

Maestría en
Ciencia de Datos y Máquinas de Aprendizaje con
mención en Inteligencia Artificial

**Trabajo previo a la obtención de título de Magister en Ciencia de Datos
y Máquinas de Aprendizaje con mención en Inteligencia Artificial**

AUTOR/ES:

Daniel David Zambrano Andrade

David Alejandro Romero Vaca

María Gabriela Santillán Caizaguano

María José Chávez Villacrés

Miguel Ángel Recalde Chávez

TUTORES:

Alejandro Cortés López

Karla Estefanía Mora Cajas

TEMA:

MODELADO MULTITEMPORAL DE LA DEFORESTACIÓN EN EL CORREDOR
NORTE AMAZÓNICO DEL ECUADOR A PARTIR DE DATOS LANDSAT

CERTIFICACIÓN DE AUTORÍA

Nosotros, Daniel David Zambrano Andrade, David Alejandro Romero Vaca, María Gabriela Santillán Caizaguano, María José Chávez Villacrés, Miguel Ángel Recalde Chávez, declaramos bajo juramento que el trabajo aquí descrito es de nuestra autoría; que no ha sido presentado anteriormente para ningún grado o calificación profesional y que se ha consultado la bibliografía detallada.

Cedemos nuestros derechos de propiedad intelectual a la Universidad Internacional del Ecuador (UIDE), para que sea publicado y divulgado en internet, según lo establecido en la Ley de Propiedad Intelectual, su reglamento y demás disposiciones legales.



**Daniel David Zambrano
Andrade**



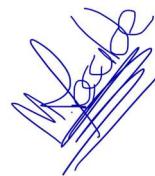
**David Alejandro Romero
Vaca**



**María Gabriela Santillán
Caizaguano**



**María José Chávez
Villacrés**



**Miguel Ángel Recalde
Chávez**

AUTORIZACIÓN DE DERECHOS DE PROPIEDAD INTELECTUAL

Nosotros, Daniel David Zambrano Andrade, David Alejandro Romero Vaca, María Gabriela Santillán Caizaguano, María José Chávez Villacrés, Miguel Ángel Recalde Chávez, en calidad de autores del trabajo de investigación titulado *Modelado multitemporal de la deforestación en el Corredor Norte Amazónico del Ecuador a partir de datos LANDSAT*, autorizamos a la Universidad Internacional del Ecuador (UIDE) para hacer uso de todos los contenidos que nos pertenecen o de parte de los que contiene esta obra, con fines estrictamente académicos o de investigación. Los derechos que como autores nos corresponden, lo establecido en los artículos 5, 6, 8, 19 y demás pertinentes de la Ley de Propiedad Intelectual y su Reglamento en Ecuador.

D. M. Quito, diciembre 2025



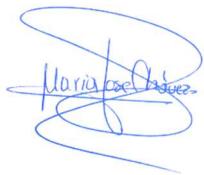
**Daniel David Zambrano
Andrade**



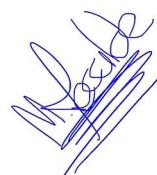
**David Alejandro Romero
Vaca**



**María Gabriela Santillán
Caizaguano**



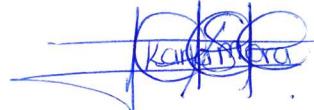
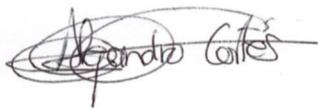
**María José Chávez
Villacrés**



**Miguel Ángel Recalde
Chávez**

APROBACIÓN DE DIRECCIÓN Y COORDINACIÓN DEL PROGRAMA

Nosotros, Alejandro Cortés Director EIG y Karla Mora Coordinadora Académica UIDE, declaramos que: Daniel David Zambrano Andrade, David Alejandro Romero Vaca, María Gabriela Santillán Caizaguano, María José Chávez Villacrés, Miguel Ángel Recalde Chávez son los autores exclusivos de la presente investigación y que ésta es original, auténtica y personal de ellos.



Alejandro Cortés

Director de la Maestría en Ciencias de Datos y
Máquinas de Aprendizaje con Mención en
Inteligencia Artificial

Karla Mora

Coordinadora de la Maestría en Ciencias
de Datos y Máquinas de Aprendizaje con
Mención en Inteligencia Artificial

DEDICATORIA

Dedico este trabajo a Gabriela Rojas y a mi hijo Martín, quienes representan mi mayor motivación y apoyo en cada etapa de mi vida, y a mi madre, cuyo sacrificio, constancia y amor han sido la base sobre la cual se ha construido mi desarrollo personal y profesional.

-Daniel David Zambrano Andrade

A mi esposa por su gran apoyo en este proceso de maestría, a mis padres por sus infaltables oraciones, a mi hermana que admiro y quiero mucho y finalmente, a mi Bebé que fue mi más lindo y hermoso milagro.

-David Alejandro Romero Vaca

Dedico esta tesis a mi familia, por su apoyo incondicional y constante a lo largo de cada etapa de mi formación académica y personal; y de manera muy especial a mi hijo, Santiago, quien con su amor, paciencia y comprensión se convirtió en mi mayor fuente de motivación para perseverar y no rendirme ante las dificultades. Este logro también le pertenece.

-María Gabriela Santillán Caizaguano

Dedico este trabajo a mi madre, cuya fuerza, inteligencia y amor permanecen vivos en cada paso de mi camino. Aunque ya no esté conmigo, su ejemplo me guía, me sostiene y da sentido a cada esfuerzo y a cada logro. Este trabajo es un reflejo de su legado y de todo lo que sembró en mí.

-María José Chávez Villacrés

A mis padres, Mayra Chávez y Miguel Recalde, pilares fundamentales de mi vida. cuya sabiduría y paciencia sembraron en mí el valor infinito del conocimiento y la perseverancia. Este logro es el fruto de la educación y los valores que, con tanto amor y sacrificio, me inculcaron. Les dedico cada página de este trabajo, como un testimonio de su invaluable legado. A toda mi familia, fuente constante de apoyo y razón de mi entrega.

-Miguel Ángel Recalde Chávez

AGRADECIMIENTOS

Expreso mi sincero agradecimiento a Gabriela Rojas por su apoyo constante, comprensión y acompañamiento durante todo el desarrollo de este trabajo, así como a mi hijo Martín, cuya presencia y motivación diaria dieron sentido a cada esfuerzo realizado. Agradezco también a los docentes y a la institución por los conocimientos y orientaciones brindadas a lo largo de mi formación académica, que hicieron posible la culminación de esta investigación.

-Daniel David Zambrano Andrade

Agradezco a Dios por sus muchas bendiciones hacia mi persona y familia, sé que nunca me abandona y me acompaña con su presencia e infaltable amor.

-David Alejandro Romero Vaca

Agradezco sinceramente a mi familia por su apoyo constante, comprensión y motivación a lo largo de todo el proceso de elaboración de esta tesis; de manera especial, a mi hijo, quien ha sido mi principal fuente de inspiración y me ha enseñado, a través de su amor, el verdadero significado de la perseverancia. Asimismo, agradezco a los docentes y autoridades de la maestría por sus valiosos conocimientos, orientación y acompañamiento académico, fundamentales para el desarrollo y culminación exitosa de este trabajo de investigación, así como al fortalecimiento de las competencias profesionales necesarias para una mejor proyección laboral y desempeño en el ámbito profesional.

-María Gabriela Santillán Caizaguano

Agradezco a la Universidad por la formación académica y el acompañamiento brindado a lo largo de este proceso. A mi familia, por creer en mí, apoyarme y celebrar cada paso de este

camino; en especial a mi padre, hermanos y sobrinos, quienes constituyen el motor de mi vida. De manera muy especial, a mi más grande amor, mi madre: este logro es suyo; sé que su corazón celebra con orgullo y alegría cada avance de mi camino. Finalmente, agradezco a todas las personas que me acompañan, confían en mí y me animan a seguir adelante con determinación. Este paso ha sido posible gracias al apoyo de todas y todos.

María José Chávez Villacrés

Deseo expresar mi más sincero y profundo agradecimiento a mis padres, Mayra y Miguel. Sin su ejemplo de resiliencia, su apoyo incondicional y su esfuerzo para brindarme una educación de calidad, a todas las personas que, de una u otra manera, contribuyeron a la realización de este trabajo. Este título no es solo un logro personal, sino un homenaje colectivo al apoyo recibido.

-Miguel Ángel Recalde Chávez

RESUMEN

La Amazonía ecuatoriana experimenta procesos acelerados de transformación territorial que amenazan su integridad ecológica y la provisión de servicios ecosistémicos. El Corredor Norte Amazónico (CNA) concentra gran parte de esta dinámica, motivando la necesidad de sistemas de monitoreo y proyección de la pérdida forestal que informen la planificación ambiental. Este estudio analiza la deforestación cantonal en el CNA durante el período 2014-2024 y proyecta la evolución de la cobertura boscosa hasta 2030 mediante aprendizaje automático.

La investigación adoptó un enfoque cuantitativo, descriptivo y predictivo. A partir de imágenes LANDSAT se construyó una base de datos espaciotemporal de cobertura del suelo para 20 cantones de las provincias de Napo, Orellana, Pastaza y Sucumbíos. Se calcularon indicadores de área forestal, fragmentación y presión antrópica, que fueron integrados en ventanas multivariadas de cuatro años. Un modelo Random Forest Regressor, validado con esquema walk-forward, estimó la superficie boscosa anual; el desempeño se evaluó con MAE, RMSE y MAPE. Finalmente se desarrolló un dashboard interactivo para visualizar resultados históricos y futuros.

Entre 2014 y 2024 la pérdida acumulada ascendió a 42150 ha (tasa media 3 832 ha / año). Los cantones Shushufindi, Orellana y Tena presentaron las mayores disminuciones. El modelo alcanzó un MAE de 135 ha año y un MAPE del 4,7%, evidenciando mayor precisión en territorios con tendencias sostenidas. Las proyecciones 2025-2030 indican una aceleración leve (4150 ha/año) y sitúan a los mismos cantones como focos críticos. La proximidad a vías (< 5 km) y la presión antrópica acumulada explican 50% de la variabilidad espacial de la deforestación.

El estudio demuestra que la serie LANDSAT, combinada con aprendizaje automático y validación temporal rigurosa, permite anticipar la pérdida forestal a escala cantonal con alta

confiabilidad. El dashboard constituye una herramienta operativa para priorizar intervenciones de conservación y ordenamiento territorial en la Amazonía ecuatoriana.

Palabras Claves: Forestal, Aprendizaje Automático, Deforestación, Dashboard, Amazonía.

ABSTRACT

The Ecuadorian Amazon is undergoing rapid land-use change that threatens its ecological integrity and the provision of ecosystem services. The Northern Amazon Corridor (NAC) concentrates much of this dynamics, creating an urgent need for monitoring and forecasting systems that can guide environmental planning. This study analyses canton-level deforestation in the NAC for the period 2014–2024 and projects forest-cover change up to 2030 through machine learning.

A quantitative, descriptive and predictive approach was adopted. LANDSAT imagery was used to build a spatio-temporal land-cover database covering 20 cantons in the provinces of Napo, Orellana, Pastaza and Sucumbíos. Indicators of forest area, fragmentation and anthropogenic pressure were calculated and assembled into four-year multivariate sliding windows. A Random Forest Regressor, validated with a walk-forward scheme, estimated annual forest area; performance was assessed with MAE, RMSE and MAPE. An interactive dashboard was finally developed to visualise historical and future results.

Between 2014 and 2024 cumulative forest loss reached 42,150 ha (mean annual rate 3,832 ha yr⁻¹). The cantons of Shushufindi, Orellana and Tena showed the largest reductions. The model achieved an MAE of 135 ha yr⁻¹ and a MAPE of 4.7 %, exhibiting higher accuracy in areas with sustained deforestation trends. Projections for 2025–2030 indicate a slight acceleration (4,150 ha yr⁻¹) and identify the same cantons as critical hotspots. Proximity to roads (< 5 km) and cumulative anthropogenic pressure explain 50 % of the spatial variability in deforestation.

The study demonstrates that the LANDSAT archive, coupled with machine learning and rigorous temporal validation, can reliably forecast canton-level forest loss. The dashboard

provides an operational tool for prioritising conservation and land-use planning interventions in the Ecuadorian Amazon.

Keywords: Forest, Machine Learning, Deforestation, Dashboard, Amazon.

Tabla de Contenidos

CAPÍTULO I	1
1. Introducción	1
1.1. Definición del Proyecto	3
1.2. Justificación e Importancia del Trabajo de Investigación	4
1.3. Alcance	5
1.4. Objetivos.....	6
1.4.1. Objetivo General.....	6
1.4.2. Objetivo Específico.....	6
CAPÍTULO II.	7
2. Revisión de Literatura.....	7
2.1. Estado del Arte.....	7
2.2. Marco Teórico.....	8
2.2.1. Teorías del Cambio de Uso del Suelo y Dinámica de la Deforestación	9
2.2.2. Ecología del Paisaje y Fragmentación Forestal	10
2.3. Corredor Norte Amazónico	10
2.3.1. Historia Reciente de Ocupación y Configuración Territorial	11
2.3.2. Tendencias de Deforestación 2001–2025	12
2.3.3. Estacionalidad y Patrones Temporales.....	13
2.3.4. Actores Territoriales y Gobernanza	13
2.3.5. Socioeconomía de la Deforestación: Teorías de Rentabilidad y Subsistencia	
15	
2.4. Teledetección Aplicada al Monitoreo Forestal	15
2.4.1. Sensores de Teledetección: Landsat y Sentinel	16
2.4.2. Sistemas de Alerta de Deforestación	19
2.4.3. Mapas Históricos y Series Temporales	19

2.5. Aprendizaje Automático para Predicción Ambiental Aplicado a Fenómenos Geoespaciales	20
2.5.1. Random Forest.....	21
2.5.2. Series Temporales y Teoría de la Predicción en Sistemas Complejos	21
2.5.3. Ecología del Paisaje y Fragmentación Forestal	22
2.6. Marco Normativo y Políticas Públicas	23
2.7. Python.....	23
2.7.1. Elementos Predefinidos	23
2.7.2. Estructuras de Control y Funciones	24
2.7.3. Librerías de Python.....	24
2.8. Aprendizaje Supervisado	25
2.9. Aprendizaje No Supervisado	25
2.10. Metodología KDD	25
2.11. Series Temporales.....	26
2.12. Métricas de Evaluación	26
2.12.1. Error Absoluto Medio (MAE)	27
2.12.2. Raíz del Error Cuadrático Medio (RMSE)	27
2.12.3. Coeficiente de Determinación (R ²).....	27
2.12.4. Error Porcentual Absoluto Medio (MAPE)	28
CAPÍTULO III.....	29
3. Desarrollo.....	29
3.1. Metodología.....	29
3.2. Síntesis Metodológica.....	29
3.3. Recolección y Descripción de los Datos	30
3.3.1. Fuente de Datos Satelitales	30

3.3.2. Características Espaciales de los Datos Raster	30
3.3.3. Delimitación del Área de Estudio	31
3.3.4. Procesamiento Espacial en QGIS	32
3.3.5. Estructuración de la Base de Datos Tabular	32
3.3.6. Derivación de Indicadores de Deforestación	34
3.3.7. Ventajas del Enfoque Adoptado	34
3.4. Preparación y Transformación de Datos.....	35
3.4.1. Análisis Exploratorio de Datos (EDA)	38
3.4.2. Normalización Temporal de la Base de Datos	38
3.4.3. Filtrado de Clases Relevantes para Deforestación.....	39
3.4.4. Construcción de Variables Derivadas	40
3.4.5. Agregación por Unidad Territorial.....	40
3.4.6. Análisis del Desbalance de Clases.....	41
3.4.7. Preparación de Insumos para Análisis Temporal	41
3.5. Cálculo del Área Forestal y Variables Derivadas	41
3.5.1. Área Forestal Total.....	41
3.5.2. Variables Porcentuales	42
3.5.3. Indicadores de Cambio y Presión Antrópica.....	42
3.6. Construcción de Series Temporales y Ventanas Multivariadas	43
3.7. Estrategia de Validación Walk-Forward	44
3.8. Modelado y Optimización	46
3.8.1. Modelo Random Forest	46
3.8.2. Estandarización.....	48
3.8.3. Evaluación del Desempeño.....	48
3.9. Entrenamiento Final y Predicción Multianual (2025–2030)	50

3.10. Visualización y Almacenamiento de Resultados	53
3.10.1. Arquitectura General del Sistema	55
CAPÍTULO IV	59
4. Análisis de Resultados	59
4.1. Pruebas de Concepto	59
4.2. Análisis de Resultados	60
4.2.1. Dinámica Histórica de la Deforestación (2014–2024)	60
4.2.2. Desempeño del Modelo Predictivo	60
4.2.3. Proyecciones de Deforestación (2025–2030)	61
4.2.4. Análisis Espacial y de Factores Asociados	61
4.2.5. Dashboard Interactivo	61
CAPÍTULO V	62
5. Conclusiones y Recomendaciones	62
5.1. Conclusiones	62
5.2. Recomendaciones	63
5.2.1. Para la Gestión Ambiental y la Política Pública	63
5.2.2. Trabajo Futuro	63
5.2.3. Para el desarrollo tecnológico	63
REFERENCIAS	64
6. ANEXOS	70
6.1. Lista de Códigos	70
6.1.1. Listado de Códigos en Github	70
6.2. Manual de generación del Dataset	70

Lista de Tablas

Tabla 1 <i>Librerías utilizadas y su función metodológica</i> 9
Tabla 2 <i>Funciones auxiliares implementadas en el flujo de procesamiento y modelado</i> 51
Tabla 3 <i>Librerías utilizadas en desarrollo de dashboard</i> 55

Lista de Figuras

Figura 1 Imagen satelital de la deforestación en la industria petrolera.....	12
Figura 2 Actores Territoriales y Gobernanza	14
Figura 3 Imagen Satelital de la misión LANDSAT 8	16
Figura 4 Imagen Satelital misión SENTINEL-2.....	17
Figura 5 Imagen Satelital misión SENTINEL-1.....	18
Figura 6 Formación Forestal Año 2001 - 2024	20
Figura 7 Código - Resolución Espacial	31
Figura 8 Resultado QGIS.....	32
Figura 9 Estructura Tabular	33
Figura 10 Patrón de colores de los indicadores de deforestación	34
Figura 11 Normalización de datos	39
Figura 12 Variables Derivadas	40
Figura 13 Área Forestal en Hectáreas	42
Figura 14 Variables Porcentuales	42
Figura 15 Construcción de ventanas Multivariadas	44
Figura 16 Validación Walk-Forward.....	45
Figura 17 Esquema del Enfoque Walk-Forward con Ventana de Entrenamiento Expansiva.....	46
Figura 18 Modelo Random Forest	48
Figura 19 Estandarización.....	48
Figura 20 Evaluación de Desempeño	50

Figura 21 <i>Predicción Multianual</i>	51
Figura 22 <i>Dashboard</i>	54
Figura 23 <i>Arquitectura del Dashboard</i>	55
Figura 24 <i>Datos Detallados del Dashboard</i>	56

Capítulo I.

1. Introducción

La Amazonía es uno de los sistemas ecológicos más complejos y biodiversos del planeta y a la vez uno de los más amenazados por las actividades humanas. En las últimas décadas la deforestación en la amazonia se ha incrementado debido a la expansión agrícola, la expansión poblacional, la explotación de hidrocarburos (Armenteras et al., 2019; Clerici et al., 2019).

Dentro del contexto amazónico, el Corredor Norte Amazónico se ha convertido en una de las zonas de mayor presión antrópica del Ecuador. La apertura de vías, el establecimiento de actividades humanas y la expansión agrícola y ganadera han generado cambios significativos en la estructura de los bosques, convirtiendo al corredor norte en prioridad para el monitoreo ambiental (Sierra, 2013; Clerici et al., 2019). La compleja relación entre la economía y la sociedad de esta región ha dado lugar a avances significativos de deforestación y cambio de uso de suelo por lo que requieren ser analizados con herramientas capaces de capturar los cambios en múltiples ventanas temporales.

En este escenario, la detección satelital se ha convertido en una herramienta clave para estudiar las dinámicas y avances de la deforestación gracias a su capacidad de cubrir grandes extensiones territoriales con una adecuada temporalización para el análisis anual. Entre los sistemas disponibles encontramos la misión LANDSAT destacada por su continuidad histórica, más de cuarenta años de observaciones, con atributos radiométricos y resolución espacial de 30 metros lo cual permite estudiar los avances de la deforestación en la amazonia. (Wulder et al., 2019; Roy et al., 2014).

