

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

**Nilton Chucos Baquerizo**

[nilton.chucos@unas.edu.pe](mailto:nilton.chucos@unas.edu.pe)

<https://orcid.org/0000-0001-5596-4454>

**Edwin Jesus Vega Ventocilla**

[edwin.vega@unas.edu.pe](mailto:edwin.vega@unas.edu.pe)

<https://orcid.org/0000-0002-3628-9016>

Universidad Nacional Agraria de la Selva  
Tingo María- Perú

### RESUMEN

En el presente artículo, en primera instancia se desarrolla el fundamento teórico e implementación de algunas técnicas de machine learning (Aprendizaje Automático), aplicadas a la clasificación de imágenes satelitales, el área estudio encuentra ubicado en la región de la amazonia peruana, fundamentalmente por el interés en abordar alguna problemática sobre la deforestación, cambios de cobertura, cambio climático, etc. La metodología general se basa en los procesos para la implementación de algoritmos de machine learning, se realizó un preprocesamiento previo con software QGIS para la descarga y corrección atmosférica, y los algoritmos se implementaron en entornos de desarrollo (IDE) como anaconda, jupyter notebook y el lenguaje de programación Python 3.0. Los resultados de clasificación basada en pixeles y no supervisada kmeans, no garantiza una clasificación robusta y tiene limitaciones en la clasificación de coberturas en imágenes satelitales ya que las firmas espectrales de cada pixel tienen una resolución espacial de 30m para imágenes Landsat, sin embargo, la clasificación supervisada, el algoritmo SVM y naive bayes presentaron mejor precisión de 0.909 en comparación con el algoritmo de árbol de decisión 0.864, considerando los pocos datos de entrenamiento y validación del modelo de clasificación de coberturas.

**Palabras clave:** algoritmos de clasificación; kmeans; imágenes satelitales; teledetección, aprendizaje no supervisado; aprendizaje supervisado.

## **Evaluation of machine learning algorithms in the classification of multispectral satellite images, case: Peruvian Amazon**

### **ABSTRACT**

In this article, in the first instance, the theoretical foundation and implementation of some machine learning techniques are developed (Automatic Learning), applied to the classification of satellite images, the study area is located in the Peruvian Amazon region, mainly due to the interest in tackling some problem about deforestation, changes in coverage, climate change, etc. The general methodology is based on the processes for the implementation of machine learning algorithms, a previous preprocessing was carried out with QGIS software for downloading and atmospheric correction, and the algorithms were implemented in development environments (IDE) such as anaconda, jupyter notebook and the Python 3.0 programming language. The results of classification based on pixels and unsupervised kmeans, do not guarantee a robust classification and have limitations in the classification of coverage in satellite images since the spectral signatures of each pixel have a spatial resolution of 30m for Landsat images, however, the supervised classification, the SVM algorithm and naive bayes presented better precision of 0.909 compared to the decision tree algorithm 0.864, considering the few training and validation data of the coverage classification model.

**Keywords:** classification algorithms; kmeans; satellite images; remote sensing; unsupervised learning; supervised learning.

Artículo recibido: 15 enero 2022

Aceptado para publicación: 08 febrero 2022

Correspondencia: [nilton.chucos@unas.edu.pe](mailto:nilton.chucos@unas.edu.pe)

Conflictos de Interés: Ninguna que declarar

## **I. INTRODUCCION**

La teledetección y la Inteligencia artificial en los últimos años están desarrollando técnicas para el procesamiento de imágenes satelitales y su interpretación en diferentes ámbitos como el estudio de coberturas, la agricultura de precisión, problemas ambientales. Esto también es posible con el uso de tecnologías y procesadores de algoritmos de machine learning con eficiencia computacional y el acceso a datos accesibles de plataformas satelitales como Landsat y Sentinel para fines de investigación.

La clasificación de coberturas basadas en píxeles también esta evolucionando, ya que actualmente se dispone de datos y resoluciones espaciales que permiten una mejor capacidad de que los algoritmos clasifiquen y utilizan en su etapa de entrenamiento datos digitales que puedan ser segmentados y/o clasificados de tal manera que se automatice la clasificación de firmas espectrales y se plasme en la interpretación de diferentes tipos de coberturas terrestres en diversas áreas de la ingeniería. Además los avances tecnológicos y la ciencia van unidos de la mano y por ello, el uso de la Teledetección y la aplicación de herramientas de inteligencia artificial están permitiendo la aplicación de técnicas de agricultura de precisión y análisis de coberturas (Utiltech, 2018).

En el presente estudio se utiliza imágenes satelitales Landsat 8, y se realiza una clasificación basada en píxeles, que contienen datos digitales de la firma espectral, aplicando técnicas de machine Learning no supervisado como el algoritmo de Kmeans y además de aprendizaje supervisado basado en que los datos se encuentren etiquetados previo al entrenamiento. Las técnicas supervisadas más empleadas e la clasificación e imágenes han sido el árbol de decisión, las redes neuronales y las máquinas de vectores de soporte (Leiva et al., 2019).

