

OBSERVATORIO REGIONAL AMAZÓNICO
MEMORIA TECNICA

Caso de Estudio 2:
Monitoreo de bosques y algoritmos de inteligencia artificial



2022



CONTENIDO

1. Caso de Estudio 2: Monitoreo de bosques y algoritmos de inteligencia artificial	3
1.1. Datos.....	3
1.2. Metodología.....	6
1.2.1.Pre-procesamiento de datos	6
1.2.2 Clasificación de imágenes.....	7
2 Conclusiones y recomendaciones.....	14
3 Bibliografía.....	14

INDICE DE FIGURAS:

Figura 1 Grillas o Tiles de Sentinel-2 en cuenca amazónica	4
Figura 2 Distribución espectral de bandas.	5
Figura 3 Resoluciones espectrales por resoluciones espaciales	5
Figura 4 Mosaicos por periodo	7
Figura 5 Mosaicos por periodo	8
Figura 6 Linear support vector machine example	11
Figura 7 Clasificación SVM	11
Figura 8 Clasificación con CART.....	12
Figura 9 Clasificación con Random Forest.....	13

1. Caso de Estudio 2: Monitoreo de bosques y algoritmos de inteligencia artificial

En este caso de estudio presenta una metodología que permite procesar y generar información acerca de la distribución espacial de bosques y otras coberturas de la tierra, definidas en el marco de las IPCC (Intergovernmental Panel on Climate Change). La metodología planteada demuestra la aplicación de la inteligencia Artificial a través del uso de algoritmos de aprendizaje automático (Random Forest), para obtener la distribución espacial de bosques y otras coberturas. Por otro lado la metodología considera como fuente de datos, imágenes satelitales del programa Sentinel-2, que constituyen en las imágenes de libre acceso con ventajosa resolución, espacial, temporal, espectral y radiométrica.

La metodología consiste en el desarrollo de 5 tareas principales, estas son; el Preprocesamiento de datos satelitales, la Recopilación de datos que sirvan para entrenar los algoritmos de inteligencia Artificial, definir un algoritmo clasificador, entrenar al algoritmo clasificador a partir de los datos recopilados, y la ejecución de la clasificación. Toda la metodología fue desarrollada en la plataforma de Google Earth Engine, permitiendo de esta manera dar el seguimiento del planteamiento metodológico.

1.1. Datos

Como fuente de datos para el presente caso de estudio se utilizaron Imágenes del Programa Sentinel-2, estas imágenes tienen una cobertura global y se pueden obtener cada 5 días, asimismo a la fecha resultan las imágenes satelitales de libre acceso con mayor resolución, espacial, espectral, temporal y radiométrica.

Sentinel-2 comprende una constelación de dos satélites en órbita polar colocados en la misma órbita heliosíncrona, en fase de 180° entre sí. Su objetivo es monitorear la variabilidad en las condiciones de la superficie terrestre y su amplio ancho de franja (290 km) y tiempo de revisita alto (10 días en el ecuador con un satélite y 5 días con 2 satélites en condiciones sin nubes, lo que da como resultado 2-3 días en latitudes medias) apoyará el seguimiento de los cambios en la superficie de la Tierra [2].

Las imágenes de Sentinel-2 están distribuidas por Grillas o Tiles que son de tamaño fijo, junto con una sola órbita. Un gránulo es la mínima partición indivisible de un producto (que contiene todas las bandas espectrales posibles). Estas grillas tienen una dimensión de 100 x 100 km² en un sistema de proyección UTM, con sistema de referencia WGS84.

A continuación, en la siguiente imagen se puede observar en polígonos blancos la distribución de las Grillas o *Tiles*, a lo largo de la cuenca amazónica (polígono en azul) y el área biogeográfica amazónica (polígono verde). Un total de 727 grillas cubren a la cuenca amazónica.