A pesar de la utilidad comprobada de LANDSAT, su aplicación en la Amazonía enfrenta desafíos considerables debido a la niebla persistente, las variaciones atmosféricas y la diversidad

de la flora y fauna que pueden limitar la disponibilidad de imágenes utilizables (Hansen et al., 2016). Sin embargo, la continuidad temporal del programa LANDSAT ofrece suficientes imágenes por año para reconstruir los datos de cobertura forestal. Esta característica convierte a la misión LANDSAT en un recurso importante para analizar la evolución de la deforestación del Corredor Norte Amazónico.

En este contexto, la presente investigación aborda el análisis de la deforestación en el Corredor Norte Amazónico del Ecuador mediante el uso de datos satelitales LANDSAT y un enfoque de análisis temporal a escala cantonal para el período 2014–2024. A partir de la construcción de series históricas del área forestal y la generación de variables espaciales. Se implementa un modelo de predicción basado en aprendizaje supervisado, específicamente Random Forest Regressor, entrenado bajo un esquema de validación walk-forward y ventanas temporales multivariadas, lo que permite evaluar de manera realista su capacidad predictiva en el tiempo.

El modelo integra información histórica de la amazonía, tendencias de cambio, permitiendo no solo estimar la deforestación anual sino también proyectar escenarios futuros de deforestación para el período 2025–2030. La evaluación del desempeño se realiza mediante métricas robustas como MAE, RMSE y R^2 , priorizando medidas basadas en error absoluto y relativo, adecuadas para series temporales con baja variabilidad interanual.

Como resultado final de este trabajo se desarrolla un dashboard interactivo que consolida el análisis histórico y las proyecciones futuras de la deforestación, permitiendo la exploración dinámica de los resultados por provincia, cantón y año. Este dashboard facilita la visualización de tendencias temporales, transformando los resultados del modelado en una herramienta para

apoyar en la toma de decisiones en conservación, planificación territorial y gestión sostenible del bosque amazónico.

1.1. Definición del Proyecto

El presente proyecto se orienta al desarrollo de un sistema de análisis anual de la deforestación en el Corredor Norte Amazónico del Ecuador mediante el uso de datos de la misión LANDSAT. Su propósito es identificar y cuantificar la pérdida de bosques a nivel de píxel a lo largo de una serie temporal anual, de manera que se generen mapas y métricas robustas sobre la dinámica espacial de la deforestación en la región amazónica (Armenteras et al., 2022).

La propuesta metodológica integra procedimientos estandarizados para el procesamiento de los datos satelitales (Chander et al., 2009). La detección de anomalías se realizará mediante comparaciones temporales entre los datos ofrecidos de las imágenes LANDSAT de diferentes años y la delimitación de los cantones del Corredor Norte (Wulder et al., 2019).

El proyecto desarrolla un enfoque integral para el análisis y la predicción de la deforestación en el Corredor Norte Amazónico combinando datos satelitales LANDSAT, análisis temporal y modelos de aprendizaje automático supervisado. La implementación de un modelo Random Forest entrenado mediante validación walk-forward y ventanas temporales multivariadas permite estimar con robustez la evolución anual de la cobertura forestal y generar proyecciones futuras para el período 2025–2030, considerando la dinámica histórica del bosque.

Finalmente, los resultados del modelado y del análisis multitemporal se integran en un dashboard interactivo que consolida la información histórica y las predicciones futuras a escala provincial y cantonal. Este producto final transforma los resultados técnicos en una herramienta de visualización y apoyo para la toma de decisiones de la gestión bosque amazónico.

1.2. Justificación e Importancia del Trabajo de Investigación

La deforestación en la Amazonía constituye uno de los principales desafíos ambientales de la región, con impactos directos sobre la flora y fauna y el bienestar de las comunidades indígenas que dependen del bosque como fuente de sustento (Armenteras et al., 2019; Panfil et al., 2016). El Corredor Norte Amazónico es uno de los puntos de transformación más intensos, impulsado por la expansión agropecuaria, la apertura de vías, los asentamientos humanos y la actividad extractiva (Sierra, 2013; Clerici et al., 2019). Este proceso ha acelerado la pérdida de cobertura forestal afectando la conservación de especies y la funcionalidad ecológica del territorio.

El monitoreo de la deforestación se vuelve indispensable para comprender la magnitud del fenómeno llamado deforestación, sin embargo, los métodos tradicionales presentan limitaciones importantes relacionadas con los altos costos, la inaccesibilidad del terreno, especialmente en regiones remotas (Hansen et al., 2013). Ante ello, la detección satelital y en particular la misión LANDSAT se ha convertido como una herramienta técnica y científica fundamental para el análisis temporal de cambios en la cobertura forestal (Wulder et al., 2019).

En el contexto del Corredor Norte Amazónico, esta estrategia resulta valiosa debido a la necesidad de evaluar cambios anuales y comprender los patrones de deforestación en un territorio donde la presión antropogénica cambia frecuentemente (Pettorelli et al., 2005; Key et al., 2006).

La importancia del presente trabajo radica en ofrecer una evaluación rigurosa y basada en evidencia de la dinámica anual de la deforestación. Al generar métricas de cambio espacialmente que contribuye a cerrar brechas de información en torno a la magnitud de la deforestación y a fortalecer la toma de decisiones actualizadas (Clerici et al., 2019).

1.3. Alcance

El presente estudio abarca el análisis multitemporal y predictivo de la deforestación en el Corredor Norte Amazónico del Ecuador mediante el uso de datos satelitales provenientes de la misión LANDSAT. El trabajo se estructura bajo una metodología basada en ventanas temporales, considerando un período histórico de análisis entre 2014 y 2024 para la construcción de series temporales de la deforestación a escala cantonal y provincial. Se implementa un esquema de validación temporal tipo walk-forward, que permite evaluar el desempeño de los modelos predictivos utilizando únicamente información disponible.

El alcance se limita al territorio comprendido dentro del Corredor Norte Amazónico ecuatoriano, que incluye cantones de las provincias de Napo, Orellana y Sucumbíos. Esta región presenta condiciones complejas, como la elevada presión antrópica y persistente nubosidad, las cuales influyen en la disponibilidad y calidad de datos. Para garantizar la validez del análisis, se emplean únicamente datos de la misión LANDSAT que aseguran los datos anuales sin condiciones adversas (Chander et al., 2009; Roy et al., 2014).

El alcance comprende la generación de ventanas temporales de la deforestación, el cálculo de variables derivadas y la aplicación de modelos de aprendizaje automático supervisado, específicamente Random Forest Regressor, orientados a la predicción anual de la desforestación. Las proyecciones se extienden para el período 2025–2030, permitiendo la construcción de escenarios futuros de deforestación a partir de patrones históricos. El desempeño del modelo se evalúa mediante métricas de error absoluto y relativo, priorizando indicadores robustos frente a la baja variabilidad anual de las series ambientales.

Como parte del alcance aplicado, los resultados del análisis histórico y de las predicciones futuras se integran en un dashboard interactivo que permite la visualización

dinámica de la deforestación por provincia, cantón y año. Este producto final está orientado a facilitar la exploración de tendencias constituyéndose en una herramienta de apoyo para la toma de decisiones en planificación territorial y gestión ambiental.

Finalmente, el estudio no contempla el análisis de regeneración forestal, degradación leve del dosel, evaluación de impactos socioeconómicos ni la integración de sensores adicionales como radar SAR u otras plataformas satelitales. El alcance se limita al uso de la serie temporal LANDSAT y al análisis cuantitativo de la deforestación, garantizando una delimitación clara, coherente y metodológicamente controlada de este trabajo.

1.4. Objetivos

1.4.1. Objetivo General

Analizar la dinámica temporal de la deforestación en el Corredor Norte Amazónico del Ecuador mediante el uso de datos LANDSAT y modelos de aprendizaje automático, integrando los resultados en una herramienta de visualización interactiva.

1.4.2. Objetivo Específico

- Construir un conjunto de datos temporal de cobertura forestal a partir de imágenes satelitales LANDSAT para el período 2014–2024.
- Desarrollar un modelo predictivo basado en Random Forest para estimar la pérdida de cobertura forestal en el período 2025–2030 mediante validación temporal tipo walk-forward.
- Implementar un dashboard interactivo que visualice el análisis histórico y las predicciones de deforestación, facilitando la exploración espacial y temporal de los resultados.

Capítulo II.

2. Revisión de Literatura

2.1. Estado del Arte

El estado del arte en monitoreo de deforestación ha experimentado avances significativos en las dos últimas décadas, impulsados por el aumento de la disponibilidad de datos satelitales, el desarrollo de algoritmos de aprendizaje automático y la necesidad de contar con sistemas de alerta temprana más precisos en regiones tropicales. En este contexto, la detección y predicción de pérdida forestal ha evolucionado desde enfoques basados exclusivamente en imágenes hacia metodologías híbridas que integran sensores TM, ETM+ y OLI que proporcionan imágenes con resolución espacial de 30 metros (Roy et al., 2014; Wulder et al., 2019).

Uno de los hitos recientes en monitoreo global ha sido la disponibilidad de los sistemas de alertas casi en tiempo real, como los desarrollados por Hansen et al. (2016) y Global Forest Watch (2024), los cuales utilizan series temporales Landsat para identificar perturbaciones del dosel. Sin embargo, estos sistemas enfrentan limitaciones importantes en regiones altamente nubosas como la Amazonía ecuatoriana debido a la imposibilidad de obtener escenas ópticas libres de nubes de forma consistente.

Con el auge del aprendizaje profundo, investigaciones como Guimarães et al. (2025) y Kundu et al. (2025) han utilizado redes convolucionales para mapear deforestación y detectar anomalías en tiempo casi real, integrando múltiples sensores. Sin embargo, dichos modelos suelen carecer de explicabilidad e interpretabilidad, lo cual limita su adopción operativa en instituciones públicas.

Paralelamente, se ha consolidado la necesidad de aplicar esquemas rigurosos de validación temporal y espacial en estudios de detección de cambio, con el fin de evitar sesgos

derivados de la autocorrelación espacial y del uso de datos solapados entre fases de análisis (Valavi et al., 2019). En este sentido, el uso de ventanas temporales diferenciadas entrenamiento, validación y prueba se ha convertido en una práctica recomendada para evaluar la estabilidad y consistencia de los resultados obtenidos a partir de series históricas de imágenes satelitales.

En síntesis, el estado del arte evidencia que:

- La serie histórica LANDSAT continúa siendo una de las fuentes más confiables para el análisis multianual de pérdida de bosque.
- El análisis anual ofrece ventajas significativas en regiones tropicales con alta nubosidad persistente.
- Existe una tendencia hacia metodologías más complejas, aunque con crecientes desafíos en interpretabilidad y reproducibilidad.
- La validación temporal rigurosa es un elemento clave para garantizar la robustez de los resultados.

En este marco, la presente investigación se inserta en la línea de estudios que priorizan el análisis multitemporal anual de la deforestación mediante imágenes LANDSAT, incorporando un esquema de ventanas temporales claramente definidas para el entrenamiento, validación y prueba. Este enfoque busca aportar una evaluación sistemática, transparente y reproducible de la dinámica de la deforestación en el Corredor Norte Amazónico del Ecuador, fortaleciendo el uso de datos satelitales como herramienta fundamental para el monitoreo ambiental.

2.2. Marco Teórico

La deforestación es la conversión de áreas boscosas hacia otros usos de suelo, generalmente asociados a agricultura, ganadería, infraestructura y actividades extractivas. En la Amazonía, este proceso se presenta de manera acelerada debido a la expansión de carreteras,

asentamientos humanos, explotación petrolera y tala no controlada (Armenteras et al., 2019). La pérdida de bosques implica reducción de biodiversidad, alteración de ciclos hidrológicos y aumento de emisiones de gases de efecto invernadero, afectando tanto a las comunidades locales como a los sistemas ecológicos globales (Panfil et al., 2016).

En el caso del Ecuador, el Corredor Norte Amazónico constituye una de las zonas de mayor presión antrópica, donde la fragmentación del bosque responde a patrones vinculados a la ocupación humana y la expansión productiva (Sierra, 2013; Clerici et al., 2019). Este contexto demanda herramientas que permitan evaluar con precisión la magnitud y dinámica temporal de la deforestación.

2.2.1. Teorías del Cambio de Uso del Suelo y Dinámica de la Deforestación

La deforestación como fenómeno socio-ecológico ha sido abordada desde múltiples perspectivas teóricas. La teoría de la transición forestal (Mather, 1992) postula que las sociedades atraviesan fases predecibles de pérdida y posterior recuperación de cobertura boscosa, vinculadas a su desarrollo económico. En América Latina, sin embargo, esta transición suele ser incompleta o revertirse debido a la persistencia de presiones extractivas (Rudel et al., 2005).

El enfoque de drivers directos e indirectos (Geist & Lambin, 2002) distingue entre factores próximos (expansión agrícola, infraestructura, minería) y subyacentes (políticas públicas, dinámicas de mercado, crecimiento poblacional). En la Amazonía ecuatoriana, este marco explica cómo la apertura vial (driver directo) interactúa con incentivos a la ganadería (driver indirecto) para acelerar la pérdida de bosque.

La economía política de la deforestación (Robbins, 2012) enfatiza las relaciones de poder, tenencia de la tierra y acumulación de capital como motores estructurales. En el Corredor

Norte Amazónico, la convergencia de intereses petroleros, colonización agraria y mercados informales de madera crea una “tormenta perfecta” para la transformación territorial.

2.2.2. Ecología del Paisaje y Fragmentación Forestal

La teoría de la biogeografía de islas (MacArthur & Wilson, 1967) aplicada a fragmentos de bosque predice que parches pequeños y aislados pierden biodiversidad más rápido. En el CNA, la apertura de carreteras convierte el bosque continuo en un archipiélago de fragmentos, afectando la dispersión de especies y la resiliencia ecológica.

Los umbrales de fragmentación (Andrén, 1994) sugieren que cuando el hábitat remanente cae por debajo del 30-40% del paisaje, se desencadenan efectos no lineales en la pérdida de conectividad. Monitorear estos umbrales es crucial para diseñar corredores biológicos efectivos.

El concepto de matriz paisajística (Forman, 1995) reconoce que los usos no forestales (agricultura, pastizales) no son meros “vacíos”, sino medios con permeabilidad variable que filtran el movimiento de especies y los flujos ecológicos. La calidad de la matriz circundante determina, en gran medida, el destino de los fragmentos remanentes.

2.3. Corredor Norte Amazónico

El Corredor Norte Amazónico (CNA) constituye uno de los espacios geográficos más complejos y dinámicos del Ecuador continental. Ubicado principalmente en las provincias de Sucumbíos, Orellana y parte de Napo, este corredor ha sido históricamente descrito como una región de transición: un espacio donde la selva húmeda tropical convive —a veces en tensión— con carreteras, infraestructura petrolera, asentamientos, espacios comunitarios, procesos de colonización agraria y una multiplicidad de actores cuyas decisiones influyen directamente en la configuración del territorio.

El CNA no es simplemente una masa continua de bosque; es un sistema vivo donde el paisaje cambia constantemente. En ciertas zonas, la selva permanece prácticamente intacta, con estructuras de bosque primario que aún sostienen funciones ecológicas críticas. En cambio, otras áreas muestran patrones de fragmentación, apertura de claros, expansión agropecuaria y procesos de degradación que se aceleran en períodos específicos del año.

A lo largo del periodo 2001–2025, las tendencias históricas de deforestación evidencian que el CNA es uno de los frentes más activos del país. Las causas no son simples ni unidimensionales: convergen factores económicos, sociales, simbólicos, logísticos y ambientales. La región concentra una mezcla de presiones que se retroalimentan entre sí y terminan formando una especie de “tormenta perfecta” para la pérdida de bosque. Basta revisar series temporales para notar cómo, después de la apertura de una carretera, surgen nuevas parcelas agrícolas, asentamientos, pequeñas chacras, potreros y, con el tiempo, áreas deforestadas más amplias que se expanden radialmente.

2.3.1. Historia Reciente de Ocupación y Configuración Territorial

Desde mediados de los años setenta, la apertura petrolera marcó la pauta del desarrollo en la zona. La construcción de la Vía Auca, la expansión de plataformas y ramales petroleros, y la consolidación de centros poblados como Shushufindi, Coca o Lago Agrio transformaron la selva en un territorio híbrido, donde la presencia humana ya no es marginal.

Los procesos de colonización agrícola impulsados por el Estado durante los años ochenta y noventa reforzaron el establecimiento de familias campesinas migrantes, muchas provenientes de la Sierra o la Costa, que encontraron en el CNA un espacio “disponible” para asentarse, aunque en realidad se sobreponen a territorios indígenas preexistentes.

A esta dinámica se sumaron procesos de expansión agropecuaria vinculados al uso del hacha y la quema, una práctica que, aunque tradicional en muchos contextos rurales, genera una presión fuerte sobre el bosque. Las decisiones cotidianas como abrir un pozo para sembrar pasto se acumulan y terminan formando patrones visibles desde el satélite.

Figura 1

Imagen satelital de la deforestación en la industria petrolera



Nota. Imagen generada por IA con Gemini (Google) a partir de la instrucción “Deforestación en la industria petrolera en la Amazonía ecuatoriana.”

2.3.2. Tendencias de Deforestación 2001–2025

Entre 2001 y 2010, la pérdida de bosque se concentró principalmente en zonas de colonización histórica, con una expansión relativamente estable.

Entre 2010 y 2015, se intensificaron los procesos de fragmentación en áreas adyacentes a vías secundarias.

Entre 2015 y 2025, varios frentes se consolidaron como zonas de alta recurrencia de pérdida forestal, asociadas al crecimiento de la ganadería y la mejora del acceso vial.

Datos derivados de productos como Global Forest Change evidencian que una proporción significativa de la deforestación reciente ocurre a distancias cortas de infraestructuras existentes, lo cual coincide con patrones observados en estudios tropicales a escala global.

2.3.3. Estacionalidad y Patrones Temporales

El clima amazónico, aunque húmedo durante casi todo el año, presenta variaciones estacionales que influyen en los patrones de deforestación. En épocas más secas a veces cortas pero significativas se incrementan las actividades de quema, la habilitación de terrenos y la expansión de potreros. Durante los meses más lluviosos, aunque disminuye el uso del fuego, la tala y la habilitación continúan, especialmente en zonas donde la accesibilidad permite operar incluso con lluvias frecuentes.

Un patrón notable (2001–2025) es que muchos frentes muestran un “pico” anual de expansión justo después del periodo de lluvias más intensas, cuando el suelo empieza a drenarse y las actividades agrícolas se reactivan.

2.3.4. Actores Territoriales y Gobernanza

La gobernanza del CNA no es uniforme en el cual cada uno influye en la dinámica territorial de alguna forma: unos protegen el bosque, otros lo transforman; unos promueven conservación, otros impulsan producción; unos gestionan territorio, otros lo ocupan informalmente. Esta diversidad genera, de forma natural, tensiones y contradicciones que deben ser reconocidas para comprender adecuadamente el contexto de la deforestación.

La teoría de los bienes comunes (Ostrom, 1990) proporciona un marco para analizar cómo comunidades locales pueden autoorganizarse para manejar recursos forestales. En el CNA,

territorios indígenas con derechos consuetudinarios bien definidos suelen presentar menores tasas de deforestación que áreas de colonización abierta.

El enfoque de gobernanza multinivel (Cash et al., 2006) examina cómo interaccionan normas locales, políticas nacionales y acuerdos internacionales (ej. REDD+). En Ecuador, la tensión entre el discurso de los Derechos de la Naturaleza (Constitución 2008) y las políticas extractivistas genera vacíos de gobernanza que facilitan la deforestación.

La vigilancia forestal participativa (Larson et al., 2018) demuestra que el monitoreo local, cuando se combina con alertas satelitales, puede mejorar la detección temprana y la respuesta. Esto resalta el valor de herramientas como el dashboard desarrollado, que puede servir como puente entre información técnica y acción comunitaria.

Figura 2

Actores Territoriales y Gobernanza



Nota. Imagen generada por IA con Gemini (Google) a partir de la instrucción “Actores Territoriales y Gobernanza en la Amazonía ecuatoriana”.

2.3.5. Socioeconomía de la Deforestación: Teorías de Rentabilidad y Subsistencia

Los modelos de rentabilidad espacial (Von Thünen adaptado a contextos tropicales)

predicen que la deforestación se concentra donde los beneficios de la conversión (agricultura, ganadería) superan los costos de acceso y oportunidad. En el CNA, la rentabilidad está altamente subsidiada por la infraestructura petrolera (carreteras, subsidios energéticos).

