



Uso de tecnologías geoespaciales

para la estimación del secuestro de
carbono de la biomasa herbácea en
terrenos rehabilitados post-minería

Jorge Eliecer Galvis Daza

AREANDINA

Fundación Universitaria del Área Andina

USO DE TECNOLOGÍAS GEOESPACIALES

PARA LA ESTIMACIÓN DEL SECUESTRO DE CARBONO
DE LA BIOMASA HERBÁCEA EN TERRENOS
REHABILITADOS POST-MINERÍA

Jorge Eliecer Galvis Daza

USO DE TECNOLOGÍAS GEOESPACIALES

PARA LA ESTIMACIÓN DEL SECUESTRO DE CARBONO
DE LA BIOMASA HERBÁCEA EN TERRENOS
REHABILITADOS POST-MINERÍA

Jorge Eliecer Galvis Daza

Galvis Daza, Jorge Eliecer

Uso de tecnologías geoespaciales para la estimación del secuestro de carbono de la biomasa herbácea en terrenos rehabilitados post-minería / autor Jorge Eliecer Galvis Daza -- Bogotá:

Fundación Universitaria del Área Andina, 2023

ISBN (Digital) 978-958-5139-94-7

Incluye índice.

1. Maquinaria minera. - 2. Minas de carbón. - 3. Protección del medio ambiente. - 4. Educación superior - Investigaciones.

Catalogación en la publicación Biblioteca Fundación Universitaria del Área Andina (Bogotá)

622.1 scdd22

Uso de tecnologías geoespaciales para la estimación del secuestro de carbono de la biomasa herbácea en terrenos rehabilitados post-minería

© Fundación Universitaria del Área Andina. Bogotá, diciembre de 2023

© Jorge Eliecer Galvis Daza

ISBN (digital): 978-958-5139-94-7

Fundación Universitaria Área Andina

Calle 71 No. 13-21, Bogotá, Colombia

Correo electrónico: publicaciones@areandina.edu.co

Proceso editorial

Dirección editorial: Omar Eduardo Peña Reina

Coordinación editorial: Camilo Andrés Cuéllar Mejía

Diseño de carátula: Proceditor Ltda.

Diagramación: Proceditor Ltda.

Todos los derechos reservados. Queda prohibida la reproducción total o parcial de esta obra y su tratamiento o transmisión por cualquier medio o método sin autorización escrita de la Fundación Universitaria del Área Andina y sus autores.

BANDERA INSTITUCIONAL

Pablo Oliveros Marmolejo †
Gustavo Eastman Vélez
Miembros Fundadores

Carlos Patricio Eastman Barona
Presidente de la Asamblea General y Consejo Superior

José Leonardo Valencia Molano
Rector Nacional y Representante Legal

Martha Patricia Castellanos Saavedra
Vicerrectora Nacional Académica

Karol Milena Pérez Calderón
Vicerrectora Nacional de Crecimiento y Desarrollo

Erika Milena Ramírez Sánchez
Vicerrectora Nacional Administrativa y Financiera

Mauricio Andrés Hernández Anzola
Vicerrector Nacional de Experiencia y Felicidad

Felipe Baena Botero
Rector - Seccional Pereira

Gelca Patricia Gutiérrez Barranco
Rectora - Sede Valledupar

María Angélica Pacheco Chica
Secretaria General

Omar Eduardo Peña Reina
Director Nacional de Investigaciones

Carmen Victoria Meza Carrillo
Decana Facultad de Ingenierías y Ciencias Básicas

Ober Adiel Romero Arias
Director programa de Ingeniería de Minas
Sede Valledupar

Camilo Andrés Cuéllar Mejía
Subdirector Nacional de Publicaciones

13	■	1.	Impactos asociados a la minería de carbón a cielo abierto
14			Minería de carbón en el departamento del Cesar
17	■	2.	Fundamentación para el desarrollo de investigaciones de captura de carbono
19	■	3.	Metodología para estimación de carbono en biomasa herbácea
19			Área de estudio
20			Metodología de modelamiento
23	■	4.	Obtención de contenido de carbono a partir de muestras tomadas en campo
23			Análisis multitemporal
27			Muestreo
29			Análisis de laboratorio
30			Conclusiones
31	■	5.	Extracción de índices de vegetación
31			Índice de vegetación
34			Conclusión
35	■	6.	Modelación de carbono en la biomasa herbácea
35			Estadística descriptiva de los datos
36			Regresión por mínimos cuadrados ordinarios (OLS)
39			K-vecinos más cercanos (K-NN)
42			Máquina vectorial de soporte (SVM)
46			Random Forest (RF)
49			Conclusiones
51	■		Referencias
55	■		Anexos
55			Anexo 1. Algoritmo para regresión por mínimos cuadrados ordinarios
57			Anexo 2. Algoritmo para regresión por K-vecinos más cercanos
60			Anexo 3. Algoritmo para regresión por máquina vectorial de soporte
63			Anexo 4. Algoritmo para regresión por Random Forest

Índice de tablas

- 28 Tabla 1. Georreferenciación y contenido de biomasa por parcela
- 30 Tabla 2. Contenido de carbono en muestras
- 31 Tabla 3. Índice de vegetación
- 32 Tabla 4. Índice de vegetación extraído para cada punto de muestreo
- 37 Tabla 5. Evaluación de modelos OLS
- 44 Tabla 6. Evaluación de modelo SVM con parámetros de cálculo por defecto
- 44 Tabla 7. Parámetros de cálculo para regresión SVM
- 44 Tabla 8. Evaluación de modelo SVM con parámetros optimizados
- 47 Tabla 9. Parámetros de cálculo RF

Índice de figuras

- 15 Figura 1. Títulos de carbón departamento del Cesar
- 15 Figura 2. Producción de carbón en el departamento del Cesar (2016-2022)
- 17 Figura 3. Acumulación de carbono terrestre
- 19 Figura 4. Localización del área de estudio
- 21 Figura 5. Esquema de metodología de trabajo
- 25 Figura 6. Análisis multitemporal
- 26 Figura 7. Áreas en proceso de rehabilitación–combinación 843
- 26 Figura 8. Pasto sembrado
- 27 Figura 9. Puntos de muestreo en campo
- 28 Figura 10. Georreferenciación de punto de muestreo
- 29 Figura 11. Procesamiento de muestras en laboratorio
- 35 Figura 12. Matriz de correlación
- 36 Figura 13. Diagrama de procesos para modelación OLS
- 38 Figura 14. Diagramas Índice de Vegetación vs. Contenido de Carbono
- 39 Figura 15. Modelación de carbono OLS
- 40 Figura 16. Diagrama de procesos para modelación K-NN
- 41 Figura 17. RMSE vs. K
- 41 Figura 18. Gráfica de residuos KNN
- 42 Figura 19. Modelación de carbono KNN
- 43 Figura 20. Diagrama de procesos para modelación K-NN
- 45 Figura 21. Gráfica de residuos SVM
- 46 Figura 22. Modelación de carbono SVM
- 47 Figura 23. Diagrama de procesos para modelación RF
- 48 Figura 24. Factor de importancia RF
- 48 Figura 25. Gráfica de residuos RF
- 49 Figura 26. Modelación de carbono RF

PRÓLOGO

La minería para el departamento del Cesar es una actividad muy importante, debido a varios factores, entre ellos el producto interno bruto. Tenemos que impulsar el producto interno bruto del departamento del Cesar, en el cual la minería corresponde a más del 40% del total. Es la actividad económica que genera el mayor número de empleos directos por empresa, permite obtener recursos económicos asociados a impuestos de industria y comercio en los municipios en donde está establecida, dinamiza la economía regional y además posibilita recursos vía regalías que permiten inversiones en proyectos de inversión pública en todo el departamento. Sin embargo, la minería también tiene consecuencias negativas para el medio ambiente y la biodiversidad. En particular, puede dejar terrenos afectados, con suelos erosionados y sin vegetación, lo que a su vez puede dificultar la recuperación de la flora y fauna locales. Afortunadamente, existen técnicas de rehabilitación y restauración de terrenos mineros que pueden ayudar a revertir estos daños, las cuales son incluidas en los planes de manejo ambiental de cada proyecto.

Uno de los principales problemas de los terrenos mineros rehabilitados es la falta de vegetación arbórea y la baja capacidad de los suelos para secuestrar carbono, lo que dificulta el cálculo de los aportes en captura de carbono. Sin embargo, gracias a los avances en tecnologías geoespaciales, ahora podemos contar con herramientas precisas para medir el secuestro de carbono en estos terrenos rehabilitados. Además, estas tecnologías permiten también conocer el potencial de la biomasa herbácea para secuestrar carbono.

Este libro explora el secuestro de carbono de la biomasa herbácea de terrenos rehabilitados post-minería mediante el uso de tecnologías geoespaciales. Aquí se presentan resultados de investigaciones y análisis que muestran cómo el uso de

estas herramientas puede ser clave para mejorar la eficiencia de la rehabilitación de los terrenos mineros y mitigar el cambio climático. Además, se incluyen propuestas y recomendaciones para la implementación de estas técnicas en proyectos de restauración en diferentes lugares del mundo.

Este libro está dirigido a investigadores, estudiantes, técnicos y profesionales que están interesados en la mitigación del cambio climático y la restauración de terrenos mineros y en la medición del secuestro de carbono por parte de coberturas vegetales asociadas a pasturas. Con su lectura, esperamos se contribuya a una gestión sostenible de los recursos naturales y a una mejor calidad de vida para las comunidades locales que dependen de ellos.

Antonio Rudas Muñoz

Director de la Maestría en Gestión Ambiental

Fundación Universitaria del Área Andina

■ 1. IMPACTOS ASOCIADOS A LA MINERÍA DE CARBÓN A CIELO ABIERTO

A nivel mundial, la minería se ha convertido en la principal fuente para la obtención de recursos minerales metálicos, no metálicos y fósiles, los cuales son empleados en la industria de la construcción, agricultura, medicina, textil, plástico y en la fabricación de una gran variedad de productos y objetos utilizados en la vida cotidiana. Durante la extracción de estos recursos se genera una serie de impactos ambientales a los componentes agua, suelo y atmósfera (Marove et al., 2022).

La primera etapa en una explotación minera a cielo abierto corresponde al descapote, el cual consiste en la remoción de la cobertura vegetal, el suelo y la roca estéril que cubre el yacimiento. En esta etapa se degradan las características físicas y químicas de suelos, limitando la instauración de vegetación nativa (Barliza et al., 2018).

A medida que se desarrollan las actividades extractivas se pueden presentar impactos al recurso hídrico (Avtar et al., 2020). Estas afecciones generalmente se asocian a la configuración hidrológica, las propiedades geológicas del mineral y las características ambientales del área intervenida (Suh et al., 2017), presentándose drenajes ácidos, contaminación por metales pesados, alta concentración de sulfuros, erosión y sedimentación de los cuerpos de agua (Emmanuel et al., 2018).

Las operaciones en minería de superficie aportan directa o indirectamente a la emisión de gases de efecto de invernadero (GEI), contribuyendo así al deterioro de la atmosfera (Dávila y Pérez, 2012). Estos gases son producidos durante el consumo de combustible de todos los equipos que intervienen en el proceso minero, las emisiones fugitivas de metano de los mantos de carbón y los gases producidos por la autocombustión del carbón (Guimar et al., 2018). El fenómeno de la autocombustión se atribuye “a la reacción química entre el carbón y el oxígeno lo que genera energía en forma de calor” (Cutiérrez et al., 2018).

Estos impactos generalmente son mitigados, compensados o corregidos simultáneamente al proceso extractivo. La rehabilitación de los suelos degradados es un proceso lento y complejo que implica “logros complejos de cobertura, densidad, composición de la comunidad, riqueza de especies y elementos estructurales” (Shackelford et al., 2018). Estas áreas rehabilitadas pueden actuar como sumideros de carbono; el establecimiento de vegetación mejora la captura de carbono en el suelo, biomasa aérea y microbiana (Tripathi et al., 2013).

Minería de carbón en el departamento del Cesar

Entre los años 2021 y 2022, el departamento del Cesar ha aportado aproximadamente el 50% del carbón que se produce en Colombia, representando el 19% del presupuesto de los municipios cesarenses y el 34% del presupuesto del gobierno departamental (Asociación Colombiana de Minería, 2022).

Existen 36 títulos de carbón activos en el Cesar, que abarcan una superficie de 109.215,09 hectáreas. Entre las principales empresas que extraen carbón en el departamento del Cesar se encuentran: Drummond, Carbones La Jagua, Colombian Natural Resources y el Consorcio Mineros Unidos. La producción de carbón en el departamento llegó a su punto más alto en el año 2016 con 52 millones de toneladas año. Con la salida de operación de la multinacional Prodeco en 2020 se registró un descenso en la producción, llegando a 24.8 millones de toneladas en el año 2022. Los datos de producción entre el año 2016 y 2022 se observan en la Figura 1.

El descenso paulatino de la producción de carbón en el departamento del Cesar supone una reducción en la emisión de gases de efecto invernadero a la atmósfera; sin embargo, aún se sigue produciendo una carga contaminante que debe ser mitigada, corregida o compensada. Por otra parte, la rehabilitación de las áreas minadas es un reto monumental que afrontan las empresas carboníferas debido a las grandes superficies a reforestar y al largo tiempo requerido para el establecimiento de un ecosistema con condiciones similares a las encontradas antes del desarrollo del proyecto minero. Este panorama abre la puerta al desarrollo de investigaciones tendientes a conocer el potencial de secuestro de carbono de estas áreas.

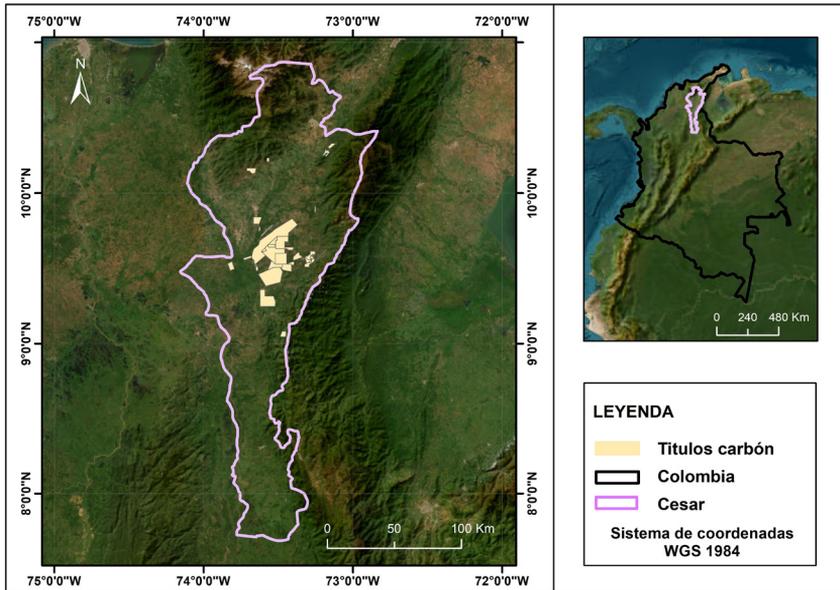


FIGURA 1
Títulos de carbón departamento del Cesar
Nota. Elaboración propia.