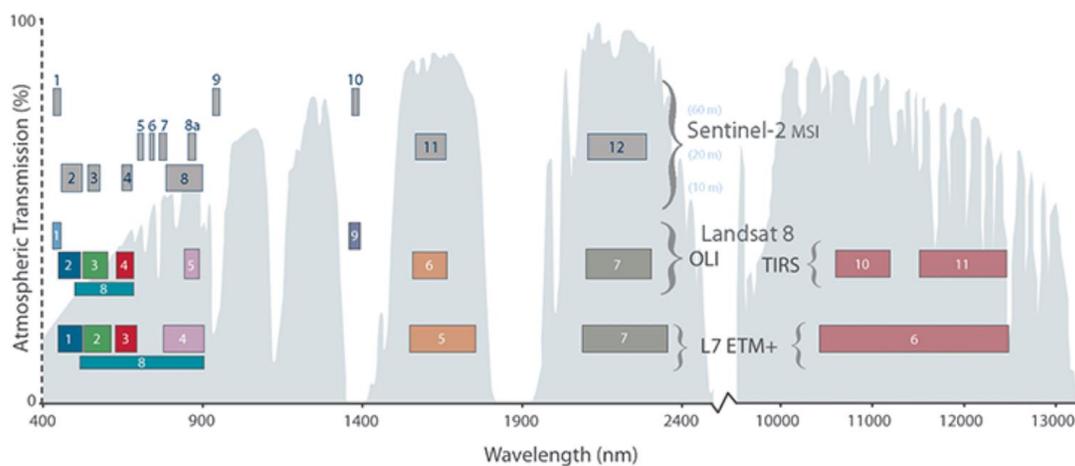
Figura 1 Grillas o Tiles de Sentinel-2 en cuenca amazónica



Fuente: Elaboración propia

Las imágenes de Sentinel-2 contienen un total de 13 bandas espectrales: cuatro bandas a 10 m, seis bandas a 20 m y tres bandas a 60 m de resolución espacial. Tal como se puede observar en la imagen a continuación las bandas espectrales se encuentran en longitudes de onda del visible, infrarrojo cercano y e infrarrojo de onda, en ese sentido deja la posibilidad de realizar comparaciones con imágenes Landsat o Spot.

Figura 2 Distribución espectral de bandas.



Fuente: Landsat Science 2022

Asimismo, resulta importante tener en cuenta que las diferentes bandas de las imágenes Sentinel-2, cuentan con diferentes resoluciones espaciales. En tal sentido se describe las diferentes longitudes de onda de las bandas por las diferentes resoluciones espaciales.

Figura 3 Resoluciones espectrales por resoluciones espaciales

10 metros de resolución espacial		20 metros de resolución espacial		60 metros de resolución espacial	
Band number	Central wavelength (nm)	Band number	Central wavelength (nm)	Band number	Central wavelength (nm)
2	490	5	705	1	443
3	560	6	740	9	945
4	665	7	783	10	1375
8	842	8b	865		
		11	1610		
		12	2190		

