

# Implementación de un Modelo Predictivo para el reconocimiento de Tratamientos Silviculturales usando AWS SageMaker

Julian David Parra Bello\*

Escuela Colombiana de Ingeniería Julio Garavito

## Resumen

El proyecto busca desarrollar una herramienta basada en Inteligencia Artificial (IA) para el Reconocimiento de Tratamientos Silviculturales en el arbolado urbano de Bogotá. El sistema analizará características dasométricas (como altura, diámetro y copa) y fitosanitarias (estado de salud, plagas y enfermedades) de los árboles, utilizando datos recopilados en campo e integrados en una plataforma tecnológica, de donde se toman los datos iniciales para la identificación de los patrones para la elección de los tratamientos, de manera se proporcionen criterios claros y basados en datos para determinar qué tratamiento silvicultural debe aplicarse en cada caso, como podas, tratamientos fitosanitarios o talas. La herramienta permitirá a los ingenieros forestales tomar decisiones informadas de manera más eficiente y fundamentada.

Palabras clave — Inteligencia Artificial, Aprendizaje automático, Arbolado Urbano, Tratamiento silvicultural, Modelo predictivo, AWS SageMaker, Manejo forestal.

**Abstract,** The project seeks to develop an Artificial Intelligence (AI)-based tool for the Recognition of Silvicultural Treatments in Bogotá's urban trees. The system will analyze dasometric characteristics (such as height, diameter and crown) and phytosanitary characteristics (health status, pests and diseases) of the trees, using data collected in the field and integrated into a technological platform, from where the initial data is taken for the identification of patterns for the choice of treatments, in order to provide clear and data-driven criteria to determine which silvicultural treatment should be applied in each case, such as pruning, phytosanitary treatments or felling. The tool will allow forestry engineers to make informed decisions more efficiently and in a well-founded manner.

**Keywords** — Artificial Intelligence, Machine Learning, Urban Trees, Silvicultural Treatment, Predictive Model, AWS SageMaker, Forest Management.

## Introducción

El arbolado urbano desempeña un papel fundamental en la sostenibilidad de las ciudades modernas al contribuir a la conservación ambiental, mejorar la calidad del aire y ofrecer espacios de bienestar a los ciudadanos en general. Sin embargo, la gestión eficiente del arbolado requiere procesos precisos y basados en datos que permitan determinar los tratamientos silviculturales más adecuados. Bogotá cuenta con alrededor de 2.2 millones de árboles, que se componen en 1.500 especies arbóreas, de tal manera que realizar la evaluación y clasificación de cada uno de estos árboles es algo logísticamente complejo y que requiere de intervención experta para la elección de un tratamiento adecuado, lo que compromete la sostenibilidad y el bienestar de la ciudad y el arbolado. Actualmente, las autoridades competentes de realizar la evaluación de estos tratamientos solo logran realizarlo a cerca de 35.000 árboles al año, lo que representa cerca del 7% del arbolado urbano.[1]

La gestión del arbolado Urbano presenta varios

TRATAMIENTO	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	TOTAL	PORCENTAJE
Tala	18.599	24.058	18.465	16.649	13.795	10.295	12.239	11.133	12.971	10.991	14.921	15.325	179.441	46,79 %
Conservar	4.381	11.175	8.229	4.018	7.256	11.763	11.245	8.524	3.624	1.741	4.719	5.996	82.671	21,56 %
Traslado	1.406	5.505	3.358	2.865	1.857	886	916	825	1.321	395	743	3.777	23.852	6,22 %
Podas de formación	2.096	1.933	928	740	1.038	1.006	195	707	2.154	695	3.269	1.478	16.229	4,23 %
Podas de estabilidad	211	174	231	409	340	584	444	299	661	327	1.400	418	5.498	1,43 %
Tratamiento integral	1.120	1.050	824	859	2.591	2.637	2.294	4.414	1.596	1.523	2.581	1.589	23.078	6,02 %
Podas de mejoramiento	565	856	882	1.952	3.052	4.682	5.962	5.231	6.764	4.101	9.953	4.841	48.841	12,74 %
Podas radicular	106	41	23	22	19	14	406	588	174	389	596	639	3.067	0,80 %
Sin tratamiento	0	110	33	19	8	14	28	0	124	150	210	127	823	0,21 %
Total	28.484	44.902	32.973	27.553	29.946	31.861	33.779	31.721	29.389	20.310	58.392	54.190	383.500	100 %

Elaboración SDA. Fuente: Sistema de Información Ambiental -SIA

Figura 1: Tipo de tratamiento silvicultural autorizado. [1]

inconvenientes entre los que se destaca la escala y complejidad debida a la gran cantidad de árboles y su diversidad de especies, haciendo que la evaluación manual sea una tarea poco eficiente, adicionalmente esta evaluación del arbolado es susceptible a la interpretación personal de cada evaluador, lo que puede llevar a diagnósticos y tratamientos inconsistentes, logrando evaluar un pequeño porcentaje del arbolado anualmente, dejando una gran cantidad de árboles sin el seguimiento y cuidado necesarios, y dichos árboles pueden afectar aquellos que presentan buenas condiciones. Estos árboles enfermos o con

\*julian.parra-b@mail.escuelaing.edu.co

daños estructurales pueden representar un peligro para la población, especialmente en áreas urbanas densamente pobladas como lo es Bogotá. Estas dificultades impiden una gestión adecuada y oportuna del arbolado, afectando su contribución al bienestar ambiental y social de la ciudadanía en general[2].

Ante este panorama, resulta evidente la necesidad de implementar soluciones innovadoras que permitan optimizar la gestión del arbolado urbano. Por tanto, en el presente proyecto se propone la implementación de un modelo predictivo, utilizando AWS SageMaker, para analizar las condiciones dasométricas y fitosanitarias de los árboles. A través del uso de inteligencia artificial, el sistema identificará patrones y recomendará tratamientos silviculturales óptimos, permitiendo mejorar la eficiencia y precisión en la planificación y ejecución de intervenciones. Esta solución representa una oportunidad significativa para fortalecer las estrategias de conservación del arbolado urbano, optimizando los recursos y contribuyendo al bienestar de la ciudadanía mediante un entorno urbano más saludable y sostenible.

## Estado del Arte

Para la realización del estado del arte se debe tener en cuenta el cuadro 1, el cual presenta el conjunto de criterios utilizados para la búsqueda exhaustiva y precisa de artículos científicos y técnicos relacionados con el desarrollo de modelos predictivos para la gestión del arbolado urbano, específicamente utilizando inteligencia artificial y plataformas como AWS SageMaker. La investigación se limita a artículos científicos y técnicos publicados en medios confiables, los cuales tienen una revisión previa certificada y demuestran rigor académico y técnico. Adicionalmente, se restringe a los últimos 5 años. Para la búsqueda se utilizaron bases de datos científicas como Scopus, ScienceDirect y ResearchGate, ya que proporcionan una gran colección de artículos en el campo de la computación y de sistemas. También se incluyó IEEEExplore, ya que es una base de datos respaldada por la IEEE (Institute of Electrical and Electronics Engineers) y es más específica enfocada en tecnología e ingeniería. Estos criterios permiten identificar los estudios más relevantes y actualizados, que contribuyen a la construcción de un estado del arte sólido y confiable.

## Arbolado Urbano en Bogotá

Bogotá, la capital de Colombia, cuenta con un patrimonio natural invaluable: su arbolado urbano. Millones de árboles de diversas especies se encuentran en las calles, parques y zonas verdes de la ciudad,

**Cuadro 1:** Selección de elementos de búsqueda

Criterio	Descripción
Base de datos	ScienceDirect, Scopus, IEEE Xplore, ResearchGate y Google Scholar.
Tipo de artículo	Artículos científicos/técnicos publicados en revistas.
Cadenas de búsqueda	Inteligencia Artificial y Arbolado", "Inteligencia Artificial y medio Ambiente", "Gestión Inteligente del Arbolado", "Modelos de automatización sustentable", "AWS SageMaker", "Modelos predictivos para arbolado", "Arbolado Urbano en Bogotá", "Gestión del Arbolado en Bogotá", etc.
Período de búsqueda	De enero de 2019 a noviembre de 2024
Idioma	Inglés/Español
Procedimiento de selección	La relevancia con el tema de investigación se juzga por el contenido del resumen, la introducción y la conclusión de cada artículo.
Esquema de clasificación	Desarrollo actual de arquitecturas tecnológicas para el análisis y gestión del arbolado.

contribuyendo a mejorar la calidad del aire, regular la temperatura, mitigar el ruido y brindar espacios de recreación y bienestar a sus habitantes.

La gestión del arbolado urbano en Bogotá es una responsabilidad compartida entre diversas entidades como lo son el Jardín Botánico de Bogotá José Celestino Mutis, que es la entidad encargada de la investigación, conservación y divulgación del conocimiento sobre la flora de Bogotá. La Secretaría Distrital de Ambiente se encarga de formular las políticas ambientales del Distrito y de velar por su cumplimiento, adicionalmente de autorizar a las entidades o ciudadanos involucrados en los procedimientos y tratamientos correspondientes de la evaluación del arbolado. En conjunto con otras entidades públicas y privadas son responsables de la gestión del arbolado en sus respectivas jurisdicciones.