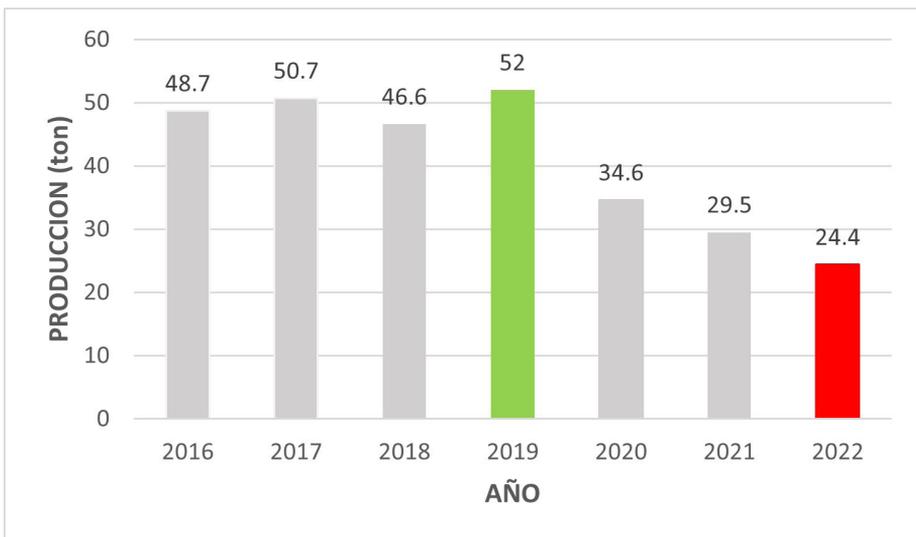


FIGURA 2
Producción de carbón en el departamento del Cesar (2016 - 2022)
Nota. Elaboración propia a partir de registros históricos de producción de carbón en Colombia presentes en el sitio web de la Unidad de Planeación Minero Energética.

■ 2. FUNDAMENTACIÓN PARA EL DESARROLLO DE INVESTIGACIONES DE CAPTURA DE CARBONO

La vegetación sembrada en zonas reforestadas cumple un rol importante en la mitigación de los impactos a la atmosfera, pues puede fungir como un potencial sumidero de carbono (Placek-Lapaj et al., 2019).

Diversos estudios realizados en todo el mundo demuestran la capacidad de secuestro de carbono de los suelos, hojarasca, biomasa leñosa y herbácea post-minería, mostrando que en zonas con plantaciones más antiguas el contenido de carbono retenido es mayor que en otras áreas (Ahirwal & Maiti, 2017; Pietrzykowski & Daniels, 2014). Adicionalmente, el carbono secuestrado varía dependiendo de variedad de la vegetación, tipo de sustrato y condiciones climáticas (Das & Maiti, 2016). En la Figura 3 se identifican las diferentes formas de acumulación de carbono terrestre.

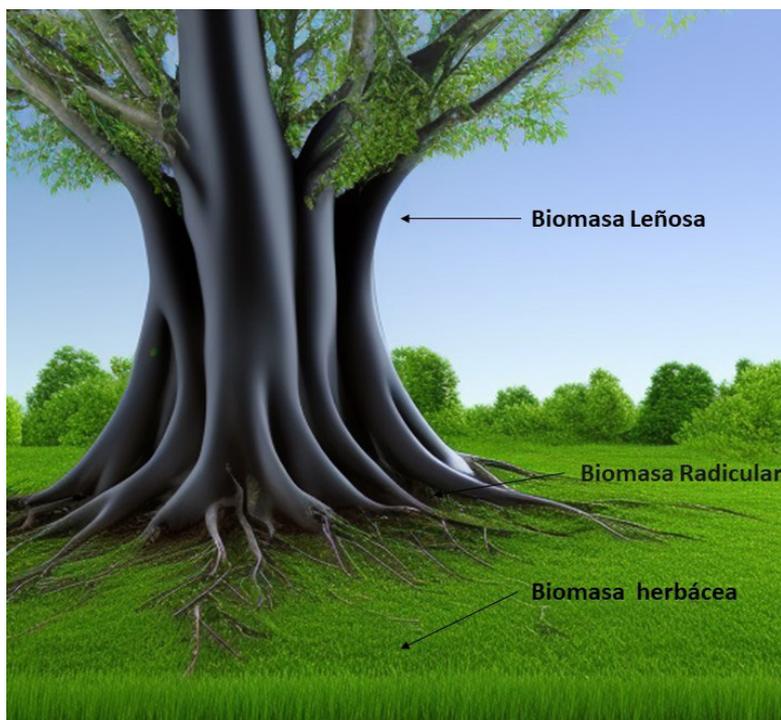


FIGURA 3
Acumulación de carbono terrestre
Nota. Creada en el sitio web BLUEWILLOW.

El fundamento del secuestro de carbono radica en que el CO₂ emanado a la atmósfera por procesos naturales o actividades antrópicas es absorbido por las plantas en el proceso de fotosíntesis. En este contexto, el bióxido de carbono es transformado en glucosa y se genera oxígeno; posteriormente se convierte en carbohidratos y tejidos que se convertirán en biomasa (FAO, 2011). Este principio es la base para la mitigación de los impactos en la atmósfera debido a la emisión de CO₂ durante el desarrollo de las operaciones mineras.

El problema ambiental ocasionado por la emisión de gases contaminantes ha captado la atención de gobiernos e investigadores a nivel mundial. Con el objetivo de reducir las emisiones GEI a la atmósfera en 1997 se adoptó el protocolo de Kyoto, en donde se compromete a los países industrializados a estabilizar las emisiones y se reconoce el papel de la biomasa en la captura y almacenamiento de carbono (Naciones Unidas, 1998).

En Colombia, por medio de la Ley 164 de 1994, el Congreso de Colombia aprueba la UNFCCC, donde el Estado colombiano se compromete a formular e implementar políticas que incluyan medidas tendientes a mitigar el cambio climático; posteriormente, el Consejo Nacional de Política Económica y Social (2011) expide el documento CONPES 3700 en el que se adopta la “Estrategia Institucional para la articulación de políticas y acciones en materia de Cambio Climático”, la cual busca garantizar que las entidades de orden nacional y territorial incorporen medidas para mitigar emisiones de gases de efecto invernadero. En el año 2017 el Ministerio de Ambiente y Desarrollo Sostenible suscribió el Plan Nacional de Cambio Climático, el cual busca integrar en las políticas que influyen en el desarrollo del país acciones de mitigación y adaptación.

En el marco de las estrategias para el desarrollo minero-energético bajo en carbono, el Ministerio de Minas y Energía (2018) expide la resolución 40807 correspondiente al Plan Integral de Gestión de Cambio Climático para el Sector Minero Energético, el cual tiene por objetivo promover el desarrollo minero con bajas emisiones de carbono, mediante la estructuración de tres componentes: mitigación, adaptación y gobernanza. En este contexto, el componente de mitigación se refiere a las acciones para promover la reducción de gas de efecto de invernadero en el sector minero-energético, mediante la contabilidad de emisiones/absorciones de GEI por compensaciones por el uso y cambio de uso del suelo.

3. METODOLOGÍA PARA ESTIMACIÓN DE CARBONO EN BIOMASA HERBÁCEA

Área de estudio

El estudio se llevó a cabo en las zonas reforestadas en el título minero 078-88, el cual se encuentra ubicado en el área de influencia de los municipios El Paso, La Jagua de Ibirico, Chiriguana, Codazzi y Becerril en el centro del departamento del Cesar (figura 4).

En el área se han reforestado más de 700 hectáreas utilizando vegetación nativa y otras especies resistentes a las características biofísicas de la zona. En la primera fase de reforestación utilizan el pasto *Andropogon gayanus* (carimagua), *Panicum máximum* (guinea) y *Dichantium aristatum* (angletón); siendo el pasto carimagua el que mejores resultados ha presentado en suelos postminería. También se siembran árboles como el *Coccoloba acuminata* (maíz tostao), *Samanea saman* (campano, genizaro), *Pseudosamanea guachapele* (igua amarillo), *Tabebuia chrysea* (cañahuate), *Crescentia cujete* (totumo), *Guayacum officianales* (guayacán), *Ceratonia siliqua* (algarrobo) (Guerra Peralta, 2017).

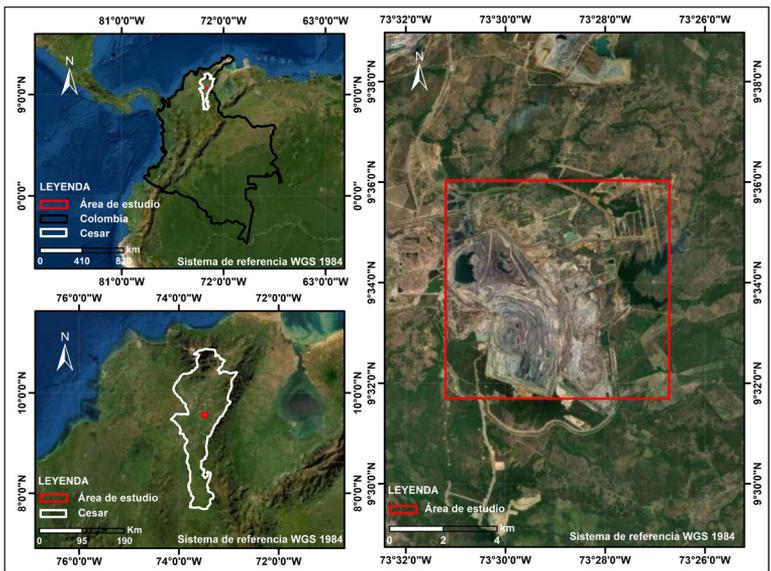


FIGURA 4
Localización del área de estudio

Nota. Elaboración propia.

Metodología de modelamiento

La modelación de carbono secuestrado por la biomasa herbácea se realiza a partir de la combinación de diversas técnicas para la captura, análisis y representación de datos. La metodología consta de tres fases: la primera corresponde a trabajo de campo, la segunda consiste en la adquisición y procesamiento de la imagen satelital y en la tercera fase se realiza la correlación de los datos de campo y sensores remotos (figura 5).

Fase I. Trabajo de campo. En esta fase se realizan visitas de campo preliminar necesario para la identificación de la biomasa, delimitación del área a trabajar y selección de puntos donde se realizó el muestreo.

Durante el trabajo de campo se tomaron 10 muestras de la biomasa herbácea, las cuales fueron adquiridas en subparcelas de 1 m², posteriormente las muestras son llevadas al laboratorio para estimar el contenido de carbono. Las muestras tomadas tienen un peso de 250 g (Yepes et al., 2011). Adicionalmente, se tomaron 11 puntos de referencia en zonas en las cuales no hay presencia de biomasa; en total 21 muestras.

Fase II. Adquisición y procesamiento de imagen satelital. Se emplea imagen de satélite sentinel-2A, las cuales se descargan desde el portal de Copernicus Open Access Hub. La imagen seleccionada para el procesamiento es del mismo mes de en que se realizó el muestreo de la biomasa.

A la escena satelital descargada se le aplica corrección atmosférica, terrenos y cirros de tal forma que se depuren los factores que puedan incidir negativamente en la calidad de los datos proporcionados por la imagen, para ello se utiliza el procesador SEN2COR (Gascon et al., 2017).

Luego, se calculan múltiples índices de vegetación (NDVI, ARVI, EVI, SAVI, MSAVI₂, OSAVI, VARI), que servirán como parámetro de entrada para el desarrollo de modelos predictivos.

Fase III. Se realiza la correlación de los datos tomados en campo con los índices de vegetación extraídos aplicando diversos métodos: Regresión por mínimos cuadrados ordinarios (OLS), Random forest (RF), Máquinas de vectores de soporte (SVM) y K-Vecinos más cercanos (K-NN). La relación entre el contenido de carbono y los índices de vegetación se realizó mediante la aplicación de la librería sklearn del lenguaje de programación Python.

Los modelos desarrollados utilizan como variable dependiente (respuesta) el contenido de carbono en la biomasa herbácea y como variable independiente (predictora) los índices de vegetación ARVI, EVI, MSAVI₂, NDVI, OSAVI, SAVI y VARI.

La evaluación de los modelos desarrollados se realiza mediante el coeficiente R² (Escribano Rodríguez et al., 2014).

Técnicas de recolección de muestras de biomasa herbácea. El muestreo se realiza durante la fase de trabajo de campo teniendo en cuenta los siguientes aspectos

- Subparcelas. Tamaño de 1 m² y se distribuyen de manera aleatoria en el área de estudio.
- Se corta todo el pasto dentro de las subparcelas de muestreo; el corte debe realizarse a nivel del suelo.
- La muestra se guarda en bolsas plásticas o de papel, se rotula y se envía a laboratorio.
- Determinación de contenido de biomasa y contenido de carbono.

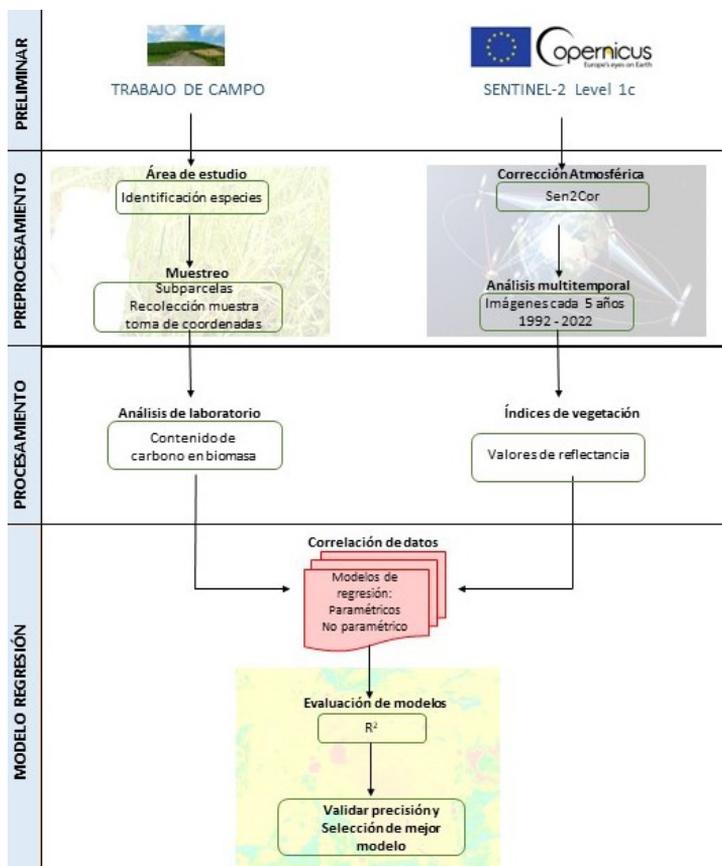


FIGURA 5
Esquema de metodología de trabajo
Nota. Elaboración propia.