Fuente: adecuado de Sentinel-2, User Handbook

1.2. Metodología

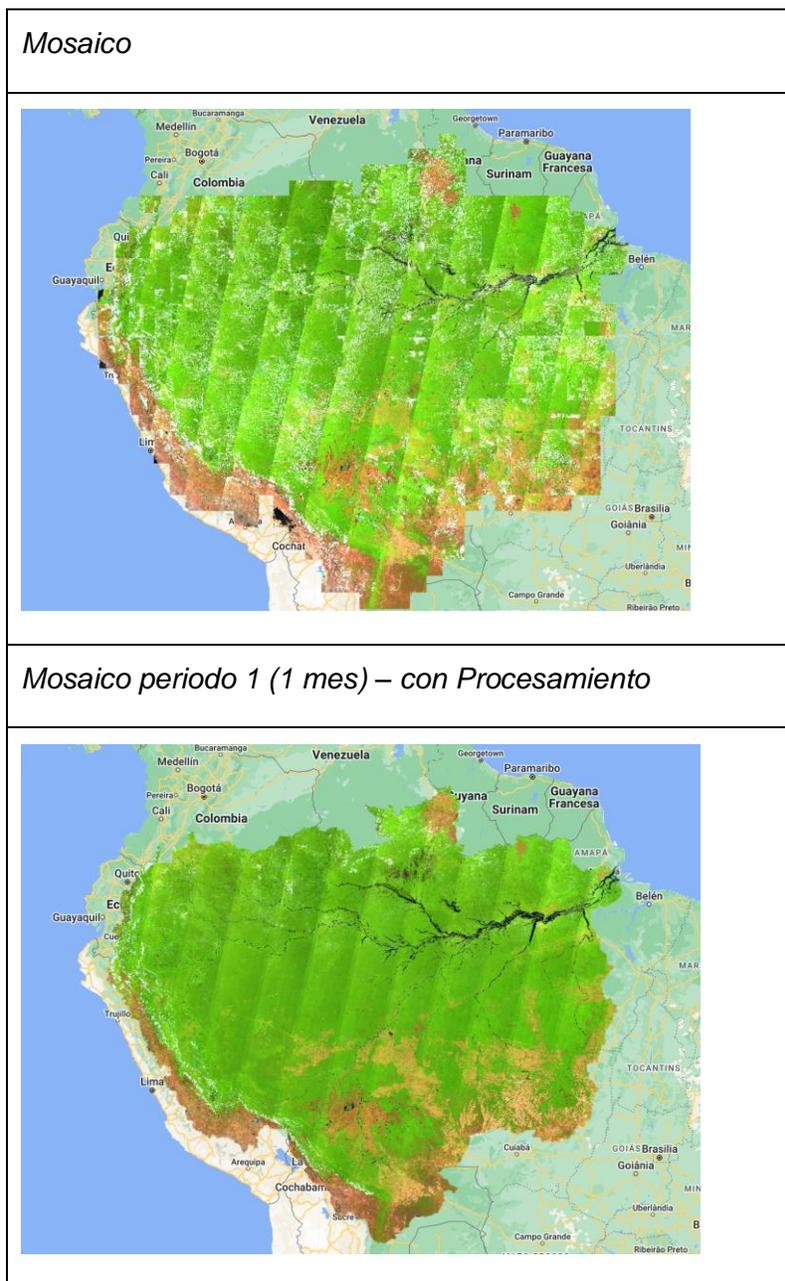
En este caso de estudio presenta una metodología que permite procesar y generar información acerca de la distribución espacial de bosques y otras coberturas de la tierra, definidas en el marco de las IPCC (Intergovernmental Panel on Climate Change). La metodología planteada demuestra la aplicación de la inteligencia Artificial a través del uso de algoritmos de aprendizaje automático (Random Forest, CART y SVM), para obtener la distribución espacial de bosques y otras coberturas. Por otro lado la metodología considera como fuente de datos, imágenes satelitales del programa Sentinel-2, que constituyen en las imágenes de libre acceso con ventajosa resolución, espacial, temporal, espectral y radiométrica

1.2.1. Pre-procesamiento de datos

Inicialmente se ha realizado las operaciones para contar con imágenes de toda el área de estudio (cuenca amazónica) con un alto nivel de corrección geométrica y con un nivel de corrección radiométrica BOA (*Bottom of Atmosphere*).

Posteriormente se ha trabajado en la construcción de mosaicos libres de perturbaciones por nubes. Los mosaicos fueron construidos para dos periodos. El primer periodo de tiempo, corresponde al periodo anterior al periodo de análisis, debido a la disponibilidad de imágenes satelitales, se evidencio que un mes, resulta ideal para este periodo. El segundo periodo corresponde al periodo de análisis de alertas de deforestación o quemas, este periodo puede ser por días, semanas o meses, de acuerdo al periodo de interés. Debido a la disponibilidad de datos entre 1 y 2 semanas resultan ideales para este periodo.