## **II. PLANTEAMIENTO DEL PROBLEMA**

Las imágenes obtenidas por satélites Landsat y Sentinel, tienen aplicaciones muy diversas en distintas áreas de conocimiento: ciencias ambientales, agricultura de precisión, teledetección, Hidrografía y muchos más, sin embargo los algoritmos de clasificación con los que se optimiza la clasificación basada en píxeles tienen limitaciones dependiendo del tipo de datos y arquitecturas además de los hiperparametros (Ramírez et al., 2020) a configurar en los algoritmos implementados, de tal manera que se pueda mejorar la precisión en la clasificación y ser mas eficientes computacionalmente ya que la cantidad de datos en imágenes satelitales implica un costo computacional para el procesamiento.

Las aplicaciones de DL en imágenes de sensores remotos son diferentes de en comparación con las imágenes naturales, las imágenes obtenidas por teledetección suelen tener patrones más complicados y diversos, así como una información espacial, temporal y espectral más rica que se puede utilizar, por lo que se imponen mayores requisitos a los métodos de procesamiento de las imágenes obtenidas por teledetección (Yuan et al., 2020). En ese sentido se plantea el objetivo siguiente:

#### **A. Objetivo General**

- Desarrollar un modelo de Machine Learning para clasificación de imágenes satelitales en la amazonia peruana.

#### **B. Objetivos Específicos**

- Evaluar y Adquirir imágenes satelitales Landsat 8 o Sentinel para su preprocesamiento considerado el área de interés en estudio.
- Realizar un preprocesamiento de las bandas espectrales y normalización de las imágenes representados en los datos de reflectancia en cada pixel.
- Realizar el entrenamiento del modelo de clasificación, configurando en el entorno de Python en Jupyter Notebook, evaluar la precisión de los modelos en cada caso de clasificación.

### **III. BASES TEORICAS**

#### **A. Resoluciones de las Imágenes de Satélite**

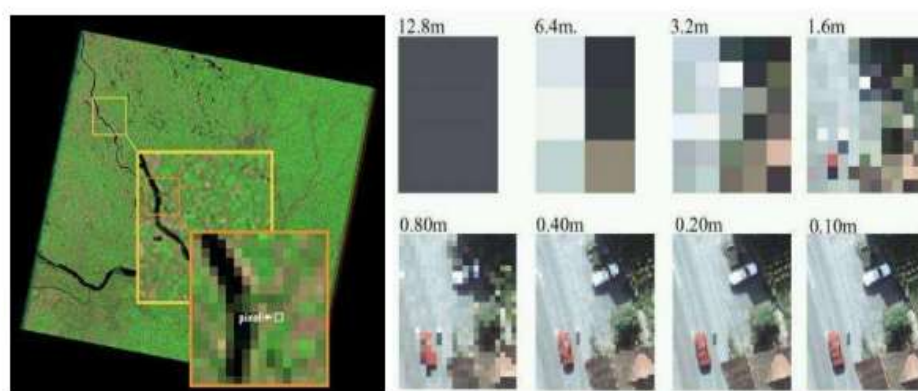
*Una imagen de satélite se caracteriza por las siguientes modalidades de resolución:*

##### **1) Resolución espacial:**

Esta determinada por el tamaño de pixel, el cual designa al objeto más pequeño que se puede distinguir en la imagen, medido en metros sobre el terreno, esto depende de la altura del sensor con respecto a la Tierra, el ángulo de visión, la velocidad de escaneado y las características ópticas del sensor (*Teledetección e Inteligencia Artificial Para Incrementar Los Beneficios de Tu Explotación Agrícola Siendo Más Respetuoso Con El Medio Ambiente - Hispaviación*, n.d.).

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 6 (térmica), Por otro lado según investigaciones revisadas se puede concluir que el nivel de detalle de las imágenes satelitales del sensor MSI es central para mejorar el análisis de mayor escala en las imágenes (Picone, 2017).

**Figura 1** Resolución espacial – Landsat 8



## 2) Resolución espectral:

La resolución espectral indica el número y ancho de las bandas en las que el sensor puede captar radiación electromagnética(Chuvieco, 1990). Por ejemplo, SPOT tiene una resolución espectral de 3, Landsat de 7. Los nuevos sensores, llamados también espectrómetros o hiperespectrales llegan a tener hasta 256 canales con un ancho de banda muy estrecho (unos pocos nm) para poder separar de forma precisa distintos objetos por su comportamiento espectral(Ingeoexpert, 2019).

## 3) Resolución radiométrica:

Resolución o sensibilidad radiométrica hace referencia al número de niveles digitales utilizados para expresar los datos recogidos por el sensor. En general, cuando mayor es el número de niveles mayor es el detalle con que se podrá expresar dicha información. Se la llama a veces también resolución dinámica, y se refiere a la cantidad de niveles de gris en que se divide la radiación recibida para ser almacenada y procesada posteriormente.

## 4) Resolución temporal:

Es la frecuencia de paso del satélite por un mismo punto de la superficie terrestre. Es decir, cada cuanto tiempo pasa el satélite por la misma zona de la Tierra. Este tipo de resolución depende básicamente de las características de la órbita.