Los lineamientos de la gestión del arbolado se encuentran en el Manual de Silvicultura Urbana[3]. Este manual es una guía detallada que ofrece los procedimientos para el manejo adecuado de los árboles. Este tipo de documento recopila información técnica y práctica sobre diversas actividades relacionadas con el arbolado, desde la selección de especies ade-

cuadas para plantación hasta la poda, el tratamiento de plagas y enfermedades, y la remoción de árboles cuando sea necesario. Dichos procedimientos se alinean con los objetivos de la Sociedad Internacional de Arboricultura (ISA)[4], que es una organización sin fines de lucro, cuyo principal objetivo es promover un mayor aprecio por los árboles y estimular la investigación, la tecnología y la práctica profesional de la arboricultura.

**Arbolado en buenas condiciones** Para determinar el buen o mal estado de un espécimen arbóreo es necesario tener en cuenta factores como el estado de las raíces, las lesiones que pueda presentar, el estado fitosanitario y la forma del mismo, ya que de esto dependerá la probabilidad de sobrevivir después de ser plantado. La elección de un árbol de buena calidad también puede reducir las fallas producidas por defectos estructurales ocasionados por fenómenos naturales como lluvias, vientos, deslizamientos, entre otras.



**Figura 2:** Factores a tener en cuenta para reconocer un árbol de buena calidad. [1]

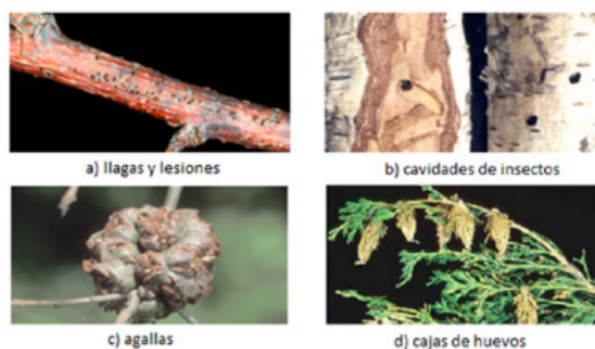
De acuerdo con la Sociedad Internacional de Arboricultura - ISA (2008), un árbol de buena calidad presenta las siguientes características:

- Un cepellón o bola de raíces de tamaño adecuado. Siempre que sea posible, debe asegurarse de que existan suficientes raíces sanas para mantener un crecimiento saludable.
- Un tronco libre de heridas mecánicas o heridas originadas por una poda incorrecta.
- Una forma fuerte, con ramas bien dispuestas y firmemente unidas.

El Estado de las raíces es uno de los factores más importantes al momento de evaluar un árbol, debido a que los problemas de raíces son difíciles de corregir y reducen la capacidad de sobrevivir y crecer. Dichas raíces no deben encontrarse aplastadas, o inmersas en una bola de raíces, ya que puede representar una falta de oxígeno para el sistema radicular. De acuerdo con su diámetro se pueden diferenciar las siguientes clases: raíces finas, flexibles, gruesas y fuertes. Las últimas proporcionan anclaje, mantienen la cohesión

del suelo y actúan como contrapeso para evitar la caída del árbol.[5]

El Estado fitosanitario se refiere a la condición de salud general, considerando la presencia o ausencia de enfermedades, plagas, daños mecánicos o estrés ambiental que puedan afectar su crecimiento, desarrollo y supervivencia. En otras palabras, es una evaluación del bienestar del árbol en un momento dado.[3]. La forma, color y estado de las hojas debe coincidir con las características propias de la especie, de lo contrario es posible que el árbol presente un estado fitosanitario negativo. Adicionalmente, se debe tener en cuenta el estado de las ramas y tronco, estos deben estar vigorosos y presentar una adecuada longitud y diámetro. No deben presentar agujeros en el tronco, ya que estos pueden indicar la presencia de insectos.



**Figura 3:** Presencia de insectos y enfermedades, en tronco y ramas. [1]

La Forma del arbolado de acuerdo con la ISA una arquitectura del árbol buena y fuerte corresponde a ramas espaciadas de manera uniforme a lo largo del tronco, las uniones de las ramas deben ser fuertes y firmes[1].

**Problemas del Arbolado Urbano** Algunos de los factores causantes de problemas en los árboles son:

- Disminución de la provisión de agua por efecto de la pavimentación de las calles, aceras y obras de drenaje para el agua lluvia.
- Condiciones cambiantes en cuanto a cantidades de radiación solar y horas de insolación, en comparación con ambientes naturales.
- Reducido espacio vital para el desarrollo de raíces y follaje.
- Intermitente corte de raíces por obras públicas para la introducción de cañerías y tuberías subterráneas.
- Contaminación del suelo, agua y aire.
- Destrucción de las partes aéreas y podas antitécnicas.
- Incremento de sales por orina.

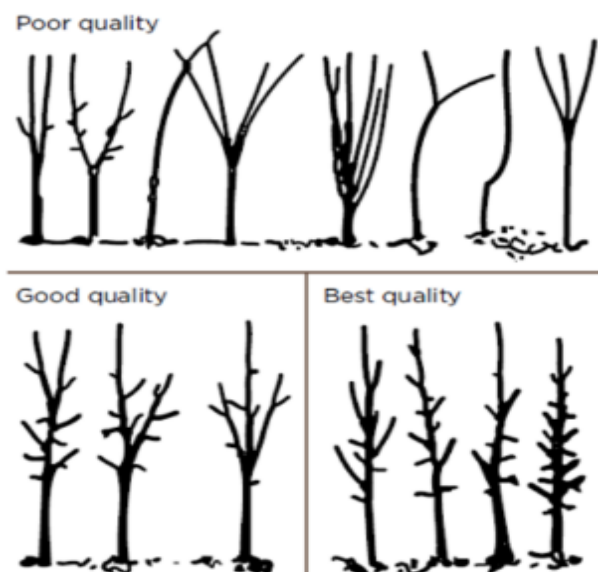


Figura 4: Comparación entre formas de árboles. [1]

- Acceso de detergentes y desinfectantes químicos que se usan en el lavado de las aceras.
- Sensible disminución de la materia orgánica del suelo.
- Cambios en el régimen de evapotranspiración.

Diversas investigaciones han revelado que los árboles urbanos, sometidos a las adversidades del entorno pavimentado, tienen una expectativa de vida significativamente reducida, a menudo inferior a los 10 años [5]. La pavimentación limita el acceso al agua y reduce el espacio para el desarrollo radicular, mientras que la contaminación y las podas inadecuadas agravan su estado de salud.

**Tratamientos Silviculturales** Según el manual de Silvicultura Urbana [3] las actividades de manejo del arbolado se realizan con la finalidad de mantener la vegetación en buenas condiciones aéreas y radicales, mejorar su aspecto y su desarrollo y manejar factores de seguridad pertinentes a

El Cuadro 2 presenta una clasificación de los diferentes tipos de tratamientos autorizados por la Secretaría Distrital de Ambiente en su labor de Evaluación del Arbolado Urbano que se realiza en la ciudad de Bogotá. Cada fila de la tabla corresponde a un tipo de tratamiento específico, y las columnas indican el término y su significado.

## Arquitectura basada en IA

La inteligencia artificial (IA) a lo largo de su desarrollo ha venido jugando un papel fundamental en la evaluación del arbolado urbano, permitiendo reconocer de patrones en los datos dasométricos y fitosanitarios de las especies arbóreas. Mediante el

uso de algoritmos avanzados, la IA puede analizar grandes volúmenes de datos para identificar características específicas de las especies arbóreas y detectar cambios en su estado de salud[7]. Por ejemplo, la automatización en la identificación de especies a través del análisis de imágenes y otros datos puede acelerar significativamente el proceso de clasificación e inventariado[8]. Además, la IA puede predecir las necesidades de intervención, como podas o tratamientos fitosanitarios, basándose en patrones históricos y condiciones actuales basándose en información climática, lo que permite a las entidades encargadas de la gestión y control del arbolado, planificar y ejecutar acciones preventivas de manera más eficiente[9].

La implementación de técnicas de Inteligencia Artificial (IA) para la gestión del arbolado urbano ha crecido significativamente en los últimos años. Más de 23 países y 85 ciudades han adoptado estrategias basadas en IA para mejorar la planificación, monitoreo y mantenimiento de los árboles en zonas urbanas. Estas iniciativas han permitido optimizar los recursos, mejorar la sostenibilidad ambiental y fomentar la resiliencia climática. La inteligencia artificial está ayudando en la gestión forestal a través de diversas herramientas y aplicaciones que permiten identificar daños forestales por factores abióticos y bióticos, valorar servicios ecosistémicos de manera rápida y estandarizada, y resolver problemas de oferta y demanda en el sector. Además, se utilizan técnicas de IA para crear gemelos digitales y ofrecer servicios técnico-científicos a administraciones y empresas, contribuyendo a afrontar el cambio global.[7]

Velásquez et al. en su artículo *Remotely Sensed Tree Characterization in Urban Areas: A Review*[10] indica que la inteligencia artificial se utiliza para reconocer árboles mediante técnicas de aprendizaje automático (ML) y aprendizaje profundo (DL). Algoritmos como las Máquinas de Vectores de Soporte (SVM) y las Redes Neuronales Convolucionales (CNN) son aplicados para la clasificación y caracterización de árboles urbanos. Estos métodos permiten trabajar con múltiples variables predictivas, mejorando significativamente los modelos de clasificación en comparación con otros clasificadores. Adicionalmente, se indica el uso de distintas arquitecturas de aprendizaje profundo (DL) utilizadas en diversas investigaciones, identificar, clasificar y analizar árboles a partir de imágenes o datos, como se muestra en el cuadro 3.