Figura 4 Mosaicos por periodo



Fuente: elaboración propia

1.2.2 Clasificación de imágenes

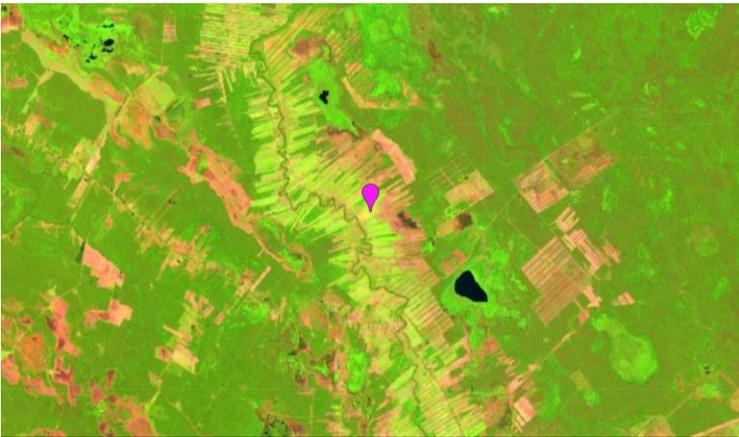
Captura de muestras

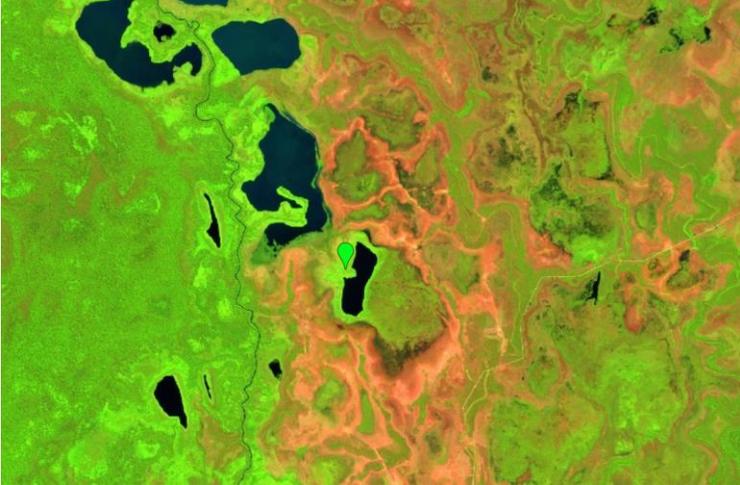
Inicialmente se realizó la definición de las clases del mapa que resultan las categorías para representar la cobertura de la tierra según las pautas del IPCC. Las clases corresponden

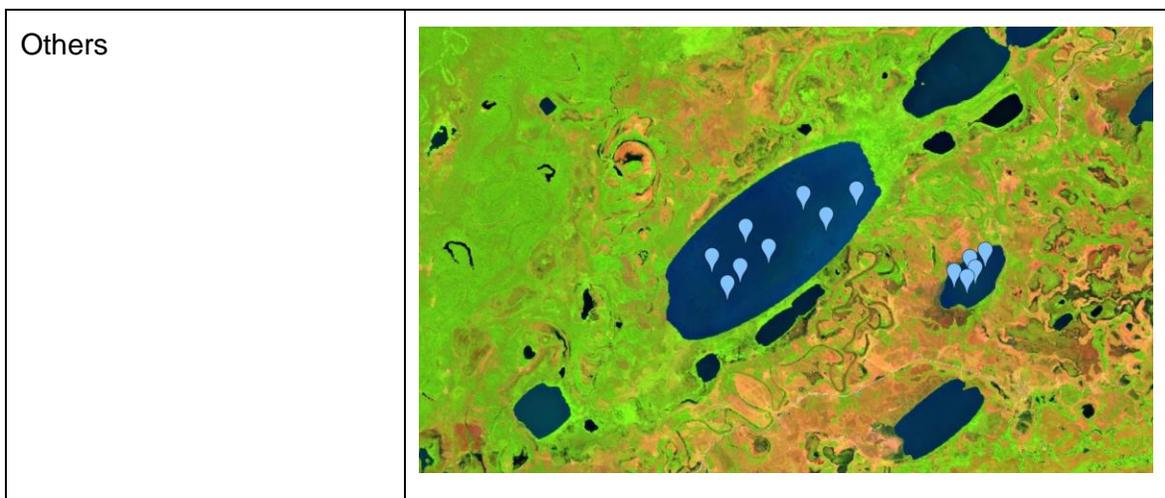
a; *Forest, Cropland, Grassland, Wetlands, Settlements y Others* (Otras coberturas). Para el caso de estudio se tomaron alrededor de 60 muestras, 17 muestras de Bosques (Forest), 2 muestras de cultivos (Cropland), 18 muestras de Herbazales (Grassland), 6 muestras de vegetación inundada (Wetlands), 3 muestras de asentamientos humanos (Settlements) y 21 muestras de otras coberturas como se cuerpos de agua (Others). Una descripción más amplia de cada una de estas clases puede encontrarse en el documento “*Good Practice Guidance for Land Use, Land-Use Change and Forestry - IPCC*”.