4. OBTENCIÓN DE CONTENIDO DE CARBONO A PARTIR DE MUESTRAS TOMADAS EN CAMPO

Análisis multitemporal

A partir del análisis de las imágenes Landsat y Sentinel-2A adquiridas con una diferencia de 5 años, se identifica cómo ha sido la evolución de las actividades extractivas y de rehabilitación en el área. En la figura 6 se plasman los detalles del análisis realizado.

Sensor y año adquisición	Imagen	Características
Landsat 5 Año 1991		En la imagen se observa la línea base antes de iniciar el proyecto minero. Predominan suelos ocupados por actividad agrícola y ganadera.
Landsat 7 Año 1996		Se observa la fase inicial de la operación minera; al norte el área de acopio del material estéril (botaderos) y en la parte inferior el pit de explotación.

Sensor y año adquisición	Imagen	Características
Landsat 7 – Año 2001		<p>Se observa la expansión del pit minero hacia el centro del título minero, la expansión de áreas de botadero hacia el este del título, construcción de vías de acarreo en dirección a los patios de acopio de carbón y se observa el inicio de la etapa de reforestación en los botaderos ubicados al norte.</p>
Landsat 7 – Año 2006		<p>En la imagen se aprecia la expansión de pit minero y los botaderos al norte y al este del título. Adicionalmente, se evidencia un aumento de las zonas reforestadas en los botaderos al norte y este del título.</p>
Landsat 7 – Año 2011		<p>Se evidencia la expansión del pit minero, la construcción de un nuevo botadero al oeste y la expansión de botadero al este hasta el límite sur del área de trabajo. De igual forma se aprecia un aumento en las zonas reforestadas.</p>
Sentinel-2A Año 2016		<p>Por su resolución espacial la imagen Sentinel-2A permite ver más detalles que la imagen Landsat. Se evidencia el inicio de reforestación del botadero ubicado al oeste del título.</p>

Sensor y año adquisición	Imagen	Características
Sentinel-2A Año 2022		La imagen muestra con mayor detalle aquellas áreas que han sido objeto de reforestación.

FIGURA 6
Análisis multitemporal
Nota. Elaboración propia.

De la información recolectada en campo y los datos proporcionados por las imágenes de satélites se identificaron las áreas intervenidas y que han sido objeto de rehabilitación. A partir de la combinación RGB843 (Figura 7) se resalta la vegetación herbácea establecida en zonas reforestadas: la 1 zona ubicada al oeste y la zona 2 ubicada al norte, corresponden a zonas antiguas con pastos bien establecidos; la zona 3 ubicada al este aún se encuentra en labores preparatorias.

El cálculo del carbono solo se realizará en las zonas 1 y 2, debido a que se encuentran en una etapa avanzada. La especie herbácea utilizada en la etapa de rehabilitación es el *Andropogon gayanus* (pasto carimagua), especie originaria de África occidental, que se caracteriza por su resistencia a suelos de baja fertilidad, ácidos bien drenados y aguante a sequías extensas (Laredo C. & Gómez S., 1982). Las características de esta especie la hacen ideal para los procesos de rehabilitación de terrenos minados (Figura 8).

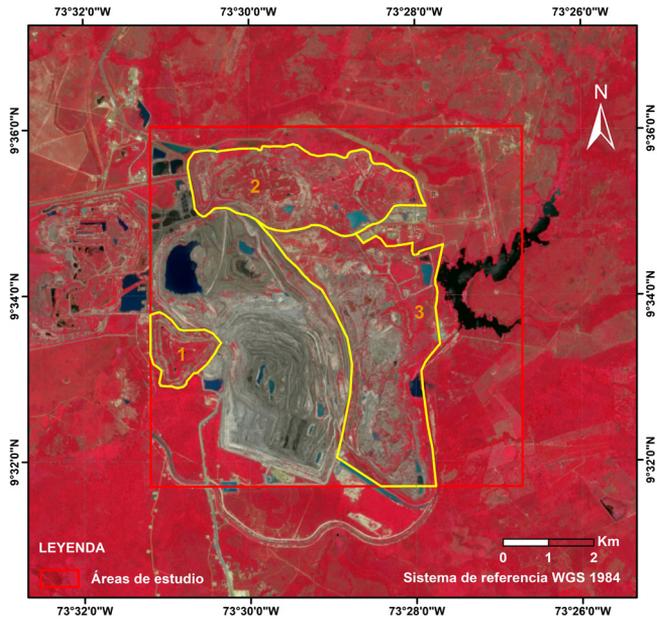


FIGURA 7
Áreas en proceso de rehabilitación-combinación 843
Nota. Elaboración propia.



FIGURA 8
Pasto sembrado
Nota. Elaboración propia.

Muestreo

Se establecieron 21 puntos de muestreo; 10 de ellos son parcelas de 1m² donde se cosechó la biomasa herbácea allí presente y 11 puntos correspondientes a suelo desnudo y agua (figura 9). Posteriormente, se georreferenciaron las parcelas utilizando los Módulos GNSS M2 de Emlid reach (figura 10), que proporcionan una precisión centimétrica en los procesos de medición (Fernández Álvarez et al., 2021).

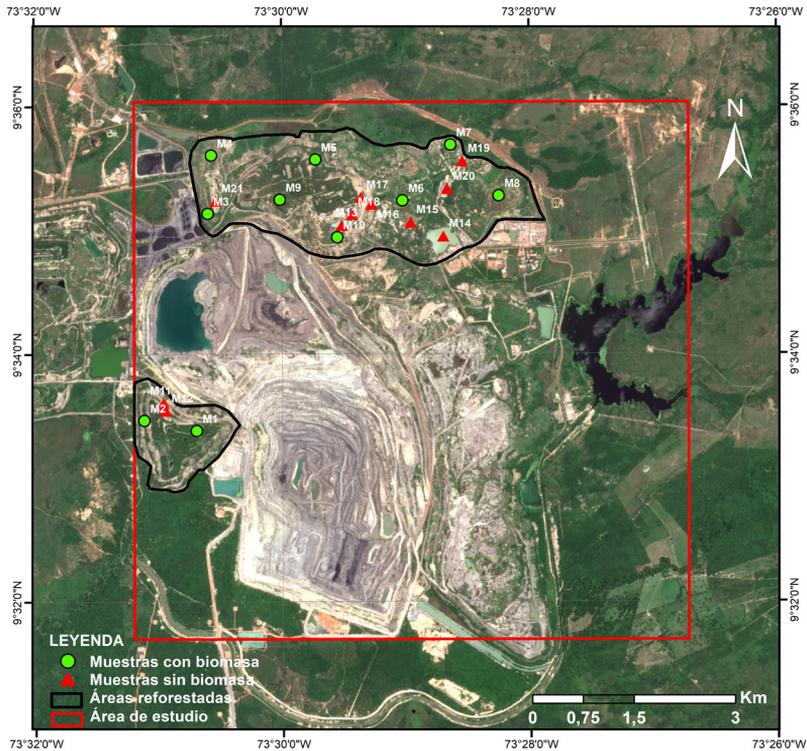


FIGURA 9

Puntos de muestreo en campo

Nota. Elaboración propia. Las muestras del 1 al 10 corresponden a las parcelas con presencia de biomasa, mientras que los valores entre 11 y 21 corresponden a los puntos donde no hay presencia de biomasa.



FIGURA 10
Georreferenciación de punto de muestreo
Nota. Elaboración propia.

Las muestras de biomasa tomadas en campo fueron llevadas al laboratorio, allí se determinó el contenido de carbono en la biomasa a partir de la materia orgánica contenida en las muestras procesadas. Los datos obtenidos se observan en la Tabla 1.

TABLA 1
Georreferenciación y contenido de biomasa por parcela

Muestra	Peso de biomasa por parcela (Kg)	Coordenadas	
		Latitud	Longitud
1	21.080	-73.5186438	9.55774462
2	5.780	-73.5115870	9.55634800
3	14.535	-73.5100234	9.58556140
4	14.280	-73.5095312	9.59340906
5	8.075	-73.5002874	9.58742947
6	12.070	-73.4954828	9.59281657
7	15.725	-73.4925590	9.58236598
8	22.355	-73.4837645	9.58727719

Muestra	Peso de biomasa por parcela (Kg)	Coordenadas	
		Latitud	Longitud
9	12.410	-73.4772657	9.59476773
10	8.755	-73.4708012	9.58790536
11	0	-73.5160359	9.56008041
12	0	-73.5157583	9.55906009
13	0	-73.4919463	9.58397406
14	0	-73.478262	9.58247888
15	0	-73.4826646	9.58436418
16	0	-73.4879326	9.58679147
17	0	-73.4893747	9.58773076
18	0	-73.4905418	9.58543923
19	0	-73.4756981	9.59258107
20	0	-73.4777094	9.58884237
21	0	-73.5089215	9.58728433

Nota. Elaboración propia.

Análisis de laboratorio

Del análisis realizado se establece que el contenido de carbono promedio en las parcelas de muestreo donde se obtuvo biomasa es de 49.83 %. En la figura 11 se observa el procesamiento de imágenes en laboratorio y los resultados obtenidos en la tabla 2.



FIGURA 11
Procesamiento de muestras en laboratorio
Nota. Elaboración propia.

TABLA 2
Contenido de carbono en muestras

Muestra	Peso muestra	Peso crisol	Peso muestra+cresol	% Materia seca	Peso muestra seca	Peso material calcinado	% Carbono
1	2	20.08	22.08	19.92	1.602	0.792	50.12 %
2	2	19.81	21.81	20.15	1.597	0.788	48.84 %
3	2	22.16	24.16	20.14	1.597	0.803	49.82 %
4	2	18.73	20.73	20.05	1.599	0.801	49.57 %
5	2	21.94	23.94	20.21	1.596	0.795	48.98 %
6	2	20.08	22.08	19.98	1.600	0.802	49.72 %
7	2	19.81	21.81	20.11	1.598	0.792	50.09 %
8	2	22.16	24.16	20.02	1.600	0.800	50.25 %
9	2	18.73	20.73	20.03	1.599	0.798	49.89 %
10	2	21.94	23.94	20.05	1.599	0.795	49.57 %

Nota. Elaboración propia.

Conclusiones

La minería a cielo abierto se caracteriza por dejar huellas visibles sobre la superficie terrestre, las cuales pueden ser mapeadas utilizando imágenes de satélite. Las características espectrales de las imágenes facilitan la diferenciación del suelo desnudo, cuerpos de agua y áreas que han sido objeto de reforestación. Mediante un análisis multitemporal, es factible identificar la evolución de las áreas objeto de explotación versus las zonas objeto de reforestación.

La especie herbácea utilizada para la regeneración de la vegetación en el área de estudio es el *Andropogon gayanus* kunth, en donde la concentración media del carbono contenido es de 49.83%, favoreciendo el secuestro de las emisiones de gases producidas durante las operaciones de extracción de mineral.

5. EXTRACCIÓN DE ÍNDICES DE VEGETACIÓN

Las imágenes satelitales juegan un importante en el desarrollo de la investigación, permiten la observación remota de la tierra a partir de la interacción de un flujo electromagnético con las coberturas sobre la superficie (Pérez Gutiérrez & Muñoz Nieto, 2002), lo que las hace ideales para el estudio de la biomasa (Song et al., 2015; Xiao et al., 2019)

Índice de vegetación

Medida utilizada para extraer información de las coberturas vegetales mediante análisis de los valores de reflectancia a diferentes longitudes de onda (Cilabert et al., 1997).

A partir del procesamiento de la imagen Sentinel-2A se extraen los índices de vegetación ARVI, EVI, MSAVI₂, NDVI, OSAVI, SAVI, VARI asociados al área de estudio. En la tabla 3 se describen las características y ventajas de los índices a utilizar.

TABLA 3
Índice de vegetación

Índice	Característica	Desventajas	Referencia
1. Índice de Vegetación de la Diferencia Normalizada (NDVI)	Utilizado para monitorear el verdor de la vegetación. Toma valores entre -1 a 1, siendo los valores cercanos a 1 indicador de vegetación saludable y abundante, a 0 presencia de agua y valores negativos suelo desnudo.	Sensible a la reflectividad del suelo; en zonas donde hay poca vegetación el suelo ejerce un efecto perturbador que varía dependiendo el tipo de suelo y porcentaje de cobertura vegetal.	(Rouse et al., 1974)
2. Índice de Vegetación Ajustado al Suelo (SAVI)	Empleado para corregir el efecto de la reflectancia del suelo en zonas áridas con cobertura vegetal escasa, como parámetro de cálculo incluye la cantidad de cubierta de vegetación verde. Toma valores entre -1 a 1.	Una mala elección del parámetro L (cantidad de cobertura vegetal) altera significativamente los resultados.	(Huete, 1988)
3. Índice de vegetación ajustado por suelo modificado (MSAVI ₂)	Este índice disminuye los efectos del suelo desnudo sobre la vegetación.	Zonas de cobertura vegetal densa presenta valores anómalos.	(Qi et al., 1994).

Índice	Característica	Desventajas	Referencia
4. Índice de Vegetación Ajustado al Suelo Optimizado (OSAVI)	Método empírico de la familia del índice savi, ayuda a disminuir los efectos de la reflectancia del suelo gracias a la inclusión de un parámetro x con un valor de 0.16 en el numerador de la expresión matemática.	Recomendado para aplicaciones agrícolas	(Rondeaux et al., 1996).
5. Índice de vegetación mejorado (EVI)	Especialmente sensible en zonas de abundante vegetación, optimiza la señal del índice captado. Se caracteriza por tener menor sensibilidad al ruido generado por la atmósfera.	Pierde precisión en zonas con escasa vegetación.	(Huete et al., 1997).
6. Índice de Vegetación Atmosféricamente Resistente (ARVI)	Índice mejorado ampliamente utilizado para atenuar la influencia de la atmósfera. Empleado en zonas con una alta presencia de aerosol atmosférico.	Para una mayor precisión del método debe conocerse previamente el modelo de aerosol y. De lo contrario el valor a utilizar es 1.	(Tanré et al., 1992).
7. Cociente simple (VARI)	Este índice se caracteriza por la detección de la vegetación solo con las bandas del rango visible	Poco sensible para suelo desnudo, agua y hielo.	(Gitelson et al., 2002).