**ResNet (Residual Network)** es una arquitectura de red neuronal profunda utilizada especialmente en tareas de visión por computadora. Se menciona en el contexto de la caracterización de árboles urbanos y se destaca por su capacidad para mejorar la precisión en la clasificación de especies de árboles. ResNet se utiliza en la caracterización de árboles urbanos



mediante la integración de datos de LiDAR y otras imágenes. Por ejemplo, se ha utilizado en combinación con algoritmos de agrupamiento para segmentar datos y en estudios que comparan su rendimiento con otras arquitecturas de redes neuronales, logrando altas precisiones en la clasificación de especies de árboles. En particular, se ha encontrado que ResNet-50 proporciona la mejor precisión en el proceso de clasificación de especies de árboles[10]

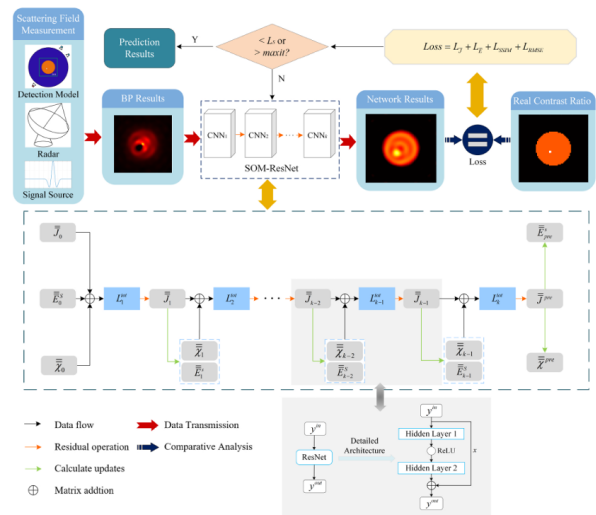
El estudio realizado por Guoyang et al. en su artículo *Tree internal defects detection method based on ResNet improved subspace optimization algorithm*[11] presenta un método novedoso para detectar defectos internos en árboles utilizando un algoritmo de optimización en subespacio mejorado llamado SOM-ResNet, que integra el aprendizaje profundo con principios físicos de dispersión inversa electromagnética. Este algoritmo mejora la precisión de detección al simular el proceso iterativo de corriente inducida y contraste, abordando efectivamente los desafíos que plantea la no linealidad y la naturaleza mal planteada de los problemas inversos. Los resultados experimentales demuestran que SOM-ResNet supera a los algoritmos existentes en la detección precisa de defectos de tamaño pequeño a mediano en árboles, lo que lo convierte en una herramienta prometedora para el monitoreo y gestión de la salud forestal.

La figura 5 presenta la arquitectura de la red neuronal SOM-ResNet, que toma como entrada medidas de un campo de dispersión y, a través de múltiples capas convolucionales y residuales, refina iterativamente sus predicciones. El objetivo principal de esta red es detectar y analizar patrones específicos en los datos, como contrastes y distribuciones de permitividad. Al combinar técnicas como mapas auto-organizativos y redes residuales, esta red neuronal logra una alta precisión en la identificación y caracterización para detectar defectos internos en árboles.

A pesar de sus éxitos, persisten desafíos en lograr una inversión de contorno perfecta y en manejar permitividades complejas. El trabajo futuro se centrará en la expansión del conjunto de datos, la optimización de la red de aprendizaje profundo y la mejora de las técnicas de inversión para contornos de árboles irregulares.[11]

**YOLO (You Only Look Once)** es una arquitectura de red neuronal profunda utilizada especialmente en tareas de visión por computadora, para detectar objetos en imágenes y videos en tiempo real. A diferencia de otros algoritmos que escanean una imagen en múltiples pasos, YOLO divide la imagen en una cuadrícula y predice los cuadros delimitadores y las probabilidades de clase para cada celda de la cuadrícula en una sola evaluación.

En el artículo *Maturity detection of 'Huping' juju-*

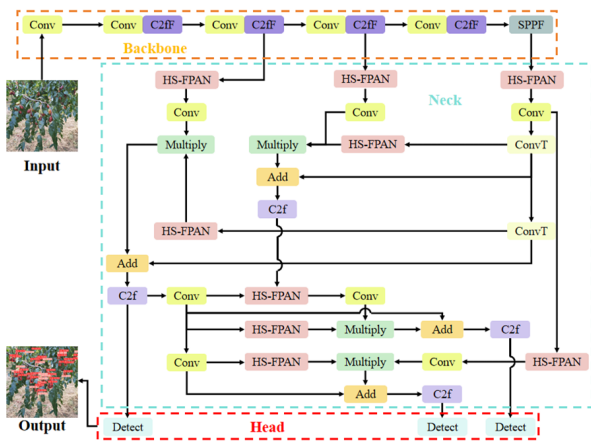


**Figura 5:** Arquitectura SOM-ResNet para detección de defectos internos en árboles. [11]

be fruits in natural environment using YOLO-FHLD se mejora la arquitectura YOLOv8n para detectar la madurez de los jujubes 'Huping'. Integrando varios módulos y técnicas, como el módulo C2ff para mejorar la extracción de características, el mecanismo de fusión de características HS-FPAN para capturar diferentes niveles de madurez en la fruta, y se ha empleado Focal Loss como función de pérdida para optimizar el rendimiento del modelo. Además, se ha utilizado una estrategia de destilación de conocimiento para mejorar aún más la precisión, que en este caso es del 85.40 % y un tamaño de modelo de solo 3.51 MB, mostrando una ventaja en la precisión de localización y reduciendo errores de clasificación y detección de fondo.

En la figura 5 se logra evidenciar la arquitectura de la red y como se combina una red troncal (backbone) que extrae características de la imagen, un cuello (neck) que integra estas características a diferentes escalas y una cabeza (head) que genera las predicciones finales. La clave de esta arquitectura reside en la estructura HS-FPAN del cuello, que permite una integración eficiente de características y una detección precisa de objetos de diversos tamaños. Esta arquitectura se destaca por su velocidad y precisión, lo que la hace ideal para aplicaciones donde la detección de objetos en tiempo real es crucial, como sistemas de vigilancia, vehículos autónomos y realidad aumentada.

Este estudio introduce nuevas ideas técnicas en el campo de la detección de objetivos agrícolas. En futuras investigaciones y trabajos, los investigadores pretenden mejorar aún más la precisión del modelo, con el fin de proporcionar soporte técnico para el posterior reconocimiento visual y posicionamiento de objetivos de la recolección inteligente.



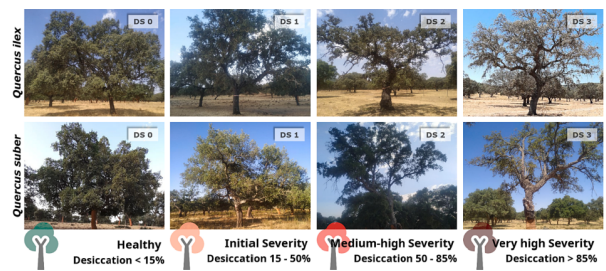
**Figura 6:** Arquitectura YOLO para detección de madurez en la fruta. [12]

## Arquitectura SIG Integrada con Modelos Predictivos

En España y Portugal se realizó un estudio sobre técnicas de teledetección, imágenes hiperespectrales y estudios ecológicos, enfocándose en el uso de sistemas de aeronaves no tripuladas y técnicas de imagen avanzadas para el monitoreo ambiental, la evaluación de la salud forestal y la detección de enfermedades en plantas, especialmente en robles, debido a la creciente amenaza que representa el declive de esta especie en la región, ya que son cruciales para la producción de corcho y la cría de cerdos ibéricos. El artículo llamado *Detection of oak decline using radiative transfer modelling and machine learning from multispectral and thermal RPAS imagery*[13], publicado en el año 2024, propone que la detección oportuna y precisa del declive forestal es esencial para planificar e implementar estrategias efectivas de restauración y reforestación, mediante el uso de la hiperautomatización.

Mediante el uso de RPAS (Sistema de Aeronaves no tripuladas) se hizo la adquisición de imágenes hiperespectrales y térmicas de alta solución en áreas específicas de Portugal y España, con el objetivo de evaluar la salud de los robles y detectar el declive forestal. Además, los RPAS permitieron la personalización de sensores y la recopilación continua de datos en áreas de difícil acceso, lo que es esencial para analizar cambios fisiológicos que indican condiciones de salud en los robles en declive.