Las muestras fueron capturadas a partir de la interpretación de los mosaicos de las imágenes Sentinel-2 elaborados anteriormente, a continuación, se presenta la muestra capturada para cada clase, en una combinación de R:11, G:8, B:4, de las bandas Sentinel-2.

Figura 5 Mosaicos por periodo

Forest	
Cropland	

<p>Grassland</p>	
<p>Wetlands</p>	
<p>Settlements</p>	



Fuente: elaboración propia

Algoritmos de aprendizaje automático (Inteligencia Artificial)

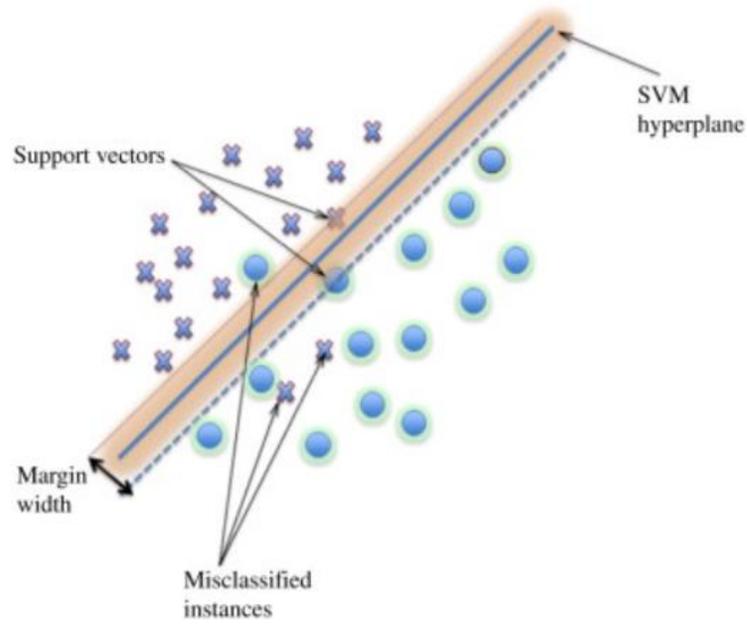
En las últimas décadas, el aprendizaje automático (Machine Learning) ha recibido una gran atención debido a su robustez en la clasificación de imágenes de teledetección [3]. Para el presente caso de estudio se ejecutarán tres algoritmos, Random Forest, CART y el SVM, el uso de cada uno de estos algoritmos demanda altas capacidades computacionales, en tal sentido se aprovecharán los beneficios de trabajar en sobre una plataforma de Big Data como Google Earth Engine.

Las muestras capturadas servirán como valores de entrada para el entrenamiento de cada uno de estos algoritmos, en tal sentido a continuación se describen algunas características de los algoritmos empleados.