El ciclo de repetición de los Landsat-1 al Landsat -3 era de 17 días. A partir del Landsat 4 en 1984 el ciclo de repetición se redujo a 15 días. SPOT permite un ciclo de repetición de entre 3 y 26 días(Cenice, 2017).

## **B. Firmas Espectrales – Reflectancia**

Las firmas espectrales muestran la variación de la radiación reflejada por los objetos en función de la longitud de onda. Este comportamiento físico puede ser tenido en cuenta en los análisis multiespectrales para reconocer elementos particulares en las imágenes aéreas. Los ojos de los seres humanos se pueden considerar como sensores remotos ya que detectan la luz reflejada por los objetos de nuestro entorno. Sin embargo, la visión humana sólo es capaz de percibir una pequeña parte del espectro electromagnético, el visible (García et al., 2012). La luz visible es sólo una de las muchas formas de radiación electromagnética que existen. Así, las ondas de radio, el calor, los rayos ultravioleta o los rayos X son otras formas comunes. En teledetección, lo normal es caracterizar a las ondas electromagnéticas por su longitud de onda en micrómetros ( $\mu\text{m}$ , 10 m) o nanómetros (nm, 10 m), es decir, por la posición que ocupan dentro del espectro electromagnético (García et al., 2012).

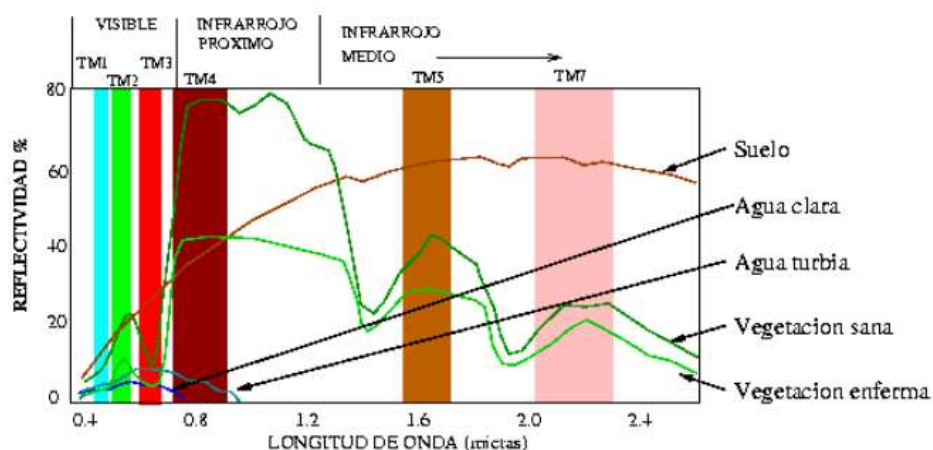
Diferentes tipos de superficies, como el agua, la tierra desnuda o la vegetación, reflejan la radiación de manera distinta en varios canales. La radiación reflejada en función de la longitud de onda se llama firma espectral de la superficie. (i.e. pixel) tiene un valor discreto en unidades de Digital Number (NASA, 2013). - firma espectral (reflectancia, longitud de onda) Interpreta: cuerpos de agua, vegetación, suelo, etc

Esto nos permitirá categorizar los diversos píxeles o firma espectral en clases asociadas de acuerdo con su mayor semejanza:

### ***Clases generales para el presente estudio con Aprendizaje supervisado:***

- Bosque
- Cultivos o Agricultura
- Área urbana
- Cuerpos de agua.

De cara a la identificación de objetos y procesos en la superficie terrestre, lo que nos interesa es la reflectividad de estos objetos respecto a las diferentes longitudes de onda. Cada tipo de material, suelo, vegetación, agua, etc. reflejará la radiación incidente de forma diferente lo que permitirá distinguirlo de los demás si medimos la radiación reflejada. Resulta así fácil por ejemplo distinguir entre suelo y vegetación, e incluso entre diferentes tipos de suelo o diferentes tipos de vegetación.

**Figura 2** *Respuestas espectrales*

### C. Bandas Espectrales Landsat 8

Landsat 8 es un satélite de observación terrestre estadounidense lanzado el 11 de febrero de 2013. Es el octavo y más reciente satélite del proyecto Landsat operado por la NASA y el Servicio Geológico de los Estados Unidos (USGS) desde 1972. 1

El satélite Landsat 8 transporta dos instrumentos OLI y TIRS, que corresponden a las siglas en inglés para Operational Land Imager (OLI) y Thermal Infrared Sensor (TIRS). El sensor OLI provee acceso a nueve bandas espectrales que cubren el espectro desde los 0.433  $\mu\text{m}$  a los 1.390  $\mu\text{m}$ , mientras que TIRS registra de 10.30  $\mu\text{m}$  a 12.50  $\mu\text{m}$ . Los dos sensores transportados por el satélite completan 11 bandas espectrales como se muestra:

**Tabla 1-** *Bandas espectrales-Landsat8.*

Bandas	Longitud de Onda	Resolución(m)
Banda 1 -Aerosol Costero	0.43- 0.45	30
Banda 2 -Azul	0.452 - 0.512	30
Banda 3 -Verde	0.533 - 0.590	30
Banda 4 -Rojo	0.636 - 0.673	30
Banda 5 – Infrarojo cercano NIR	0.851 - 0.879	30
Banda 6 Infrarojo onda corta 1	1.566 - 1.651	30
Banda 7 - Infrarojo onda corta 2	2.107 - 2.294	30
Banda 8 – Pancromatica	0.503 - 0.676	15
Banda 9 - Cirrus	1.363 - 1.384	30
Banda 10 – TIR 1	10.60 - 11.19	100
Banda 11 – TIR2	11.50 - 12.51	100



RGB. Esta combinación suele llamarse “color natural” pues involucra a las tres bandas visibles y se le asigna a cada una de ellas su verdadero color, resultando una combinación que se aproxima a los colores naturales de la escena. Podemos generarla utilizando la combinación de bandas 4,3,2 en Landsat 8 y 4,3,2 en Sentinel 2.

#### **D. Algoritmos de Machine Learning**

La clasificación de aprendizaje automático se ha convertido en un foco principal de la teledetección. Los algoritmos de aprendizaje automático generalmente son capaces de modelar complejas firmas de clase, pueden aceptar una variedad de datos de predicción de entrada, y no hacen suposiciones sobre la distribución de datos, no son paramétricas. Una amplia gama de estudios generalmente ha encontrado que estos métodos tienden a producir una mayor precisión en comparación con clasificadores paramétricos tradicionales, especialmente para datos complejos con un espacio de características de alta dimensión, es decir, muchas variables de predicción. Los enfoques de aprendizaje automático han llegado a ser ampliamente aceptados, como lo demuestra su uso en el mapeo de la cobertura terrestre. A pesar de la creciente aceptación de los clasificadores de aprendizaje automático, parece que los métodos paramétricos todavía se usan comúnmente en los artículos de aplicación y siguen siendo uno de los principales estándares para los experimentos de clasificación de evaluación comparativa. Por ejemplo, encontraron que el clasificador paramétrico de máxima verosimilitud (ML) era el método más utilizado, empleado en el 32% de los artículos, aunque los métodos de aprendizaje automático se encontraron rutinariamente con precisiones notablemente más altas que ML.

Hay una amplia gama de artículos que han investigado el aprendizaje automático en teledetección, esto incluye una serie de excelentes artículos de revisión para métodos específicos, como la revisión de Mountrakis de máquinas de vectores de soporte (SVM), y el artículo de sobre Bosques Aleatorios (RF), así como artículos que revisan el proceso de clasificación y sus complejidades.

#### **E. Aprendizaje NO Supervisado**

Para una imagen satelital dada, si falta información de referencia sobre el terreno, se pueden aplicar técnicas no supervisadas para clasificar la imagen de satélite en distintas regiones de cobertura terrestre. En técnicas no supervisadas, los conjuntos de datos sin conocimiento previo de las etiquetas se subdividen en grupos o clústeres basados en



algunos atributos, y el analista los compara con las clases de información. El conocimiento extraído del conjunto de datos tiene la forma de centros de clúster óptimos. El objetivo principal es garantizar que las distancias entre las distancias intra-clusters sean mínimas y que las distancias entre los inter-clusters sean máximas.

### 1) Algoritmo Kmeans

Un método no paramétrico popular es k-means. La función objetivo de k-means es la suma del error al cuadrado, que debe minimizarse. De hecho, el algoritmo k-means es una técnica paramétrica en el sentido de que el número de clusters se conoce a priori (Senthilnath et al., 2015).

$$J(k) = \sum_{k=1}^K \sum_{x_i \in c_k} \|x_i - c_k\|^2$$

Donde, la distancia métrica en el espacio euclidiano,  $C_k$  es el centro en la mayoría de las aplicaciones, se minimiza directamente la función objetivo anterior para obtener los centros del cluster en la agrupación k-means .

## IV. METODOLOGIA DE CLASIFICACION CON MACHINE LEARNING

### A. Área de Estudio de la clasificación

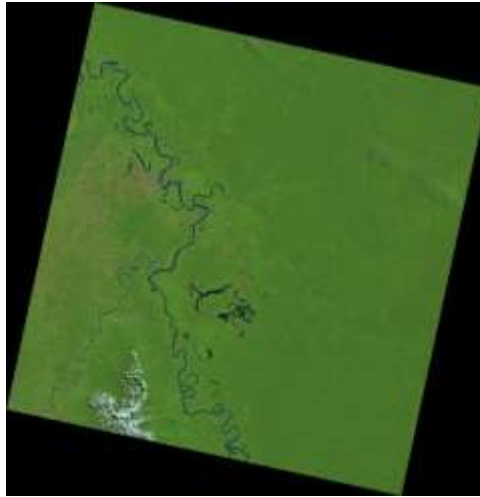
La región sobre la cual se desea hacer un análisis de los algoritmos anteriormente descritos corresponde a la zona de la selva peruana alrededores de la provincia de coronel Portillo, Departamento de Ucayali– Perú (Figura 3). Esta es un segmento de una escena LANDSAT 7 ETM+ composición de 11 bandas ) con una extensión de 225 Km<sup>2</sup> (500\*500 pixeles).