Se hizo uso de inteligencia artificial y Machine Learning para identificar y delinear en las imágenes las copas de los árboles de manera independiente. Este procesamiento de imágenes se logró mediante métodos de segmentación basados en objetos y la utilizando el clasificador de distancia. Los modelos de automatizados ayudan a discriminar entre diferentes etapas del declive de los robles, evaluando la incidencia y severidad de la enfermedad, los cuales



**Figura 7:** La imagen presenta un estudio visual sobre los efectos de la desecación en dos especies de robles: *Quercus ilex* y *Quercus suber*. Esta desecación, que se refiere a la pérdida excesiva de agua en las plantas, es un problema común en muchas regiones, especialmente en periodos de sequía prolongada[13].

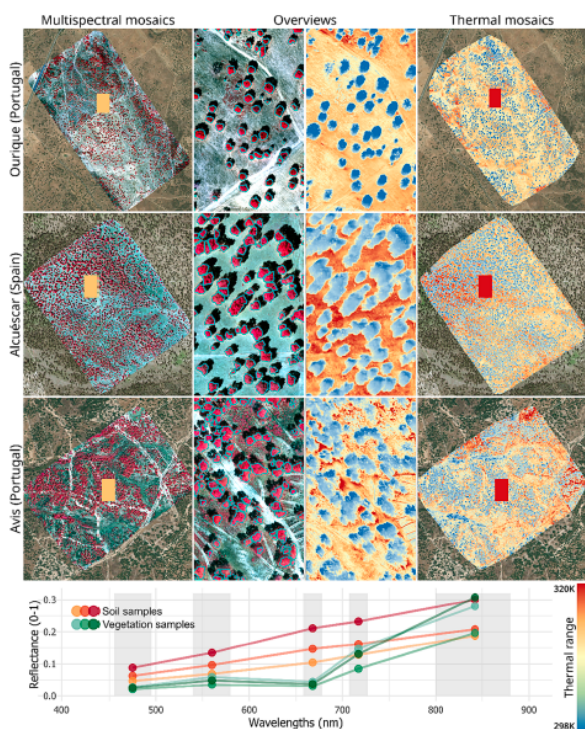
mostraron capacidades de predicción que permiten identificar los estados de salud de las plantas y mejorar los métodos de detección temprana de enfermedades.

El documento concluye que los modelos de aprendizaje automático (ML) han demostrado una capacidad justa a buena para discriminar entre diferentes etapas del declive de los robles, siendo más efectivos en la identificación de etapas avanzadas de la enfermedad. Además, se subraya la necesidad de mejorar los métodos de detección temprana, ya que los modelos tienen dificultades para identificar las etapas iniciales del declive, lo cual es crítico para detener la propagación de la enfermedad. La combinación de datos de diversas fuentes y la capacidad de los modelos para adaptarse y mejorar continuamente hacen de esta tecnología una herramienta indispensable para monitorear la salud de los bosques a gran escala y tomar decisiones informadas para su conservación.

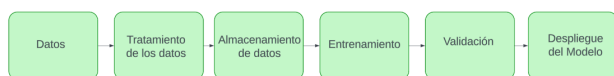
## Metodología

El objetivo principal del proyecto es desarrollar una herramienta de inteligencia artificial para gestionar el arbolado urbano de Bogotá. Esta herramienta analizará datos de cada árbol (altura, diámetro, estado fitosanitario, etc.) y recomendará el tratamiento silvicultural más adecuado (poda, tratamiento fitosanitario o tala), basado en criterios del set de datos proporcionado e información complementaria para la selección del tratamiento adecuado. Por esta razón se pretende desarrollar la metodología presentada en la figura 9, la cual representa el flujo de trabajo que se tendrá en cuenta para la elaboración del modelo usando AWS SageMaker.

AWS SageMaker es un servicio de aprendizaje automático en la nube. Está diseñada para simplificar el proceso de creación, entrenamiento e implementación de modelos de machine learning, desde la preparación de los datos hasta la implementación del



**Figura 8:** Estas son imágenes se capturaron en diferentes longitudes de onda de luz, lo que permitió identificar detalles como la salud de las hojas y la temperatura de las plantas, analizando sus patrones de radiación[13].



**Figura 9:** Metodología de desarrollo de la solución.

modelo en producción. Este servicio se encarga de la infraestructura subyacente, como los servidores, los frameworks de aprendizaje automático (TensorFlow, PyTorch, etc.) y las herramientas de desarrollo, permitiendo escalar los recursos computacionales según las necesidades del proyecto. Adicionalmente, se integra fácilmente con otros servicios de AWS, como Amazon S3 para almacenamiento, Amazon EC2 para computación y Amazon RDS para bases de datos. Dicho servicio será utilizado en las etapas que componen el proyecto de la siguiente manera:

## Datos

Para el desarrollo del proyecto se cuenta con un data set proporcionado por el Área de Silvicultura de la Secretaría Distrital de Ambiente, con información recopilada entre el 2010 y 2024 de los registros realizados por conceptos de evaluación en la ciudad de Bogotá, en total son 312 mil registros categorizados de la siguiente manera.

La figura 9 contiene los atributos o características que se registran para cada árbol en un inventario

Item	Descripción	Valores
ID	Identificador único del árbol	Númerico
ESPECIE	Especie Botánica del árbol	N/A
ALTURA	Altura del árbol en metros	Decimal
DENSIDAD	Densidad del árbol	Denso, Medio, Ralo, Muy Ralo
ESTADO COPA	Estado General de la copa	Bueno, Regular, Malo
ESTADO FUSTE	Estado General del fuste	Bueno, Regular, Malo
ESTADO RAIZ	Estado General de la raíz	Bueno, Regular, Malo
ESTADO SANITARIO	Estado de salud del árbol	Bueno, Regular, Malo
UBICACION	Ubicación geográfica del árbol	Coordenadas
TRATAMIENTOS	Tratamiento Silvicultural realizado	Tala, conservación, traslado, poda de formación, poda de control de altura, poda de realce, poda de aclareo, poda de equilibrio, poda de mejoramiento, poda de estructura, poda sanitaria, poda radicular, tratamiento integral, tratamiento especial.

**Figura 10:** Set de Datos.

forestal urbano. Cada fila representa un árbol individual, y cada columna corresponde a una propiedad específica de ese árbol. Dicha información específica y clasificada por cada árbol es fundamental para la gestión y el mantenimiento del arbolado urbano. Ya que determinar qué tratamientos son necesarios para cada árbol, basados en su estado de salud y otras características. Adicionalmente, apoya la toma de decisiones relacionadas con la gestión del arbolado urbano, como la planificación de plantaciones, la remoción de árboles peligrosos, etc.

## Tratamiento de los datos

Los datos recolectados se limpian, transforman y preparan para ser utilizados en el modelo. Esto incluye tareas como la eliminación de valores faltantes, la normalización de datos, y la codificación de variables categóricas. Datos como la ubicación son eliminados del data set, ya que esto no proporciona una variable a considerar para el análisis de los datos, ya que solo se van a tener en cuenta las condiciones dasométricas y fitosanitarias del espécimen.

En los datos originales es necesario eliminar aquellos registros que presentan información incompleta, como lo es el caso de la columna de estado fitosanitario, ya que solo 305.418 filas tenían este dato, con lo cuales es necesario descartar aquellos registros incompletos.

La normalización de los datos es indispensable en el preprocesamiento de datos para el entrenamiento de modelos de aprendizaje automático. Consiste en transformar los datos numéricos a una escala común, en este caso se hará uso de una escala numérica. Estos datos son normalizados bajo los criterios presentados en el Cuadro 4, basándose en la información suministrada y la clasificación dada por la entidad de cada registro.

En la figura 11 se evidencia una muestra de los datos originales proporcionados y en la figura 12 los datos ya normalizados bajo los criterios del Cuadro 4.



ID_DETALLE	NOMCIENTIFICO	ALTURATOTAL	VOLUMENCOMER	DENSIDAD	ESTADO_COPA	ESTADO_FUSTE	ESTADO_RAIZ	
0	266	Sambucus nigra	3.0	0.000	Medio	Bueno	Bueno	No apreciable
1	267	Tecoma stans	2.5	0.000	Denso	Bueno	Bueno	No apreciable
2	268	Fraxinus chinensis	4.5	0.000	Ralo	Regular	Bueno	No apreciable
3	269	Acacia melanoxylon	3.0	0.003	Muy Ralo	Malo	Regular	Raíces Descubiertas
4	270	Acacia decurrens	9.5	0.665	Medio	Regular	Regular	Raíces Descubiertas
5	271	Acacia calamifolia	2.3	0.001	Muy Ralo	Regular	Regular	Raíces Descubiertas
6	272	Sambucus nigra	3.5	0.008	Medio	Regular	Regular	Raíces Descubiertas
7	273	Sambucus nigra	2.0	0.001	Denso	Bueno	Regular	Raíces Descubiertas
8	274	Prunus capuli	3.5	0.000	Denso	Regular	Bueno	No apreciable
9	275	Sambucus nigra	1.6	0.000	Medio	Regular	Regular	Raíces Descubiertas

Figura 11: Datos Originales

	ID_DETALLE	TRATAMIENTO	ALTURA	TOTAL	DENSIDAD	ESTADO_COPA	ESTADO_FUSTE	EST_RAIZ_GEN	NIVEL_RIESGO	ESTADO_FITO
0	266	1	3.0		1	0	0	0	1	0
1	267	1	2.5	0	0	0	0	0	1	0
2	268	9	4.5	3	1	0	0	0	1	1
3	269	0	3.0	2	2	1	1	1	1	2
4	270	0	9.5	1	1	1	1	1	0	1
5	271	0	2.3	2	1	1	1	1	1	1
6	272	0	3.5	1	1	1	1	1	1	1
7	273	0	2.0	0	0	1	1	1	1	1
8	274	9	3.5	0	1	0	0	0	1	1
9	275	0	1.6	1	1	1	1	1	1	

Figura 12: Datos Normalizados

## Almacenamiento de los datos

Los datos preparados se almacenan en un formato adecuado para su uso en el entrenamiento del modelo. esto permite que los algoritmos de optimización pueden converger más rápidamente hacia la solución óptima, reduciendo el tiempo de entrenamiento. Los modelos entrenados con datos normalizados tienden a ser más estables y menos sensibles a cambios en los datos de entrada, al permitir tener todas las características en una misma escala que permita hacer la comparación entre las variables.

