystem. Acta Prataculturae Sinica, 23(3):20-29. DOI: <https://doi.org/10.11686/cyxb20140303>.
- Wongchai, W; Onsree, T; Sukkam, N; Promwungkwa, A; Tippayawong, N. 2022. Machine learning models for estimating above ground biomass of fast-growing trees. Expert Systems with Applications 199:117186. DOI: <https://doi.org/10.1016/j.eswa.2022.117186>.
- Xu, X; Cao, M; Li, K 2007. Temporal-spatial dynamics of carbon storage vegetation in China. Progress in Geography 26(6):1-16. DOI: <https://doi.org/10.11820/dlkxjz.2007.06.001>.