**Figura 3.** Área de Estudio – Pucallpa -.Perú



Se escogió esta zona ya que se pretende realizar una clasificación inicial con un algoritmo no supervisado que permita discriminar zonas urbanas, cultivos, cuerpos de agua y bosque. El cual se puede observar en las imágenes adquiridas. Además, posterior estudio en el cual se puede abordar problemáticas de deforestación, minería ilegal, contaminación ambiental.

**Figura 4.** *Área específica – Pucallpa -.Perú*

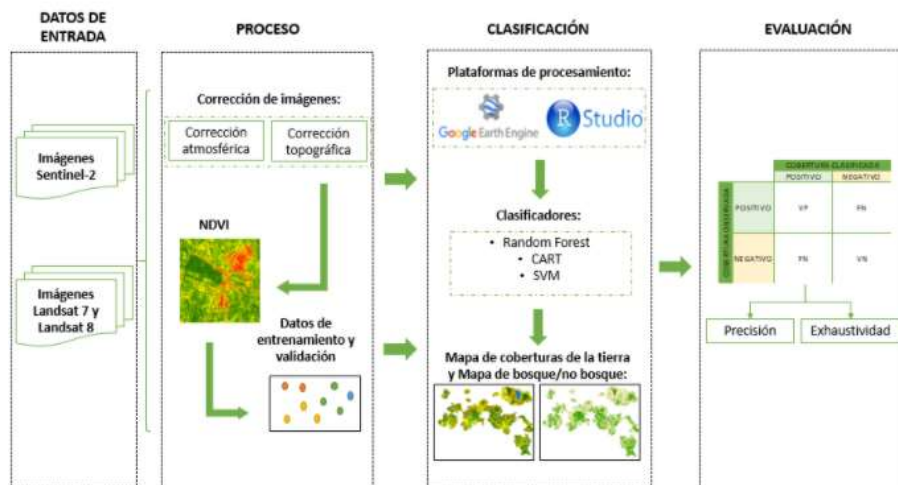


*Fuente: Landsat LC08\_L1TP\_006066\_20170722\_20170728\_01\_T1*

Para la descarga de la imagen a procesar se utilizó la plataforma Google Earth Engineering, en la que se realizó una búsqueda con los criterios de coordenadas lat. -8.4072 y Lon. -74.5093. Sensor Landsat 8 OLI/TIRS C! Level-1, porcentaje de nubosidad menor de 20%, además se dispone de su metadata, en el cual se tiene información sobre el tamaño en superficie, latitud y longitud máxima y mínima, radiancia máxima por cada banda, reflectancia máxima y mínima por banda, sistema de georreferenciación WGS84, y demás datos e las 11 bandas.

#### **B. Proceso de Clasificación con Machine Learning**

La metodología aplicada en el presente estudio se observa en la Fig 4. El cual consistió en utilizar como datos de entrada 3 colecciones imágenes satelitales Landsat 8 procesadas sobre la plataforma de Google Earth Engine (GEE) y Jupyter Notebook en donde se aplica la corrección atmosférica y topográfica para la eliminación de nubes y sombras, además de calcular el NDVI para cada imagen, clasificar las coberturas, y posteriormente evaluar el mapa resultante, para el presente estudio no fue posible y se analizo las imágenes seleccionadas e el cual las correcciones mencionadas no interfieran en la clasificación no supervisada.

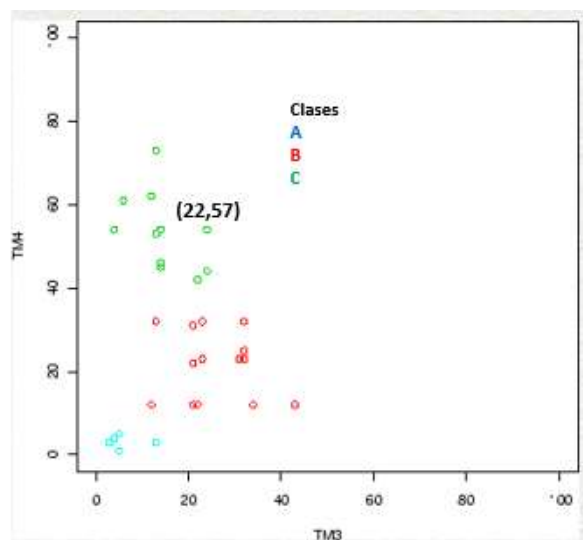
**Figura 4 . Proceso de clasificación -Etapas**

Fuente: <https://ingeoexpert.com>

Google Earth Engine es una herramienta de procesamiento de datos desarrollada por Google, y en la actualidad es una de las más robustas e importantes, gracias a la velocidad en los procesos, la cantidad de datos disponibles y a los mecanismos de análisis de datos espaciales.

### C. Representación de datos de Clasificación

El modelo inicial es la que se representa las dimensiones como los datos de las firmas espectrales en cada una de las bandas (8), el cual serán representados en arreglos o matrices con datos para cada banda entonces las variables predictoras son las bandas, la variable a predecir las clases están definidas son 4 según la imagen procesada los mismos que no son etiquetadas.