## Entrenamiento del modelo

Durante este proceso, el modelo aprende a identificar patrones en los datos. En este caso se crean los trabajos de entrenamiento en SageMaker usando el algoritmo RandomForestClassifier, que es un algoritmo de aprendizaje supervisado muy popular para problemas de clasificación, donde se generan muchos árboles de decisión, cada uno construido con un subconjunto aleatorio de los datos y características para tomar decisiones más precisas y robustas.[14]

## Validación del modelo

La validación de modelos es una etapa crucial en el desarrollo de cualquier modelo de machine learning. Permite evaluar su desempeño en datos no vistos previamente y garantizar que generalice bien a nuevos datos. Amazon SageMaker ofrece una variedad de herramientas y técnicas para validar modelos de manera efectiva. En este caso se calculan las métricas

de desempeño promedio en todos los pliegues del modelo para obtener una estimación más robusta del rendimiento del modelo. SageMaker facilita la implementación de la validación cruzada a través de sus API y notebooks, separando una parte de los datos para evaluar el modelo al final del entrenamiento, estos datos también se almacenarán adecuadamente.

## Despliegue del modelo

SageMaker genera los contenedores que almacenan el modelo de machine Learning entrenado y lo despliega como un Endpoint, que son puntos de acceso HTTP que permiten realizar predicciones individuales o por lotes, donde se define la configuración como el tipo de instancia, el número de instancias y la configuración del modelo, de esta manera el servicio crea toda la infraestructura correspondiente y despliega el modelo. Esto permite una alta escalabilidad en la arquitectura, permitiendo ajustar el número de instancias del endpoint y manipularlo a la demanda del mismo, garantizando una alta disponibilidad y tolerancia a fallos, además de optimizar el costo y el rendimiento del endpoint.[15]

## Implementación

### Arquitectura del Diseño

La arquitectura que se muestra en la figura 13 ilustra el flujo de trabajo para el desarrollo y despliegue del modelo de machine learning utilizando Amazon SageMaker.

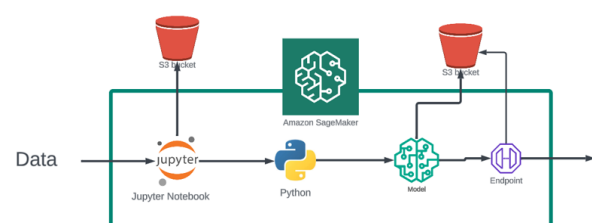


Figura 13: Arquitectura del Diseño

En este caso, los datos de entrenamiento y prueba se almacenan en buckets de Amazon S3. S3 es un servicio de almacenamiento de objetos escalable y duradero, ideal para grandes conjuntos de datos. Estos datos se manipulan en los Jupyter notebooks donde se puede escribir y ejecutar código Python y el uso del framework scikit-learn para preparar los datos, entrenar modelos, por otro lado, estos notebooks se conectan a SageMaker para acceder a los recursos computacionales y a los servicios de machine learning. El modelo entrenado se despliega como un



endpoint en SageMaker y Se envían nuevas muestras de datos al endpoint para obtener predicciones.

Esta arquitectura permite escalar fácilmente los recursos computacionales para entrenar modelos grandes y servir a muchas solicitudes de inferencia, además de proporcionar una interfaz intuitiva y herramientas preconstruidas para acelerar el desarrollo de modelos, permitiendo la gestión completa del ciclo de vida del modelo, desde el tratamiento de los datos, pasando por el entrenamiento y llegando hasta el despliegue y el monitoreo.

## Desarrollo del Notebook

Un Notebook de SageMaker es un entorno de desarrollo interactivo, basado en JupyterLab, que se ejecuta en la nube de Amazon Web Services (AWS), proporcionando un espacio de trabajo completo para explorar datos, construir modelos y realizar experimentos.[16]

El notebook se encarga de todo el flujo del proceso, realizando la importación de las librerías necesarias, que en este caso son sagemaker, ipykernel, sklearn, boto3 y pandas principalmente.

---

```
import sagemaker
import ipykernel
from sklearn.model_selection import
    train_test_split
import boto3
import pandas as pd
```

---

La configuración del Sagemaker consiste en establecer la conexión necesaria para interactuar con el servicio de Amazon SageMaker creando un cliente Boto3. Al crear la sesión se gestionan detalles como la región de AWS en la que se está trabajando y el perfil de usuario. Adicionalmente, se define el nombre del bucket de Amazon S3, previamente creado, que se utilizará para almacenar objetos en la nube.[17]

---

```
sm_boto3 = boto3.client("sagemaker")
sess = sagemaker.Session()
region = sess.boto_session.region_name
bucket = 'bucketprojectsagemaker' #S3 Instance
```

---

El proceso de extracción de datos consiste en la lectura de un documento de csv proporcionado para sus análisis.

---

```
df = pd.read_csv("./datasets/info.csv")
#Ver datos
df.head(10)
```

---

Posterior a la extracción se realiza la normalización y categorización de datos, para estandarizar las muestras y hacer uso de los algoritmos para el entrenamiento.

---

#Eliminar datos Nulos

```
df= df[['ID_DETALLE', 'TRATAMIENTO',
        'ALTURATOTAL', 'DENSIDAD', 'ESTADO_COPA',
        'ESTADO_FUSTE', 'EST_RAIZ_GEN',
        'NIVEL_RIESGO', 'ESTADO_FITO']].dropna()
df.info()
```

```
densidad_type = {'Denso':0, 'Medio':1, 'Muy
    Ralo':2, 'Ralo':3}
estado_types = {'Bueno':0, 'Regular':1, 'Malo':2,
    'Suprimido':4}
riesgo_types = {'Moderada':0, 'Ninguna':1,
    'Severa':2}
```

```
df['DENSIDAD'] =
    df['DENSIDAD'].map(densidad_type).astype(int)
df['ESTADO_COPA'] =
    df['ESTADO_COPA'].map(estado_types).astype(int)
df['ESTADO_FUSTE'] =
    df['ESTADO_FUSTE'].map(estado_types).astype(int)
df['EST_RAIZ_GEN'] =
    df['EST_RAIZ_GEN'].map(estado_types).astype(int)
df['ESTADO_FITO'] =
    df['ESTADO_FITO'].map(estado_types).astype(int)
df['NIVEL_RIESGO'] =
    df['NIVEL_RIESGO'].map(riesgo_types).astype(int)
```

```
tratamientos_types = {'Tala':0,
    'Conservacion':1, 'Traslado':2, 'Poda
    Formacion':3, 'Poda Control de Altura':4,
    'Poda Realce':5, 'Poda Aclareo':6, 'Poda
    Equilibrio':7, 'Poda Mejoramiento':8, 'Poda
    Estructural':9, 'Poda Saneamiento':10,
    'Poda Radicular':11, 'Tratamiento
    Integral':12, 'Tratamiento Especial':13}
```

```
df['TRATAMIENTO'] =
    df['TRATAMIENTO'].map(tratamientos_types).astype(int)
```

---

Para la selección del conjunto de datos de entrenamiento y prueba se decide usar el 80 % del set de datos para el entrenamiento y el 20 % restante como set de pruebas. Primero se deciden las variables a tener en cuenta y la columna a evaluar que en este caso es 'TRATAMIENTO'.

