

# Evaluación de algoritmos de machine learning en la clasificación de imágenes satelitales multiespectrales, caso: Amazonia Peruana.

Erik J. Olmedo Zambrano

October 2023

## Abstract

En este artículo se aborda la implementación de técnicas de machine learning en la clasificación de imágenes satelitales de la región amazónica peruana. El objetivo principal es abordar problemáticas como la deforestación y el cambio climático. Para ello, se realizó un preprocesamiento con QGIS y se implementaron algoritmos en entornos de desarrollo como anaconda, jupyter notebook y Python 3.0. Los resultados obtenidos muestran que la clasificación basada en píxeles y no supervisada kmeans presenta limitaciones en la clasificación de coberturas. En cambio, la clasificación supervisada con los algoritmos SVM y naive bayes obtuvo una precisión de 0.909, en comparación con el algoritmo de árbol de decisión que obtuvo una precisión de 0.864. A pesar de los pocos datos de entrenamiento y validación del modelo, se puede concluir que la implementación de técnicas de machine learning en la clasificación de imágenes satelitales puede ser una herramienta útil para abordar problemáticas ambientales en la región amazónica peruana.

## 1 Introducción

En los últimos años, la Teledetección y la Inteligencia Artificial han avanzado significativamente en el procesamiento e interpretación de imágenes satelitales en diversas áreas de la ingeniería, como la agricultura de precisión y la detección de problemas ambientales. Gracias al acceso a plataformas satelitales como Landsat y Sentinel, y al uso de tecnologías y procesadores de algoritmos de Machine Learning, se ha logrado una mayor eficiencia computacional en la clasificación de coberturas basadas en píxeles. Actualmente, se dispone de datos y resoluciones espaciales que permiten una mejor capacidad para que los algoritmos clasifiquen y utilicen datos digitales segmentados y/o clasificados. Además, el uso de herramientas de inteligencia artificial está permitiendo la aplicación de técnicas de análisis de coberturas y agricultura de precisión. En este estudio, se utiliza la clasificación basada en píxeles con imágenes satelitales Landsat 8, empleando técnicas tanto supervisadas como no supervisadas, tales

como el algoritmo de Kmeans, árboles de decisión, redes neuronales y máquinas de vectores de soporte.

## 2 Planteamiento del Problema

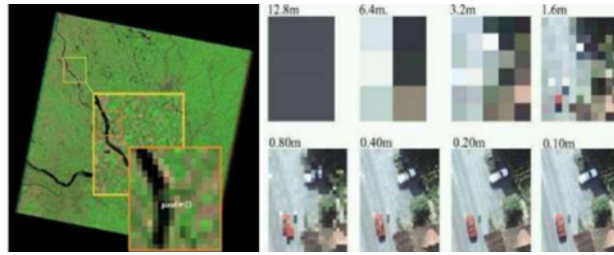
La utilización de imágenes satelitales en distintas áreas de conocimiento es cada vez más común, sin embargo, los algoritmos de clasificación basados en píxeles presentan limitaciones que afectan la precisión y eficiencia computacional del procesamiento de datos. Es por eso que se hace necesario evaluar y mejorar los algoritmos implementados, considerando factores como el tipo de datos, arquitecturas y hiperparámetros configurados. En el caso de las imágenes obtenidas por teledetección, se presentan patrones más complejos y una información espacial, temporal y espectral más rica, lo que impone mayores requerimientos a los métodos de procesamiento. En este contexto, se plantea el objetivo de optimizar la clasificación de imágenes satelitales mediante el uso de algoritmos de Deep Learning, con el fin de mejorar la precisión y eficiencia computacional en su procesamiento.

## 3 Bases Teóricas

### 3.1 Resoluciones de las Imágenes de Satélite

- **Una imagen de satélite se caracteriza por las siguientes modalidades de resolución:**

La resolución espacial de una imagen satelital se determina por el tamaño de píxel, que representa el objeto más pequeño que se puede distinguir en la imagen. Esta medida se expresa en metros sobre el terreno y depende de la altura del sensor, el ángulo de visión, la velocidad de escaneo y las características ópticas del sensor. Por ejemplo, las imágenes Landsat TM tienen una resolución espacial de 30x30 m en las bandas 1,2,3,4,5 y 7, y de 120x120m en la banda 6 térmica. Por otro lado, la resolución espectral indica el número y ancho de las bandas en las que el sensor puede captar radiación electromagnética. Los nuevos sensores hiperespectrales pueden tener hasta 256 canales con un ancho de banda muy estrecho para separar objetos por su comportamiento espectral. En cuanto a la resolución radiométrica, se refiere al número de niveles digitales utilizados para expresar los datos recogidos por el sensor. Cuanto mayor sea este número, mayor será el detalle con el que se puede expresar la información. Finalmente, la resolución temporal se refiere a la frecuencia con la que el satélite pasa por un mismo punto de la superficie terrestre. Esto depende de las características de la órbita del satélite.