**Figura 5: Distribución de los pixels ((b1,b2....bn)**

#### D. Entorno de Desarrollo del Modelo de Machine Learning

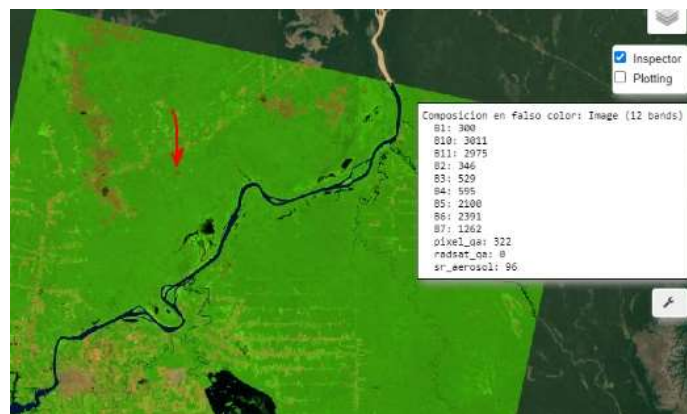
En cada etapa del proceso de clasificación, se utilizó entorno de desarrollo y de programación, en primer lugar, en la etapa de adquisición la plataforma Earth Engine que mediante una suscripción permite acceder a código Python (Varo et al., 2014) para el procesamiento y adquisición de meta data de la imagen satelital. Posteriormente para la visualización y selección de bandas se utilizó el entorno de Anaconda y Jupyter Notebook, en el cual se realizó la normalización de datos y ploteo de las diferentes bandas (11) y combinación para el análisis de la imagen satelital previo a la clasificación con algoritmos de machine learning.

También se utilizó software especializado de sistemas de Información Geográfica (GIS) como QGIS y GRASSGIS para la etapa de análisis de datos espectrales, para generar un stack de bandas y la visualización de resultados de clasificación de los algoritmos aplicados.

#### E. Visualización y Preprocesamiento

Inicialmente se analiza las bandas (11) de la imagen, y podemos realizar la visualización RGB con los canales rojo, verde y azul. Y con el inspector podemos visualizar los datos de las firmas espectrales en cualquier pixel de la imagen como se muestra en la figura 6.

**Figura 6.** Área específica – Pucallpa -. Perú – GRASS GIS



También es posible realizar el cálculo de los índices espectrales como NDVI Índice de vegetación, esto con el fin de definir y analizar las clases que podemos inferir en la clasificación con algoritmos de machine learning, se realizó un corte del área de interés con las librerías utilizadas, considerando las clases que intervienen: Bosques, Urbano, Agrícola y cuerpos de agua., además se normalizo los datos mediante el proceso en Python para efectos de mejorar los sobreajustes que pueda generar el modelo con los datos atípicos.

## V. RESULTADOS

### A. Aprendizaje No supervisado

Para una imagen satelital dada, si falta información de referencia sobre el terreno, se pueden aplicar técnicas no supervisadas para clasificar la imagen de satélite en distintas regiones de cobertura terrestre.

#### Ploteo de la Imagen en RGB.

Se procedió a la visualización del área de estudio en Jupyter notebook, para lo cual se utilizó las librerías pandas, gdal, y numpy básicamente y para el ploteo matplotlib, luego se se creo un raster en un arreglo donde están los datos de las firmas espectrales.

Se normaliza a valores entre 0 y 1, además se aplica para las 6 bandas seleccionadas y el RGB se realiza un stack entre las bandas 4, 5 y 6 para su ploteo como vemos en la Figura7.

**Figura 7.** Ploteo RGB – bandas 4,5,6



#### Clasificación con Algoritmo Kmeans

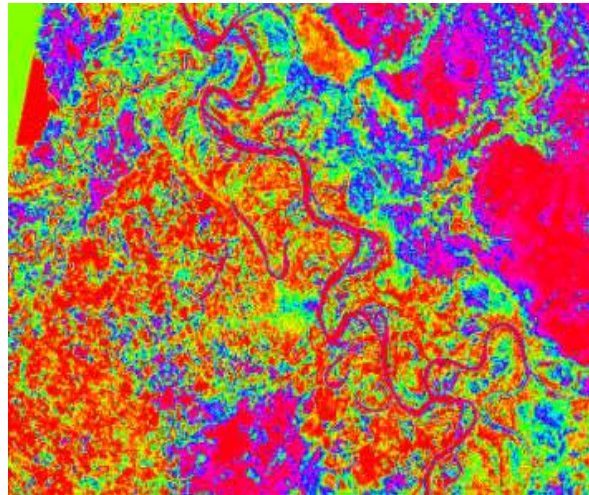
*Se realizo la clasificación con la librería rasterio y sklearn.cluster Kmeans, para lo cual se paso el método fit siguiente:*

```
kmeans = KMeans(n_clusters=4) # max_iter=300 (por defecto)
kmeans.fit(mat)
print(kmeans.predict(datos))
```

Previo analisis se realizo pruebas para K=3,4,5, analizando resultados en base a las clases definidas en la etapa previa, se define k=4 la que mejor clasificacion dispone en base a la revision bibliografica de tal manera que el algortimo no se sobreajuste.