Fuente. Elaboración propia.

Los índices de vegetación extraídos son la base para el desarrollo de los modelos de estimación de carbono en la biomasa herbácea (tabla 4).

TABLA 4
Índice de vegetación extraído para cada punto de muestreo

ID	arvi	evi	msavi2	ndvi	osavi	Savi	vari
1	0.61262	0.000590	1.36675	0.639758	0.764832	1.14714	0.354798
2	0.33012	0.000403	1.19063	0.47224	0.527481	0.791134	0.111861
3	0.54033	0.000523	1.33193	0.627223	0.712221	1.06821	0.225313
4	0.52859	0.000472	1.32809	0.606342	0.706618	1.059820	0.242117
5	0.35524	0.000321	1.22476	0.55205	0.568348	0.85243	0.032967
6	0.52608	0.000483	1.32601	0.648332	0.7036	1.05529	0.147541
7	0.55994	0.000523	1.3426	0.640046	0.728013	1.09191	0.237838
8	0.70347	0.000632	1.4059	0.718613	0.827973	1.24185	0.393891

ID	arvi	evi	msavi2	ndvi	osavi	Savi	vari
9	0.53594	0.000561	1.32779	0.595506	0.706184	1.05917	0.284483
10	0.40367	0.000437	1.24898	0.602108	0.598713	0.897989	-0.008282
11	-0.0067	0.000106	0.62005	0.087703	0.063867	0.095793	-0.041927
12	-0.0394	0.000059	0.68503	0.196392	0.101961	0.152931	-0.146154
13	-0.2185	-0.002888	-0.76527	-0.44763	-0.387831	-0.581664	0.118659
14	-0.2991	-0.002125	-2.15028	-0.62494	-0.570993	-0.856333	0.132867
15	-0.2574	-0.000675	-0.96263	-0.49260	-0.423011	-0.634382	0.13597
16	-0.0665	0.000043	0.66251	0.209155	0.088452	0.13267	-0.178867
17	-0.0414	0.000061	0.71955	0.242401	0.123322	0.184972	-0.178961
18	-0.1303	-0.000006	0.47525	0.104849	-0.012226	-0.018338	-0.170543
19	-0.0728	0.000036	0.64199	0.1994	0.076425	0.114632	-0.182734
20	-0.0091	0.000087	0.71465	0.187255	0.120248	0.180356	-0.108629
21	-0.0727	0.000039	0.65263	0.216216	0.082633	0.123938	-0.197647

Nota. Elaboración propia.

Los valores espectrales asociados a los puntos de muestreo varían dependiendo el índice calculado. Para las zonas donde hay pastos sembrados el valor varía dependiendo la densidad de la biomasa, con valores de NDVI entre 0.3 y 0.5 para áreas con vegetación escasa y mayor a 0.5 para zonas con vegetación densa (Toosi et al., 2022). Lo que es consistente con los datos obtenidos de la imagen de satélite, que muestra que la mayor parte de los puntos de muestreo tiene una cobertura vegetal densa.

Los puntos de muestreo 11, 12, 13, 17, 18, 19, 20 y 21 se ubican sobre suelo desnudo, de los cuales los puntos 11, 12, 19 y 20 se encuentran sobre las vías o en zonas aledañas a ellas, mientras que 16, 17, 18 y 21 se encuentran sobre los taludes. Adicionalmente, los puntos 13, 14 y 15 se encuentran sobre el agua.

Los valores espectrales de los puntos sobre el terreno desnudo presentan los valores más bajos. Diversos autores han concluido en líneas de investigación similares. (Kaufman & Tanré, 1992) describen que los valores asociados al suelo desnudo oscilan entre -0.08 a 0.08, y (Toosi et al., 2022) sostienen que los valores de NDVI menores a 0.1 corresponden a áreas áridas.

Adicionalmente se puede observar que los valores asociados a los cuerpos de agua adoptan valores negativos (Santiago et al., 2009). De los datos extraídos de la imagen Sentinel-2A se observan valores negativos que varían dependiendo del índice de vegetación evaluado, solo se exceptúa el VARI que toma valores positivos.

Conclusión

No existe el Índice de Vegetación perfecto, cada uno de ellos tiene un conjunto de características que los hace útiles bajo ciertas características del área de estudio. Los resultados obtenidos presentarán variaciones dependiendo del índice utilizado.

6. MODELACIÓN DE CARBONO EN LA BIOMASA HERBÁCEA

El análisis de los datos del área de estudio es desarrollado a partir de la construcción de modelos de regresión, el cual tiene como “objetivo predecir o aproximar el valor de una característica (variable dependiente) en función de los valores que pueden tomar en conjunto otra serie de características (variables independientes)” (Moral Peláez, 2016).

Estadística descriptiva de los datos

Inicialmente se analiza la relación entre los índices de vegetación extraídos de la imagen de satélite con los valores de contenido de carbono a partir del coeficiente de correlación de Pearson, el cual muestra un panorama general de los modelos a generar.

El diagrama de la matriz de correlación de la Figura 12 muestra una alta correlación entre los índices de vegetación ARVI, OSAVI y el SAVI con el contenido de carbono de las muestras, con un valor de 0.94, 0.86 y 0.86 respectivamente, mientras el EVI es el índice con menor correlación con el contenido de carbono (0.52). Adicionalmente, existe una alta correlación entre algunos predictores; entre el SAVI–OSAVI muestra una relación del 1 y entre el NDVI–SAVI la relación es de 0.98.

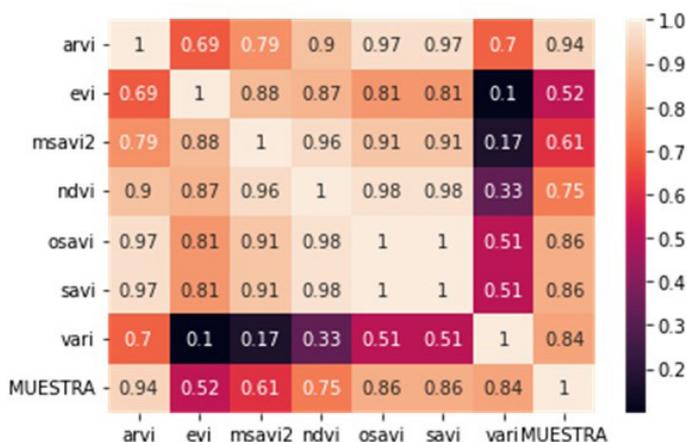


FIGURA 12
Matriz de correlación
Nota. Elaboración propia.

Los valores obtenidos muestran preliminarmente al ARVI como el índice de vegetación que obtendría mejores resultados al modelar el carbono contenido en la biomasa herbácea, al utilizar el modelo de regresión por mínimos cuadrados ordinarios.

Regresión por mínimos cuadrados ordinarios (OLS)

Este método tiene por objetivo modelar una variable dependiente relacionada con una variable explicativa, generando una línea recta ajustada (Hanke, 2006). En la Figura 13 se ilustra el proceso para el modelamiento utilizando la regresión OLS.

En la Figura 14 se observa la relación entre los índices de vegetación extraída de la imagen de satélite y el contenido de carbono en la biomasa herbácea. Se puede apreciar que la mejor relación entre puntos y la línea de tendencia se presenta con los índices ARVI, OSAVI y SAVI.

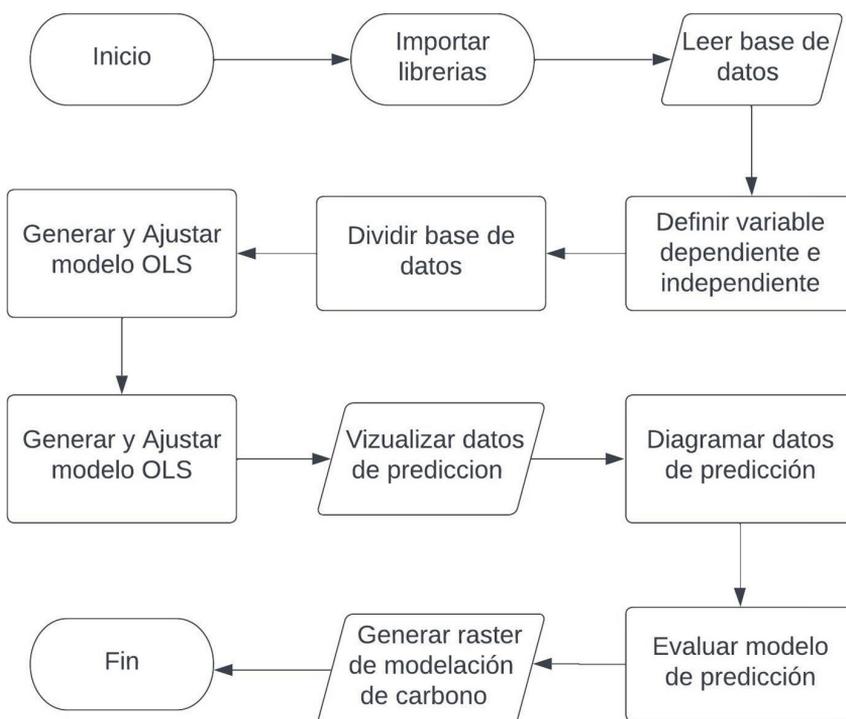


FIGURA 13

Diagrama de procesos para modelación OLS

Nota. Elaboración propia. El código Python para la generación del modelo OLS se observa en el Anexo 1.

La evaluación de los modelos desarrollados utilizando la regresión OLS, muestra que los ARVI presentaron el mejor ajuste ($R^2 = 0.88$, $RMSE=1.41$, $MAE=1.19$) y OSAVI ($R^2 = 0.72$, $RMSE=2.43$, $MAE=2.15$), mientras que los índices que presentaron la menor relación lineal fueron el EVI ($R^2 = 0.31$, $RMSE=3.41$, $MAE=3.50$) y el MSAVI2 ($R^2 = 0.35$, $RMSE=3.57$, $MAE=3.07$). La evaluación desarrollada muestra gran similitud con lo evidenciado en la matriz de correlación, mostrando los índices ARVI, SAVI y OSAVI como los de mayor relevancia para la estimación del carbono y el EVI como el de menor importancia. Adicionalmente, es importante resaltar que el MSAVI2 no muestra un buen rendimiento para la estimación. En la Figura 14 se observa la relación entre los índices de vegetación extraídos de la imagen de satélite y el contenido de carbono en la biomasa herbácea, donde se aprecia que la mejor línea de tendencia con mejor ajuste se presenta con los índices ARVI, OSAVI y SAVI.

TABLA 5
Evaluación de modelos OLS

	ARVI	EVI	MSAVI2	NDVI	OSAVI	SAVI	VARI
MAE	1.19	3.50	3.07	1.86	2.15	2.08	1.41
RMSE	1.41	3.49	3.57	2.12	2.43	2.35	2.00
R^2	0.88	0.31	0.35	0.55	0.72	0.70	0.68

Nota. Elaboración propia.

La Figura 15 muestra el contenido de carbono: en tonos rojos las áreas sin presencia de biomasa como lo son el suelo desnudo y depósitos de agua; en tonos verdes zonas con biomasa. La máxima concentración de carbono capturado presenta un valor de 10.93 Kg C/m², siendo mayor en aquellas áreas de mayor densidad de biomasa; es decir, en las zonas de reforestación más antiguas.

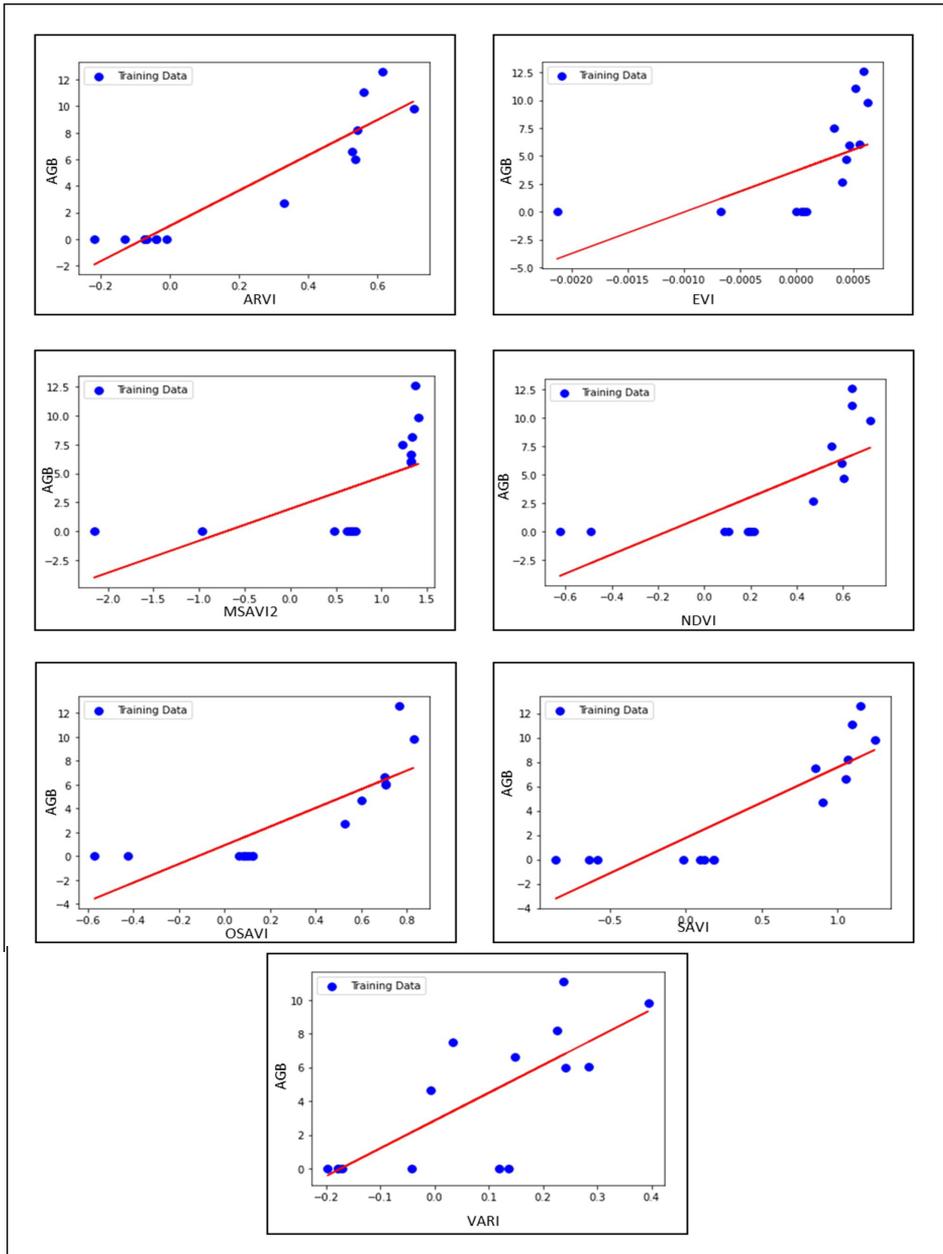


FIGURA 14
Diagramas Índice de Vegetación vs. Contenido de Carbono
Nota. Elaboración propia.