---

```
X = df[['ALTURATOTAL', 'DENSIDAD',
        'ESTADO_COPA', 'ESTADO_FUSTE',
        'EST_RAIZ_GEN', 'NIVEL_RIESGO',
        'ESTADO_FITO']].values
y = df[['TRATAMIENTO']].values
```

```
# 80% -> Data Training | 20% -> Data Testing
X_train, X_test, y_train, y_test =
    train_test_split(X, y, test_size=0.2,
        random_state=0)
```

```
features = list({'ALTURATOTAL', 'DENSIDAD',
    'ESTADO_COPA', 'ESTADO_FUSTE',
```

```

        'EST_RAIZ_GEN', 'NIVEL_RIESGO',
        'ESTADO_FITO'})

label = 'TRATAMIENTO'
trainX = pd.DataFrame(X_train)
trainX[label] = y_train

testX = pd.DataFrame(X_test)
testX[label] = y_test

```

Una vez se realiza la partición de los datos se almacenan en el Bucket S3 para la correspondiente trazabilidad y manejo de la información, permitiendo el acceso rápido y seguro a la información a través del notebook.

```

#Creacion de CSV
trainX.to_csv("train-V-1.csv", index=False)
testX.to_csv("test-V-1.csv", index=False)

#Almacenar datos en el S3
sk_prefix =
    "sagemaker/tratamientos/sklearncontainer"
trainpath = sess.upload_data(
    path = "train-V-1.csv", bucket=bucket,
    key_prefix=sk_prefix
)

testpath = sess.upload_data(
    path = "test-V-1.csv", bucket=bucket,
    key_prefix=sk_prefix
)

```

En la etapa de Creación de Script para SageMaker, el notebook crea un archivo *script.py* que contiene la lógica para entrenar el modelo de clasificación y luego evaluarlo, haciendo uso de las bibliotecas de Scikit-learn, definiendo las variables de entorno que proporcionan las rutas a los directorios donde se almacenan el modelo, los datos de entrenamiento y los datos de prueba, respectivamente. En la figura 14 se muestra la consola de Amazon SageMaker, específicamente la sección de detalles de un trabajo de entrenamiento que ha sido completado satisfactoriamente, mostrando detalles como la duración del entrenamiento, los recursos utilizados, y la configuración del entorno de entrenamiento.

Configuración del trabajo			
Nombre del trabajo RF-custom-sklearn-2024-12-07-01-12-16-244	Estado <span>Completed</span> <a href="#">Ver historial</a>	Series temporales de métricas de SageMaker Desahabilitado	
ARN arn:aws:sagemaker-us-east-1:445567096458:training-job/RF-custom-sklearn-2024-12-07-01-12-16-244	Hora de creación Dec 07, 2024 01:12 UTC	Tiempo de entrenamiento (segundos) 99	ARN del rol de IAM arn:aws:iam::445567096458:role/service-role/SageMaker-SageMaker_Pipelines
	Hora de la última modificación Dec 07, 2024 01:14 UTC	Tiempo facturable (segundos) 35	
		Algoritmo de entrenamiento de spot administrado 64%	
		Origen/nodo principal del trabajo de ajuste -	
Algoritmo			
ARN del algoritmo -	Tamaño del volumen adicional (GiB) 50	Tiempo de espera máximo para entrenamientos de spot administrados 7200	Clave de cifrado de volumen -
Imagen de entrenamiento 68531568378.dkr.ecr.us-east-1.amazonaws.com/sagemaker-sklearn0.23-1-cpu-py3	Tiempo de ejecución máximo (s) 3600	Entrenamiento de spot administrado Habilitado	
Modo de entrada			

Figura 14: Detalles del trabajo de entrenamiento

Para la configuración de entrenamiento del modelo se utiliza un estimador de scikit-learn que es un objeto que encapsula la lógica necesaria para entrenar el modelo de aprendizaje automático, de la siguiente manera:

- `entry_point`: Ruta al script de Python que contiene la lógica de entrenamiento.
- `role`: Rol de IAM en AWS para acceder a recursos (buckets S3).
- `instance_count`: Número de instancias de cómputo para el entrenamiento.
- `instance_type`: Tipo de instancia de cómputo (ml.m5.large).
- `framework_version`: Versión de scikit-learn a utilizar.
- `base_job_name`: Nombre base para los trabajos de entrenamiento.
- `hyperparameters`: Hiperparámetros del modelo (número de estimadores).
- `use_spot_instances`: Indica si usar instancias spot (más económicas).
- `max_wait`: Tiempo máximo de espera para completar el entrenamiento.
- `max_run`: Tiempo máximo de ejecución para el entrenamiento.

```
from sagemaker.sklearn.estimator import SKLearn
```

```
FRAMEWORK_VERSION = "0.23-1"
```

```

sklearn_estimator = SKLearn(
    entry_point="script.py",
    role="arn:aws:iam:::role/service-role/SageMaker",
    instance_count=1,
    instance_type="ml.m5.large",
    framework_version=FRAMEWORK_VERSION,
    base_job_name="RF-custom-sklearn",
    hyperparameters={
        "n_estimators": 100,
        "random_state": 0,
    },
    use_spot_instances = True,
    max_wait = 7200,
    max_run = 3600
)

```

```

# launch training job, with asynchronous call
sklearn_estimator.fit({"train": trainpath,
    "test": testpath}, wait=True)

```

En la figura 15 se muestra la configuración de los datos de salida y los hiperparámetros del modelo, que se configuran antes de comenzar el entrenamiento y afectan directamente el rendimiento del modelo.

Luego del entrenamiento se extrae la ubicación del modelo entrenado alojado en el bucket de S3.

<b>Configuración de datos de salida</b>	
ruta de salida de S3 s3://sagemaker-us-east-1-445567096458/	
<b>Hiperparámetros</b>	
Clave	Valor
n_estimators	100
random_state	0
sagemaker_container_log_level	20
sagemaker_job_name	"RF-custom-sklearn-2024-12-07-01-12-16-244"
sagemaker_program	"script.py"
sagemaker_region	"us-east-1"
sagemaker_submit_directory	"S3://sagemaker-us-east-1-445567096458/RF-custom-sklearn-2024-12-07-01-12-16-244/source/source.tar.gz"

Figura 15: Hiperparámetros y Datos de Salida del entrenamiento

```
sklearn_estimator.latest_training_job.wait(logs="None")
artifact = sm_boto3.describe_training_job(
    TrainingJobName =
        sklearn_estimator.latest_training_job.name
)["ModelArtifacts"]["S3ModelArtifacts"]
print("Model artifact " + artifact)
```

Creación un modelo de SageMaker a partir del artifact

```
from sagemaker.sklearn.model import SKLearnModel
from time import gmtime, strftime

model_name = "Custom-sklearn-model-" +
    strftime("%Y-%m-%d-%H-%M-%S", gmtime())
model = SKLearnModel(
    name=model_name,
    model_data=artifact,
    role="arn:aws:iam::role/service-role/SageMaker",
    entry_point="script.py",
    framework_version=FRAMEWORK_VERSION,
)
model
```

Para el Despliegue del modelo se crea un nombre único para un endpoint, posteriormente se indica la creación de la instancia para el endpoint y el tipo de instancia

```
endpoint_name = "Custom-sklearn-model-" +
    strftime("%Y-%m-%d-%H-%M-%S", gmtime())
print("EndpointName={}".format(endpoint_name))

predictor = model.deploy(
    initial_instance_count=1,
    instance_type="ml.m4.xlarge",
    endpoint_name=endpoint_name
)
endpoint_name
```

El endpoint creado muestra su estado en la consola de Amazon SageMaker indicando el correcto despliegue del modelo, como se ve en la figura 16.

De esta manera, el Endpoint del modelo se encuentra disponible para la realización de pruebas y validación de los datos, la información detallada del endpoint se presenta como en la figura 17

Amazon SageMaker AI >

Puntos de conexión

Puntos de conexión

🔍

Buscar puntos de conexión

Actualizar punto de enlace

Acciones ▾

Crear punto de enlace

< 1 >

Nombre	ARN	Hora de creación	Estado	Actualización más reciente
<div><div>○</div><div><div>Custom-sklearn-model-2024-12-07-01-15-15</div></div></div>	arn:aws:sagemaker:us-east-1:445567096458:endpoint/Custom-sklearn-model-2024-12-07-01-15-15	6/12/2024, 20:15:18	<div>InService</div>	6/12/2024, 20:18:51

Figura 16: Endpoint del Modelo

Amazon SageMaker AI > Puntos de conexión > Custom-sklearn-model-2024-12-07-01-15-15

Custom-sklearn-model-2024-12-07-01-15-15

Eliminar

Resumen del punto de conexión

Nombre	Estado	Tipo
Custom-sklearn-model-2024-12-07-01-15-15	InService	En tiempo real
ARN	Hora de creación	Última actualización
arn:aws:sagemaker:us-east-1:445567096458:endpoint/Custom-sklearn-model-2024-12-07-01-15-15	Fri Dec 06 2024 20:15:18 GMT-0500 (hora estándar de Colombia)	Fri Dec 06 2024 20:18:51 GMT-0500 (hora estándar de Colombia)
URL	Registros del contenedor del modelo	Alarmas
https://runtime.sagemaker.us-east-1.amazonaws.com/endpoints/Custom-sklearn-model-2024-12-07-01-15-15/invocations	/aws/sagemaker/endpoints/Custom-sklearn-model-2024-12-07-01-15-15	0 alarmas
Obtenga más información sobre la API		

Figura 17: Información detallada del Endpoint del Modelo

## Script de Entrenamiento en Python

El archivo *script.py* creado desde el Notebook, contiene la lógica para entrenar el modelo de clasificación y luego evaluarlo, haciendo uso de las bibliotecas de Scikit-learn. Este código implementa un flujo de trabajo básico de aprendizaje automático utilizando un modelo de Random Forest para entrenar y evaluar un clasificador, además de guardar el modelo entrenado para usos futuros.