**Figura 8.** *Clasificación No Supervisada con K=4 - Clases*



Es necesario la interpretación experticia para valorar la clasificación ya que no se dispone datos de entrenamiento y validación para calcular la precisión del modelo Kmeans.

El algoritmo de K-means para la clasificación de imágenes destaca por la sencillez y velocidad de su algoritmo, sin embargo, presenta una serie de limitaciones entre las cuales requiere que se indique de antemano el número de clústeres que se van a crear. Una de las dificultades de este método es que no es tan autoorganizado ya que se deben definir parámetros basado en la experiencia como numero de agrupaciones deseadas “K”(Gutiérrez et al., 2005). Se han desarrollado varias estrategias para ayudar a identificar potenciales valores óptimos de K como el método del codo de manera orientativa y sujeto a prueba y error.

### **B. Aprendizaje Supervisado**

En técnicas supervisadas, se utilizan muestras representativas de diferentes tipos de interés de cobertura superficial (denominadas muestras de entrenamiento). La información numérica en todas las bandas espectrales para los píxeles que comprenden estas muestras se usa para "entrenar" a la computadora para reconocer muestras espectralmente similares para cada clase. Cada píxel de la imagen se compara con estas firmas y se etiqueta como la clase a la que más se "parece" en términos de distancias(Bundle & Rosebrock, 2017) Entre las diversas técnicas supervisadas, las redes neuronales artificiales(Negnevitsky, 2005) se han utilizado ampliamente para la clasificación de imágenes y, en algunos casos, han superado a los clasificadores estadísticos. Entre los clasificadores supervisados tenemos a support vector

machine(Support Vector Machines)(Haykin, 2009), decision tree, random forest, bayes, neural networks, entre otros.

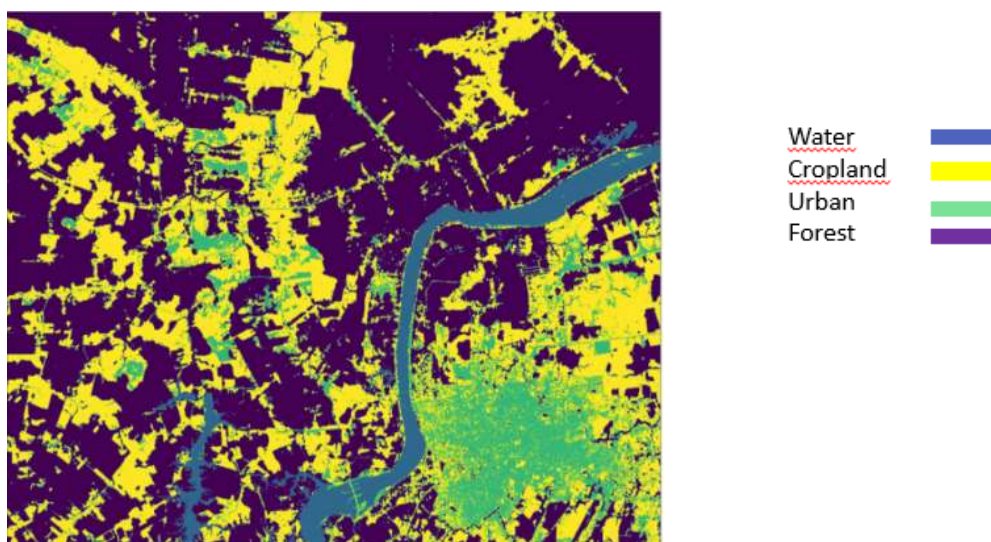
Para el estudio se utilizó un stack de imágenes satelitales ya procesadas, en el cual se tienen etiquetadas las clases.

- Cuerpos de Agua,-Cultivos,Urbano,Bosques

Los datos utilizados para el entrenamiento y pruebas del modelo, consiste en una imagen vectorial shape que esta etiquetada en 4 clases, con un total de 110 puntos evaluados y procesados en QGIS, antes de su implementación con Python

- 1) SVM - Support Vector Machine
- 2) SVM realiza la clasificación mediante la construcción de un hiperplano N-dimensional El hiperplano deseado es también la bisectriz de la línea entre los puntos más cercanos

**Figura 9.** Clasificación Aprendizaje Supervisado



Según los resultados en las pruebas, el algoritmo tiene una precisión de 0.909, es un nivel aceptable, pero es necesario más datos de entrenamiento para su generalización en el mapa.

#### 1) **Árbol de decisión – Decision Tree**

Se procedió a utilizar las librerías de sklearn, y se invocó al método fit instanciado con las variables de entrenamiento.

```
from sklearn.tree import DecisionTreeClassifier as DT
instancia_dt.fit(Xtrain, ytrain)
```



Según los resultados en las pruebas, el algoritmo tiene una precisión de **0.864**, es un nivel aceptable, pero es necesario más datos de entrenamiento para su generalización en el mapa.