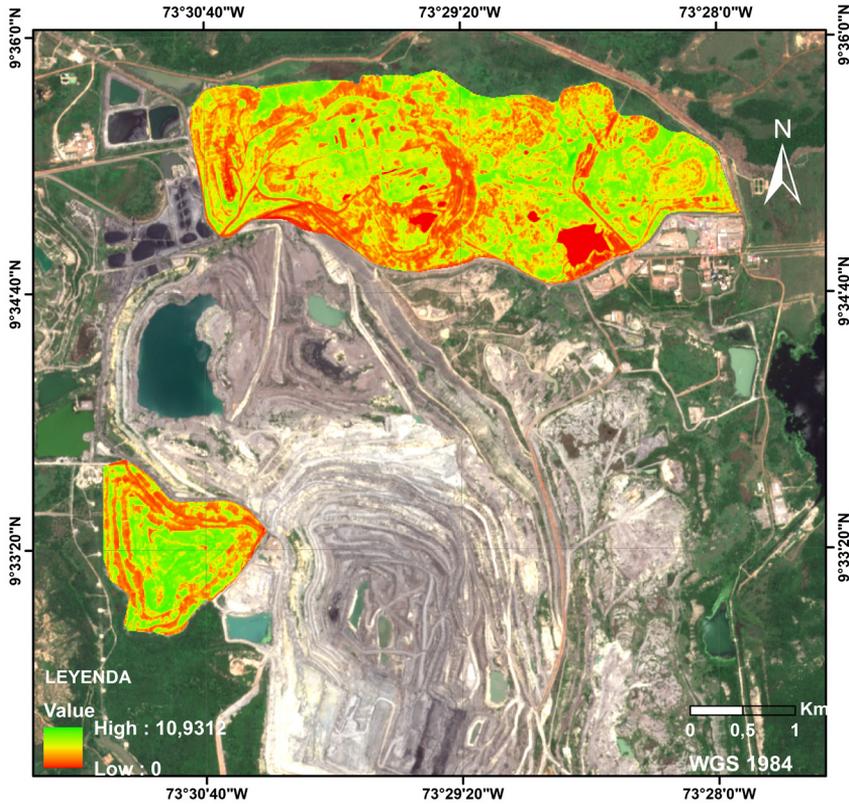


FIGURA 15
Modelación de carbono OLS
Nota. Elaboración propia.

K-vecinos más cercanos (K-NN)

Es un método supervisado no paramétrico empleado para predecir valores tanto en problemas de clasificación como de regresión; el algoritmo utiliza el promedio de los datos de entrenamiento más cercanos para predecir un valor (Rezaei & Jabbari, 2022). El proceso para la modelación del carbono en la biomasa herbácea utilizando la regresión K-NN se observa en la Figura 16; el algoritmo en Python del modelo regresión K-NN se observa en el Anexo 2.

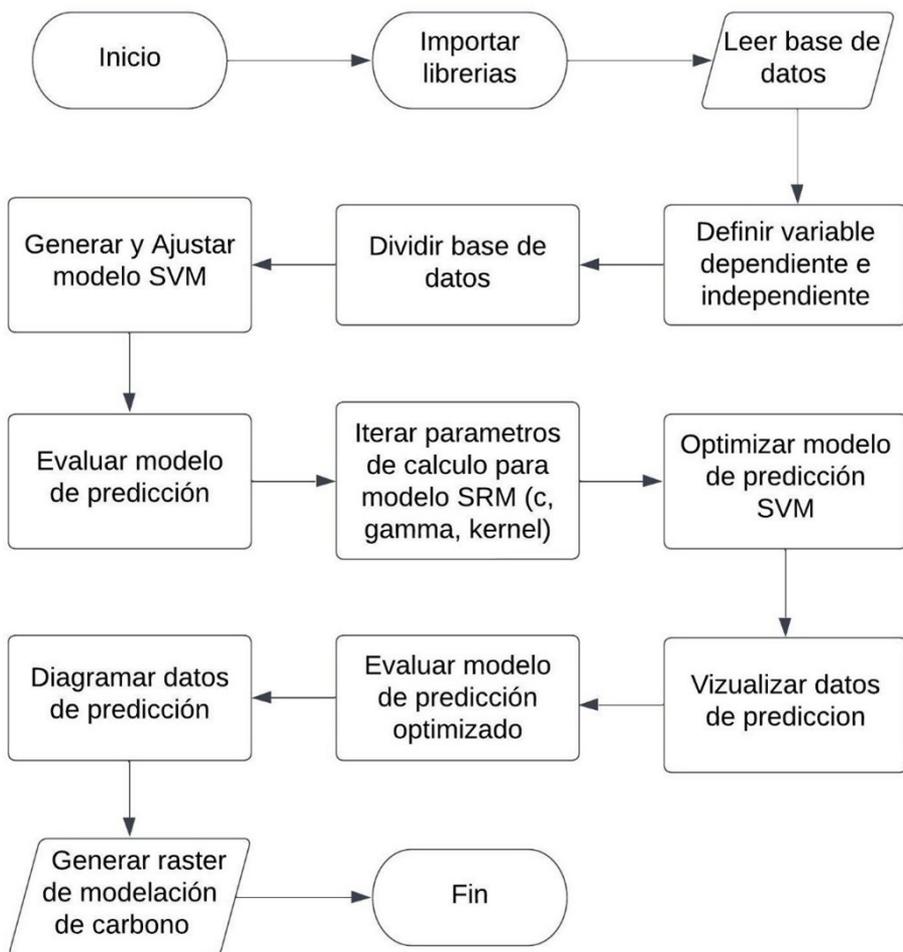


FIGURA 16
 Diagrama de procesos para modelación K-NN
 Nota. Elaboración propia.

A partir de los resultados iniciales, se aplicó el método del codo para determinar el Coeficiente K más idóneo para la estimación del carbono en la biomasa, siendo el valor de $K=1$ y $K=2$ con valores de error cuadrático medio de 1.60 y 1.65 respectivamente (Figura 17).

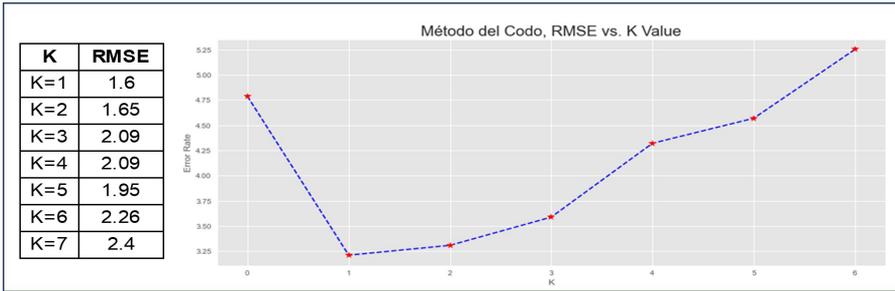


FIGURA 17

RMSE vs. K

Nota. Elaboración propia.

Finalmente, el modelo predictivo para estimar el carbono en la biomasa se generó en función de los datos arrojados por la matriz de correlación, utilizando el índice ARVI. La gráfica de residuos resultante (Figura 18) muestra una dispersión adecuada de puntos sobre la recta, lo que nos permite inferir que los modelos generados pueden estimar de manera adecuada el contenido de carbono.

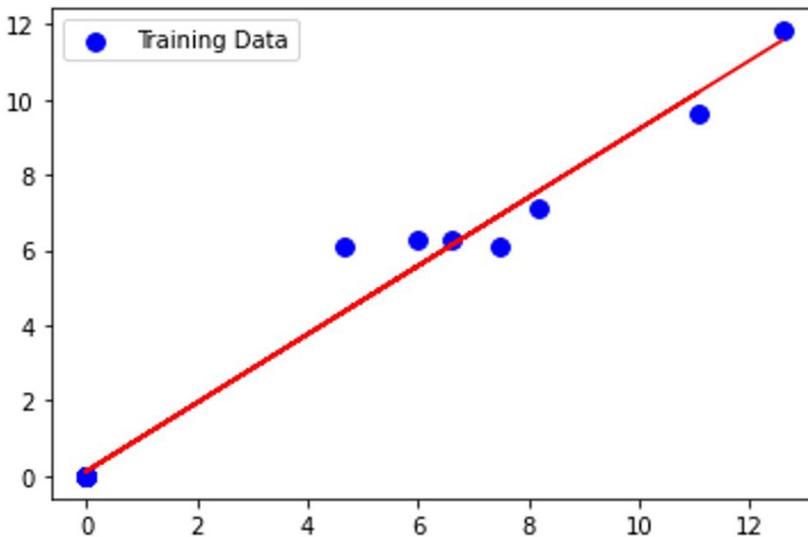


FIGURA 18

Gráfica de residuos KNN

Nota. Elaboración propia.

La distribución del contenido de carbono muestra gran semejanza al modelo desarrollado empleando la regresión OLS, siendo diferente la cantidad de carbono secuestrado por metro cuadrado que pasa de 10.93 kg C/m² a 11.80 kg C/m² en las

áreas de mayor acumulación (Figura 19). La evaluación del modelo muestra valores de $R^2=0.87$, $RMSE=1.51$ un $MAE=1.14$ reflejando un buen ajuste de los datos.

Máquina vectorial de soporte (SVM)

Es un algoritmo de aprendizaje automático que utiliza el algoritmo de clasificación máquina vectorial de soporte, empleado para predecir una variable dependiente. Este algoritmo modela datos de entrenamiento en una región llamada hiperplano, la cual se encuentra limitada por una región tubular llamada ϵ -tuve (Hawkins, 2015). Los valores por fuera de la función de predicción son penalizados y reciben el nombre de vector de soporte (Vapnik, 1998). El proceso para la modelación del carbono en la biomasa herbácea utilizando la regresión SVM se observa en la figura 20; el algoritmo en Python del modelo regresión SVM en el Anexo 3.

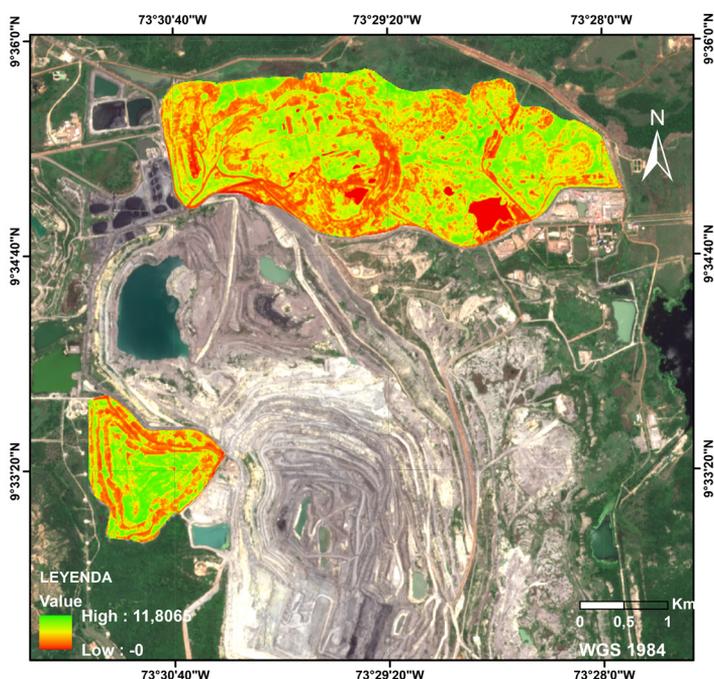


FIGURA 19
Modelación de carbono KNN
Nota. Elaboración propia.

La aplicación de este método de regresión se aplicó de manera separada a cada uno de los índices de vegetación. El algoritmo inicialmente se corrió con los parámetros

por defecto, mostrando que el mejor índice para estimar el contenido de carbono era el ARVI, mostrando valores de $R^2=0.77$, $RMSE=2.16$ y $MAE=1.27$. La evaluación para los modelos realizados se observa en la Tabla 6.

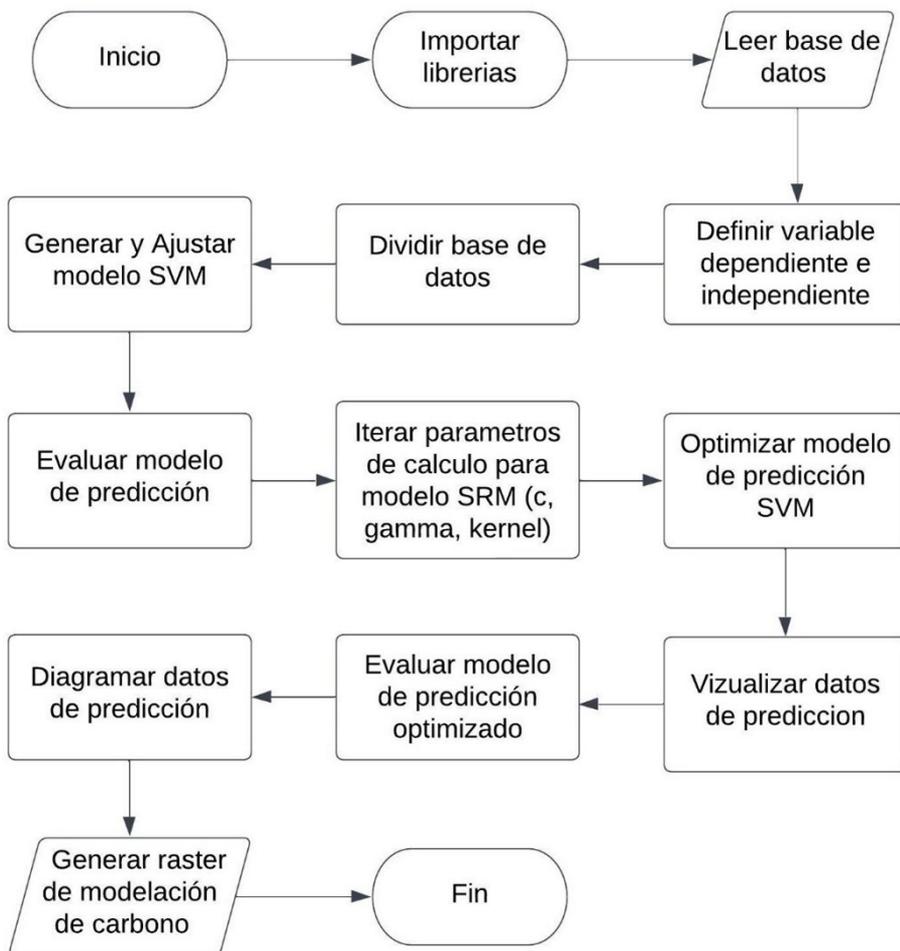


FIGURA 20
Diagrama de procesos para modelación K-NN
Nota. Elaboración propia.