SVM Support Vector Machine

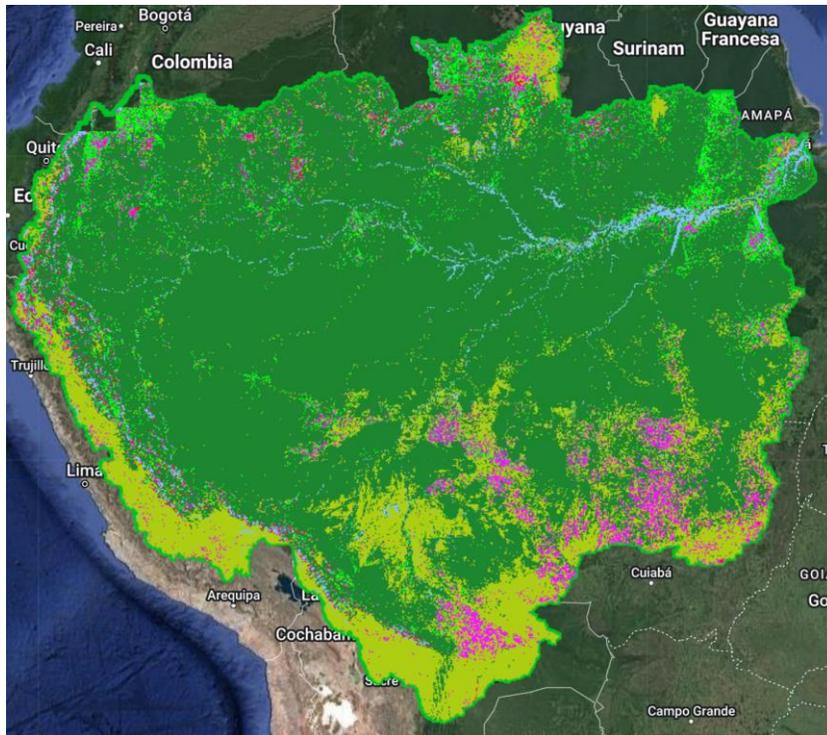
SVM es un algoritmo de aprendizaje automático o un algoritmo de clasificación en términos de teledetección, este algoritmo es uno de los más sólidos en aprendizaje automático y uno de los más populares en la investigación reciente de detección remota. En [4] se considera que, en su forma simple, SVM es un clasificador binario lineal que asigna una muestra de prueba dada a una categoría o clase de una de las dos etiquetas probables. Aunque este algoritmo es uno de los más preferidos, el uso de este no es una tarea sencilla, en este sentido en [5] se propone una alternativa sencilla para ajustar los parámetros de este algoritmo, y obtener resultados razonables. Para este caso de estudio se utilizaron los parámetros de los valores definidos por defecto en GEE, para la ejecución de SVM.

Figura 6 Linear support vector machine example



Fuente: [5]

Figura 7 Clasificación SVM

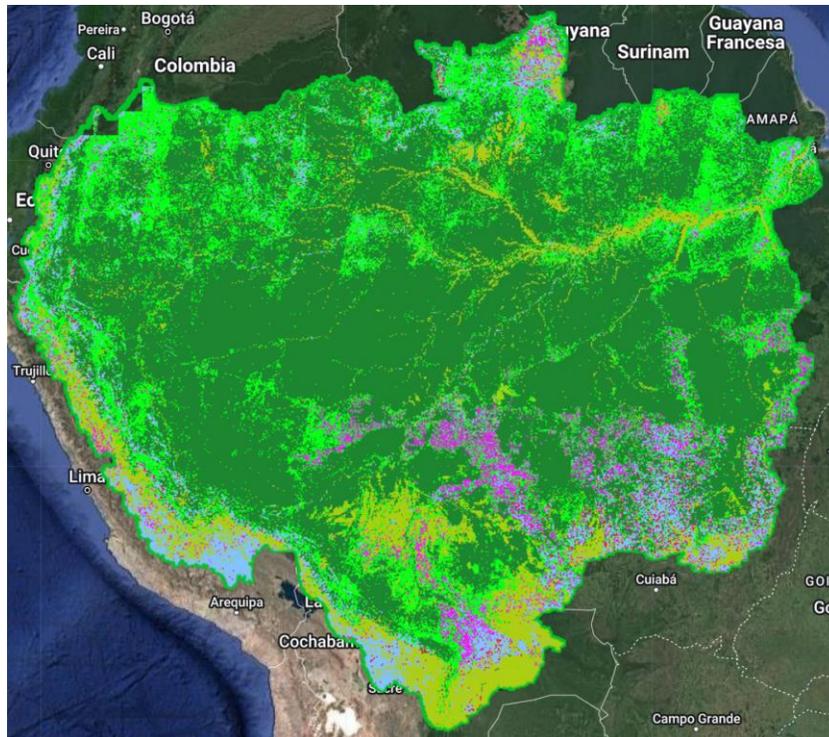


Fuente: Elaboración propia

CART classification

Al igual que SVM, CART es un algoritmo de aprendizaje automático o un algoritmo de clasificación en términos de teledetección, algunos autores señalan que CART es un término que se refiere a clasificaciones y regresiones de árboles. La estructura de este algoritmo es un árbol binario que crece de acuerdo a los requerimientos de clasificación, CART fue introducido por [5], en 1984, y el uso con parámetro por defecto se encuentra en GEE, para realizar la clasificación con CART se asumieron estos valores por defecto de GEE.