El script realiza el cargue de los archivos en el modelo desde su ubicación en el S3, dicha información se registra en variables globales. Los datos se dividen en conjuntos de entrenamiento (X\_train, y\_train) y prueba (X\_test, y\_test).

```
train_df = pd.read_csv(os.path.join(args.train,
    args.train_file))
test_df = pd.read_csv(os.path.join(args.test,
    args.test_file))

features = list(train_df.columns)
label = features.pop(-1)

X_train = train_df[features]
X_test = test_df[features]
y_train = train_df[label]
y_test = test_df[label]
```

Hace uso de las librerías Scikit-learn, para crear el modelo y calcular métricas, Joblib, para guardar y cargar el modelo entrenado, Pandas y NumPy para manipular datos del modelo. El código se encarga de definir los hiperparámetros para el modelo y los directorios de los archivos de prueba y entrenamiento, con los cuales se entrena un modelo Random Forest con parámetros como el número de árboles y la profundidad del mismo.

```
model = RandomForestClassifier(n_estimators=100,
                              max_depth=2, random_state=0)
model.fit(X_train, y_train)
```

Se generan predicciones (`y_pred_test`) para los datos de prueba y se calcula la precisión del modelo (`test_acc`) y un reporte de clasificación (`test_rep`) con métricas como precisión, recuperación y F1score.

```
y_pred_test = model.predict(X_test)
test_acc = accuracy_score(y_test, y_pred_test)
test_rep = classification_report(y_test,
                                y_pred_test)
```

Los resultados del modelo se imprimen, incluyendo el número total de filas en el conjunto de prueba, la precisión del modelo y los detalles del desempeño por clase en el reporte de clasificación, tal como se ve en la figura 18.

```
Training RandomForest Model
/miniconda3/lib/python3.7/site-packages/sklearn/metrics/_classification.py:1221: UndefinedMetricWarning: Precision a
ware pef(average, modifier, msg_start, len(result))
RESULTS FOR TESTING DATA
Total Rows are: 61084
[TESTING] Model Accuracy is: 0.5143736494008251
[TESTING] Testing Report:
      precision    recall  f1-score   support
0      0.55      0.93      0.69      28220
1      0.38      0.67      0.49      7721
2      0.00      0.00      0.00      3695
4      0.00      0.00      0.00      539
5      0.00      0.00      0.00      1781
6      0.00      0.00      0.00      795
7      0.00      0.00      0.00      989
9      0.00      0.00      0.00      9308
10     0.00      0.00      0.00      174
11     0.00      0.00      0.00      958
12     0.00      0.00      0.00      6963
13     0.00      0.00      0.00      21
accuracy      0.51      0.51      0.51      61084
macro avg     0.08      0.13      0.10      61084
weighted avg  0.30      0.51      0.38      61084
2024-12-07 01:14:29,502 sagmaker-containers INFO Reporting training SUCCESS
```

Figura 18: Resultados de la ejecución del script.py

## Resultados y Evaluación



Figura 19: Resultados del modelo

Los resultados del modelo de IA desarrollado para reconocer tratamientos silviculturales en árboles urbanos de Bogotá muestran un desempeño promotor. El modelo logra identificar correctamente los tratamientos en un 65 % de los casos, lo cual es un buen comienzo. El uso de recursos como memoria, CPU y disco es eficiente, lo que indicando que el modelo está optimizado.

A pesar de los resultados positivos, existen oportunidades para mejorar el rendimiento del modelo. Por lo tanto, es necesario recopilar más datos de alta calidad, los cuales se pueden validar con expertos en el

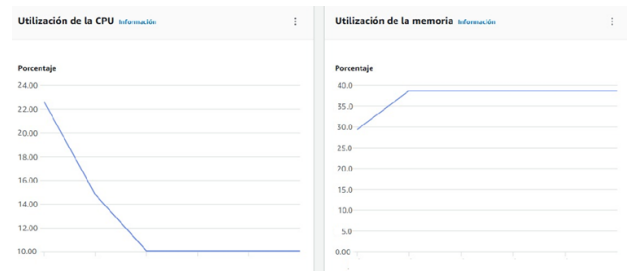


Figura 20: Uso de la CPU y memoria

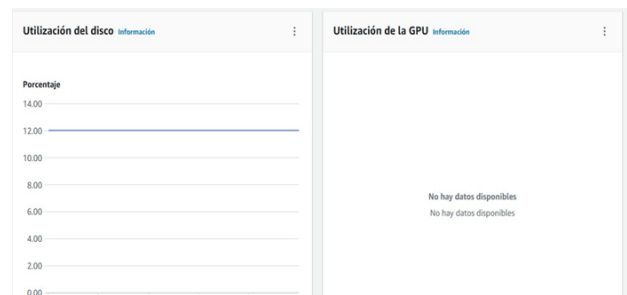


Figura 21: Utilización del disco

área mencionada, adicionalmente se pueden obtener mejores resultados ajustando los hiperparámetros y explorar diferentes arquitecturas de redes neuronales. Además, es importante analizar los errores que comete el modelo para identificar patrones y mejorar su precisión.

## Conclusión

Los resultados obtenidos indican que el modelo presenta una buena precisión para determinar el tratamiento silvicultural más adecuado según las características dasométricas y fitosanitarias evaluadas. No obstante, se identificaron áreas de oportunidad para optimizar el rendimiento del modelo, tales como la realización de pruebas más exhaustivas y la exploración de otros algoritmos que contribuyan a su mejora.

Además, surge la posibilidad de mejorar el servicio integrándolo al sistema actual de la entidad, permitiendo la validación de datos por parte de ingenieros. Esto facilitaría la optimización del modelo y una evaluación más sólida de la escalabilidad del sistema.

La implementación de este sistema entrenado con inteligencia artificial no solo representa un avance en la automatización de decisiones silviculturales, sino que también ofrece un enfoque basado en datos que fortalece la gestión sostenible del arbolado urbano. Asimismo, se plantea como una herramienta clave para respaldar la planificación estratégica de intervenciones, contribuyendo a la conservación y al manejo eficiente de los recursos forestales.



Finalmente, se destaca la importancia de realizar colaboraciones interdisciplinarias entre expertos en inteligencia artificial, ingenieros forestales y entidades gubernamentales o privadas. Esto garantizará la continuidad del desarrollo del sistema, su adaptación a nuevas necesidades y su alineación con las políticas de sostenibilidad urbana y medioambiental.

El trabajo futuro de este proyecto se puede enfocar en optimizar el modelo de inteligencia artificial mediante pruebas más exhaustivas y la exploración de nuevos algoritmos que mejoren su precisión y rendimiento. Además, de ampliar el conjunto de datos, incluyendo una mayor diversidad de especies arbóreas y condiciones fitosanitarias, así como el uso de datos sintéticos para cubrir escenarios poco frecuentes. Esto permitirá una mayor generalización y robustez en las predicciones del sistema.

Otro aspecto clave sería la integración del sistema con plataformas y herramientas existentes de gestión forestal, facilitando su uso por parte de ingenieros y entidades responsables del manejo del arbolado urbano. También se puede trabajar en interfaces intuitivas y mecanismos que permitan la retroalimentación continua para ajustar y mejorar el modelo, y a su vez realizar estudios pilotos para evaluar el impacto del sistema en diferentes ciudades y explorar su escalabilidad a nivel nacional o internacional.

Asimismo, se plantea incorporar de nuevos factores de análisis, como variables ambientales y tecnologías avanzadas, incluyendo imágenes satelitales y datos LiDAR, para enriquecer los criterios de decisión. Con la finalidad de automatizar la recopilación de datos mediante dispositivos IoT o drones, y generar reportes en tiempo real que optimicen la toma de decisiones. La divulgación del sistema y la capacitación de los usuarios finales también serán fundamentales para garantizar su adopción y sostenibilidad a largo plazo.