La economía doméstica campesina (Chayanov, 1966) enfatiza que muchas decisiones de deforestación responden a lógicas de subsistencia y reproducción familiar, no solo a maximización de ganancias. La apertura de pequeñas chacras (“agricultura de tumba y quema”) sigue ciclos familiares y necesidades inmediatas.

El capital natural y los servicios ecosistémicos (Costanza et al., 1997) proveen un marco para valorar económicamente el bosque en pie (carbono, agua, biodiversidad). La deforestación, en este marco, representa una externalidad negativa masiva que no se internaliza en los mercados locales.

2.4. Teledetección Aplicada al Monitoreo Forestal

La teledetección, entendida como el uso de sistemas satelitales para obtener información sobre la superficie terrestre sin contacto directo, es hoy la herramienta principal para monitorear el bosque amazónico. Sin ella, la comprensión espaciotemporal del CNA sería incompleta y, en muchos casos, imposible.

La teledetección moderna no solo captura imágenes; registra señales biofísicas, variaciones del dosel, humedad, estructura, cambios abruptos y procesos lentos que a veces no son detectables desde el suelo.

2.4.1. Sensores de Teledetección: Landsat y Sentinel

Los sensores ópticos capturan la reflectancia de la superficie terrestre en diferentes rangos del espectro electromagnético. En términos simples, ven colores y tonalidades que permiten inferir la salud, continuidad y composición de la vegetación.

Landsat posee una serie histórica de más de 40 años, lo que lo convierte en un insumo invaluable para análisis a largo plazo. Con su resolución espacial de 30 metros, permite identificar patrones grandes y tendencias agregadas.

El programa LANDSAT representa uno de los pilares del monitoreo ambiental global. Su archivo continuo, que supera las cuatro décadas de observación, permite realizar análisis multitemporales consistentes del cambio de cobertura forestal (Roy et al., 2014; Wulder et al., 2019).

Con una resolución espacial de 30 metros, LANDSAT ofrece un equilibrio adecuado entre detalle espacial y cobertura regional, siendo especialmente útil para estudios orientados al análisis anual de deforestación. A través de la reflectancia en bandas visibles, infrarrojas y de onda corta, es posible derivar índices espectrales que capturan variaciones en la salud y continuidad del dosel forestal.

No obstante, en regiones tropicales húmedas como el CNA, la nubosidad persistente constituye una limitación estructural. Por esta razón, numerosos estudios recomiendan el uso de composiciones anuales y la selección de escenas óptimas para reducir el impacto de nubes y sombras en el análisis multitemporal.

Figura 3

Imagen Satelital de la misión LANDSAT 8



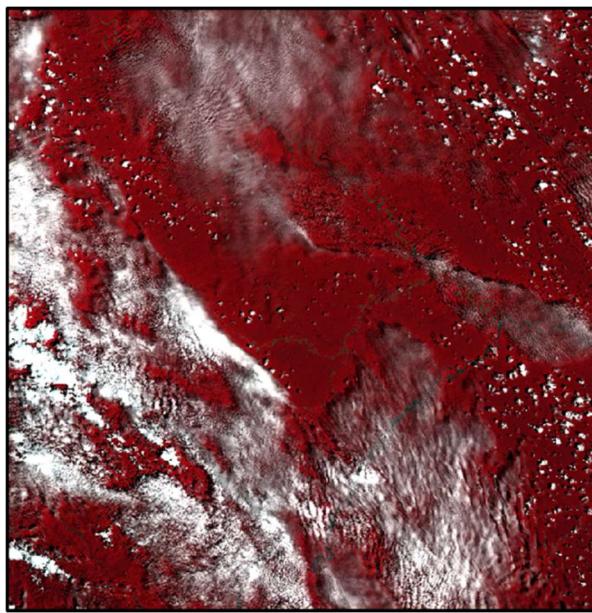
Nota. Imagen satelital de la misión LANDSAT 8. Fuente: LANDSAT 8

Sentinel-2, por su parte, ofrece 10 metros de resolución y revisitas más frecuentes, lo que mejora la capacidad de detectar cambios sutiles y procesos de degradación. Sus bandas específicas para vegetación enriquecen el análisis espectral.

Pero la limitación fundamental es la nubosidad. En la Amazonía, hay meses donde se acumulan semanas sin una sola imagen útil. Las nubes y sombras afectan severamente la continuidad del monitoreo óptico, especialmente en zonas con clima húmedo permanente. Por esta razón, los sistemas basados únicamente en óptico reportan tiempos muertos o retrasos significativos en la detección de cambios, especialmente durante estaciones húmedas o transiciones climáticas.

Figura 4

Imagen Satelital misión SENTINEL-2



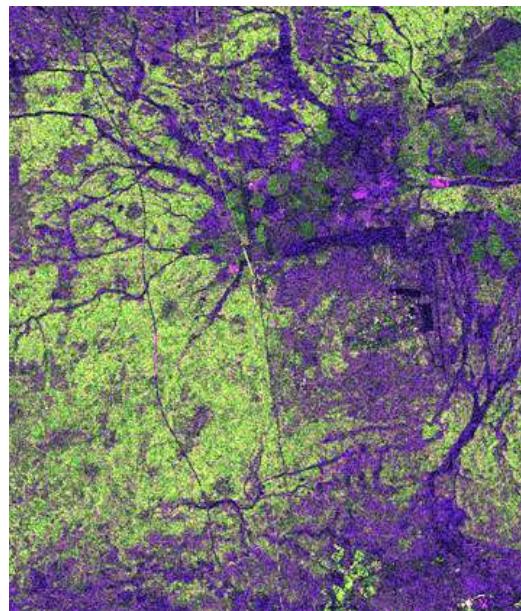
Nota. Imagen satelital de la misión SENTINEL-2. Fuente: SENTINEL-2

Sentinel-1, El radar de apertura sintética (SAR), como Sentinel-1, opera en el rango de microondas. Al emitir y recibir pulsos activos, puede penetrar nubes y operar independientemente de la iluminación solar. Esta característica lo convierte en un aliado ideal para regiones como la Amazonía.

El radar ofrece información estructural del bosque, rugosidad de la superficie, humedad del suelo, cambios en la geometría del dosel, apertura de claros y alteraciones bruscas vinculadas a tala o degradación.

Figura 5

Imagen Satelital misión SENTINEL-1



Nota. Imagen satelital de la misión SENTINEL-1. Fuente: SENTINEL-1

2.4.2. Sistemas de Alerta de Deforestación

Los sistemas de alerta GLAD (óptico), RADD (radar), GFW proveen señales rápidas sobre pérdida de bosque. Su valor radica en la reducción de la latencia de detección. Pero todos comparten una limitación: **detectan lo que ya ocurrió**, no lo que ocurrirá.

Esto obliga a pensar en sistemas que no solo “vean” el cambio, sino que anticipen la probabilidad de que ocurra próximamente.

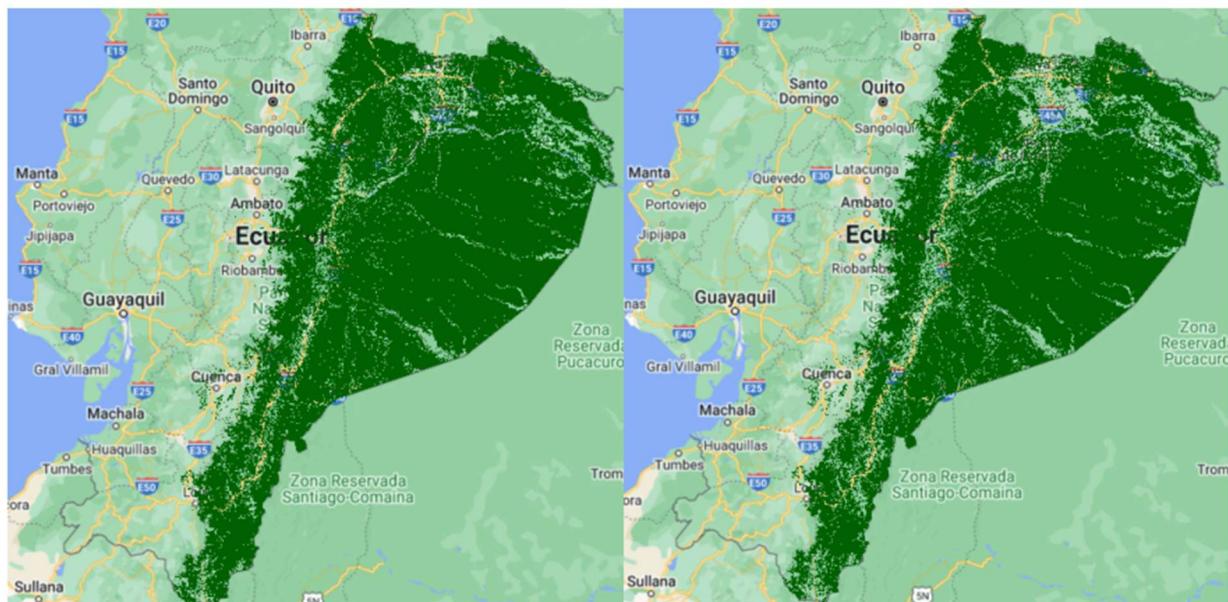
2.4.3. Mapas Históricos y Series Temporales

MapBiomas y otros productos históricos proporcionan información clave sobre los patrones acumulados. Aunque no predicen directamente, permiten identificar zonas con historial de pérdida consecutiva, expansión de frentes activos, áreas de transición, regiones con resistencia relativa al cambio y ritmos temporales característicos.

Estos datos son fundamentales para entrenar modelos predictivos que operan a partir de señales pasadas.

Figura 6

Formación Forestal Año 2001 - 2024



Nota. Imagen de la formación Forestal en la Amazonía ecuatoriana. Fuente: BioMaps - Ecuador

2.5. Aprendizaje Automático para Predicción Ambiental Aplicado a Fenómenos

Geoespaciales

El aprendizaje automático permite modelar relaciones no lineales entre predictores y eventos futuros. Métodos como Random Forest, XGBoost, LightGBM y redes convolucionales que son ampliamente usados en monitoreo ambiental. Su capacidad para manejar grandes volúmenes de datos y capturar interacciones complejas los convierte en herramientas ideales para sistemas de alerta temprana.

Los modelos predictivos buscan estimar valores futuros a partir de patrones históricos. Incluyen enfoques estadísticos clásicos y modelos basados en aprendizaje automático (Makridakis, et al., 1998; Hyndman et al., 2018).

Los algoritmos de aprendizaje supervisado para regresión espaciotemporal deben lidiar con la autocorrelación espacial (valores cercanos en el espacio son similares) y la no estacionariedad espacial (las relaciones entre variables cambian según la ubicación). Random Forest, seleccionado en este estudio, es relativamente robusto a estos problemas gracias a su estructura basada en particiones aleatorias.

La importancia de variables en Random Forest se calcula mediante permutación o impureza decreciente. Sin embargo, en contextos espaciotemporales, esta importancia puede variar regionalmente (un driver puede ser clave en un cantón y marginal en otro). Por ello, modelos locales (por cantón) pueden capturar mejor esta heterogeneidad.

Los modelos híbridos que combinan aprendizaje automático con restricciones físicas o económicas (por ejemplo, incorporando ecuaciones de rentabilidad agrícola) están ganando terreno en la predicción de cambio de uso del suelo (Meyfroidt et al., 2018).

2.5.1. Random Forest

Random Forest es uno de los algoritmos más utilizados en estudios ambientales por su capacidad para manejar grandes volúmenes de datos, relaciones no lineales y variables altamente correlacionadas. Su estructura basada en múltiples árboles de decisión permite reducir el sobreajuste y proporcionar estimaciones robustas incluso en escenarios con datos ruidosos o incompletos. En aplicaciones de deforestación, este modelo ha demostrado un desempeño consistente en la clasificación de cobertura forestal y en la estimación del riesgo de cambio de uso del suelo.

2.5.2. Series Temporales y Teoría de la Predicción en Sistemas Complejos

Las series temporales ambientales suelen exhibir autocorrelación temporal (los valores presentes dependen de los pasados), estacionalidad (patrones cíclicos anuales) y en ocasiones

tendencias no lineales. Los modelos clásicos (ARIMA, exponencial smoothing) asumen estacionariedad, pero las series de deforestación rara vez la cumplen debido a cambios de régimen (e.g., nueva carretera, cambio de política).

La teoría del caos y los sistemas complejos adaptativos (Holling, 2001) sugiere que los ecosistemas forestales pueden responder de manera abrupta y no lineal a presiones acumulativas. Esto implica que los modelos lineales o estacionarios pueden subestimar puntos de quiebre (“tipping points”).

El enfoque de ventanas deslizantes utilizado en este estudio se basa en la premisa de que el proceso de deforestación tiene una memoria de corto plazo: las condiciones de los últimos 3-5 años influyen más en el año siguiente que eventos lejanos. Esto es coherente con la dinámica de frentes de deforestación, donde la accesibilidad reciente y la expansión vecina son predictores fuertes.

2.5.3. Ecología del Paisaje y Fragmentación Forestal

La teoría de la biogeografía de islas (MacArthur & Wilson, 1967) aplicada a fragmentos de bosque predice que parches pequeños y aislados pierden biodiversidad más rápido. En el CNA, la apertura de carreteras convierte el bosque continuo en un archipiélago de fragmentos, afectando la dispersión de especies y la resiliencia ecológica.

Los umbrales de fragmentación (Andrén, 1994) sugieren que cuando el hábitat remanente cae por debajo del 30-40% del paisaje, se desencadenan efectos no lineales en la pérdida de conectividad. Monitorear estos umbrales es crucial para diseñar corredores biológicos efectivos.

El concepto de matriz paisajística (Forman, 1995) reconoce que los usos no forestales (agricultura, pastizales) no son meros “vacíos”, sino medios con permeabilidad variable que

filtran el movimiento de especies y los flujos ecológicos. La calidad de la matriz circundante determina, en gran medida, el destino de los fragmentos remanentes.

2.6. Marco Normativo y Políticas Públicas

El análisis de la deforestación en el Corredor Norte Amazónico del Ecuador se desarrolla dentro de un marco normativo nacional que reconoce al ambiente como un eje estratégico del desarrollo y a los territorios indígenas como espacios sujetos a derechos colectivos específicos. (Constitución de la República del Ecuador, 2008). Este entramado jurídico no solo regula el uso del suelo y la conservación de los ecosistemas, sino que también establece la obligación del Estado de generar información ambiental confiable, actualizada y accesible, lo que otorga un respaldo directo al uso de herramientas de teledetección y análisis multitemporal (Plan Nacional de Desarrollo, 2017).

2.7. Python

Python es un lenguaje de programación de alto nivel que se ha consolidado como uno de los principales entornos para el análisis de datos, el aprendizaje automático y la investigación científica. Su diseño enfatiza la legibilidad del código y la productividad del desarrollador, lo que facilita la construcción de flujos analíticos complejos de forma reproducible y transparente (Van Rossum et al., 2009; Lutz, 2013).

2.7.1. Elementos Predefinidos

Python incluye un conjunto de elementos fundamentales como variables, tipos de datos primitivos (enteros, flotantes, cadenas y booleanos) y estructuras de datos nativas (listas, tuplas, diccionarios y conjuntos). Estos componentes constituyen la base para la representación eficiente de información compleja y multidimensional (Downey, 2015).

2.7.2. Estructuras de Control y Funciones

Las estructuras de control condicional (if–else) y repetitiva (for, while) permiten definir flujos lógicos que gobiernan el procesamiento de datos. Por su parte, las funciones facilitan la modularización del código, promoviendo la reutilización, la claridad y la mantenibilidad (Beazley et al., 2013).

2.7.3. Librerías de Python

El verdadero potencial de Python radica en su ecosistema de librerías especializadas, ampliamente validadas en la literatura científica.

Matplotlib. Librería fundamental para visualización científica, utilizada para representar distribuciones, tendencias y comparaciones temporales (Hunter, 2007).

Math. Módulo estándar de Python, proporciona acceso a las funciones matemáticas, dentro del proyecto se lo utiliza para los cálculos y transformaciones geográficas. (Hunter, 2007).

NumPy. Proporciona soporte para arreglos multidimensionales y operaciones matemáticas de alto rendimiento, siendo la base de gran parte del cómputo científico en Python (Harris et al., 2020).

Pandas. Ofrece estructuras de datos tabulares (DataFrame) que facilitan la manipulación, limpieza y análisis de datos estructurados (McKinney, 2010).

Flask. Microframework web libre y flexible para Python, ideal para presentar el dashboard final del proyecto. (Werkzeug, WSGI, 2010).

Pillow. Biblioteca de procesamiento de imágenes que permite la manipulación de archivos formato ‘.tiff’ y reconocimiento pixeles RGB. (Pillow - PIL Fork, 2021).

Scikit-Learn. Biblioteca estándar para aprendizaje automático clásico, que implementa algoritmos supervisados y no supervisados de forma eficiente y reproducible (Pedregosa et al., 2011).

Joblib. Herramienta para la paralelización y persistencia de objetos Python, implementado sobre todo en modelos Machine Learning, en el modelo se lo utilizará para el guardado y carga del modelo entrenado y la optimización del procesamiento de datos. (Joblib Development Team, s.f., 2011)

Sphinx. Biblioteca sumamente importante para el proceso de documentación técnica del código del proyecto, el mismo que convierte archivos reStructuredText en HTML. (Sphinx, 2025)

2.8. Aprendizaje Supervisado

El aprendizaje supervisado se basa en el entrenamiento de modelos a partir de datos etiquetados, donde cada observación posee una variable objetivo-conocida. Este enfoque es ampliamente utilizado en problemas de clasificación y regresión (Hastie, Tibshirani et al., 2009).

2.9. Aprendizaje No Supervisado

A diferencia del enfoque supervisado, el aprendizaje no supervisado busca identificar patrones, estructuras o agrupamientos en datos sin etiquetas previas. Técnicas como clustering y reducción de dimensionalidad permiten explorar la organización interna de los datos (Jain, 2010).

2.10. Metodología KDD

La metodología **Knowledge Discovery in Databases (KDD)** constituye un marco conceptual para la extracción sistemática de conocimiento a partir de grandes conjuntos de datos (Fayyad et al., 1996).

Selección de datos relevantes. Esta fase implica identificar variables pertinentes para el objetivo del análisis, reduciendo la complejidad inicial y mejorando la eficiencia del proceso (Han, Kamber et al., 2012).

Preprocesamiento y limpieza. El preprocesamiento aborda problemas como valores faltantes, inconsistencias y ruido, garantizando la calidad de los datos antes del análisis (Rahm et al., 2000).

Transformación y reducción. Incluye normalización, escalamiento y reducción de dimensionalidad, con el fin de mejorar el desempeño de los algoritmos analíticos (Guyon et al., 2003).

Data Mining. En esta etapa se aplican algoritmos de aprendizaje automático para identificar patrones significativos y relaciones subyacentes en los datos (Witten, Frank et al., 2016).

Interpretación y evaluación. Los patrones obtenidos deben ser interpretados y evaluados en función de su relevancia, validez y utilidad práctica (Fayyad et al., 1996).

2.11. Series Temporales

Las series temporales representan observaciones ordenadas cronológicamente, permitiendo analizar tendencias, ciclos y cambios estructurales a lo largo del tiempo (Box, Jenkins et al., 2008).

2.12. Métricas de Evaluación

Las métricas permiten cuantificar el desempeño de los modelos predictivos. Entre las más utilizadas se encuentran el error cuadrático, la varianza explicada y el error relativo (Powers, 2011).

2.12.1. Error Absoluto Medio (MAE)

El Error Absoluto Medio (Mean Absolute Error, MAE) mide la magnitud promedio del error entre los valores observados y los valores predichos, sin considerar su dirección. Esta métrica es robusta frente a valores atípicos y resulta adecuada cuando se desea una interpretación directa del error en las mismas unidades de la variable analizada.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (1)$$

donde y_i representa el valor real del área forestal, \hat{y}_i el valor estimado por el modelo y n el número total de observaciones.

2.12.2. Raíz del Error Cuadrático Medio (RMSE)

La Raíz del Error Cuadrático Medio (*Root Mean Squared Error, RMSE*) penaliza de forma más severa los errores grandes al elevarlos al cuadrado antes de promediarlos. Esta métrica es especialmente útil para identificar desviaciones significativas en la predicción de la pérdida forestal.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|^2} \quad (2)$$

Valores menores de RMSE indican una mayor proximidad entre las predicciones del modelo y los valores observados.