TABLA 6
Evaluación de modelo SVM con parámetros de cálculo por defecto

	ARVI	EVI	MSAVI ₂	NDVI	OSAVI	SAVI	VARI
MAE	1.27	2.34	1.48	1.40	1.36	1.36	2.33
RMSE	2.16	3.03	2.42	2.35	2.44	2.34	3.11
R ²	0.77	0.55	0.71	0.73	0.73	0.73	0.52

Nota. Elaboración propia.

A partir de los resultados iniciales se optimiza el modelo, mostrando los mejores C, kernel y gamma para estimar el contenido de carbono (Tabla 7).

TABLA 7
Parámetros de cálculo para regresión SVM

	ARVI	EVI	MSAVI ₂	NDVI	OSAVI	SAVI	VARI
C	10	100	100	100	10	100	100
gamma	1	1	1	1	1	0.1	0.1
Kernel	Rbf	rbf	rbf	Rbf	Rbf	rbf	rbf

Nota. Elaboración propia.

Al evaluar los modelos utilizando los parámetros optimizados C, gamma y kernel se observa que los índices MSAVI₂ Y OSAVI presentaron un mejor ajuste estadístico. El MSAVI₂ presentó valores de R²=0.90, RMSE=1.43 y MAE=0.88; el índice OSAVI exhibió valores de R²=0.90, RMSE=1.46 y MAE=0.94 (Tabla 8). Para modelar el contenido de carbono en la biomasa se emplea el índice con valores más bajos, es decir, el MSAVI₂.

TABLA 8
Evaluación de modelo SVM con parámetros optimizados

	ARVI	EVI	MSAVI ₂	NDVI	OSAVI	SAVI	VARI
MAE	0.96	4.14	0.88	1.09	0.94	0.94	1.95
RMSE	1.60	4.86	1.43	1.65	1.46	1.49	2.65
R ²	0.87	0.16	0.90	0.87	0.90	0.89	0.66

Nota. Elaboración propia.

Al diagramar los datos de entrenamiento del modelo con el índice MSAVI2 se observa poca dispersión de los residuos, indicando una buena modelación de los datos (Figura 21).

El valor máximo de carbono en la biomasa herbácea es de 14.28 kg C/m². En el modelo desarrollado se observa presencia de carbono en áreas donde hay presencia de suelo desnudo, lo que lleva a una sobreestimación de los valores estimados (Figura 22).

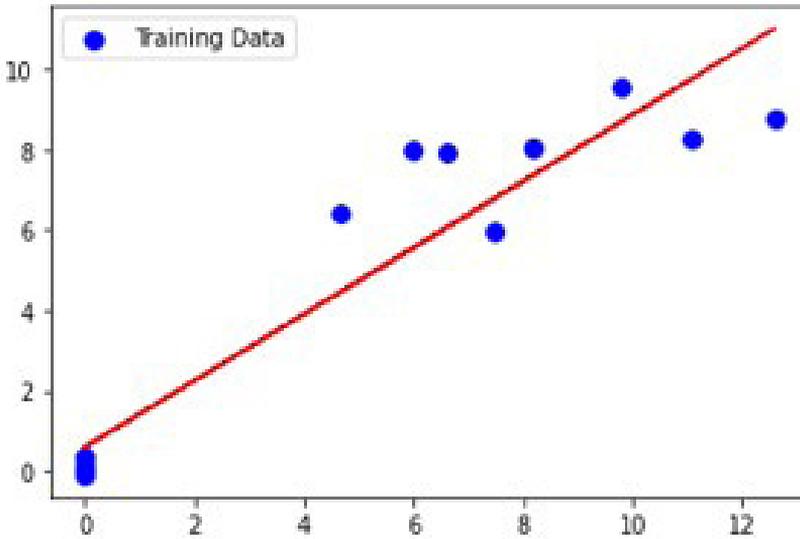


FIGURA 21
Gráfica de residuos SVM
Nota. Elaboración propia.

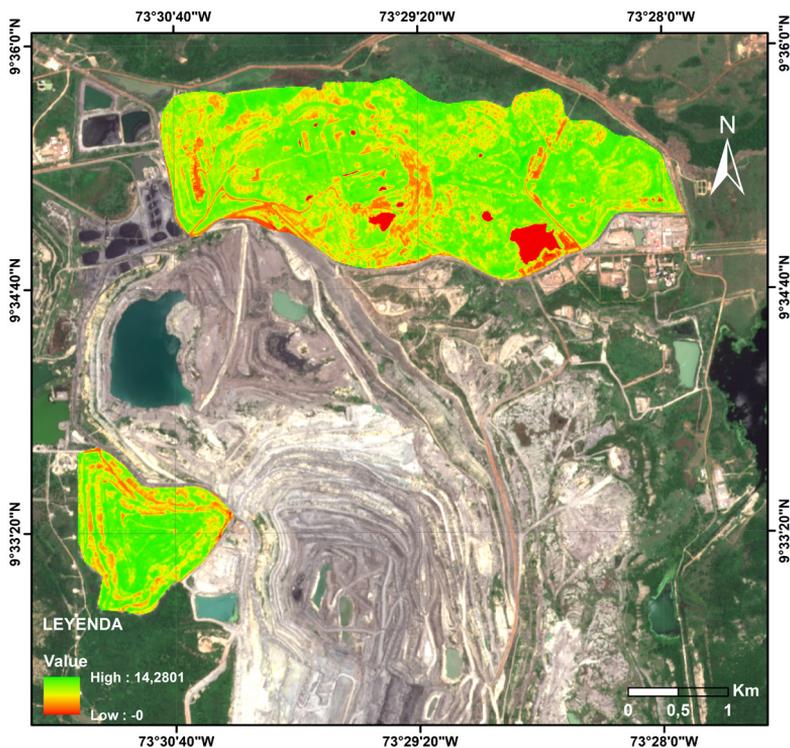


FIGURA 22
Modelación de carbono SVM
Nota. Elaboración propia.

Random Forest (RF)

Es un método de aprendizaje automático que permite crear modelos y generar predicciones. El algoritmo crea muchos árboles, cada uno de los cuales trabaja con un subconjunto de datos ligeramente distintos seleccionados aleatoriamente y obtiene una estimación (cada árbol es un modelo que genera su predicción). Luego, las respuestas de todos los árboles son analizadas tomando el dato más frecuente para dar la respuesta final (Rodrigo, 2020). El proceso para la modelación del carbono en la biomasa herbácea utilizando la regresión RF se observa en la Figura 23; el algoritmo en Python del modelo regresión RF en el Anexo 4.

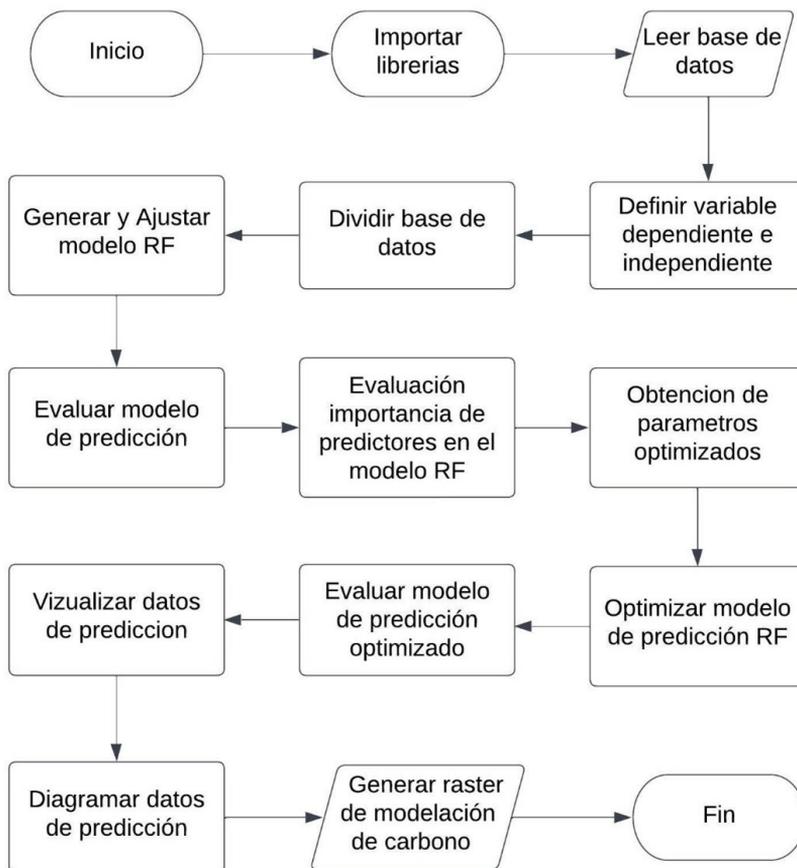


FIGURA 23
 Diagrama de procesos para modelación RF
 Nota. Elaboración propia.

La estimación de carbono utilizando el modelo Random Forest se realiza empleando un modelo general, donde se utilizan todos los índices de vegetación. Se corrió el modelo con los parámetros max_depth, max_features, n_estimato y random_state por defecto optimizados, mostrando valores de $R^2=0.93$, $RMSE=0.71$ y $MAE=0.59$.

TABLA 9
 Parámetros de cálculo RF

Max Depth	Max features	n estimators	Random state
5	log2	10	0

Nota. Elaboración propia.

Se estimó el factor de importancia para determinar cuál de los índices de vegetación tiene más incidencia en el modelo, donde se pudo determinar que el MSAVI2 tiene un mayor peso en la estimación de carbono (Figura 24).

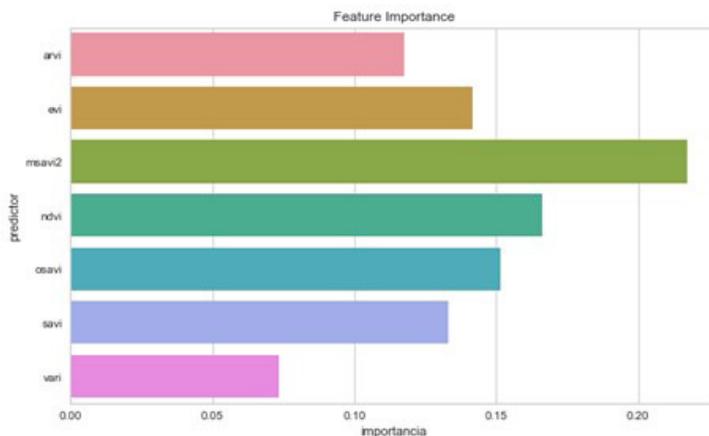


FIGURA 24
Factor de importancia RF
Nota. Elaboración propia.

Una vez determinado el factor de importancia se aplica la regresión Random forest para determinar el contenido de carbono empleando como variable independiente el índice MSAVI2. Al evaluar el modelo se observan valores de $R^2=0.935$, $RMSE=0.78$ y $MAE=0.585$. La Figura 25 muestra la distribución de los residuos, donde se aprecia poca dispersión de los datos.

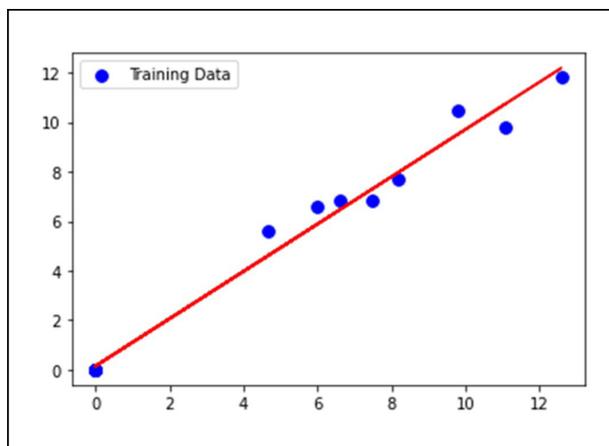


FIGURA 25
Gráfica de residuos RF
Nota. Elaboración propia.

El carbono estimado presenta un valor máximo de 13.04 kg C/m^2 . En la Figura 26 se identifican con facilidad las áreas con presencia de agua, suelo desnudo de aquellas que contienen biomasa.

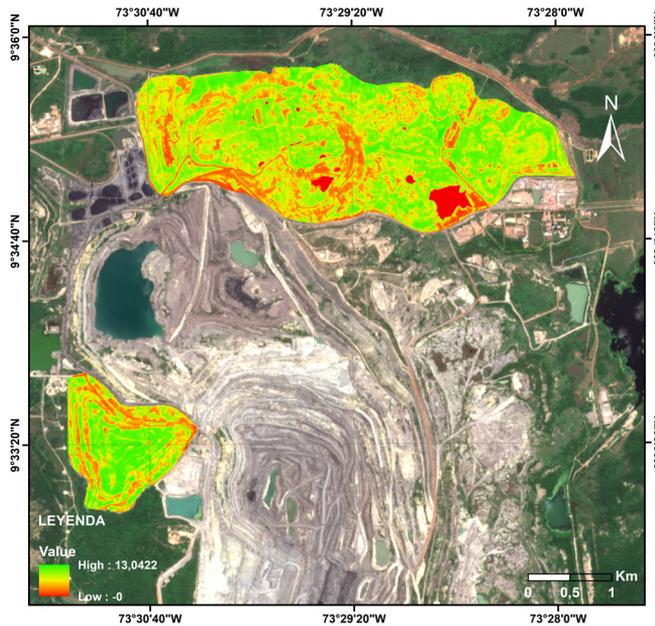


FIGURA 26
Modelación de carbono RF
Nota. Elaboración propia.

Conclusiones

A partir de una correcta delimitación de la superficie que contiene la biomasa herbácea mediante los índices de vegetación, el contenido de carbono en la biomasa y el uso de modelos de regresión (OLS, SVM, K-NN y RF), fue posible modelar el carbono secuestrado en la biomasa herbácea de los terrenos rehabilitados en el título minero 078-88.

La correcta ponderación de los parámetros asociados a los modelos de regresión de máquinas vectoriales de soporte, K-vecinos más cercanos y Random Forest mejoran significativamente la cuantificación del carbono secuestrado en la biomasa herbácea.

La mejor relación lineal de datos se presentó utilizando el modelo de regresión Random Forest en combinación con el índice de vegetación MSAVI2. Al evaluar el

modelo se obtienen valores de $R^2=0.935$, $RMSE=0.78$ y $MAE=0.585$, para una máxima concentración de carbono secuestrado de 13.04 kg C/m^2 .