## 2) Naive bayes– Clasificador de Bayes

De igual manera se procedió a utilizar las librerías de sklearn, y se invocó al método fit instanciado con las variables de entrenamiento.

```
from sklearn.naive_bayes import GaussianNB as GNB
```

```
instancia_nb = GNB()  
instancia_nb.fit(Xtrain, ytrain)
```

Según los resultados en las pruebas, el algoritmo tiene una precisión de **0.909**, similar al algoritmo SVM.

Para todos los algoritmos el cálculo del nivel de precisión se determinó a través una matriz de confusión de los datos predichos con los datos de prueba (22 datos).

## VI. CONCLUSIONES

La información espectral contenida en los píxeles no es suficiente en la mayoría de los casos. La combinación con otras fuentes de datos puede mejorar las clasificaciones

Las imágenes satelitales requieren un preprocesamiento de correcciones radiométricas y/ atmosféricas previo a la clasificación con algoritmos de Machine Learning.

La Clasificación No Supervisada Kmeans, no garantiza una clasificación Robusta y tiene limitaciones en la aplicación de Imágenes satelitales, esto debido a la asignación inicial aleatoria de centroides, outliers y los recursos utilizados en el entrenamiento del modelo.

En la clasificación Supervisada, el algoritmo SVM y Naive Bayes presentaron mejor precisión de 0.909 en comparación con el algoritmo de árbol de decisión 0.864, considerando los pocos datos e entrenamiento y validación del modelo de clasificación.

En base a la revisión bibliográfica, el clasificador máxima probabilidad y las redes neuronales convolucionales y SVM son los clasificadores más procesamiento de Imágenes Satelitales y la teledetección.

## REFERENCIAS BIBLIOGRAFICAS

Bundle, S., & Rosebrock, A. (2017). *Deep Learning for Computer Vision* (Primera). PYIMAGESEARCH. <https://www.pyimagesearch.com/deep-learning-computer-vision-python-book/>

- Cenice. (2017). *Resolución de las imágenes satélite*.  
[http://concurso.cnice.mec.es/cnice2006/material121/unidad1/i\\_resolucion.htm](http://concurso.cnice.mec.es/cnice2006/material121/unidad1/i_resolucion.htm)
- Chuvieco, E. (1990). Fundamentos de teledetección espacial. In Ediciones RIALP (Ed.),  
(*Fundamentals of remote sensing from space* (2da ed.). Segunda Edición.
- García, M. L., Brondo, J. A. É., & Pérez, M. A. (2012). *Satélites para la Gestión del Territorio* (1ra edición, pp. 1–66). 2012.  
<http://onlinelibrary.wiley.com/doi/10.1002/cbdv.200490137/abstract>
- Gutiérrez, M. A., William, J., & Bedoya, B. (2005). *Clasificación NO supervisada de coberturas vegetales sobre imágenes satelitales de sensores remotos: Landsat - ETM +*.
- Haykin, S. (2009). *Neural Networks and Learning Machines* (Pearson (Ed.); Third Edit).
- Ingeoexpert. (2019). *Teledetección y procesamiento de Imágenes satelitales*.
- Leiva, I. G., Díaz, P., Vicente, J., & Muñoz, R. (2019). *Técnicas y usos en la clasificación automática de imágenes*. 1–14.
- Negnevitsky, M. (2005). *Artificial Intelligence a guide to intelligent systems* (Pearson (Ed.); Edition, S).
- Picone, N. (2017). *Comparación De Imágenes Satelitales Sentinel 2 Y Landsat 8. October*, 1–10.
- Ramírez, M., Martínez, L., Montilla, M., Sarmiento, O., Lasso, J., & Díaz, S. (2020). Obtención de coberturas del suelo agropecuarias en imágenes satelitales Sentinel-2 con la inyección de imágenes de dron usando Random Forest en Google Earth Engine. *Revista de Teledetección*, 56, 49. <https://doi.org/10.4995/raet.2020.14102>
- Teledetección e inteligencia artificial para incrementar los beneficios de tu explotación agrícola siendo más respetuoso con el medio ambiente - Hispaviación*. (n.d.). Retrieved March 9, 2022, from <https://www.hispaviacion.es/teledeteccion-e-inteligencia-artificial-para-incrementar-los-beneficios-de-tu-explotacion-agricola-siendo-mas-respetuoso-con-el-medio-ambiente/>
- Utiltech, S. (2018). *Teledetección e inteligencia artificial para incrementar los beneficios de tu explotación agrícola siendo más respetuoso con el medio ambiente*. Hispaaviacion. <https://www.hispaviacion.es/teledeteccion-e-inteligencia-artificial-para-incrementar-los-beneficios-de-tu-explotacion-agricola-siendo-mas-respetuoso-con-el-medio-ambiente/>

- Varo, A. M., Luengo, I. G., & Sevilla, P. G. (2014). *Introducción a la programación con Python 3* (Universidad Jaume I (Ed.); Primera).
- Yuan, Q., Shen, H., Li, T., Li, Z., Li, S., Jiang, Y., Xu, H., Tan, W., Yang, Q., Wang, J., Gao, J., & Zhang, L. (2020). Deep learning in environmental remote sensing: Achievements and challenges. *Remote Sensing of Environment*, 241(March 2019), 111716. <https://doi.org/10.1016/j.rse.2020.111716>