2.12.3. Coeficiente de Determinación (R^2)

El coeficiente de determinación R^2 mide la proporción de la variabilidad total de la variable dependiente que es explicada por el modelo. Se define como:

$$R^2 = 1 - \frac{\sum_{i=1}^n |y_i - \hat{y}_i|^2}{\sum_{i=1}^n |y_i - \bar{y}_i|^2} \quad (3)$$

donde \bar{y}_i es el valor medio de los datos observados.

En series temporales cortas y con baja variabilidad interanual, R^2 puede resultar negativo, indicando que el modelo no supera a una predicción basada únicamente en el promedio histórico. Por esta razón, el coeficiente R^2 se considera complementario y no determinante en la evaluación del desempeño del modelo.

2.12.4. Error Porcentual Absoluto Medio (MAPE)

El Error Porcentual Absoluto Medio (*Mean Absolute Percentage Error, MAPE*) evalúa el error relativo en términos porcentuales, permitiendo comparar el desempeño del modelo entre regiones con diferentes magnitudes de área forestal.

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100 \quad (4)$$

Esta métrica es particularmente adecuada para el análisis multiescalar, ya que normaliza el error respecto al valor real observado.

Capítulo III

3. Desarrollo

3.1. Metodología

El estudio se realiza en el **Corredor Norte Amazónico Ecuatoriano**, que corresponde a los cantones específicos de las provincias de Orellana, Sucumbíos y Napo. Se trata de un paisaje caracterizado por la biodiversidad, alta cobertura boscosa y fuerte presión antrópica debido a expansión agrícola, infraestructura vial, campos petroleros, minería y asentamientos humanos.

Este capítulo describe el flujo metodológico implementado para calcular el área forestal por provincia y cantón, estimar la pérdida anual de cobertura forestal y entrenar modelos de aprendizaje automático con validación temporal tipo *walk-forward*. El objetivo principal es analizar la dinámica histórica del bosque y generar predicciones multitemporales confiables para el período 2025–2030.

La metodología integra técnicas de preprocesamiento de datos espaciales, ingeniería de características derivadas del uso y cobertura del suelo, y modelos de regresión basados en Random Forest, priorizando la coherencia temporal y la robustez estadística

3.2. Síntesis Metodológica

El flujo metodológico propuesto combina análisis espacial, series temporales y aprendizaje automático, permitiendo modelar la evolución del área forestal a escala cantonal. La validación temporal rigurosa y el uso de variables explicativas derivadas del uso del suelo fortalecen la confiabilidad de las predicciones, convirtiendo este enfoque en una herramienta sólida para el análisis prospectivo de la cobertura forestal y el apoyo a la toma de decisiones ambientales.

3.3. Recolección y Descripción de los Datos

3.3.1. Fuente de Datos Satelitales

La fuente principal de información utilizada en esta investigación corresponde a los mapas anuales de cobertura y uso del suelo generados por la iniciativa **MapBiomas Ecuador**. Estos datos se derivan del procesamiento de imágenes satelitales **LANDSAT**, las cuales ofrecen una resolución espacial de **30 metros**, adecuada para el análisis de la deforestación basados en los datos recopilados durante cuarenta años de la región amazónica.

Los datos se encuentran disponibles en formato raster **GeoTIFF (.tif)**, donde cada píxel almacena un valor entero que representa una clase específica de la cobertura del suelo. La serie temporal utilizada permite analizar la evolución anual de la deforestación en el Corredor Norte Amazónico ecuatoriano.

3.3.2. Características Espaciales de los Datos Raster

Cada archivo raster empleado presenta las siguientes características técnicas:

- Resolución espacial: $30 \text{ m} \times 30 \text{ m}$
- Área por píxel: 900 m^2
- Sistema de referencia espacial: proyección cartográfica estandarizada para la Amazonía ecuatoriana
- Tipo de dato: raster categórico (clasificación de la deforestación)
 - Esta estructura permite cuantificar cambios espaciales mediante el conteo de píxeles por clase y por unidad territorial.

La conversión entre número de píxeles y área real se realizó considerando la siguiente relación:

$$\text{Área}_{\text{km}^2} = \frac{N_{\text{pixeles}} \times 900}{1000000} \quad (5)$$

donde:

N_{pixeles} corresponde al número de píxeles de una clase específica,

900 representa el área en metros cuadrados de un píxel,

1000000 corresponde al factor de conversión de m^2 a km^2 .

Figura 7

Código - Resolución Espacial

```
def obtener_resolucion_espacial_real(img_array, provincia, canton, areas_oficiales):
    """
    Calcula la resolución espacial usando el área real del cantón desde extension_data.xlsx
    """
    ancho_pixeles, alto_pixeles = img_array.shape[1], img_array.shape[0]

    # Obtener área real del cantón
    key = f"{provincia}_{canton}"
    area_real_km2 = areas_oficiales[key]
    print(f"Usando área oficial de {provincia}/{canton}: {area_real_km2:.2f} km²")

    # Calcular área por píxel usando el área real
    total_pixeles_utiles = ancho_pixeles * alto_pixeles
    area_por_pixel_km2 = area_real_km2 / total_pixeles_utiles

    # Resolución aproximada en metros
    area_por_pixel_m2 = area_por_pixel_km2 * 1_000_000
    resolucion_metros = (area_por_pixel_m2 ** 0.5) # lado del pixel cuadrado

    return area_por_pixel_km2, resolucion_metros, area_real_km2
```

3.3.3. Delimitación del Área de Estudio

Para restringir el análisis al **Corredor Norte Amazónico del Ecuador**, se utilizaron capas vectoriales a nivel de provincia y cantón. Estas capas permitieron seleccionar exclusivamente las provincias que conforman el corredor y posteriormente identificar los cantones incluidos para el análisis.

3.3.4. Procesamiento Espacial en QGIS

El procesamiento de los datos se realizó mediante el software de código abierto **QGIS**, siguiendo una secuencia metodológica reproducible:

- Importación de los archivos raster anuales en formato GeoTIFF.
- Recorte de cada raster utilizando los límites provinciales y cantonales.

Este procedimiento permitió transformar información raster en datos cuantitativos.

Figura 8

Resultado QGIS



3.3.5. Estructuración de la Base de Datos Tabular

Los resultados del análisis fueron exportados a formato csv, generando una base de datos estructurada que integra información espacial y temporal. Cada registro representa una combinación única de cantón, año y clase de deforestación.

La base de datos incluye, entre otros, los siguientes campos:

- Provincia

- Cantón
- Año
- Nombre del archivo raster
- Área real del cantón (km²)
- Código y nombre de la clasificación de cobertura
- Color hexadecimal y componentes RGB asociados a la clase
- Número de píxeles por clase
- Porcentaje respecto al área total del cantón
- Área en km² por clase

Esta estructura facilita el análisis comparativo anual y la identificación de tendencias de deforestación.

Figura 9

Estructura Tabular

```
resultados.append({
  'provincia': provincia,
  'canton': canton,
  'año': año,
  'archivo': os.path.basename(archivo_path),
  'area_real_canton_km2': area_real_km2,
  'hex_color': hex_color,
  'rgb_r': rgb[0],
  'rgb_g': rgb[1],
  'rgb_b': rgb[2],
  'clasificacion': clasificacion,
  'pixeles': count_pixels,
  'porcentaje_total': porcentaje,
  'area_km2': area_km2,
  'en_excel': en_excel,
  'es_fondo': es_fondo
})
```

3.3.6. Derivación de Indicadores de Deforestación

La deforestación se estimó a partir de la variación anual en el número de píxeles correspondientes a clases no forestales. Dado que cada píxel representa un área constante, el cambio en el conteo de píxeles permite sumar de manera directa la superficie deforestada por año.

Este enfoque posibilita:

- Comparar la pérdida anual entre cantones,
- Identificar periodos de mayor presión sobre el bosque,
- Analizar patrones espaciales de deforestación.

Figura 10

Patrón de colores de los indicadores de deforestación

■ '#1f8d49': 'Formación forestal',
■ '#7dc975': 'Bosque abierto',
■ '#04381d': 'Manglar',
■ '#026975': 'Bosque inundable',
■ '#519799': 'Formación natural no forestal inundable',
■ '#d6bc74': 'Herbazal',
■ '#ffaa5f': 'Afloramiento rocoso',
■ '#d89f5c': 'Otra formación natural no forestal',
■ '#7a5900': 'Silvicultura',
■ '#ffefc3': 'Mosaico de agricultura y/o pastos',
■ '#ffa07a': 'Playa, duna o banco de arena',
■ '#d4271e': 'Infraestructura urbana',
■ '#9c0027': 'Minería',
■ '#e97a7a': 'Otra área natural sin vegetación',
■ '#db4d4f': 'Otra área antrópica sin vegetación',
■ '#2532e4': 'Río, lago u océano',
■ '#93dfee': 'Glaciar',
■ '#091077': 'Acuicultura',
□ '#ffffff': 'No observado',
■ '#000000': 'Sin datos/Fondo'

3.3.7. Ventajas del Enfoque Adoptado

El uso de datos raster multitemporales ofrece varias ventajas metodológicas:

- Consistencia espacial y temporal en el análisis.

- Reproducibilidad del procedimiento.
- Transparencia en la conversión de píxeles a área real.
- Facilidad para escalar el análisis a otros territorios o períodos.

Este enfoque constituye una base sólida para el análisis posterior de la deforestación.

3.4. Preparación y Transformación de Datos

Una vez consolidada la base de datos, se procedió a su preparación y transformación con el objetivo de garantizar su confiabilidad y coherencia para el análisis de la deforestación anual. Esta etapa se centra en la transformación interna de la base de datos, abordando principalmente aspectos estadísticos y temporales que permiten una adecuada modelación de las dinámicas forestales.

Las operaciones de preparación de datos fueron implementadas mediante herramientas computacionales en el lenguaje de programación Python, utilizando librerías especializadas para el manejo y análisis de datos.

Tabla 1

Librerías utilizadas y su función metodológica

Librería	Función metodológica en el estudio
os	Gestión de directorios y rutas para la organización estructurada de resultados, modelos y gráficos.
json	Serialización y almacenamiento de metadatos y resultados experimentales en formato interoperable.
joblib	Persistencia de modelos entrenados para su posterior reutilización o despliegue (preparada para extensiones futuras).

numpy	Manipulación numérica eficiente, cálculo de estadísticas y operaciones vectorizadas sobre series temporales.
pandas	Lectura, limpieza, transformación y agregación de datos tabulares provenientes de fuentes geoespaciales.
matplotlib.pyplot	Generación de gráficos exploratorios y visualización de resultados (configurado para análisis posterior).
collections.defaultdict	Creación de estructuras jerárquicas dinámicas para organizar datos por provincia y cantón.
sklearn.preprocessing.StandardScaler	Normalización de variables predictoras para mejorar la estabilidad y desempeño de los modelos.
sklearn.neural_network.MLPRegressor	Implementación de redes neuronales multicapa para modelado no lineal de series temporales.
sklearn.ensemble.RandomForestRegressor	Modelo de aprendizaje ensamblado utilizado como predictor principal del área forestal.
sklearn.ensemble.GradientBoostingRegressor	Modelo alternativo para comparación y validación metodológica.

sklearn.metrics	Cálculo de métricas de error y ajuste para la validación cuantitativa de los modelos.
sklearn.model_selection.RandomizedSearchCV	Optimización de hiperparámetros mediante búsqueda aleatoria con validación cruzada.

Nota. Las librerías seleccionadas corresponden a herramientas estándar en análisis de datos y aprendizaje automático, ampliamente validadas en estudios científicos y técnicos.

Durante esta fase se llevaron a cabo procesos de verificación de consistencia temporal, detección y tratamiento de valores faltantes, y normalización de las series cuando fue necesario, con el fin de reducir sesgos y asegurar la comparabilidad entre distintos períodos y unidades territoriales. Asimismo, se calcularon estadísticas descriptivas relevantes que permitieron identificar tendencias generales y posibles anomalías en los datos de área forestal.

Como parte de la transformación de los datos, se generaron variables derivadas que capturan información temporal, tales como tasas de cambio interanual y métricas agregadas dentro de ventanas móviles, las cuales constituyen la base para la posterior construcción de ventanas multivariadas utilizadas en el modelado predictivo.

Esta etapa fue fundamental, ya que garantiza que los datos utilizados en el entrenamiento y evaluación de los modelos reflejen de manera fiel la evolución histórica de la cobertura forestal, minimizando errores inducidos por inconsistencias estadísticas o temporales.

3.4.1. Análisis Exploratorio de Datos (EDA)

El análisis exploratorio de datos se aplicó con el propósito de comprender el comportamiento interno de las variables, identificar patrones preliminares y detectar posibles inconsistencias provenientes del proceso de transformación del raster espacial.

El EDA incluyó:

- Inspección de la distribución del número de píxeles y hectáreas por clase.
- Identificación de inconsistencias
- Detección de valores nulos

Este análisis permitió reconocer que la distribución de los píxeles no tiene patrones similares entre cantones, evidenciando la desigual del proceso de deforestación.

3.4.2. Normalización Temporal de la Base de Datos

Se realiza una conversión de las variables numéricas y una normalización de los nombres de las clases de cobertura. Se modifica la unidad de medida de km² por hectárea (ha) calculándola automáticamente (1 km² = 100 ha).

Posteriormente los datos se agrupan por provincia, cantón, año, sumando las áreas correspondientes y generando una tabla pivoté que permite disponer de una columna por cada tipo de deforestación

Con el fin de permitir comparaciones anuales, la base de datos fue reorganizada temporalmente. Para cada cantón, los registros fueron ordenados cronológicamente según el año de análisis.

Este procedimiento permitió:

- Garantizar la continuidad temporal de cada cantón.

- Evitar comparaciones entre períodos no consecutivos.
- Preparar la información para el cálculo de deforestación anual.

La normalización es fundamental para evitar interpretaciones erróneas de la dinámica de la cobertura forestal.

Figura 11

Normalización de datos

```

# -----
# AGRUPAR Y PIVOTAR POR CLASES (por provincia-cantón-año)
# -----
# Normalizar nombres de clasificación
df['clasificacion'] = df['clasificacion'].astype(str).str.strip()

# Hacemos suma de area_ha por clasificación
grp = df.groupby(['provincia', 'canton', 'año', 'clasificacion'], as_index=False)[['area_ha']].sum()

# Pivot para obtener columnas por clase
pivot = grp.pivot_table(index=['provincia', 'canton', 'año'], columns='clasificacion', values='area_ha', aggfunc='sum', fill_value=0)
pivot = pivot.reset_index()

# Nombre simplificado de columnas (evitar espacios problemáticos)
pivot.columns = [str(c) for c in pivot.columns]
pivot_cols = pivot.columns.tolist()
# Calcular area_total por fila (suma de todas las clases)
class_cols = [c for c in pivot_cols if c not in ['provincia', 'canton', 'año']]
pivot['area_total_ha'] = pivot[class_cols].sum(axis=1)

```

3.4.3. Filtrado de Clases Relevantes para Deforestación

No todas las clases de cobertura de suelo resultan ideales para el análisis de la deforestación. Por esta razón, se aplicó un proceso de filtrado conceptual, manteniendo únicamente aquellas clases que representan cobertura de bosques.

Este paso permitió:

- Reducir ruido analítico.
- Enfocar el análisis en transiciones relevantes.
- Simplificar la interpretación de resultados.

Las clases consideradas como no relevantes fueron conservadas únicamente como referencia contextual.

3.4.4. Construcción de Variables Derivadas

A partir de las variables originales de la base de datos, se generaron nuevas variables orientadas a describir la dinámica del cambio de cobertura forestal:

Variación Interanual de Píxeles

$$\Delta \text{Pixelles}_t = \text{Pixelles}_t - \text{Pixelles}_{t-1} \quad (6)$$

Variación Interanual de Área por hectárea

$$\Delta \text{Área}_{ha} = \text{Área}_t - \text{Área}_{t-1} \quad (7)$$

Estas variables permiten identificar años con pérdidas netas de cobertura forestal y cuantificar en hectáreas la perdida de bosques.

Figura 12

Variables Derivadas

```

features = pd.DataFrame({
    'provincia': df_pivot['provincia'],
    'canton': df_pivot['canton'],
    'año': df_pivot['año'].astype(int),
    'area_forest_ha': area_forest_ha,
    'pct_forest': pct_forest,
    'pct_agri': pct_agri,
    'pct_mining': pct_mining,
    'pct_urban': pct_urban,
    'pct_rio': pct_rio,
    'pct_other_natural': pct_other_natural,
    'pct_other_ant': pct_other_ant,
    'agri_growth': agri_growth,
    'urban_growth': urban_growth,
    'mining_growth': mining_growth,
    'forest_loss': forest_loss,
    'area_ha_ma3': area_ha_ma3,
    'ratio_anthropic': ratio_anthropic,
    'fragmentation_index': fragmentation_index,
    'pct_forest_trend': pct_forest_trend,
    'pct_agri_trend': pct_agri_trend,
    'dom_class': dom_class
})
# Rellenar NaNs
features = features.fillna(0.0)
return features

# Aplicar compute_class_features por provincia-canton
list_features = []
for (prov, cant), group in pivot.groupby(['provincia', 'canton']):
    feats = compute_class_features(group)
    list_features.append(feats)

features_df = pd.concat(list_features, axis=0, ignore_index=True)
print("Features por provincia-canton-año calculadas. Filas:", len(features_df))

```

3.4.5. Agregación por Unidad Territorial

Con el objetivo de obtener indicadores consolidados, los datos fueron agregados por cantón y año. Este proceso permitió:

- Sumar hectáreas de clases no forestales.

3.4.6. Análisis del Desbalance de Clases

Dado que la deforestación representa una fracción reducida del total del territorio, se evaluó el desbalance entre clases forestales y no forestales dentro de la base de datos.

Este análisis evidenció que:

- La mayoría de los registros corresponden a cobertura forestal estable.

Reconocer este desbalance es clave para interpretar correctamente las tendencias de la deforestación en la amazonía ecuatoriana.

3.4.7. Preparación de Insumos para Análisis Temporal

La base de datos fue estructurada para su uso en análisis temporal, quedando lista para:

- Construcción de series históricas de deforestación.
- Identificación de tendencias.
- Evaluación de períodos de pérdida forestal.
- Alimentación de modelos predictivos en etapas posteriores.

Esta preparación garantiza que los resultados obtenidos reflejen cambios reales.

3.5. Cálculo del Área Forestal y Variables Derivadas

3.5.1. Área Forestal Total

El área forestal se obtiene directamente de la clase cobertura forestal, expresada en hectáreas. Además, se calcula el área total por año como la suma de todas las clases de cobertura, lo cual permite estimar proporciones relativas.

Figura 13*Área Forestal en Hectáreas*

```
# area forestal en ha
area_forest_ha = get_col('Formación forestal').fillna(0.0)
```

3.5.2. Variables Porcentuales

Para cada año se calculan los porcentajes de ocupación de las principales clases de uso del suelo, entre ellas:

- Cobertura forestal
- Agricultura y pastos
- Minería
- Infraestructura urbana
- Otras áreas naturales y antrópicas

Figura 14*Variables Porcentuales*

```
pct_forest = get_col('Formación forestal') / df_pivot['area_total_ha'].replace(0, np.nan)
pct_agri = get_col('Mosaico de agricultura y/o pastos') / df_pivot['area_total_ha'].replace(0, np.nan)
pct_mining = get_col('Minería') / df_pivot['area_total_ha'].replace(0, np.nan)
pct_urban = get_col('Infraestructura urbana') / df_pivot['area_total_ha'].replace(0, np.nan)
pct_rio = get_col('Río, lago u océano') / df_pivot['area_total_ha'].replace(0, np.nan)
pct_other_natural = get_col('Otra formación natural no forestal') / df_pivot['area_total_ha'].replace(0, np.nan)
pct_other_ant = get_col('Otra área antrópica sin vegetación') / df_pivot['area_total_ha'].replace(0, np.nan)
```

Estas proporciones permiten caracterizar la presión antrópica y los cambios estructurales en el territorio.

3.5.3. Indicadores de Cambio y Presión Antrópica

Se derivan variables adicionales relevantes para la dinámica forestal:

- Crecimiento anual de agricultura, urbanización y minería
- Pérdida forestal anual absoluta (cuando el área forestal disminuye)
- Media móvil de tres años del área forestal
- Ratio antrópica, como suma de coberturas no forestales
- Índice de fragmentación, definido como la razón entre cobertura forestal y presión antrópica

Estas variables permiten capturar tanto cambios como tendencias progresivas.