### 3.2 Firmas Espectrales – Reflectancia

La utilización de firmas espectrales permite analizar la variación de la radiación reflejada por los objetos en función de la longitud de onda, lo que resulta muy útil en análisis multispectrales para identificar elementos específicos en imágenes aéreas. A pesar de que los ojos humanos detectan la luz reflejada por los objetos, sólo pueden percibir una pequeña parte del espectro electromagnético, el visible. Por ello, en teledetección se caracteriza a las ondas electromagnéticas por su longitud de onda en micrómetros o nanómetros. La radiación reflejada en función de la longitud de onda se conoce como firma espectral de la superficie, lo que nos permite categorizar los píxeles o firma espectral en diferentes clases según su semejanza, como bosques, cultivos, áreas urbanas o cuerpos de agua. La reflectividad de los objetos respecto a las diferentes longitudes de onda permite identificarlos y distinguirlos entre sí, lo que resulta muy útil en la identificación de objetos y procesos en la superficie terrestre. El satélite Landsat 8 es una herramienta valiosa para la observación terrestre y la teledetección. Su capacidad para capturar imágenes en 11 bandas espectrales, proporciona información detallada sobre la cobertura terrestre. Los algoritmos de aprendizaje automático se han convertido en una herramienta útil para la clasificación de imágenes satelitales, ya que pueden modelar firmas de clase complejas y aceptar una variedad de datos de predicción de entrada. Aunque los métodos paramétricos todavía se utilizan comúnmente en los artículos de aplicación, los enfoques de aprendizaje automático han demostrado producir resultados más precisos para datos complejos con un espacio de características de alta dimensión. En ausencia de información de referencia sobre el terreno, las técnicas no supervisadas pueden utilizarse para clasificar imágenes satelitales en distintas regiones de cobertura terrestre. El conocimiento extraído del conjunto de datos tiene la forma de centros de clúster óptimos, lo que permite compararlos con las clases de información y garantizar que las distancias intra-clusters sean mínimas y que las distancias entre los inter-clusters sean máximas.

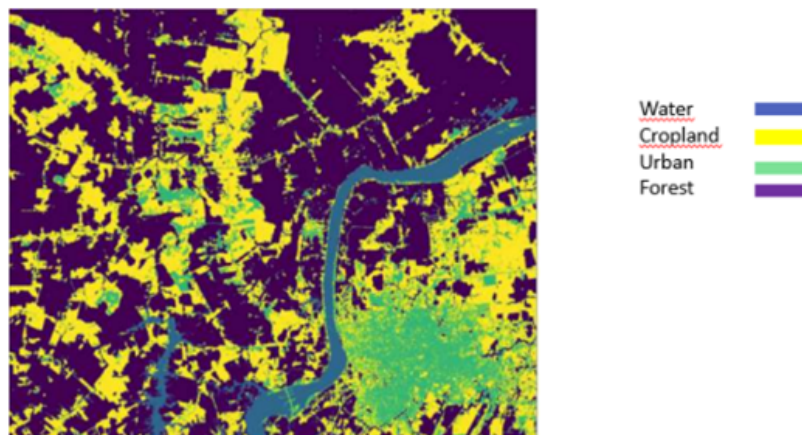
## 4 Metodologia De Clasificacion Con Machine Learning

El presente estudio tiene como objetivo realizar una clasificación inicial de la zona de la selva peruana alrededores de la provincia de coronel Portillo, De-

partamento de Ucayali, utilizando algoritmos de machine learning. Se pretende discriminar zonas urbanas, cultivos, cuerpos de agua y bosque, con el fin de abordar problemáticas de deforestación, minería ilegal y contaminación ambiental. Para ello, se utilizó la plataforma Google Earth Engine y Jupyter Notebook para el procesamiento de las imágenes satelitales Landsat 8, aplicando corrección atmosférica y topográfica para la eliminación de nubes y sombras, además de calcular el NDVI para cada imagen. Se utilizó software especializado de sistemas de Información Geográfica (GIS) como QGIS y GRASSGIS para la etapa de análisis de datos espectrales. Además, se realizó una visualización RGB con los canales rojo, verde y azul y se calculó el índice espectral NDVI para definir y analizar las clases que intervienen en la clasificación. Los datos fueron normalizados para mejorar los sobreajustes que pueda generar el modelo con los datos atípicos.

## 5 Resultados

En el campo de la teledetección, existen dos técnicas principales para la clasificación de imágenes satelitales: el aprendizaje no supervisado y el supervisado. En la primera, se aplican técnicas para clasificar la imagen en distintas regiones de cobertura terrestre sin información previa de referencia sobre el terreno. En la segunda, se utilizan muestras representativas de diferentes tipos de interés de cobertura superficial para entrenar a la computadora para reconocer muestras espectrales similares para cada clase. Entre las diversas técnicas supervisadas, las redes neuronales artificiales y los clasificadores estadísticos como SVM, decision tree, random forest, bayes y otros son ampliamente utilizados. Es necesario tener en cuenta que estos algoritmos requieren datos de entrenamiento y validación para calcular su precisión. En general, el aprendizaje supervisado tiene una precisión mayor que el no supervisado, pero se necesita más datos de entrenamiento para su generalización en el mapa.



## 6 Conclusión

En el procesamiento de imágenes satelitales, la información espectral contenida en los píxeles no es suficiente para lograr una clasificación precisa. Es necesario combinar esta información con otras fuentes de datos y realizar un preprocesamiento de correcciones radiométricas y/o atmosféricas antes de aplicar algoritmos de Machine Learning. En cuanto a la clasificación no supervisada, el algoritmo Kmeans presenta limitaciones en la aplicación de imágenes satelitales debido a la asignación inicial aleatoria de centroides y posibles outliers. Por otro lado, la clasificación supervisada con algoritmos como SVM y Naive Bayes han demostrado mayor precisión en comparación con el algoritmo de árbol de decisión. Es importante destacar que la cantidad de datos de entrenamiento y validación del modelo de clasificación influye en su precisión. Sin embargo, según la revisión bibliográfica, los clasificadores máxima probabilidad, redes neuronales convolucionales y SVM son los más utilizados y efectivos en el procesamiento de imágenes satelitales y teledetección.

## 7 Link

<https://github.com/joelinho10/Inteligencia-Artificial.git>