## Referencias

- [1] Secretaria Distrital de Ambiente Bogotá. *Cartilla Silvicultura Urbana*. <https://www.ambientebogota.gov.co/es/documentos-silvicultura-urbana>. 2021.
- [2] Kathleen L. Wolf et al. "Urban Trees and Human Health: A Scoping Review". En: *International Journal of Environmental Research and Public Health* 17.12 (#jun# de 2020), pág. 4371. ISSN: 1660-4601. DOI: 10.3390/ijerph17124371. URL: <http://dx.doi.org/10.3390/ijerph17124371>.
- [3] Jardín Botánico de Bogotá José Celestino Mutis - Alcaldía Mayor de Bogotá. *Manual de Silvicultura Urbana para Bogotá*. Compendio de investigaciones y recomendaciones para la arborización urbana en Bogotá. Jardín Botánico de Bogotá José Celestino Mutis, Alcaldía Mayor de Bogotá. 2011. URL: <https://www.anla.gov.co/eureka/manuales-guias-y-programas/manuales/24-manual-de-silvicultura-urbana-para-bogota>.
- [4] Sociedad Internacional De Arboricultura. *Sociedad Internacional De Arboricultura*. <https://www.isahispana.com/>. 2024.
- [5] Alcaldía Mayor de Bogotá - Secretaría Distrital de Ambiente - Jardín Botánico de Bogotá José Celestino Mutis. *Arbolado Urbano de Bogotá: Identificación, descripción y bases para su manejo*. <https://www.ambientebogota.gov.co/documents/10184/626774/Arbolado.pdf/c687bdec-d212-4001-9bea-3edf227557e4>. 2010.
- [6] Leonardo Correa Franco. *Manual de silvicultura urbana para Medellín: gestión, planeación y manejo de la infraestructura verde*. es. 2015.
- [7] Fernando Pérez Rodríguez Francisco Rodríguez Puerta et al. "La inteligencia artificial ya es una realidad en el sector forestal: SECCIÓN APUNTES - REVISTA FORESTA". En: 81(2021) (dic. de 2021), págs. 22-23. URL: [https://www.researchgate.net/publication/358262839-La\\_inteligencia\\_artificial\\_ya\\_es\\_una\\_realidad\\_en\\_el\\_sector\\_forestal\\_SECCION\\_APUNTES\\_-\\_REVISTA\\_FORESTA](https://www.researchgate.net/publication/358262839-La_inteligencia_artificial_ya_es_una_realidad_en_el_sector_forestal_SECCION_APUNTES_-_REVISTA_FORESTA).
- [8] Nadina J. Galle et al. "Mapping the diversity of street tree inventories across eight cities internationally using open data". En: *Urban Forestry & Urban Greening* 61 (#jun# de 2021), pág. 127099. ISSN: 1618-8667. DOI: 10.1016/j.ufug.2021.127099. URL: <http://dx.doi.org/10.1016/j.ufug.2021.127099>.
- [9] Mohammadhossein Ghahramani et al. "Leveraging artificial intelligence to analyze citizens' opinions on urban green space". En: *City and Environment Interactions* 10 (#apr# de 2021), pág. 100058. ISSN: 2590-2520. DOI: 10.1016/j.cacint.2021.100058. URL: <http://dx.doi.org/10.1016/j.cacint.2021.100058>.
- [10] Luisa Velasquez-Camacho et al. "Remotely Sensed Tree Characterization in Urban Areas: A Review". En: *Remote Sensing* 13.23 (dic. de 2021), pág. 4889. ISSN: 2072-4292. DOI: 10.3390/rs13234889. URL: <http://dx.doi.org/10.3390/rs13234889>.

- [11] Guoyang Liu et al. "Tree internal defects detection method based on ResNet improved subspace optimization algorithm". En: *NDT and E International* 147 (2024). ISSN: 0963-8695. DOI: 10.1016/j.ndteint.2024.103183. URL: <http://dx.doi.org/10.1016/j.ndteint.2024.103183>.
- [12] Haixia Sun et al. "Maturity detection of 'Huping' jujube fruits in natural environment using YOLO-FHLD". En: *Smart Agricultural Technology* 9 (dic. de 2024), pág. 100670. ISSN: 2772-3755. DOI: 10.1016/j.atech.2024.100670. URL: <http://dx.doi.org/10.1016/j.atech.2024.100670>.
- [13] A. Hornero et al. "Detection of oak decline using radiative transfer modelling and machine learning from multispectral and thermal RPAS imagery". En: *International Journal of Applied Earth Observation and Geoinformation* 127 (mar. de 2024), pág. 103679. ISSN: 1569-8432. DOI: 10.1016/j.jag.2024.103679. URL: <http://dx.doi.org/10.1016/j.jag.2024.103679>.
- [14] Amazon Web Services. *Model training - Amazon SageMaker AI*. 2024. URL: <https://docs.aws.amazon.com/sagemaker/latest/dg/train-model.html>.
- [15] Amazon Web Services. *Deploy models for inference - Amazon SageMaker AI*. 2024. URL: <https://docs.aws.amazon.com/sagemaker/latest/dg/deploy-model.html>.
- [16] Amazon Web Services. *Create, store, and share features with Feature Store - Amazon SageMaker AI*. 2024. URL: <https://docs.aws.amazon.com/sagemaker/latest/dg/feature-store.html>.
- [17] Amazon Web Services. *Recommendations for choosing the right data preparation tool in SageMaker AI - Amazon SageMaker AI*. 2024. URL: <https://docs.aws.amazon.com/sagemaker/latest/dg/data-prep.html>.

**Cuadro 2:** *Tratamientos Silviculturales.* [3][6]

Tratamiento	Descripción
Tala	Corte y remoción completa de un árbol. Se realiza por diversas razones, como enfermedades graves, riesgo de caída, obstrucción de servicios públicos o como parte de un proyecto de construcción.
Conservación	Conjunto de acciones encaminadas a proteger y preservar los árboles y los bosques, garantizando su supervivencia a largo plazo.
Traslado	Proceso de remover un árbol de un lugar a otro, generalmente para salvarlo de una obra o para mejorar su ubicación. Requiere de técnicas especializadas y cuidados posteriores para garantizar la supervivencia del árbol.
Poda de formación	Se realiza en árboles jóvenes para guiar su crecimiento y desarrollo, favoreciendo la formación de una estructura fuerte y equilibrada.
Poda de control de altura	Se realiza para limitar la altura de un árbol, por ejemplo, para evitar que interfiera con líneas eléctricas o con la vista.
Poda de realce	Consiste en eliminar las ramas bajas de un árbol para aumentar el espacio libre debajo de la copa y facilitar el paso de personas o vehículos.
Poda de aclareo	Se realiza para reducir la densidad de la copa de un árbol, mejorando la circulación del aire y la penetración de la luz.
Poda de equilibrio	Busca restablecer el equilibrio entre las diferentes partes de la copa de un árbol, corrigiendo desequilibrios causados por el crecimiento natural o por podas anteriores.
Poda de mejoramiento	Es una poda integral que combina diferentes técnicas para mejorar la apariencia y la salud de un árbol.
Poda de estructura	Se enfoca en fortalecer la estructura del árbol, eliminando ramas débiles o mal posicionadas que puedan causar daños.
Poda sanitaria	Consiste en eliminar las ramas enfermas, muertas o dañadas, para prevenir la propagación de enfermedades y mejorar la salud del árbol.
Poda radical	Se realiza en las raíces del árbol, generalmente para reducir su tamaño o para evitar daños a infraestructuras.
Tratamiento Integral	Abarca un conjunto de medidas para mejorar la salud y el vigor de un árbol, incluyendo podas, fertilización, riego y control de plagas y enfermedades.
Tratamiento Especial	Se refiere a tratamientos específicos para solucionar problemas particulares de un árbol, como la aplicación de fungicidas, insecticidas o tratamientos para heridas.

**Cuadro 3:** Arquitecturas de Deep Learning para Caracterización de Árboles Urbanos.[10]

Arquitectura	Descripción
PSPNet	Red neuronal que captura información contextual a diferentes escalas, ideal para segmentar imágenes y entender la relación entre los píxeles.
Faster R-CNN	Variante de R-CNN, diseñada para detectar objetos en imágenes con alta precisión, lo que permite identificar y localizar árboles en imágenes aéreas o terrestres.
ResNet	Red neuronal residual que permite aprender características muy profundas, lo que la hace adecuada para tareas complejas como la clasificación y segmentación de imágenes de alta resolución.
SegNet	Especializada en segmentación de imágenes, es decir, en clasificar cada píxel de una imagen en una determinada categoría (árbol, suelo, etc.).
VGG16	Arquitectura clásica de CNN, utilizada como punto de referencia en muchas investigaciones y ampliamente utilizada para tareas de clasificación.
YOLO	Red neuronal conocida por su velocidad y precisión en la detección de objetos en tiempo real, ideal para aplicaciones donde se requiere una respuesta rápida.
DCNN	Término general que engloba a todas las redes neuronales convolucionales profundas, utilizadas para extraer características de imágenes.
PointNet	Diseñada para procesar datos de puntos 3D, como los obtenidos por sensores LiDAR, lo que permite crear modelos 3D detallados de árboles.
Bayesian Network	Modelo gráfico probabilístico utilizado para representar relaciones entre variables y hacer inferencias probabilísticas, útil para modelar la incertidumbre en la clasificación de árboles.
Other	Categoría que incluye otras arquitecturas o combinaciones de arquitecturas no mencionadas anteriormente.

**Cuadro 4:** Normalización de los datos

Criterio	Normalización
Densidad	Denso:0 Medio:1 Muy Ralo:2 Ralo:3
Estado de Copa	Bueno:0  Regular:1 Malo:2 Suprimido:3
Estado de Fuste	Bueno:0  Regular:1 Malo:2 Suprimido:3
Estado Raíz General	Bueno:0  Regular:1 Malo:2 Suprimido:3
Estado Fito-sanitario	Bueno:0  Regular:1 Malo:2 Suprimido:3
Nivel de Riesgo	Moderado:0  Ninguno:1 Severo:2
Tratamientos	Tala:0 Conservación:1 Traslado:2 Poda Formación:3 Poda Control de Altura:4 Poda Realce:5 Poda Aclareo:6 Poda Equilibrio:7 Poda Mejoramiento:8 Poda Estructural:9 Poda Saneamiento:10 Poda Radicular:11 Tratamiento Integral:12 Tratamiento Especial:13