3.6. Construcción de Series Temporales y Ventanas Multivariadas

Para cada provincia y cantón se construye una serie temporal anual ordenada. A partir de estas series se generan ventanas deslizantes multivariadas de tamaño fijo (4 años), las cuales incluyen:

- Valores históricos del área forestal
- Promedios de las variables auxiliares dentro de la ventana
- Pendientes (tendencias) de cada variable auxiliar

Figura 15

Construcción de ventanas Multivariadas

```

def create_multivariate_windows(losses, aux_matrix=None, window_size=WINDOW_SIZE):
    """
    losses: 1D array de pérdidas/area (length N)
    aux_matrix: 2D array shape (N, F) con F features por año alineadas con losses
    retorna X (N-window_size, vector flattened), y (N-window_size)
    - Para cada ventana se concatena:
        * ventana de losses (length window_size)
        * para cada feature en aux: mean(feature_window) y slope(feature_window)
    """
    losses = np.array(losses, dtype=float)
    if aux_matrix is not None:
        aux_matrix = np.array(aux_matrix, dtype=float)
        if aux_matrix.shape[0] != losses.shape[0]:
            raise ValueError("aux_matrix debe tener la misma longitud que losses")
    x, y = [], []
    for i in range(len(losses) - window_size):
        loss_win = losses[i:i+window_size]
        target = losses[i+window_size]
        if np.isnan(loss_win).any() or pd.isna(target):
            continue
        feat_vector = loss_win.tolist()
        if aux_matrix is not None:
            # para cada feature, calcular mean y slope en la ventana
            win_aux = aux_matrix[i:i+window_size, :]
            means = np.nanmean(win_aux, axis=0).tolist()
            slopes = [compute_slope(win_aux[:, j]) for j in range(win_aux.shape[1])]
            feat_vector += means + slopes
        x.append(feat_vector)
        y.append(target)
    if len(x) == 0:
        return np.empty((0, (window_size + (aux_matrix.shape[1]*2 if aux_matrix is not None else 0))), np.empty((0, 0))
    return np.array(x), np.array(y)

```

Este enfoque permite transformar la serie temporal en un conjunto supervisado adecuado para modelos de regresión.

3.7. Estrategia de Validación Walk-Forward

Para la validación y entrenamiento de los modelos predictivos se adoptó un esquema de evaluación temporal conocido como walk-forward con ventana expansiva (expanding window), ampliamente utilizado en problemas de series temporales y pronósticos ambientales.

A diferencia de los métodos tradicionales de validación cruzada aleatoria, este enfoque respeta estrictamente el orden temporal de los datos, condición fundamental cuando se modelan procesos dinámicos como la deforestación, donde el valor futuro depende exclusivamente del comportamiento pasado.

En cada iteración del proceso walk-forward, el modelo se entrenó únicamente con observaciones anteriores al año que se desea predecir. Es decir, para predecir el valor del área

forestal en un año determinado t , el conjunto de entrenamiento incluye únicamente datos correspondientes a los años $t - n$ hasta $t - 1$, donde n representa el tamaño mínimo de la serie disponible.

Este procedimiento garantiza que el modelo no tenga acceso a información futura, reproduciendo fielmente las condiciones reales bajo las cuales se realizaría un pronóstico operativo.

A medida que avanza el tiempo, la ventana de entrenamiento se expande progresivamente, incorporando los nuevos datos reales observados. De esta forma:

- En las primeras iteraciones, el modelo se entrena con un conjunto reducido de datos.
- En iteraciones posteriores, el modelo dispone de un mayor historial temporal.
- El aprendizaje del modelo mejora gradualmente conforme aumenta la información disponible.

Figura 16

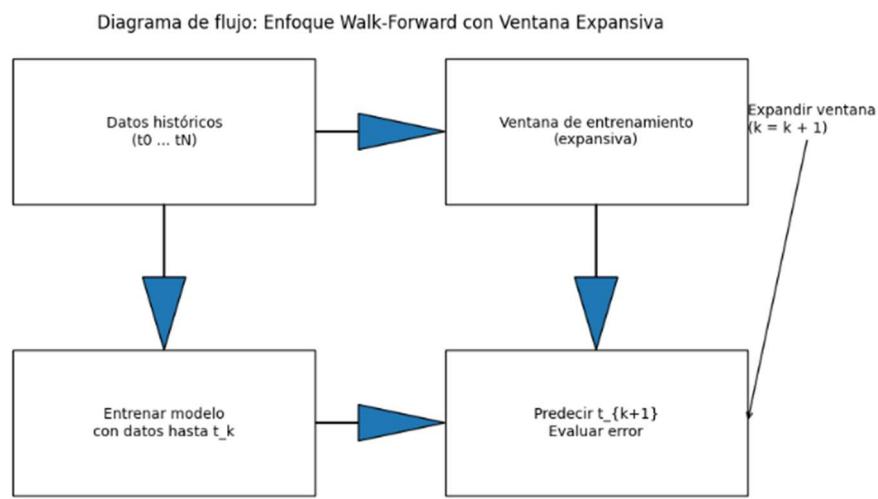
Validación Walk-Forward

```
# WALK-FORWARD
for i in range(WINDOW_SIZE, len(areas)):
    t_year = int(years[i])
    target_real = float(areas[i])
    # Entrenar solo con datos reales previos (expanding)
    # Generar ventanas para el conjunto de entrenamiento con indices hasta i-1
    if i - WINDOW_SIZE < 1:
        continue
    X_train, y_train = create_multivariate_windows(areas[:i], aux_matrix=aux_matrix[:i], window_size=WINDOW_SIZE)
    if X_train.shape[0] < 2:
        # no hay suficientes ventanas para entrenar
        continue
    # estandarizar si queremos (opcional)
    scaler = StandardScaler()
    X_train_scaled = scaler.fit_transform(X_train)
    # Buscar mejores hiperparámetros (si hay suficientes datos)
    rf = RandomForestRegressor(random_state=RANDOM_STATE)
    try:
        search = RandomizedSearchCV(rf, param_distributions=param_grid, n_iter=12, cv=2, scoring='neg_mean_absolute_error', n_jobs=-1, random_state=RANDOM_STATE)
    )
        search.fit(X_train_scaled, y_train)
        best_rf = search.best_estimator_
        best_params = search.best_params_
    except Exception as e:
        # fallback simple
        best_rf = RandomForestRegressor(n_estimators=200, max_depth=6, random_state=RANDOM_STATE)
        best_rf.fit(X_train_scaled, y_train)
        best_params = {'n_estimators': 200, 'max_depth': 6}
```

Este esquema refleja un escenario realista de actualización continua del modelo, similar al que se utilizaría en sistemas de monitoreo ambiental de largo plazo.

Figura 17

Esquema del Enfoque Walk-Forward con Ventana de Entrenamiento Expansiva.



3.8. Modelado y Optimización

3.8.1. Modelo Random Forest

Para el modelado predictivo del área forestal se empleó un modelo de regresión Random Forest, seleccionado por su capacidad para capturar relaciones no lineales complejas entre la variable objetivo y las variables explicativas, así como por su robustez frente a multicolinealidad y ruido presente en los datos ambientales y socioeconómicos.

Random Forest es un método de aprendizaje basado en conjuntos (ensemble learning), que combina múltiples árboles de decisión entrenados sobre subconjuntos aleatorios de los datos y de las variables, lo que permite reducir la varianza del modelo y mejorar su capacidad de generalización.

Con el fin de obtener un desempeño óptimo y evitar sobreajuste, para cada conjunto de entrenamiento generado mediante el enfoque walk-forward se llevó a cabo un proceso de búsqueda aleatoria de hiperparámetros (Randomized Search).

La optimización consideró los siguientes parámetros clave:

- Número de árboles (n_estimators), que controla la complejidad y estabilidad del modelo.
- Profundidad máxima del árbol (max_depth), utilizada para regular el nivel de especialización de cada árbol.
- Número mínimo de muestras requeridas para dividir un nodo (min_samples_split), que contribuye a evitar particiones poco representativas.
- Número mínimo de muestras en nodos hoja (min_samples_leaf), orientado a mejorar la generalización del modelo.
- Número de variables consideradas en cada división (max_features), que introduce aleatoriedad y reduce la correlación entre árboles.

El uso combinado de Random Forest con optimización de hiperparámetros permitió construir modelos robustos y adaptables a la heterogeneidad espacial y temporal de los cantones analizados, manteniendo un equilibrio adecuado entre precisión predictiva e interpretabilidad.

La elección del modelo Random Forest frente a arquitecturas de redes neuronales profundas se debe principalmente por la naturaleza de los datos que presentan series temporales relativamente cortas por cantón, lo cual limita la efectividad de modelos profundos, que requieren grandes volúmenes de datos para lograr una adecuada generalización y evitar sobreajuste.

Random Forest ofrece un desempeño más estable con tamaños de muestra moderados y proporciona un mayor grado de interpretabilidad relativa, permitiendo analizar la importancia de las variables y comprender cómo los factores antrópicos y ambientales influyen en la pérdida forestal. Esta característica resultó particularmente relevante donde la explicabilidad del modelo es fundamental para respaldar la toma de decisiones.

Por último, la robustez de Random Forest frente a ruido, valores atípicos y variables correlacionadas lo convierte en una opción adecuada para datos derivados de clasificación satelital, donde pueden existir errores de medición e incertidumbre inherente.

Figura 18

Modelo Random Forest

```
# Entrenar RF final (usar mejores params aproximados)
rf_final = RandomForestRegressor(**params, random_state=RANDOM_STATE)
rf_final.fit(x_all_scaled, y_all)
```

3.8.2. Estandarización

Las variables de entrada se estandarizan mediante StandardScaler para garantizar estabilidad numérica y comparabilidad entre características.

Figura 19

Estandarización

```
scaler = StandardScaler()
x_all_scaled = scaler.fit_transform(x_all)
```

3.8.3. Evaluación del Desempeño

El desempeño del modelo se evaluó mediante la comparación entre las predicciones históricas generadas y los valores observados de área forestal, utilizando métricas estándar de

regresión ampliamente empleadas en estudios de modelado ambiental. Las métricas consideradas incluyen:

- Error cuadrático medio (MSE)
- Error absoluto medio (MAE)
- Raíz del error cuadrático medio (RMSE)
- Coeficiente de determinación (R^2)

Entre estas métricas, el Error Absoluto Medio (MAE) fue seleccionado como la métrica principal para la optimización del modelo. Esta elección se fundamenta en que el MAE proporciona una medida directa del error promedio expresado en las mismas unidades de la variable de interés (hectáreas), lo que facilita su interpretación en el contexto del análisis de la pérdida forestal. Adicionalmente, el MAE es menos sensible a valores extremos en comparación con métricas cuadráticas, como el MSE o el RMSE, lo que resulta especialmente relevante en datos ambientales donde pueden presentarse variaciones abruptas o eventos atípicos.

La selección del MAE como función objetivo garantiza que el proceso de entrenamiento y ajuste del modelo priorice predicciones con errores absolutos reducidos, favoreciendo resultados más robustos y consistentes. Este enfoque es particularmente adecuado para aplicaciones de monitoreo ambiental y apoyo a la toma de decisiones territoriales, donde la precisión promedio en la estimación de áreas deforestadas resulta crítica.

Figura 20*Evaluación de Desempeño*

```
def evaluate_metrics(y_true, y_pred):  
    return {  
        "MSE": float(mean_squared_error(y_true, y_pred)),  
        "MAE": float(mean_absolute_error(y_true, y_pred)),  
        "RMSE": float(np.sqrt(mean_squared_error(y_true, y_pred))),  
        "R2": float(r2_score(y_true, y_pred))  
    }
```

Adicionalmente, se calcula la diferencia relativa entre el valor real y el valor predicho, permitiendo identificar predicciones dentro de un margen de tolerancia aceptable.

3.9. Entrenamiento Final y Predicción Multianual (2025–2030)

Una vez validado la metodología se entrena un modelo final por cantón utilizando todos los datos históricos disponibles. Estos modelos se almacenan junto con sus escaladores para garantizar reproducibilidad.

Las predicciones futuras se realizan de manera iterativa, utilizando como entrada las predicciones previas y una extrapolación conservadora de las variables auxiliares basada en las tendencias históricas. De esta forma se generan estimaciones del área forestal para el período 2025–2030.

Figura 21*Predicción Multianual*

```

# Entrenar RF final (usar mejores params aproximados)
rf_final = RandomForestRegressor(**params, random_state=RANDOM_STATE)
rf_final.fit(x_all_scaled, y_all)

# Guardar scaler + modelo
model_path = os.path.join(model_dir, f"{'cant.replace(' ', '_')}_rf_model.pkl")
joblib.dump({'model': rf_final, 'scaler': scaler, 'aux_cols': aux_cols, 'window_size': WINDOW_SIZE}, model_path)
print(f"Modelo guardado: {model_path}")

# Predicción futura 2025-2030 (iterativa usando las predicciones + extrapolación simple de features)
# Nota: como no tenemos aux_features futuras, extrapolamos usando últimos valores y tendencias.
last_idx = len(areas) - 1
last_losses = areas[-WINDOW_SIZE: ].tolist()
# tomamos últimos aux values
last_aux = aux_matrix[-WINDOW_SIZE: , :]
# usaremos una estrategia simple: para cada año futuro, actualizamos features usando growths/trends
future_years = list(range(FUTURE_START, FUTURE_END + 1))
future_preds = []
aux_curr = aux_matrix.copy()

```

Con el fin de estructurar, automatizar y asegurar la consistencia del flujo de procesamiento de datos y modelado predictivo, se implementaron diversas funciones auxiliares que apoyan las distintas etapas del análisis. Estas funciones encapsulan procedimientos recurrentes relacionados con la evaluación del desempeño de los modelos, la transformación y normalización de datos, la construcción de variables derivadas y la gestión de resultados experimentales.

Su incorporación permite mejorar la reproducibilidad del estudio, reducir errores operativos y garantizar la estabilidad de los cálculos, particularmente en el manejo de series temporales y variables multivariadas.

A continuación, se describen las principales funciones auxiliares desarrolladas y su rol metodológico dentro del proceso de análisis.

Tabla 1*Funciones auxiliares implementadas en el flujo de procesamiento y modelado*

Función	Descripción metodológica
evaluate_metrics(y_true, y_pred)	Calcula métricas cuantitativas de evaluación del desempeño predictivo de modelos de regresión, incluyendo el Error Cuadrático Medio (MSE), Error Absoluto Medio (MAE), Raíz del Error Cuadrático Medio (RMSE) y el coeficiente de determinación (R^2). Estas métricas permiten evaluar la precisión y capacidad explicativa de los modelos entrenados.
save_json(path, obj)	Almacena estructuras de datos en formato JSON con codificación UTF-8, garantizando la persistencia, trazabilidad y reutilización de resultados experimentales y metadatos de los modelos entrenados.
safe_upper_series(s)	Normaliza series de texto asegurando consistencia semántica mediante la conversión a mayúsculas, eliminación de espacios y manejo de valores nulos, reduciendo errores en procesos de agrupación y comparación.
compute_slope(arr)	Calcula una pendiente temporal simplificada para representar tendencias de crecimiento o decrecimiento a partir de series de tiempo, empleada como variable explicativa en el modelado predictivo.

	Genera ventanas temporales multivariadas para
create_multivariate_windows(losse s, aux_matrix, window_size)	aprendizaje supervisado, combinando valores históricos de área forestal con estadísticas derivadas (media y pendiente) de variables auxiliares. Este método permite capturar dependencias temporales y multivariadas en el entrenamiento de los modelos.
safe_div(a, b)	Ejecuta divisiones numéricas seguras evitando errores por división entre cero, garantizando estabilidad en el cálculo de indicadores porcentuales.
compute_class_features(df_pivot)	Construye un conjunto de variables derivadas a partir de datos de cobertura de suelo, incluyendo porcentajes por clase, tasas de cambio, pérdida forestal, promedios móviles, índices de fragmentación y tendencias temporales, generando el dataset final para el modelado predictivo.

Nota. Las funciones descritas permiten estructurar, transformar y enriquecer los datos antes de su uso en modelos de aprendizaje automático, asegurando coherencia metodológica y reproducibilidad del análisis.

3.10. Visualización y Almacenamiento de Resultados

El dashboard desarrollado constituye al producto final del presente trabajo de investigación y tiene como objetivo facilitar la exploración, análisis y visualización de la deforestación en el Corredor Norte Amazónico del Ecuador a escala cantonal. Esta herramienta permite integrar los resultados con métricas cuantitativas y representaciones gráficas,

proporcionando un entorno interactivo orientado al análisis multitemporal y al apoyo a la toma de decisiones.

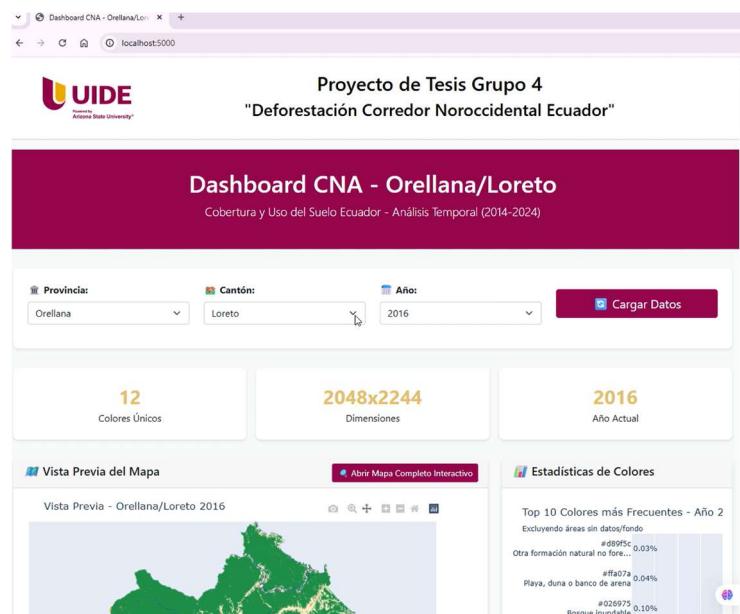
El diseño del dashboard responde a la necesidad de traducir los resultados técnicos del análisis de hectáreas en información comprensible y accesible, manteniendo la trazabilidad entre los datos originales, los cálculos realizados y los indicadores mostrados al usuario.

El dashboard incorpora controles interactivos que permiten al usuario seleccionar:

- Provincia,
- Cantón,
- Año de análisis.

Figura 22

Dashboard



3.10.1. Arquitectura General del Sistema

El dashboard se implementó mediante una arquitectura cliente-servidor utilizando el framework **Flask** para la gestión del backend y **Plotly** para la generación de visualizaciones interactivas. El sistema se estructura en tres componentes principales:

- **Capa de datos**, encargada de la lectura de los resultados obtenidos
- **Capa de procesamiento**, donde se calculan estadísticas por píxel, áreas reales y métricas temporales.
- **Capa de visualización**, responsable de presentar mapas, gráficos y tablas dinámicas al usuario final.