Figura 8 Clasificación con CART



Fuente: Elaboración propia

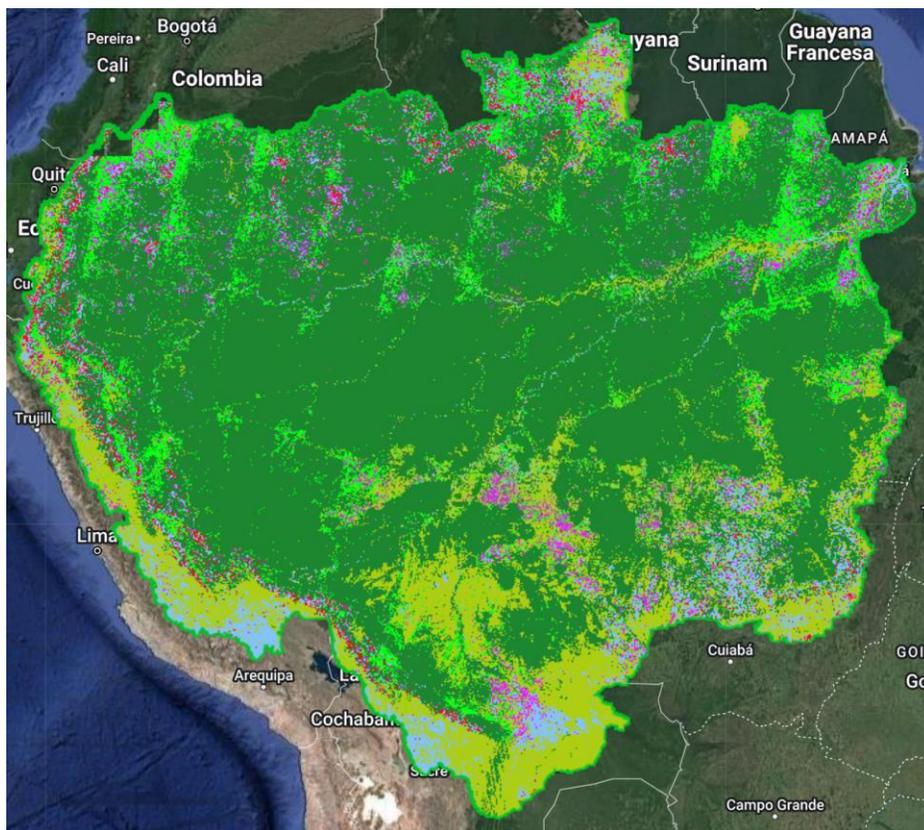
Random Forest clasificación

Random forest resulta el algoritmo preferido de las investigaciones recientes de teledetección, en los últimos años este algoritmo ha demostrado su capacidad para clasificar diferentes objetos o fenómenos a diferentes escalas (mapeo global, mapeo local, etc.), como se menciona en [6] y otras investigaciones, el trabajo con este algoritmo ha ido acompañado de una mejora computacional. Además de algunas de las investigaciones, existe una gran cantidad de investigaciones destacadas y recientes que han obtenido

buenos resultados utilizando este algoritmo, entre las cuales podemos mencionar [7], [8], [9], [10], finalmente podemos recomendar las publicaciones [11] y [12], en las que se detalla más extensamente la aplicación de este algoritmo en teledetección

Con base en la extensa bibliografía reciente que avala la utilidad de Random forest, realizar clasificaciones con este algoritmo fue uno de los principales objetivos de este caso de estudio, se ha realizado un ajuste a estos parámetros de Random Forest en cuanto al número de árboles de decisión.