El uso combinado de las herramientas de la teledetección y los Sistemas de Información Geográfica facilita la generación de cartografía para el seguimiento y control de áreas que cumplen la función de captura de carbono. Lo que permitirá a la empresa aportar con las medidas de reducción de emisiones de gases de efecto invernadero.

Para tener una mayor comprensión de la cantidad de carbono almacenado en los terrenos recuperados por parte de empresas mineras, se deben elaborar estudios más detallados que incluyan modelos de captura de carbono en los árboles, raíces y el suelo, lo que permitirá diseñar estrategias efectivas de conservación y restauración. De esta manera, el sector minero puede aportar al compromiso adquirido por Colombia en el Convenio Marco de las Naciones Unidas para el Cambio Climático, en el cual se comprometió a elaborar, actualizar y publicar datos de emisiones y absorción de gases de efecto invernadero.

En Colombia no se han definido metas que obliguen a las empresas del sector minero a medir y reducir las emisiones de gases de efecto invernadero producidas durante la extracción de minerales. Por tal motivo, los datos obtenidos en la presente investigación servirán de referencia para estudios tendientes a la mitigación de emisiones de carbono por parte del sector minero.

■ REFERENCIAS

- Ahirwal, J., & Maiti, S. K. (2017). Assessment of carbon sequestration potential of revegetated coal mine overburden dumps: A chronosequence study from dry tropical climate. *Journal of Environmental Management*, 201, 369–377. <https://doi.org/10.1016/j.jenvman.2017.07.003>
- Asociación Colombiana de Minería. (2022, March). *Cesar recibirá durante el 2022, más de 475 mil millones en regalías por actividad minera - Asociación Colombiana de Minería*. <https://acmineria.com.co/girasmineras-cesar/>
- Avtar, R., Komolafe, A. A., Kouser, A., Singh, D., Yunus, A. P., Dou, J., Kumar, P., Das Gupta, R., Johnson, B. A., Vuong, H., Minh, T., Kumar Aggarwal, A., & Kurniawan, A. (2020). Assessing sustainable development prospects through remote sensing: A review. *Remote Sensing Applications: Society and Environment*, 20, 100402. <https://doi.org/10.1016/j.rsase.2020.100402>
- Barliza, J. C., Peláez, J. D. L., & Campo, J. (2018). Recovery of biogeochemical processes in restored tropical dry forest on a coal mine spoil in La Guajira, Colombia. *Land Degradation & Development*, 29(9), 3174–3183. <https://onlinelibrary.wiley.com/doi/epdf/10.1002/ldr.3069>
- Das, R., & Maiti, S. K. (2016). Estimation of carbon sequestration in reclaimed coalmine degraded land dominated by *Albizia lebbek*, *Dalbergia sissoo* and *Bambusa arundinacea* plantation: a case study from Jharia Coalfields, India. *International Journal of Coal Science and Technology*, 3(2), 246–266. <https://doi.org/10.1007/S40789-016-0131-4/FIGURES/12>
- Dávila, E. Á., & Pérez, A. (2012). Bosques y minería responsable en Antioquia. *Revista Ambiental ÉOLO*, 12(17), 1–14. <http://revistaeolo.fconvida.org/index.php/eolo/article/view/818>
- Emmanuel, A. Y., Jerry, C. S., & Dzigbodi, D. A. (2018). Review of Environmental and Health Impacts of Mining in Ghana. *Journal of Health & Pollution*, 8(17), 52. <https://doi.org/10.5696/2156-9614-8.17.43>
- Escribano Rodríguez, J. A., Díaz-Ambrona, C. G. H., & Tarquis Alfonso, A. M. (2014). Selección de índices de vegetación para la estimación de la producción herbácea en dehesas. *Pastos*, 44(2), 6–18. <http://polired.upm.es/index.php/pastos/article/view/3108>
- FAO. (2011). *La gestión de los bosques ante el cambio climático*. <https://www.fao.org/3/i1960s/i1960s00.pdf>

- Fernández Álvarez, H., Álvarez-Narciandi, G., García-Fernández, M., Laviada, J., Álvarez López, Y., & Las-Heras Andrés, F. (2021). A Portable Electro-magnetic System Based on mm-Wave Radars and GNSS-RTK Solutions for 3D Scanning of Large Material Piles. *Sensors*, 21(3), 757. <https://doi.org/10.3390/S21030757>
- Gascon, F., Bouzinac, C., Thépaut, O., Jung, M., Francesconi, B., Louis, J., Lonjou, V., Lafrance, B., Massera, S., Gaudel-Vacaresse, A., Languille, F., Alhammoud, B., Viallefont, F., Pflug, B., Bieniarz, J., Clerc, S., Pessiot, L., Trémas, T., Cadau, E., ... Fernandez, V. (2017). Copernicus Sentinel-2A Calibration and Products Validation Status. *Remote Sensing*, 9(6), 584. <https://doi.org/10.3390/RS9060584>
- Gitelson, A. A., Stark, R., Grits, U., Rundquist, D., Kaufman, Y., & Derry, D. (2002). Vegetation and soil lines in visible spectral space: a concept and technique for remote estimation of vegetation fraction. *Int. j. Remote Sensing*, 23(13), 2537–2562. <https://doi.org/10.1080/01431160110107806>
- Guimar, M., Aes Da Silva, ~, Rosa Costa Muniz, A., Hoffmann, R., Carlos, A., Lisb, L., & Oa, ^ (2018). *Impact of greenhouse gases on surface coal mining in Brazil*. <https://doi.org/10.1016/j.jclepro.2018.05.076>
- Gutiérrez, G. A., Cotes, D. A., & Bastidas, M. J. (2018). Estudio de la Combustión Espontánea del Carbón durante el Acopio en la Mina El Hatillo, ubicada en Cesar, Colombia. *Información Tecnológica*, 29(6), 287–294. <https://doi.org/10.4067/S0718-07642018000600287>
- Hanke, J. E. (2006). Pronósticos en los negocios. *Pearson Educación de México, SA de CV*, 556P. https://books.google.com/books/about/Pronósticos_en_los_negocios_8a_ed.html?hl=es&id=JqqwtAEACAAJ
- Huete, A. R., Liu, H. Q., Batchily, K., & Van Leeuwen, W. (1997). A comparison of vegetation indices over a global set of TM images for EOS-MODIS. *Remote Sensing of Environment*, 59(3), 440–451. [https://doi.org/10.1016/S0034-4257\(96\)00112-5](https://doi.org/10.1016/S0034-4257(96)00112-5)
- Kaufman, Y. J., & Tanré, D. (1992). Atmospherically Resistant Vegetation Index (ARVI) for EOS-MODIS. *IEEE Transactions on Geoscience and Remote Sensing*, 30(2), 261–270. <https://doi.org/10.1109/36.134076>
- Laredo C., M. A., & Gómez S., J. (1982). Valor nutritivo de pastos tropicales, 4: Pasto Carimagua-1(*Andropogon gayanus* Kunth) anual y estacional. *Revista ICA*, 17(1), 29–36. <https://repository.agrosavia.co/handle/20.500.12324/35352>

- Marove, C. A., Sotozono, R., Tangviroon, P., Tabelin, C. B., & Igarashi, T. (2022). Assessment of soil, sediment and water contaminations around open-pit coal mines in Moatize, Tete province, Mozambique. *Environmental Advances*, 8. <https://doi.org/10.1016/J.ENVADV.2022.100215>
- Moral Peláez, I. (2016). Modelos de regresión: lineal simple y regresión logística. *SEDEN*, 195–214.
- Naciones Unidas. (1998). *PROTOCOLO DE KYOTO DE LA CONVENCIÓN MARCO DE LAS NACIONES UNIDAS SOBRE EL CAMBIO CLIMÁTICO*. <https://unfccc.int/resource/docs/convkp/kpspan.pdf>
- Pérez Gutiérrez, C., & Muñoz Nieto, Á. L. (coord. . (2002). *Teledetección : nociones y aplicaciones*. Universidad de Salamanca. <https://dialnet.unirioja.es/servlet/libro?codigo=5256>
- Pietrzykowski, M., & Daniels, W. L. (2014). Estimation of carbon sequestration by pine (*Pinus sylvestris* L.) ecosystems developed on reforested post-mining sites in Poland on differing mine soil substrates. *Ecological Engineering*, 73, 209–218. <https://doi.org/10.1016/J.ECOLENG.2014.09.058>
- Placek-Lapaj, A., Grobelak, A., Fijalkowski, K., Singh, B. R., & Almás, Á. R. (2019). Post – Mining soil as carbon storehouse under polish conditions. *Journal of Environmental Management*, 48, 307–314. <https://doi.org/doi.org/10.1016/j.jenvman.2019.03.005>
- Qi, J., Chehbouni, A., Huete, A. R., Kerr, Y. H., & Sorooshian, S. (1994). A modified soil adjusted vegetation index. *Remote Sensing of Environment*, 48(2), 119–126. [https://doi.org/10.1016/0034-4257\(94\)90134-1](https://doi.org/10.1016/0034-4257(94)90134-1)
- Rezaei, N., & Jabbari, P. (2022). K-nearest neighbors in R. *Immunoinformatics of Cancers*, 181–190. <https://doi.org/10.1016/B978-0-12-822400-7.00006-3>
- Rodrigo, J. A. (2020, October). *Random Forest con Python*. Association for Computing Machinery. <https://doi.org/10.1145/2939672.2939785>
- Rondeaux, G., Steven, M., & Baret, F. (1996). Optimization of soil-adjusted vegetation indices. *Remote Sensing of Environment*, 55(2), 95–107. [https://doi.org/10.1016/0034-4257\(95\)00186-7](https://doi.org/10.1016/0034-4257(95)00186-7)
- Santiago, M. M., LastNameAlves Da Silva, H., Domiciano, J., Tiago, G., & LastNameLastNameDe Oliveira, H. (2009). Análise da Cobertura Vegetal Através dos Índices de Vegetação (NDVI, SAVI e IAF) no Entorno da Barragem do Botafogo-PE. *XIV Simpósio Brasileiro de Sensoriamento Remoto*, 3003–3009.

- Shackelford, N., Miller, B. P., & Erickson, T. E. (2018). Restoration of Open-Cut Mining in Semi-Arid Systems: A Synthesis of Long-Term Monitoring Data and Implications for Management. *Land Degradation & Development*, 29(4), 994–1004. <https://doi.org/10.1002/LDR.2746>
- Song, C., Chen, J. M., Hwang, T., Hakkenberg, C., & Li, J. (2015). Ecological characterization of vegetation using multisensor remote sensing in the solar reflective spectrum. En Prasad Thenkabail. In *Land Resources Monitoring, Modeling, and Mapping with Remote Sensing* (pp. 533–575). United States: Geological Survey. <https://doi.org/10.1201/b19322>
- Suh, J., Kim, S. M., Yi, H., & Choi, Y. (2017). An Overview of GIS-Based Modeling and Assessment of Mining-Induced Hazards: Soil, Water, and Forest. *International Journal of Environmental Research and Public Health* 2017, 14(12), 1463. <https://doi.org/10.3390/IJERPH14121463>
- Tanré, D., Holben, B. N., & Kaufman, Y. J. (1992). Atmospheric Correction Algorithm for NOAA-AVHRR Products: Theory and Application. *IEEE Transactions on Geoscience and Remote Sensing*, 30(2), 231–248. <https://doi.org/10.1109/36.134074>
- Toosi, A., Javan, F. D., Samadzadegan, F., Mehravar, S., Kurban, A., & Azadi, H. (2022). Citrus orchard mapping in Juybar, Iran: Analysis of NDVI time series and feature fusion of multi-source satellite imageries. *Ecological Informatics*, 70, 101733. <https://doi.org/10.1016/J.ECOINF.2022.101733>
- Tripathi, N., Singh, R. S., & Nathanail, P. (2013). Mine spoil acts as a sink of carbon dioxide in Indian dry tropical environment. *Science of the Total Environment*, 468, 1162–1171. <https://doi.org/10.1016/j.scitotenv.2013.09.024>
- Vapnik, V. (1998). The Support Vector Method of Function Estimation. In: Suykens, J.A.K., Vandewalle, J. (Eds), *Nonlinear Modeling* (pp. 55-85). Springer, Boston, MA, https://doi.org/10.1007/978-1-4615-5703-6_3
- Xiao, J., Chevallier, F., Gomez, C., Guanter, L., Hicke, J. A., Huete, A. R., Ichii, K., Ni, W., Pang, Y., Rahman, A. F., Sun, G., Yuan, W., Zhang, L., & Zhang, X. (2019). Remote sensing of the terrestrial carbon cycle: A review of advances over 50 years. *Remote Sensing of Environment*, 233, 111383. <https://doi.org/10.1016/J.RSE.2019.111383>

■ ANEXOS

Anexo 1. Algoritmo para regresión por mínimos cuadrados ordinarios

```
# Importing the libraries
import numpy as np
import pandas as pd
import requests, io
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
# scikit-learn modules
from sklearn.model_selection import train_test_split
from sklearn.metrics import mean_squared_error
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score
from sklearn import metrics
#Load data
df = pd.read_table('C:/Users/USER/Documents/jupyter/BDD20/20.txt')
df.head()
#Load info data
(df.info())
#define the target(y) and learning variable(x)
ARVI = df.iloc[:, 0]
AGB = df.iloc[:, -1]
#Splitting the dataset
```

```

ARVI_train, ARVI_test, AGB_train, AGB_test = train_test_split(ARVI, AGB)
#OLS model
olsreg = LinearRegression()
#Fitting the model
olsreg.fit(ARVI_train.values.reshape(-1,1), AGB_train.values)
ols = olsreg.fit(ARVI_train.values.reshape(-1,1), AGB_train.values)
#Prediction on the testing data
y_pred = olsreg.predict(ARVI_test.values.reshape(-1,1))
y_pred
#Create first simple correlation matrix
corrMatrix = df.corr()
sns.heatmap(corrMatrix, annot=True)
plt.show()
#Draw training, testing and prediction data
plt.scatter(ARVI_train, AGB_train, label='Training Data', s=60, c='blue')
plt.scatter(ARVI_test, AGB_test, label='Testing Data', s=60, c='green')
plt.legend()
plt.plot(ARVI_train.values.reshape(-1,1), y_pred, color='red')
plt.show()
#Regression model evaluation
print ("Mean Squared Error:" , metrics.mean_squared_error(AGB_test.values.
reshape(-1,1), y_pred))
print ("Root Mean Squared Error:" , np.sqrt(metrics.mean_squared_error(A-
GB_test.values.reshape(-1,1), y_pred)))
print('Coefficient of determination: %.2f%' r2_score(AGB_test.values.resha-
pe(-1,1), y_pred))
#Raster generation of the model
inpRaster = 'C:/Users/USER/Documents/jupyter/BDD20/ARVI.tif' #define path
to the input raster (on which we will make predic