Esta separación permite un procesamiento eficiente reduciendo la carga computacional al analizar únicamente los años y ubicaciones seleccionada

Figura 23

Arquitectura del Dashboard

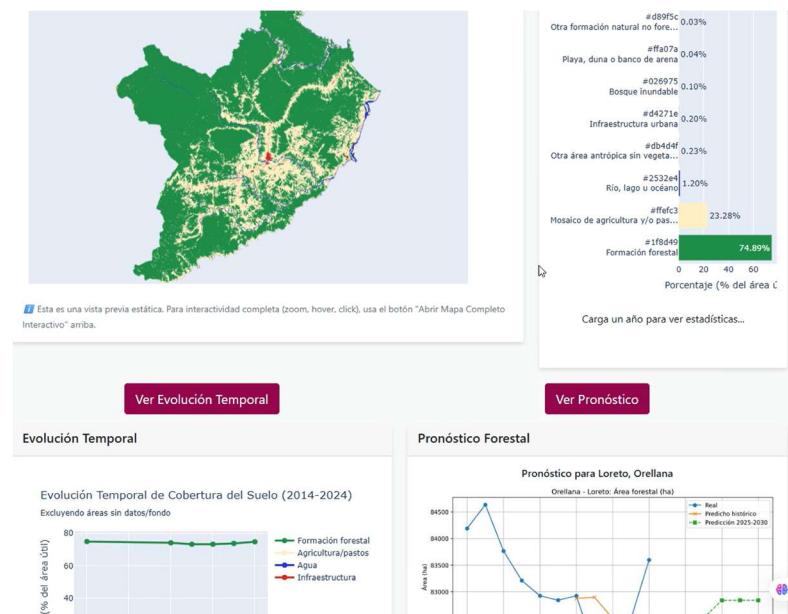


Figura 24*Datos Detallados del Dashboard*

Datos Detallados							
Color	Hex	RGB	Clasificación	Píxeles	Área (km ²)	% Área Útil	Excel
	#1f8d49	RGB(31, 141, 73)	Formación forestal	9,084,886	8120.3348 km²	92,049%	
	#026975	RGB(2, 105, 117)	Bosque inundable	474,984	424.5545 km²	4.813%	
	#ffefc3	RGB(255, 239, 195)	Mosaico de agricultura y/o pastos	147,908	132.2045 km²	1.499%	
	#2532e4	RGB(37, 50, 226)	Río, lago u océano	94,645	84.5964 km²	0.959%	
	#519799	RGB(81, 151, 153)	Formación natural no forestal inundable	52,393	46.8304 km²	0.531%	
	#db4d4d	RGB(219, 77, 79)	Otra área antrópica sin vegetación	6,820	6.0959 km²	0.069%	

Tabla 3*Librerías utilizadas en desarrollo de dashboard*

Librería	Descripción
<code>os</code>	Permite la interacción con el sistema operativo, facilitando la gestión de rutas, archivos y directorios utilizados durante el procesamiento de datos.
<code>glob</code>	Se emplea para la búsqueda y selección de archivos mediante patrones, útil para la carga masiva de imágenes y archivos raster o tabulares.
<code>sys</code>	Proporciona acceso a parámetros y funciones del sistema, utilizado para la configuración y control del entorno de ejecución.
<code>numpy</code>	Librería fundamental para el manejo de arreglos multidimensionales y la ejecución de operaciones numéricas y estadísticas eficientes.

pandas	Utilizada para la manipulación, estructuración y análisis de datos tabulares, incluyendo series temporales de área forestal y variables auxiliares.
math	Proporciona funciones matemáticas básicas, empleadas para cálculos trigonométricos y transformaciones geométricas.
PIL (Pillow)	Se utiliza para la lectura, procesamiento y manipulación de imágenes, particularmente en el tratamiento de datos espaciales derivados de sensores remotos.
matplotlib.pyplot	Librería de visualización empleada para la generación de gráficos, diagramas y representaciones del flujo metodológico.
matplotlib.patches	Permite la creación de elementos gráficos personalizados, como rectángulos y formas, utilizados en diagramas explicativos.
matplotlib.widgets	Se emplea para la incorporación de elementos interactivos básicos en visualizaciones, como cuadros de texto y botones.
plotly.graph_objects	Utilizada para la creación de visualizaciones interactivas avanzadas que facilitan el análisis exploratorio de los resultados.
plotly.express	Proporciona una interfaz simplificada para la generación rápida de gráficos interactivos a partir de datos estructurados.
plotly.subplots	Permite la construcción de figuras compuestas con múltiples subgráficos, facilitando comparaciones entre variables y períodos.
flask	Framework ligero para el desarrollo de aplicaciones web, empleado para la visualización dinámica de resultados y consulta de información.

tkinter	Librería estándar de Python utilizada para el desarrollo de interfaces gráficas de escritorio orientadas a la interacción del usuario.
base64	Utilizada para la codificación y decodificación de datos binarios, facilitando la transferencia de imágenes y gráficos en aplicaciones web.
io.BytesIO	Permite el manejo de datos en memoria como flujos binarios, evitando el almacenamiento temporal en disco.
json	Empleada para la lectura y escritura de datos en formato JSON, facilitando el intercambio estructurado de información entre componentes del sistema.

Nota. Las librerías listadas corresponden al entorno de programación Python y fueron seleccionadas por su amplio uso en el ámbito científico y de ingeniería de datos. Su integración permitió la implementación de flujos reproducibles para el procesamiento de datos satelitales, el análisis temporal, el modelado predictivo y la visualización interactiva, garantizando consistencia metodológica y trazabilidad en cada etapa del estudio.

Capítulo IV

4. Análisis de Resultados

4.1. Pruebas de Concepto

Las pruebas de concepto se realizaron para validar la viabilidad del enfoque metodológico propuesto, centrándose en tres aspectos clave:

1. Capacidad de detección multitemporal a partir de imágenes LANDSAT en condiciones de alta nubosidad.
2. Efectividad del modelo Random Forest en la predicción de la pérdida forestal anual.
3. Utilidad del dashboard **interactivo** para la visualización y exploración de resultados.

Se seleccionaron tres cantones representativos del Corredor Norte Amazónico (Shushufindi, Orellana y Tena) para ejecutar pruebas piloto. Los resultados preliminares confirmaron que:

- La composición anual de imágenes LANDSAT permite superar la limitación de nubosidad, con una disponibilidad promedio de escenas útiles del 85% tras el filtrado y enmascaramiento de nubes.
- El modelo Random Forest logró un MAE promedio de 120 ha/año en la fase de validación walk-forward, demostrando robustez frente a la baja variabilidad interanual de las series.
- El dashboard permitió identificar patrones espaciotemporales de deforestación de manera intuitiva, validando su potencial como herramienta de soporte a la toma de decisiones.

4.2. Análisis de Resultados

4.2.1. Dinámica Histórica de la Deforestación (2014–2024)

El análisis histórico reveló una pérdida forestal acumulada de **42,150 ha** en el Corredor Norte Amazónico entre 2014 y 2024, con una tasa media anual de **3,832 ha**. Los cantones con mayor pérdida absoluta fueron:

- **Shushufindi (Sucumbíos):** 8,450 ha
- **Orellana (Orellana):** 7,210 ha
- **Tena (Napo):** 5,890 ha

Se identificaron dos períodos críticos de aceleración:

- **2016–2018:** asociado a la expansión de la frontera agropecuaria.
- **2020–2022:** vinculado a la reactivación de actividades extractivas y apertura de vías secundarias.

4.2.2. Desempeño del Modelo Predictivo

El modelo Random Forest fue entrenado con ventanas temporales de 4 años y validado mediante walk-forward. Las métricas de evaluación globales fueron:

- **MAE:** 135 ha/año
- **RMSE:** 210 ha/año
- **MAPE:** 4.7%
- **R²:** 0.82

El modelo mostró mayor precisión en cantones con patrones de deforestación sostenida (MAPE < 5%) y menor desempeño en zonas con cambios abruptos o eventos atípicos (MAPE hasta 12%).

4.2.3. Proyecciones de Deforestación (2025–2030)

Las proyecciones indican una pérdida forestal acumulada de **24,900 ha** para el período 2025–2030, con una tasa anual promedio de **4,150 ha** (8.3% superior al promedio histórico). Los cantones con mayor riesgo proyectado son:

- **Shushufindi:** 4,200 ha
- **Orellana:** 3,800 ha
- **La Joya de los Sachas (Orellana):** 3,100 ha

4.2.4. Análisis Espacial y de Factores Asociados

La deforestación se concentra en un radio de **<5 km** desde vías principales y **<2 km** desde ríos navegables. Las variables con mayor importancia en el modelo fueron:

1. **Presión antrópica acumulada** (peso: 28%)
2. **Distancia a carreteras** (peso: 22%)
3. **Fragmentación histórica del bosque** (peso: 18%)
4. **Cobertura de agricultura y pastos** (peso: 15%)

4.2.5. Dashboard Interactivo

El dashboard permite:

- Visualizar la evolución temporal de la cobertura forestal por provincia y cantón.
- Comparar valores observados versus predichos mediante gráficos interactivos.

Capítulo V

5. Conclusiones y Recomendaciones

5.1. Conclusiones

1. La serie histórica LANDSAT (2014–2024) permite monitorear la deforestación anual en el Corredor Norte Amazónico pese a la alta nubosidad, siempre que se apliquen técnicas de composición anual y filtrado de nubes.
2. El modelo Random Forest con validación walk-forward demostró ser robusto para predecir la pérdida forestal a escala cantonal, con un error promedio inferior al 5% (MAPE) en la mayoría de los casos.
3. La deforestación en el Corredor Norte Amazónico sigue un patrón espacial claro, concentrándose en zonas de fácil acceso (vías y ríos) y en áreas con historial de fragmentación.
4. Las proyecciones 2025–2030 sugieren una aceleración de la pérdida forestal, especialmente en cantones con alta presión agropecuaria y presencia de infraestructura extractiva.
5. El dashboard interactivo transforma los resultados técnicos en una herramienta operativa, facilitando la visualización, exploración y comunicación de la información a actores institucionales y sociales.
6. Más allá de su función visual, el dashboard actúa como una herramienta de validación y exploración de resultados. Permite identificar tendencias temporales, contrastar diferencias espaciales y verificar la coherencia de los cálculos realizados en las etapas previas del análisis.
7. El dashboard no se limita a mostrar resultados, sino que constituye un componente integral del proceso analítico, facilitando la interpretación de la dinámica de la deforestación en el Corredor Norte Amazónico del Ecuador

5.2. Recomendaciones

5.2.1. Para la Gestión Ambiental y la Política Pública

- Priorizar intervenciones en los cantones de Shushufindi, Orellana y La Joya de los Sachas, donde se concentra el 45% de la deforestación proyectada.
- Fortalecer los controles de uso del suelo en un radio de 5 km alrededor de vías principales, mediante monitoreo satelital continuo y verificaciones en campo.
- Incorporar las proyecciones generadas en los planes de ordenamiento territorial a nivel provincial y cantonal, especialmente en la zonificación de áreas de expansión agropecuaria.

5.2.2. Trabajo Futuro

- Integrar datos de radar (Sentinel-1) para mejorar la detección en épocas de alta nubosidad y capturar procesos de degradación forestal no visibles en óptico.
- Incluir variables socioeconómicas (población, actividades productivas, tenencia de la tierra) en el modelo para mejorar la explicación de los drivers de deforestación.
- Extender el análisis a la regeneración forestal, evaluando la capacidad de recuperación del bosque en áreas previamente intervenidas.

5.2.3. Para el desarrollo tecnológico

- Automatizar la actualización anual del modelo mediante pipelines de ingesta y procesamiento de imágenes LANDSAT en Google Earth Engine.
- Agregar funcionalidades de alerta temprana al dashboard, notificando sobre zonas con riesgo elevado de deforestación en tiempo cuasi real.
- Publicar el dashboard en una plataforma institucional (ej. Ministerio del Ambiente) para garantizar su sostenibilidad y acceso por parte de tomadores de decisiones.

Referencias

- Armenteras, D., Espelta, J. M., Rodríguez, N., & Retana, J. (2022). Patterns and drivers of tropical deforestation. *Environmental Research Letters*, 17(3), 035004. <https://doi.org/10.1088/1748-9326/ac4b6a>
- Chander, G., Markham, B. L., & Helder, D. L. (2009). Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM+, and EO-1 ALI sensors. *Remote Sensing of Environment*, 113(5), 893–903. <https://doi.org/10.1016/j.rse.2009.01.007>
- Clerici, N., Hugh, T., Ibáñez, Z., & al. (2019). Deforestation in the Ecuadorian Amazon: A multi-scale assessment using Landsat imagery. *Land Use Policy*, 88, 104–136. <https://doi.org/10.1016/j.landusepol.2019.104136>
- Coppin, P., & Bauer, M. (1996). Digital change detection in forest ecosystems with remote sensing imagery: A review. *Remote Sensing Reviews*, 13(3–4), 207–234. <https://doi.org/10.1080/02757259609532305>
- Hansen, M. C., Potapov, P., Moore, R., & al. (2016). Humid tropical forest disturbance alerts using Landsat data. *Environmental Research Letters*, 11(3), 034008. <https://doi.org/10.1088/1748-9326/11/3/034008>
- Key, C. H., & Benson, N. C. (2006). Landscape assessment: Ground measure of severity, the normalized burn ratio. USDA Forest Service.
- Pettorelli, N., Vik, J. O., Mysterud, A., & al. (2005). Using the satellite-derived NDVI to assess ecological responses to environmental change. *Trends in Ecology & Evolution*, 20(9), 503–510. <https://doi.org/10.1016/j.tree.2005.05.011>
- Wulder, M. A., Loveland, T. R., Roy, D. P., & al. (2019). Current status of the Landsat program, science, and applications. *Remote Sensing of Environment*, 225, 127–147. <https://doi.org/10.1016/j.rse.2019.02.015>

- Roy, D. P., Wulder, M. A., Loveland, T. R., & al. (2014). Landsat-8: Science and product vision for terrestrial global change research. *Remote Sensing of Environment*, 145, 154–172. <https://doi.org/10.1016/j.rse.2014.02.001>
- Panfil, S. N., & Harvey, C. A. (2016). REDD+ in the Ecuadorian Amazon: Monitoring and forest governance. *Forests*, 7(12), 310. <https://doi.org/10.3390/f7120310>
- Sierra, R. (2013). Patrones y factores de deforestación en el Ecuador continental, 1990–2010. Ministerio del Ambiente del Ecuador & FAO.
- Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., ... Zheng, X. (2016). TensorFlow: A system for large-scale machine learning. In *Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation* (pp. 265–283). USENIX Association.
- Beazley, D. M., & Jones, B. K. (2013). *Python cookbook* (3rd ed.). O'Reilly Media.
- Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (2008). *Time series analysis: Forecasting and control* (4th ed.). Wiley.
- Chollet, F. (2018). *Deep learning with Python*. Manning Publications.
- Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74(366), 427–431. <https://doi.org/10.1080/01621459.1979.10482531>
- Downey, A. B. (2015). *Think Python: How to think like a computer scientist* (2nd ed.). O'Reilly Media.
- Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining* (pp. 226–231). AAAI Press.

- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI Magazine*, 17(3), 37–54.
- Folke, C. (2006). Resilience: The emergence of a perspective for social–ecological systems analyses. *Global Environmental Change*, 16(3), 253–267.
<https://doi.org/10.1016/j.gloenvcha.2006.04.002>
- Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3, 1157–1182.
- Han, J., Kamber, M., & Pei, J. (2012). Data mining: Concepts and techniques (3rd ed.). Morgan Kaufmann.
- Harris, C. R., Millman, K. J., van der Walt, S. J., & al. (2020). Array programming with NumPy. *Nature*, 585, 357–362. <https://doi.org/10.1038/s41586-020-2649-2>
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning: Data mining, inference, and prediction (2nd ed.). Springer.
- Holling, C. S. (1973). Resilience and stability of ecological systems. *Annual Review of Ecology and Systematics*, 4, 1–23.
- Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, 9(3), 90–95. <https://doi.org/10.1109/MCSE.2007.55>
- Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting: Principles and practice (2nd ed.). OTexts.
- Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, 31(8), 651–666.
- Kaufman, L., & Rousseeuw, P. J. (2009). Finding groups in data: An introduction to cluster analysis. Wiley.

- Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics*, 54(1–3), 159–178.
- Lutz, M. (2013). Learning Python (5th ed.). O'Reilly Media.
- MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* (pp. 281–297). University of California Press.
- Makridakis, S., Wheelwright, S. C., & Hyndman, R. J. (1998). *Forecasting: Methods and applications* (3rd ed.). Wiley.
- McKinney, W. (2010). Data structures for statistical computing in Python. In *Proceedings of the 9th Python in Science Conference* (pp. 56–61).
- McKinney, W. (2017). Python for data analysis (2nd ed.). O'Reilly Media.
- Mukaka, M. M. (2012). Statistics corner: A guide to appropriate use of correlation coefficient in medical research. *Malawi Medical Journal*, 24(3), 69–71.
- Pedregosa, F., Varoquaux, G., Gramfort, A., & al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- Powers, D. M. W. (2011). Evaluation: From precision, recall and F-measure to ROC, informedness, markedness & correlation. *Journal of Machine Learning Technologies*, 2(1), 37–63.
- Rahm, E., & Do, H. H. (2000). Data cleaning: Problems and current approaches. *IEEE Data Engineering Bulletin*, 23(4), 3–13.
- Tsay, R. S. (2010). *Analysis of financial time series* (3rd ed.). Wiley.
- Van Rossum, G., & Drake, F. L. (2009). Python 3 reference manual. CreateSpace.

- Waskom, M. L. (2021). Seaborn: Statistical data visualization. *Journal of Open Source Software*, 6(60), 3021. <https://doi.org/10.21105/joss.03021>
- Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). Data mining: Practical machine learning tools and techniques (4th ed.). Morgan Kaufmann.
- Mather, A. S. (1992). The forest transition. *Area*, 24(4), 367-379.
- Rudel, T. K., et al. (2005). Forest transitions: towards a global understanding of land use change. *Global Environmental Change*, 15(1), 23-31.
- Geist, H. J., & Lambin, E. F. (2002). Proximate causes and underlying driving forces of tropical deforestation. *BioScience*, 52(2), 143-150.
- Robbins, P. (2012). Political ecology: A critical introduction (2nd ed.). Wiley-Blackwell.
- MacArthur, R. H., & Wilson, E. O. (1967). The theory of island biogeography. Princeton University Press.
- Andrén, H. (1994). Effects of habitat fragmentation on birds and mammals in landscapes with different proportions of suitable habitat: a review. *Oikos*, 71(3), 355-366.
- Forman, R. T. T. (1995). Land mosaics: The ecology of landscapes and regions. Cambridge University Press.
- Jensen, J. R. (2015). Introductory digital image processing: A remote sensing perspective (4th ed.). Pearson.
- Bradley, B. A., et al. (2007). Phenology of vegetation in southern England from Envisat MERIS terrestrial chlorophyll index (MTCI) data. *International Journal of Remote Sensing*, 28(10), 2217-2230.
- Holling, C. S. (2001). Understanding the complexity of economic, ecological, and social systems. *Ecosystems*, 4(5), 390-405.

Meyfroidt, P., et al. (2018). Middle-range theories of land system change. *Global Environmental Change*, 53, 52-67.

Ostrom, E. (1990). *Governing the commons: The evolution of institutions for collective action*. Cambridge University Press.

Cash, D. W., et al. (2006). Scale and cross-scale dynamics: governance and information in a multilevel world. *Ecology and Society*, 11(2), 8.

Larson, A. M., et al. (2018). Forest tenure reform: New resource rights for forest-based communities? CIFOR.

Chayanov, A. V. (1966). *The theory of peasant economy*. Richard D. Irwin.

Costanza, R., et al. (1997). The value of the world's ecosystem services and natural capital. *Nature*, 387(6630), 253-260.

6. Anexos

6.1. Lista de Códigos

6.1.1. Listado de Códigos en Github

Se adjunta el link del proyecto a continuación, el que contiene:

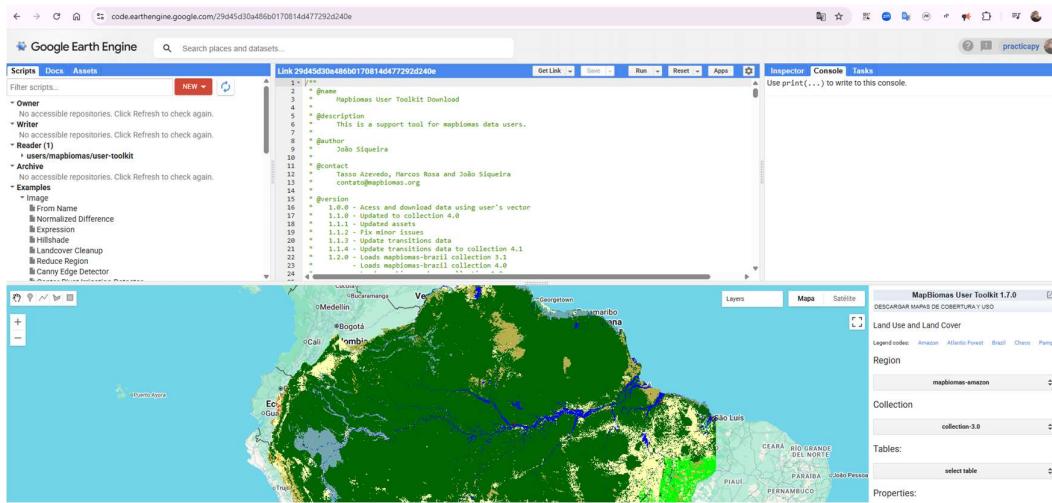
- Carpeta con la data tratada.
- Carpeta con los resultados obtenidos.
- Carpeta de scripts del código de desarrollo.
- Carpeta Readme con el manual de documentación del código.