Figura 9 Clasificación con Random Forest



Fuente: Elaboración propia

Investigaciones mencionadas anteriormente sugieren que a mayor número de árboles definidos para RF es posible obtener mejores resultados, sin embargo se debe tener en cuenta que este hecho no se cumple en todos los casos y sobre todo también se debe tener en cuenta que aumentar el número de árboles RF, multiplica exponencialmente las necesidades de tener altas capacidades computacionales, pero debido al trabajo sobre la

plataforma de Google Earth Engine, se cuenta con las capacidades computacionales requeridas

2 Conclusiones y recomendaciones

- Se presenta ejemplos de aplicación de algoritmos de aprendizaje automático (Inteligencia artificial), para el monitoreo de cobertura de bosques y los cambios de cobertura que pueda tener. Estos algoritmos demandan altas capacidades computacionales que fueron cubiertas por la plataforma Google Earth Engine.
- Todo el caso de estudio se puede ejecutar a través de un código JavaScript, sobre la plataforma de Google Earth Engine, de tal forma se puede evaluar de forma práctica todo lo descrito. Asimismo, se puede redefinir fechas y espacios geográficos de análisis, como también aumentar las muestras necesarias que permitan obtener mejores resultados.
- A partir del presente caso de estudio y la ejecución del script, se puede realizar una comparación de los resultados de los algoritmos (SVM, CART y Random Forest), basados en una interpretación visual. Sin lugar a duda para tener una evaluación cuantitativa del rendimiento de cada algoritmo, es preciso realizar cálculos estadísticos, estos cálculos pueden ser incorporados dentro del Script desarrollado.

3 Bibliografía

1. Noel Gorelick, Matt Hancher, Mike Dixon, Simon Ilyushchenko, David Thau, Rebecca Moore, Google Earth Engine: Planetary-scale geospatial analysis for everyone, Remote Sensing of Environment.
2. SENTINEL-2 User Handbook
3. Adugna, Tesfaye, Wenbo Xu, and Jinlong Fan. 2022. "Comparison of Random Forest and Support Vector Machine Classifiers for Regional Land Cover Mapping Using Coarse Resolution FY-3C Images" Remote Sensing
4. Mountrakis G., Im J., Ogole, C., Support vector machines in remote sensing: A review [A], Journal of Photogrammetry and Remote Sensing ISPRS [J], 2011
5. Breiman L., Friedman J., Stone C., Olshen R. Classification and Regression Trees [S], Taylor & Francis, 1984.
6. Gong P., Wang J., Yu L., Zhao Y., et al. Finer resolution observation and monitoring of global land cover: first mapping results with Landsat TM and

- ETM+ data, International Journal of Remote Sensing Taylor & Francis [J], 2013
7. Shao Y., Lunetta R., Comparison of support vector machine, neural network, and CART algorithms for the land-cover classification using limited training data points [A], Journal of Photogrammetry and Remote Sensing ISPRS [J], 2012
 8. Carrasco L.; O'Neil A.W.; Morton R.D.; Rowland, C.S. Evaluating Combinations of Temporally Aggregated Sentinel-1, Sentinel-2 and Landsat 8 for Land Cover Mapping with Google Earth Engine [A]. Remote Sensing MDPI [J], 2019.
 9. Pelletier C., Valero S., Inglada J., Champion N., Dedieu G., Assessing the robustness of Random Forests to map land cover with high resolution satellite image time series over large areas [A]. Remote Sensing of Environment ELSEVIER [J], 2016.
 10. J. Chen, J. Chen, A. Liao, X. Cao, L. Chen, X. Chen, C. He, G. Han, et al. Global land cover mapping at 30m resolution: A POK-based operational approach [A]. Journal of Photogrammetry and Remote Sensing [J]. 2015 .
 11. Gislason P.O., Benediktsson J.A., Sveinsson J. Random forests for land cover classification [A]. Pattern Recognition Letters ELSEVIER [J], 2006.
 12. Belgiu M., Drăgut L. Random forest in remote sensing: a review of applications and future directions [A]. Photogram. Remote Sensing ISPRS [J], 2016