```

```
outraster = 'C:/Users/USER/Documents/jupyter/BDD2o/AGB_OLS_ARVI.tif'
#spath to the output raster—where it will be saved

df = pd.read_table('C:/Users/USER/Documents/jupyter/BDD2o/2o.txt') #apath
to the dataset with the training data

data = df.iloc[:, 0].values #List of predictors
label = df.iloc[:, 1].values

ds = gdal.Open(inpRaster, gdal.GA_ReadOnly) #with gdal module we will open
the input raster

rows = ds.RasterYSize
cols = ds.RasterXSize
bands = ds.RasterCount
geo_transform = ds.GetGeoTransform()
projection = ds.GetProjectionRef()
array = ds.ReadAsArray()
ds = None
array = np.stack(array,axis=2)
array = np.reshape(array, [rows*cols,bands])
test = pd.DataFrame(array, dtype='float32')
outRaster = 'C:/Users/USER/Documents/jupyter/BDD2o/AGB_OLS_ARVI.tif'
```

Anexo 2. Algoritmo para regresión por K-vecinos más cercanos

```
# Importing the libraries
import numpy as np
import pandas as pd
import requests, io
import matplotlib.pyplot as plt
import seaborn as sns
```

```
%matplotlib inline
# scikit-learn modules
from sklearn.model_selection import train_test_split
from sklearn.metrics import mean_squared_error
from sklearn import preprocessing
from sklearn.metrics import mean_squared_error, r2_score
from sklearn import metrics
from sklearn.neighbors import KNeighborsRegressor
from sklearn import neighbors
from sklearn.metrics import mean_squared_error
from yellowbrick.regressor import ResidualsPlot
#Load data
df = pd.read_table('C:/Users/USER/Documents/jupyter/BDD2o/2o.txt')
df.head()
#Load info data
(df.info())
#define the target(y) and learning ARVlable(x)
ARVI = df.iloc[:, 0]
AGB = df.iloc[:, -1]
#Splitting the dataset
ARVI_train, ARVI_test, AGB_train, AGB_test = train_test_split(ARVI, AGB, test_
size = 0.20, random_state = 0)
#KNN model
knnreg = KNeighborsRegressor(n_neighbors=2)
#Fitting the model
knnreg.fit(ARVI_train.values.reshape(-1,1), AGB_train)
mknn = knnreg.fit(ARVI_train.values.reshape(-1,1), AGB_train)
#Prediction on the testing data
AGB_pred = knnreg.predict((ARVI_train.values.reshape(-1,1)))
```

```

AGB_pred
#Elbow method RMSE vs K Value
rmse_val = []
for K in range(7):
    K = K+1
    modelknn = neighbors.KNeighborsRegressor(n_neighbors = K)
    modelknn.fit(ARVI_train.values.reshape(-1,1), AGB_train)
    pred=modelknn.predict(ARVI_test.values.reshape(-1,1))
    error = sqrt(mean_squared_error(AGB_test,pred))
    rmse_val.append(error) #store rmse values
    print('Valor de RMSE para k= ', K, 'es:', error)
#Draw training, testing and prediction data
plt.scatter(AGB_train, AGB_PRED_train, label='Training Data', s=60, c='blue')
plt.legend()
plt.plot(AGB_train.values.reshape(-1,1), y_pred, color='red')
plt.show()
#Regression model evaluation
print ("Mean Squared Error:" , metrics.mean_squared_error(AGB_train.values.
reshape(-1,1), AGB_pred))
print ("Root Mean Squared Error:" , np.sqrt(metrics.mean_squared_error(A-
GB_train.values.reshape(-1,1), AGB_pred)))
print('Coefficient of determination: %.2f%' r2_score(AGB_train.values.resha-
pe(-1,1), AGB_pred))
#Raster generation of the model
from osgeo import gdal
inpRaster = 'C:/Users/USER/Documents/jupyter/BDD2o/ARVI.tif'
outraster = 'C:/Users/USER/Documents/jupyter/BDD2o/AGB_KNN_ARVI.tif'
df = pd.read_table('C:/Users/USER/Documents/jupyter/BDD2o/2o.txt')
data = df.iloc[:, 0].values

```

```
label = df.iloc[:, 1].values
ds = gdal.Open(inpRaster, gdal.GA_ReadOnly)
rows = ds.RasterYSize
cols = ds.RasterXSize
bands = ds.RasterCount
geo_transform = ds.GetGeoTransform()
projection = ds.GetProjectionRef()
array = ds.ReadAsArray()
ds = None
array = np.stack(array,axis=2)
array = np.reshape(array, [rows*cols,bands])
test = pd.DataFrame(array, dtype='float32')
outRaster = 'C:/Users/USER/Documents/jupyter/BDD2o/AGB_KNN_ARVI.tif'
```

Anexo 3. Algoritmo para regresión por máquina vectorial de soporte

```
# Importing the libraries
import numpy as np
import pandas as pd
import requests, io
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
# scikit-learn modules
from sklearn.model_selection import train_test_split
from sklearn.metrics import mean_squared_error
from sklearn.svm import SVR
```

```
from sklearn.metrics import mean_squared_error, r2_score
from sklearn import metrics
from yellowbrick.regressor import ResidualsPlot
from sklearn.model_selection import cross_val_score
from sklearn.model_selection import GridSearchCV
#Load data
df = pd.read_table('C:/Users/USER/Documents/jupyter/BDD2o/2o.txt')
df.head()
#Load info data
(df.info())
#define the target(y) and learning MSAVI2able(x)
MSAVI2 = df.iloc[:, 2]
AGB = df.iloc[:, -1]
#Splitting the dataset
rand_state=0
MSAVI2_train, MSAVI2_test, AGB_train, AGB_test = train_test_split(MSAVI2,
AGB, test_size = 0.20, random_state = rand_state)
#SVR model
svrreg = SVR()
#Fitting the model
svrreg.fit(MSAVI2_train.values.reshape(-1,1), AGB_train)
msvr = svrreg.fit(MSAVI2_train.values.reshape(-1,1), AGB_train)
#Prediction on the testing data
AGB_train_pred = svrreg.predict((MSAVI2_train.values.reshape(-1,1)))
AGB_train_pred
#Regression model evaluation
print ("Mean Squared Error:", metrics.mean_squared_error(AGB_train.values.
reshape(-1,1), AGB_pred))
```

```

print ("Root Mean Squared Error:" , np.sqrt(metrics.mean_squared_error(A-
GB_train.values.reshape(-1,1), AGB_pred)))

print('Coefficient of determination: %.2f%' r2_score(AGB_train.values.reshape(-1,1), AGB_pred))

#Tuning hyperparameters

my_param_grid = {'C': [1,10,100], 'gamma': [1,0.1,0.01], 'kernel': ['rbf']}

#Import gridsearchCV

GridSearchCV(estimator=SVR(),param_grid= my_param_grid, refit = True, ver-
bose=3, cv=5)

#Define grid

grid = GridSearchCV(estimator=SVR(),param_grid= my_param_grid, refit =
True, verbose=2, cv=5)

#Iteration for parameter adjustment

grid.fit(MSAVI2_train.values.reshape(-1,1), AGB_train)

#the best parameter

grid.best_params_

#the best estimator

SVR(C=100, gamma=1)

#Optimized prediction data

AGB_train_pred_optimized = grid.predict(MSAVI2_train.values.reshape(-1,1))

AGB_train_pred_optimized

#Regression model evaluation

print ("Mean Squared Error:" , metrics.mean_squared_error(AGB_train.values.
reshape(-1,1), AGB_train_pred_optimized))

print ("Root Mean Squared Error:" , np.sqrt(metrics.mean_squared_error(A-
GB_train.values.reshape(-1,1), AGB_train_pred_optimized)))

print('Coefficient of determination: %.2f%' r2_score(AGB_train.values.reshape(-1,1), AGB_train_pred_optimized))

#Draw training, testing and prediction data

plt.scatter(AGB_train, AGB_train_pred_optimized_train, label='Training Data',
s=60, c='blue')

```

```
plt.legend()
plt.plot(AGB_train.values.reshape(-1,1), AGB_train_pred_optimized, color='red')
plt.show()
#Raster generation of the model
from osgeo import gdal
inpRaster = 'C:/Users/USER/Documents/jupyter/BDD2o/MSAVI2.tif'
outraster = 'C:/Users/USER/Documents/jupyter/BDD2o/AGB_SVM_MSAVI2.tif'
df = pd.read_table('C:/Users/USER/Documents/jupyter/BDD2o/2o.txt')
data = df.iloc[:, 0].values
label = df.iloc[:, 1].values
ds = gdal.Open(inpRaster, gdal.GA_ReadOnly)
rows = ds.RasterYSize
cols = ds.RasterXSize
bands = ds.RasterCount
geo_transform = ds.GetGeoTransform()
projection = ds.GetProjectionRef()
array = ds.ReadAsArray()
ds = None
array = np.stack(array,axis=2)
array = np.reshape(array, [rows*cols,bands])
test = pd.DataFrame(array, dtype='float32')
outRaster = 'C:/Users/USER/Documents/jupyter/BDD2o/AGB_SVM_MSAVI2.tif'
```

Anexo 4. Algoritmo para regresión por Random Forest

```
# Importing the libraries
import numpy as np
```

```
import pandas as pd
import requests, io
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
# scikit-learn modules
from sklearn.model_selection import train_test_split
from sklearn.metrics import mean_squared_error
from sklearn.metrics import mean_squared_error, r2_score
from sklearn import metrics
from sklearn.ensemble import RandomForestRegressor
from sklearn.model_selection import train_test_split
from yellowbrick.regressor import ResidualsPlot
from sklearn.model_selection import cross_val_score
from sklearn.model_selection import GridSearchCV
#Load data
df = pd.read_table('C:/Users/USER/Documents/jupyter/BDD2o/2o.txt')
df.head()
#Load info data
(df.info())
#define the target(y) and learning IVable(x)
IV = df.drop(columns = "MUESTRA")
AGB = df['MUESTRA']
#Splitting the dataset
rand_state=0
IV_train, IV_test, AGB_train, AGB_test = train_test_split(IV, AGB, test_size =
0.20, random_state = rand_state)
#Random forest model
```

```

rfreg = RandomForestRegressor(n_estimators = 100, max_depth= None, n_
jobs= -1, random_state = 0)

#Fitting the model
rfreg.fit(IV_train, AGB_train)

rf = rfreg.fit(IV_train, AGB_train)

#Prediction on the testing data
AGB_train_pred = rfreg.predict((IV_train))

AGB_train_pred

#Regression model evaluation
print ("Mean Squared Error:" , metrics.mean_squared_error(AGB_train.values.
reshape(-1,1), AGB_pred))

print ("Root Mean Squared Error:" , np.sqrt(metrics.mean_squared_error(A-
GB_train.values.reshape(-1,1), AGB_pred)))

print("Coefficient of determination: %.2f% r2_score(AGB_train.values.resha-
pe(-1,1), AGB_pred))

#Importance of predictors in the model
importancia_predictores = pd.DataFrame(
    {'predictor': df.drop('MUESTRA', axis = 1).columns,
    'importancia': rfreg.feature_importances_}
)

print("Importancia de los predictores en el modelo")
print("-----")
importancia_predictores.sort_values('importancia', ascending=False)

#Feature importance
plt.figure(figsize=(10,6))
plt.title('Feature Importance')
sns.barplot(y='predictor', x='importancia', data=importancia_predictores)
plt.show()

#Tuning hyperparameters

```

```

my_param_grid = {'n_estimators': [10,100,500], 'max_features':['sqrt','log2'],
'max_depth':[5,10,20]}

#Import gridserchCV

GridSearchCV(estimator=RandomForestRegressor(random_state=rand_state),param_grid= my_param_grid, refit = True, verbose=2, cv=5)

grid = GridSearchCV(estimator=RandomForestRegressor(random_state=rand_state),param_grid= my_param_grid, refit = True, verbose=2, cv=5)

#Iteration for parameter adjustment

grid.fit(IV_train,AGB_train)

#The best parameter

grid.best_params_

#The best estimator

grid.best_estimator_

#Optimized prediction data

AGB_train_pred_optimized = grid.predict((IV_train))

AGB_train_pred_optimized

#Iteration for parameter adjustment

grid.fit(IV_train.values.reshape(-1,1), AGB_train)

#the best parameter

grid.best_params_

#the best estimator

SVR(C=100, gamma=1)

#Optimized prediction data

AGB_train_pred_optimized = grid.predict(IV_train.values.reshape(-1,1))

AGB_train_pred_optimized

#Regression model evaluation

print ("Mean Squared Error:" , metrics.mean_squared_error(AGB_train, AGB_train_pred_optimized))

print ("Root Mean Squared Error:" , np.sqrt(metrics.mean_squared_error(AGB_train, AGB_train_pred_optimized)))

```

```

print('Coefficient of determination: %.2f' % r2_score(AGB_train, AGB_train_
pred_optimized))

#Regression model evaluation

print ("Mean Squared Error:" , metrics.mean_squared_error(AGB_train.values.
reshape(-1,1), AGB_train_pred_optimized))

print ("Root Mean Squared Error:" , np.sqrt(metrics.mean_squared_error(A-
GB_train.values.reshape(-1,1), AGB_train_pred_optimized)))

print('Coefficient of determination: %.2f' % r2_score(AGB_train.values.reshape(-1,1), AGB_train_pred_optimized))

#Draw training, testing and prediction data

plt.scatter(AGB_train, AGB_train_pred_optimized_train, label='Training Data',
s=60, c='blue')

plt.legend()

plt.plot(AGB_train.values.reshape(-1,1), AGB_train_pred_optimized, color='red')

plt.show()

#Raster generation of the model

from osgeo import gdal

inpRaster = 'C:/Users/USER/Documents/jupyter/BDD2o/MSAVI2.tif'
outraster = 'C:/Users/USER/Documents/jupyter/BDD2o/AGB_RF_MSAVI2.tif'
df = pd.read_table('C:/Users/USER/Documents/jupyter/BDD2o/2o.txt')
data = df.iloc[:, 0].values
label = df.iloc[:, 1].values

ds = gdal.Open(inpRaster, gdal.GA_ReadOnly)
rows = ds.RasterYSize
cols = ds.RasterXSize
bands = ds.RasterCount
geo_transform = ds.GetGeoTransform()
projection = ds.GetProjectionRef()
array = ds.ReadAsArray()
ds = None

```

```
array = np.stack(array,axis=2)
array = np.reshape(array, [rows*cols,bands])
test = pd.DataFrame(array, dtype='float32')
outRaster = 'C:/Users/USER/Documents/jupyter/BDD2o/AGB_RF_MSAV1z.tif'
```