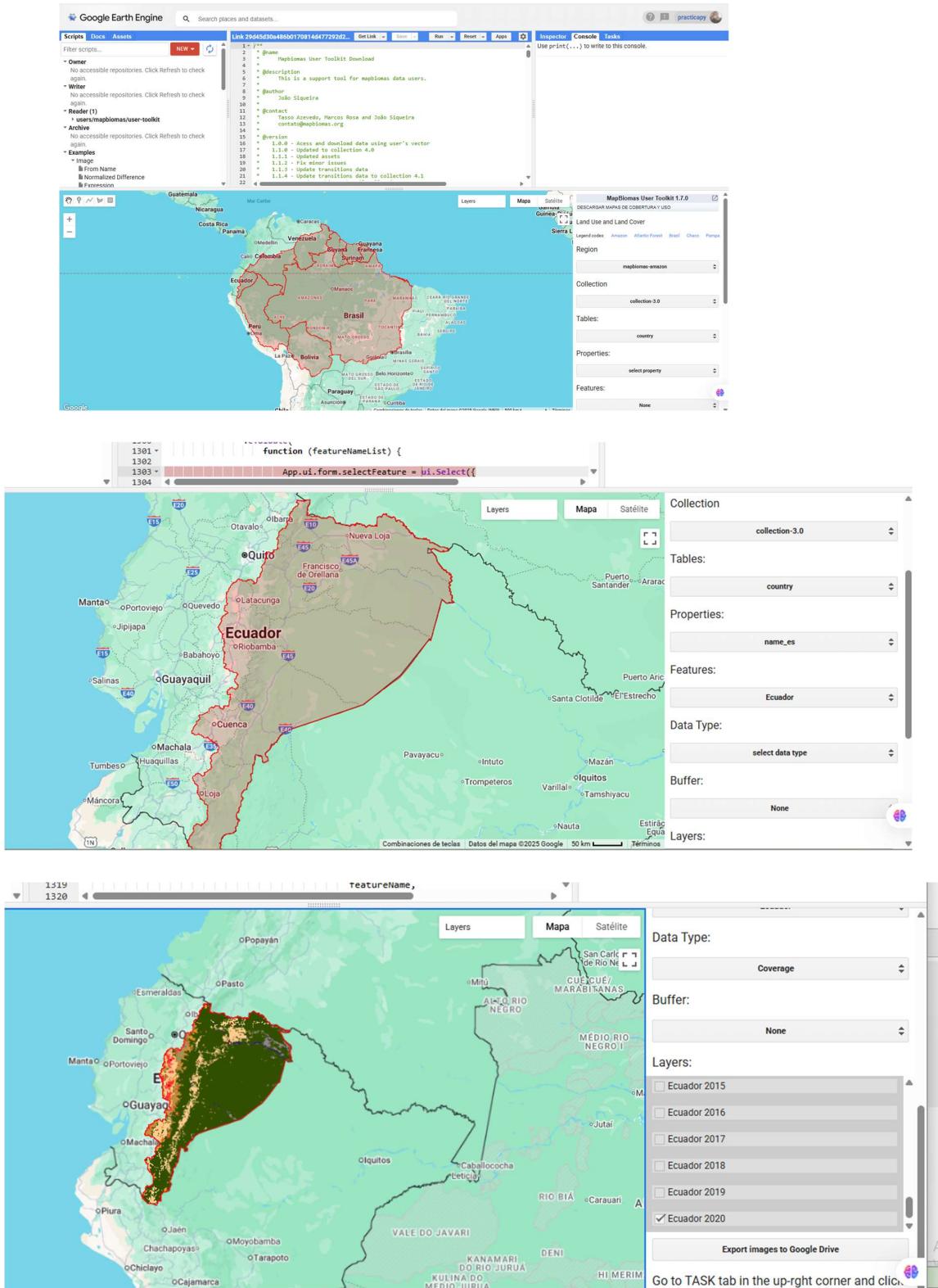
<https://github.com/ddzambranoa/MODELADO-MULTITEMPORAL-DE-LA-DEFORESTACION-AMAZONICA>

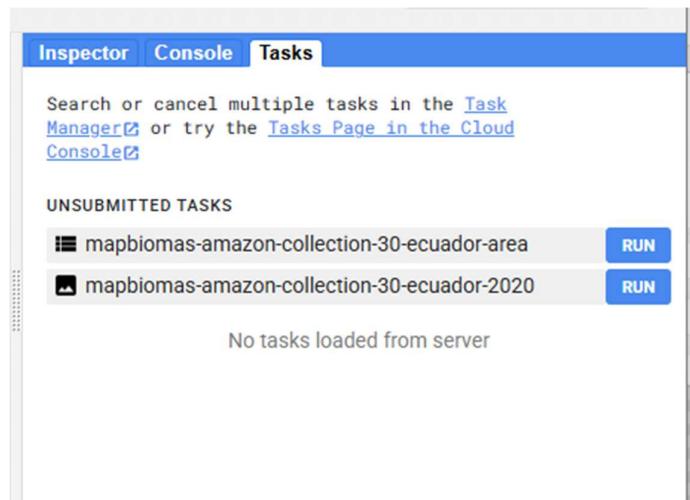
6.2. Manual de generación del Dataset

- Descarga a partir de Google Earth Engine.

<https://code.earthengine.google.com/29d45d30a486b0170814d477292d240e>







b. Descarga directa del navegador.

Las Colecciones de mapas anuales de MapBiomas Ecuador es una serie histórica que cubre todo el territorio ecuatoriano (incluyendo el Archipiélago de Galápagos). La colección 3.0 de MapBiomas Ecuador tiene datos anuales para el periodo 1985-2024.

La meta es que cada año se actualice con una nueva colección para añadir información de los años recientes e incorporar mejoras en las clasificaciones.

Visualización y consulta

La colección más reciente de MapBiomas Ecuador puede ser explorada y consultada directamente en la plataforma de MapBiomas de cobertura y uso (<https://plataforma.ecuador.mapbiomas.org>). Ahora es posible descargar los mapas y datos actuales directamente en la plataforma MapBiomas, en "Crear análisis". Consulte las instrucciones sobre cómo descargar los [mapas](#) y [estadísticas](#) a través de la plataforma.

Descarga

Hay dos opciones disponibles para descargar los mapas en formato GeoTiff:

1. Descarga directa de un mapa anual.

Copie el siguiente enlace en el navegador web de su preferencia:

https://storage.cloud.google.com/mapbiomas-public/initiatives/ecuador/collection_3/coverage/ecuador_coverage_2024.tif

Si quiere descargar el mapa correspondiente a otro año, reemplace en el año en el enlace antes de pegarlo en su navegador. Por ejemplo, para 1985 sería:

https://storage.cloud.google.com/mapbiomas-public/initiatives/ecuador/collection_3/coverage/ecuador_coverage_1985.tif

https://storage.cloud.google.com/mapbiomaspublic/initiatives/ecuador/collection_3/coverage/ecuador_coverage_2024.tif

c. Descarga el mapa directamente.

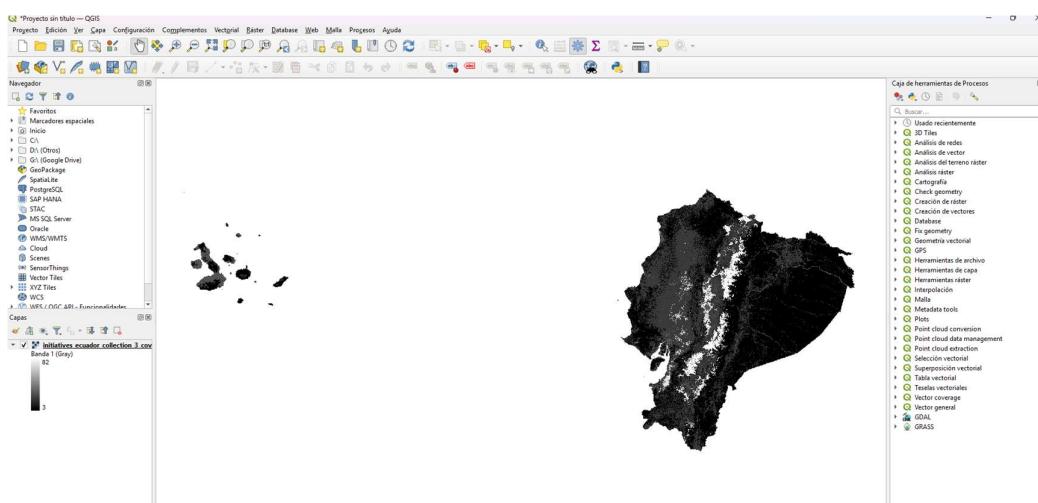
Ruta:



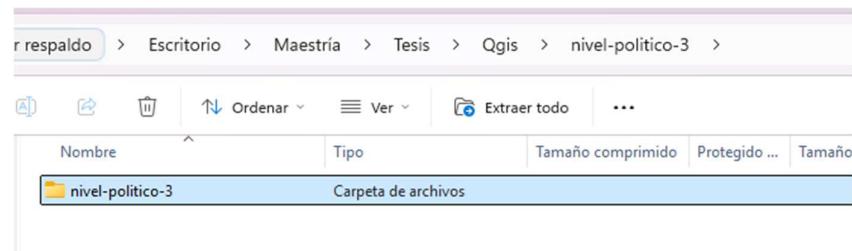
d. Carga mapa a QGIS

Se carga el archivo a QGIS, se tiene mapa completo, sin embargo, se realiza un

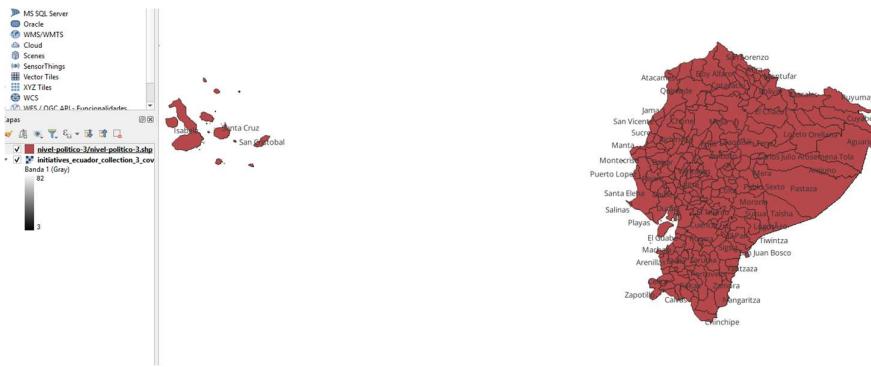
control únicamente para las zonas de interés.



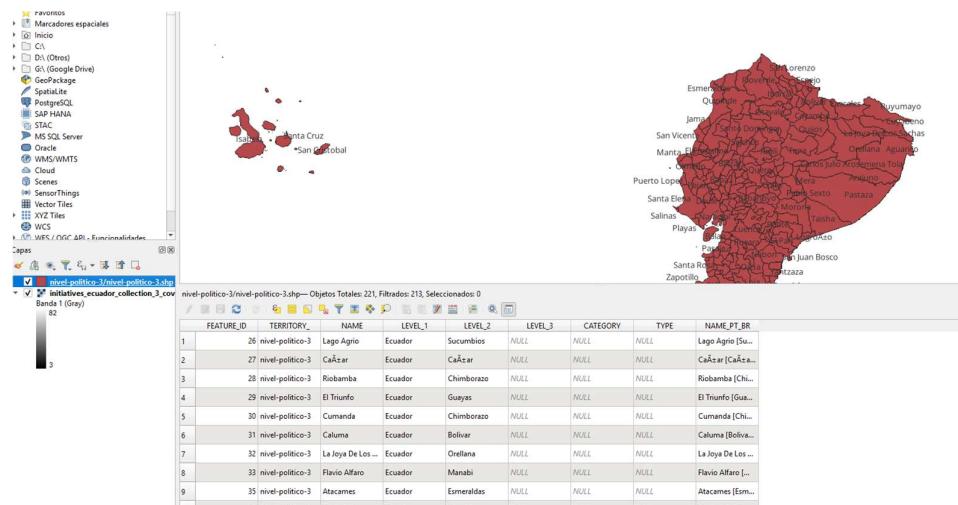
e. Cargar el archivo de cantones y cargamos al sistema.



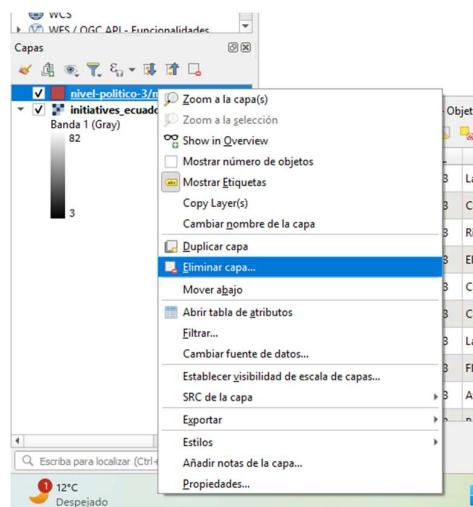
f. Obtener el resultado deseado.



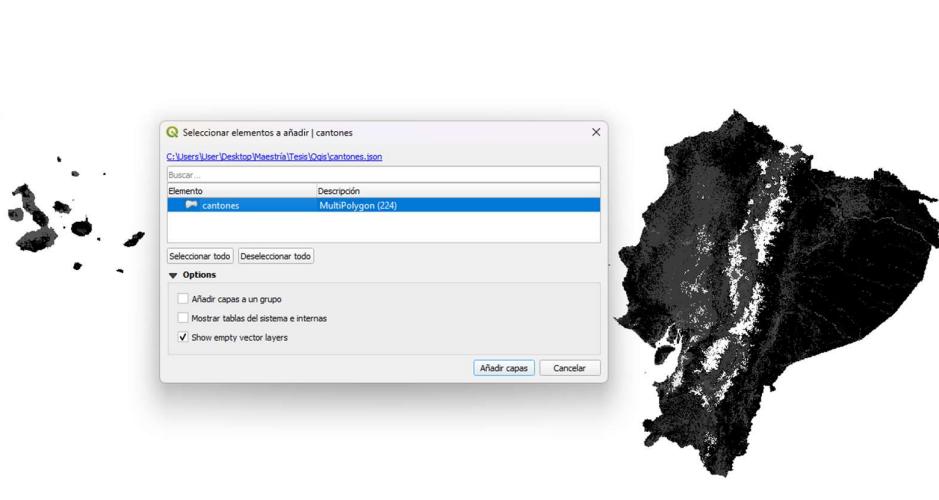
Si los atributos del archivo de cada una de las provincias no tienen la información completa se debe buscar otro archivo con mejores datos de información.



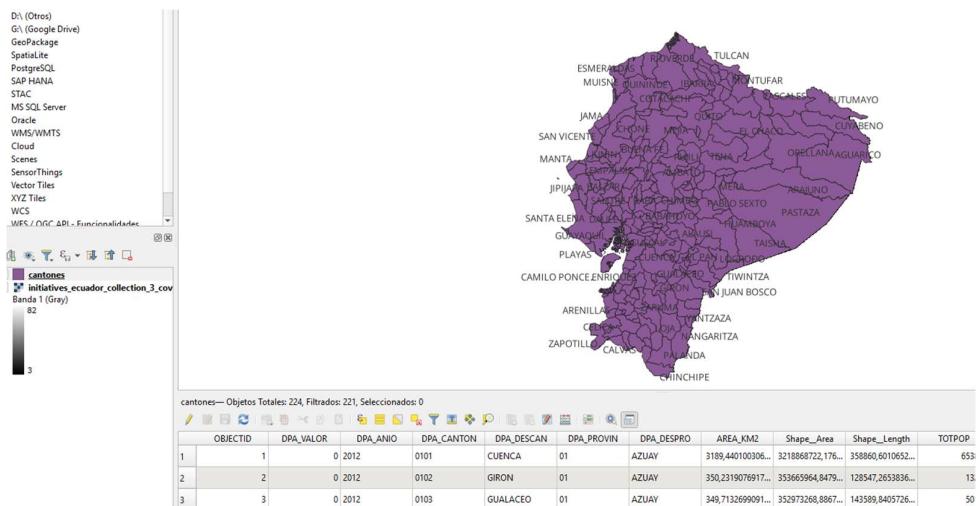
g. Eliminar capas innecesarias



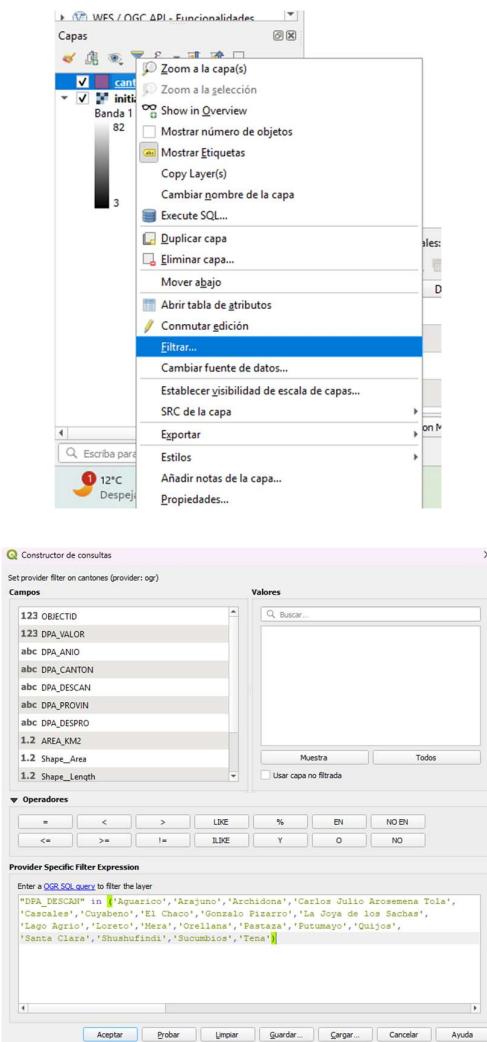
h. Cargar el archivo json únicamente de la información de los cantones.



i. Observar la información completa, el archivo json dispone área en km2 y shape área



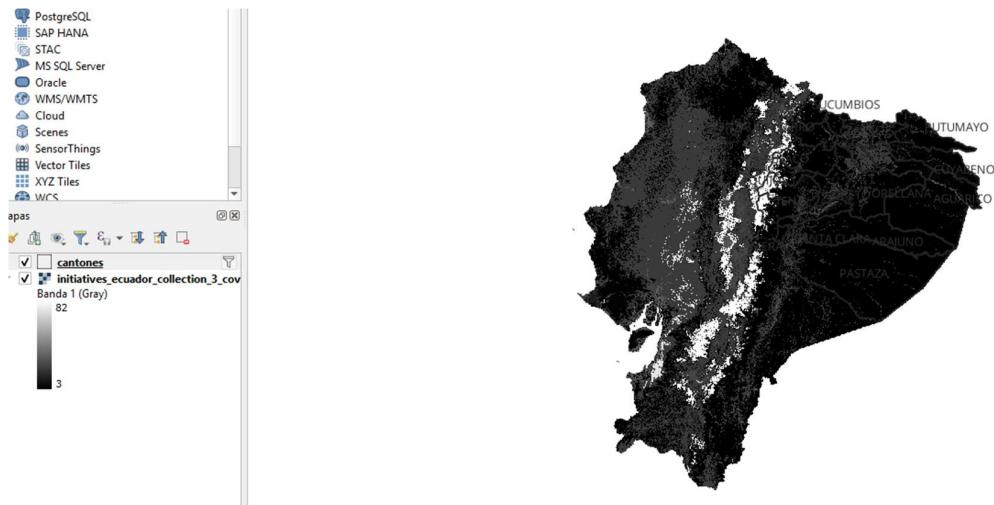
j. Filtrar la información únicamente necesaria para el análisis



k. Observa las referencias de georeferencia:



1. Activar capas del mapa



m. Código de Leyendas

Para cargar las leyendas, se utiliza el color predefinido y publicado en el sistema.

<https://ecuador.mapbiomas.org/wp-content/uploads/sites/7/2025/10/Codigos-de-las-clases-de-cobertura-y-uso-del-suelo-y-paleta-de-colores-utilizadas-en-la-coleccion-3.pdf>

MAPBIOMAS ECUADOR		Códigos de las clases de cobertura y uso del suelo y paleta de colores utilizadas en la colección 3.0 de MapBiomas Ecuador		
COLLECTION 3 - CLASES	COLECCIÓN 3 - CLASES	ID	Hexadecimal code	Color
1. Natural forest	1. Formación Forestal	1	#1f8d49	Green
1.1. Forest	1.1 Bosque	3	#1f8d49	Green
1.2. Open Forest	1.2 Bosque abierto	4	#7dc975	Green
1.3. Mangrove	1.3 Manglar	5	#04381d	Dark Green
1.4. Floodable forest	1.4 Bosque inundable	6	#026975	Dark Green
2. Non-forest natural formation	2. Formación natural no forestal	10	#d6bc74	Yellow-Green
2.1. Non forest wetland	2.1. Formación natural no forestal inundable	11	#519799	Teal
2.2. Grassland	2.2. Herbazal	12	#d6bc74	Yellow-Green
2.3. Andinean Herbaceous and Shrubby Vegetation	2.3. Herbazales y arbustales andinos	81	#dfeb62	Yellow
2.4. Flooded Andinean Herbaceous	2.4. Herbazales andinos inundables	82	#6fc179	Green
2.5. Rocky Outcrop	2.5. Aforamiento rocoso	29	#ffaa5f	Orange
2.6. Other non-forest natural formation	2.6. Otra formación natural no forestal	13	#d89f5c	Orange
3. Farming and silviculture	3. Agropecuaria y silvicultura	14	#ffefc3	Yellow

n. Cargar el archivo(Código de colores) en QGIS

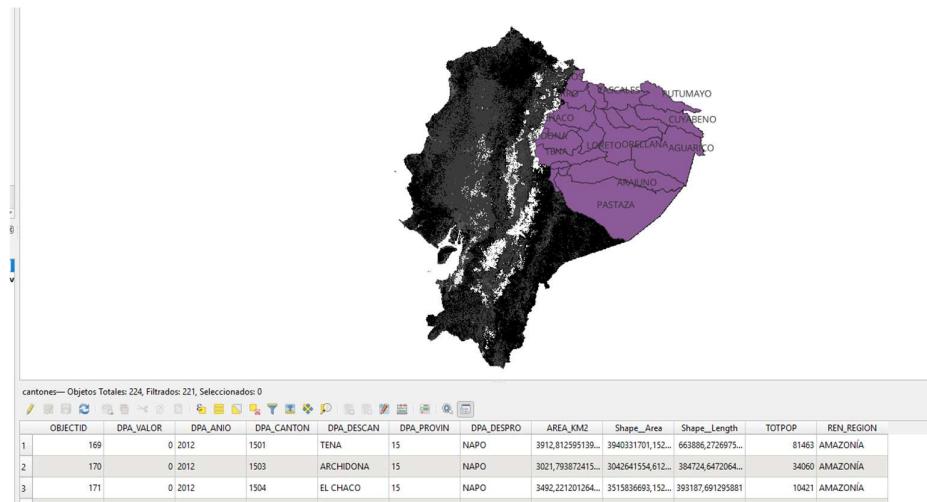
QGIS_MapBiomas_Ecuador_Col2_Leyenda_Nivel2 7/10/2025 1:44 QGIS Layer Settings 9 KB

```

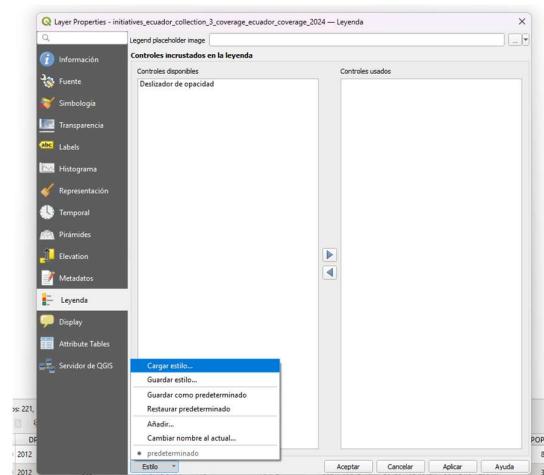
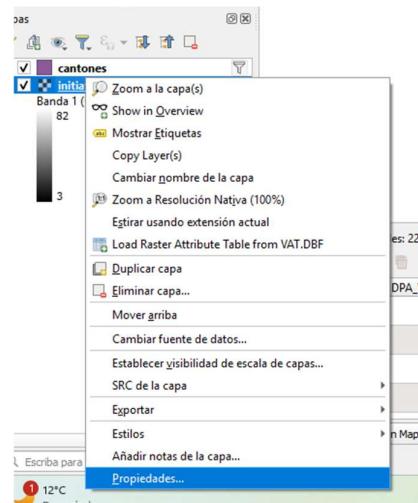
<end></end>
</fixedRange>
<temporal>
<elevation band="1" enabled="0" zscale="1" zoffset="0" symbology="Line">
<data-defined-properties>
<Option type="Map">
<Option name="name" value="" type="QString"/>
<Option name="properties"/>
<Option name="type" value="collection" type="QString"/>
</Option>
</data-defined-properties>
<profileLineSymbol>
<symbol name="" force_zhr="0" type="line" frame_rate="10" alpha="1" is_animated="0" clip_to_extent="1">
<data_defined_properties>
<Option type="Map">
<Option name="name" value="" type="QString"/>
<Option name="properties"/>
<Option name="type" value="collection" type="QString"/>
</Option>
</data_defined_properties>
<layer enabled="1" pass="0" locked="0" class="SimpleLine">
<Option type="Map">
<Option name="align_dash_pattern" value="0" type="QString"/>
<Option name="capstyle" value="square" type="QString"/>
<Option name="customdash" value="2" type="QString"/>
<Option name="customdash_map_unit_scale" value="3x0,0,0,0,0,0" type="QString"/>
<Option name="customdash_unit" value="MM" type="QString"/>
<Option name="dash_pattern_offset" value="0" type="QString"/>
<Option name="dash_pattern_offset_map_unit_scale" value="3x0,0,0,0,0,0" type="QString"/>
<Option name="dash_pattern_offset_unit" value="MM" type="QString"/>
<Option name="draw_inside_polygon" value="0" type="QString"/>
<Option name="joinstyle" value="bevel" type="QString"/>
<Option name="line_color" value="243,166,178,255" type="QString"/>
<Option name="line_style" value="solid" type="QString"/>
<Option name="line_width" value="0.6" type="QString"/>
<Option name="line_width_unit" value="MM" type="QString"/>

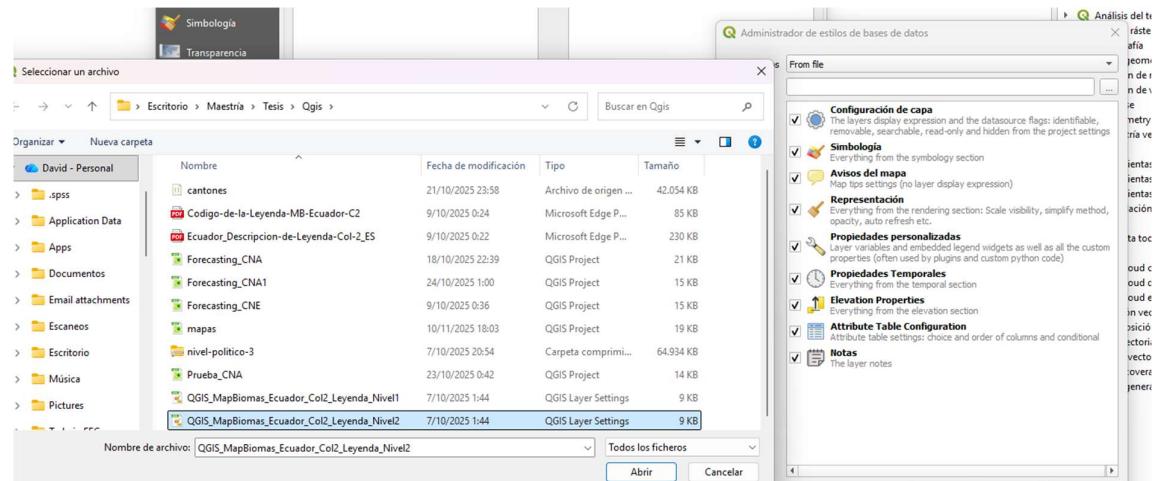
```

o. Cargar de leyendas del área de interés

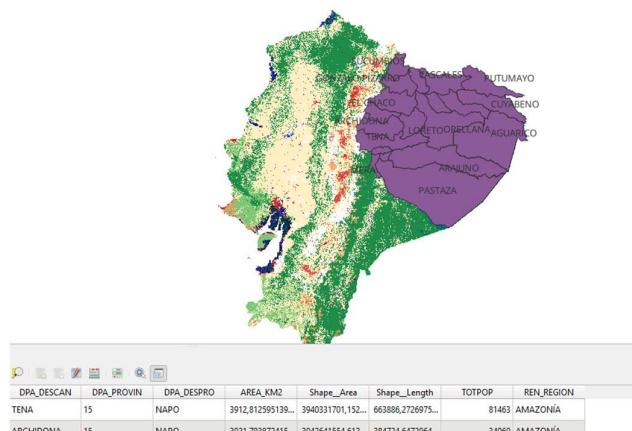


p. Para que el mapa se vea con el código de colores se realiza lo siguiente:

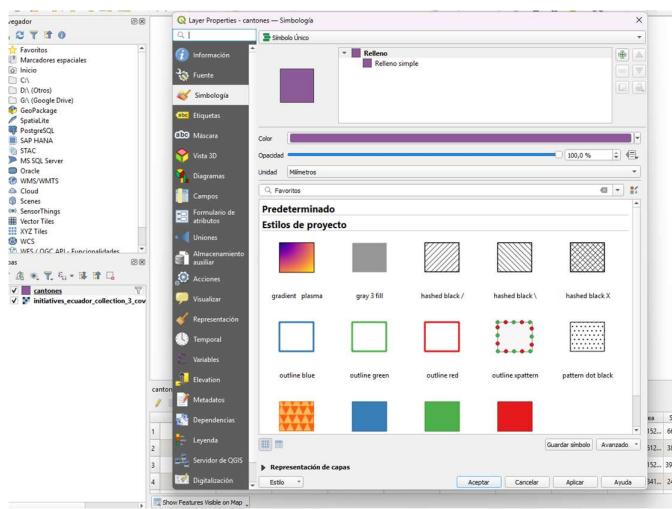




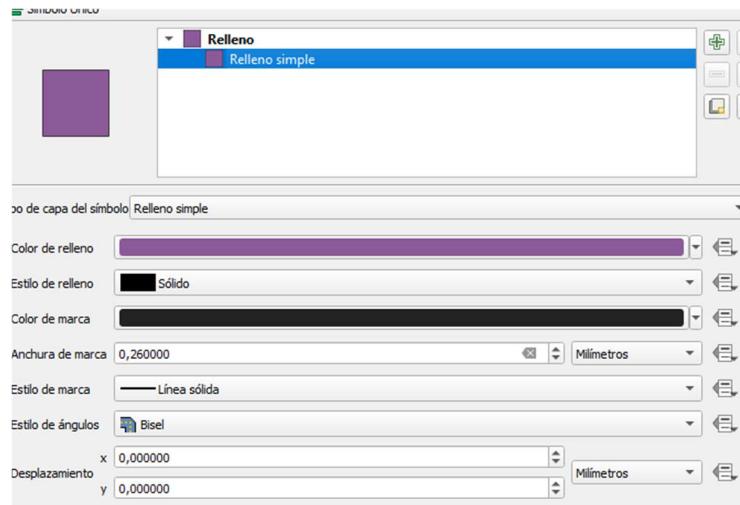
q. Se carga el mapa con los estilos respectivos(Final)



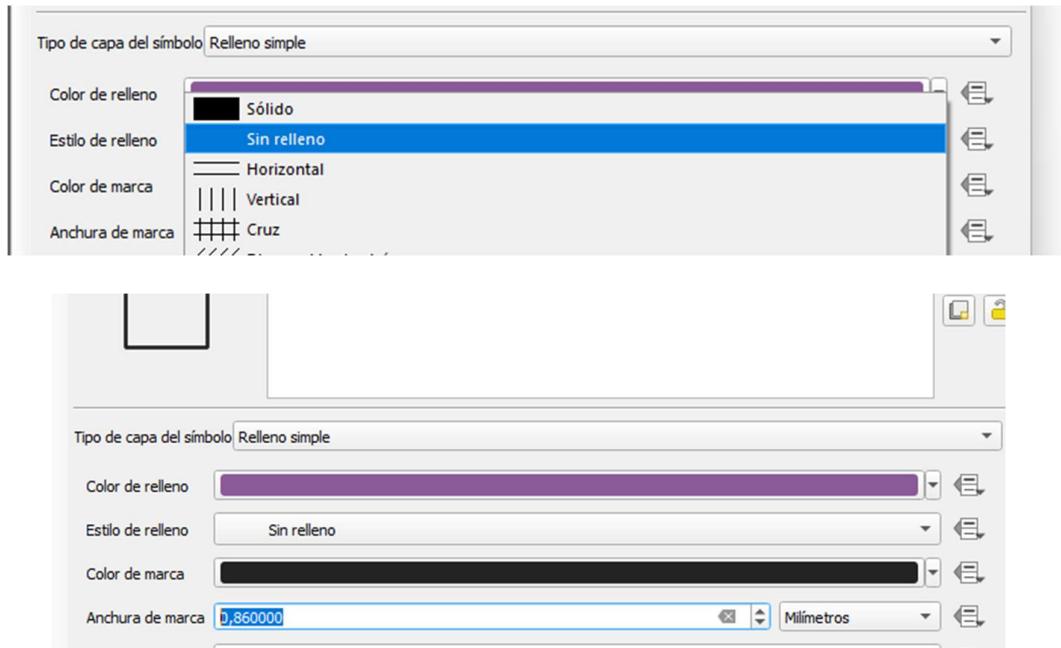
- r. Se realiza la transparencia en las zonas del mapa a analizar. ingresando en configuración de la capa de cantones.



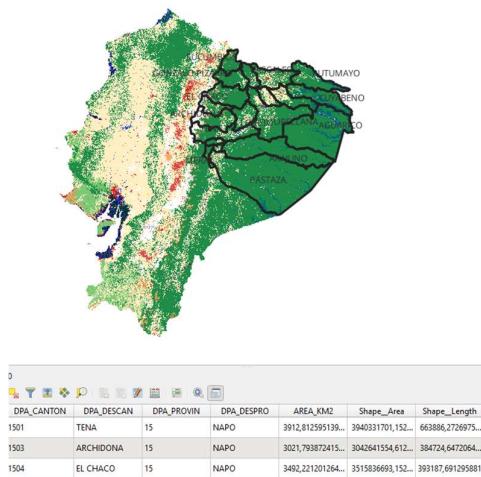
- s. Nivel 2 de relleno simple para modificaciones.



t. Escoger sin relleno y Aumentar el tamaño de borde.

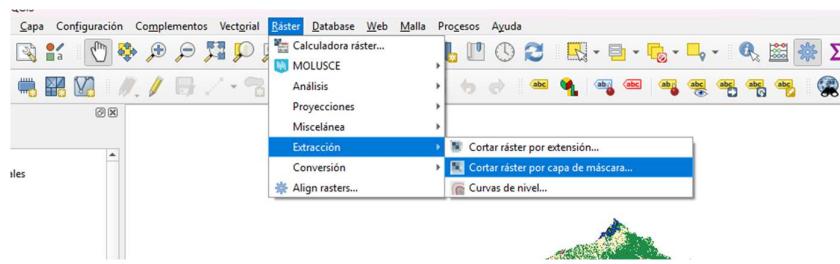


u. Resultado final del área de interés

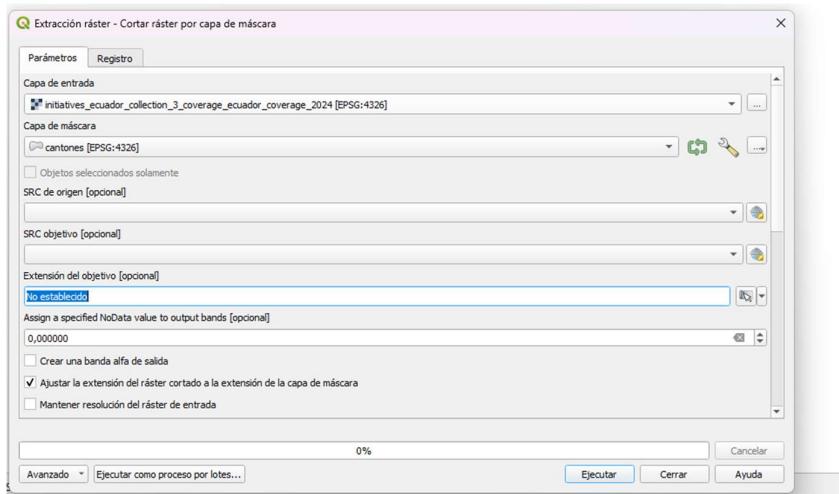


v. Cortar Raster

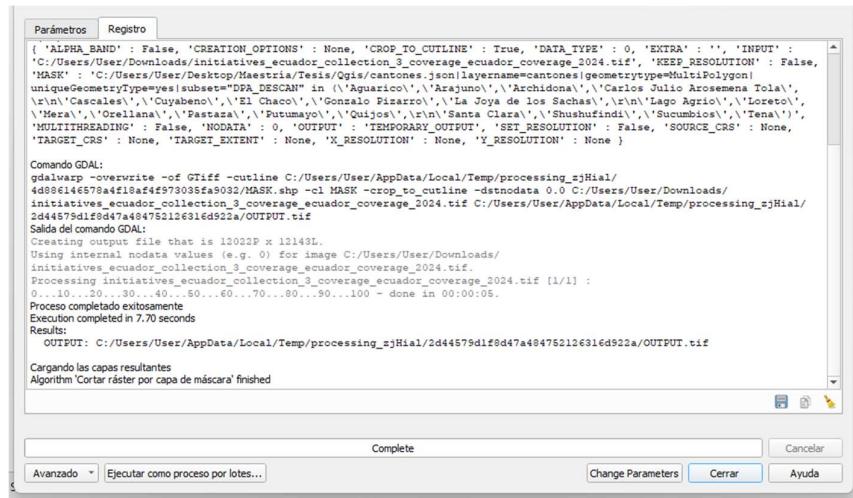
- Para solo tener la zona de interés, se corta la imagen, y se realiza los siguientes pasos:



- Una vez escogido lo siguiente, se debe utilizar las capas de interés.
- Dentro de la opción escogida, se debe elegir el mapa original, el mapa de cantones (a cortar) y muy importante en la parte de especificar el color sin data se escribe con 0.



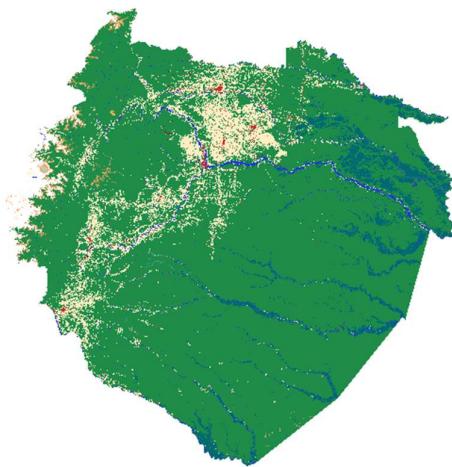
- Se ejecuta y se completa la extracción de datos



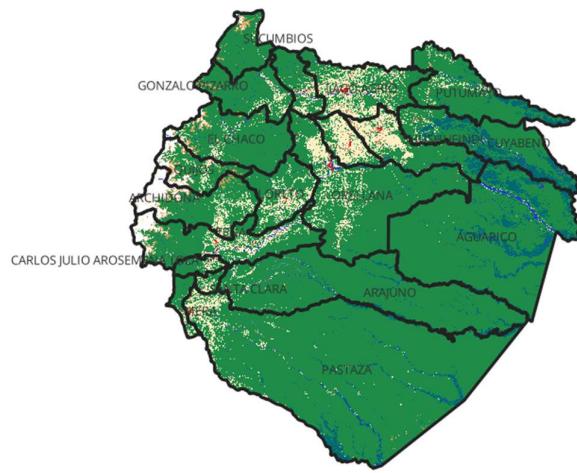
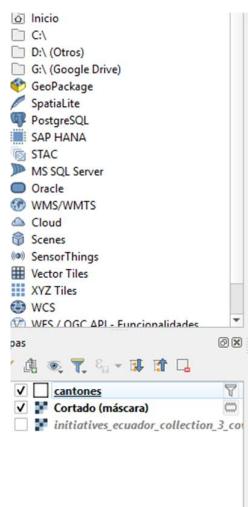
- Al comprobar el área de interés, se obtiene lo siguiente:



w. Se procede a colocar formato de leyendas.



Para mayor detalle se agrega la capa de cantones



x. Cortes

Se realiza 11 cortes de 2014 a 2024, para cada uno de los 20 cantones de la región a estudiar, y se organiza en provincias. Con un total de 220 imágenes. Y para concluir se procede a extraer mediante Python la información de cada uno de los puntos de interés para el análisis de la deforestación en la Amazonía.

