



PROYECTO I+D+I-XVII-2023-048
“ANÁLISIS DE PROBLEMAS TERRITORIALES EN EL ECUADOR
UTILIZANDO MACHINE LEARNING Y SENSORES REMOTOS DESDE UNA
PERSPECTIVA DE CAMBIO CLIMÁTICO (APTECUCC)”

Entregable V.1.4 Informe con los resultados de la validación de la clasificación de US.

Actividad T.1.6. Desarrollo y validación (matriz de confusión y estadísticos de acuerdo/desacuerdo) de un proceso automático de clasificación de imágenes de satélite (Clasificación basada en píxeles usando ML) para identificar US, integrando la base de datos de asentamientos informales (M2).

1. Clasificación de US a partir de imágenes de satélite

En el proceso de clasificación de imágenes se realizaron las siguientes etapas:

1.1 Codificación de clases

Cada clase tendrá asignado un número para su representación. Esto es un requisito en los algoritmos de clasificación supervisada.

Class	1990	2000	2008	2018
0	SIN INFORMACIÓN	SIN INFORMACIÓN	SIN INFORMACIÓN	SIN INFORMACIÓN
1	BOSQUE	BOSQUE	BOSQUE	BOSQUE
2	CUERPO DE AGUA	CUERPO DE AGUA	CUERPO DE AGUA	CUERPO DE AGUA
3	TIERRA AGROPECUARIA	TIERRA AGROPECUARIA	TIERRA AGROPECUARIA	TIERRA AGROPECUARIA
4	VEGETACION ARBUSTIVA Y HERBACEA	VEGETACIÓN ARBUSTIVA Y HERBACEA	VEGETACIÓN ARBUSTIVA Y HERBACEA	VEGETACIÓN ARBUSTIVA Y HERBACEA
5	ZONA ANTRÓPICA y OTRAS TIERRAS (Áreas sin cobertura vegetal)	ZONA ANTRÓPICA y OTRAS TIERRAS (Áreas sin cobertura vegetal)	ZONA ANTRÓPICA	ZONA ANTRÓPICA OTRAS TIERRAS (Áreas sin cobertura vegetal)

Nivel I	Nivel II	Definición Operativa
Bosque	Bosque Nativo	Ecosistema arbóreo, primario o secundario, regenerado por sucesión natural, se caracteriza por la presencia de árboles de diferentes especies nativas, edades y portes variados, con uno o más estratos.
	Plantación Forestal	Masa arbórea establecida antrópicamente con una o más especies forestales.
Vegetación Arbustiva y Herbácea	Vegetación Arbustiva	Áreas con un componente substancial de especies leñosas nativas no arbóreas. Incluye áreas degradadas en transición a una cobertura densa del dosel.
	Páramo	Vegetación tropical altoandino caracterizada por especies dominantes no arbóreas que incluyen fragmentos de bosque nativo propios de la zona.
	Vegetación Herbácea	Áreas constituidas por especies herbáceas nativas con un crecimiento espontáneo, que no reciben cuidados especiales, utilizados con fines de pastoreo esporádico, vida silvestre o protección.
Tierra Agropecuaria	Cultivo Anual	Comprende aquellas tierras dedicadas a cultivos agrícolas, cuyo ciclo vegetativo es estacional, pudiendo ser cosechados una o más veces al año.
	Cultivo Semipermanente	Comprenden aquellas tierras dedicadas a cultivos agrícolas cuyo ciclo vegetativo dura entre uno y tres años.
	Cultivo Permanente	Comprenden aquellas tierras dedicadas a cultivos agrícolas cuyo ciclo vegetativo es mayor a tres años, y ofrece durante éste periodo varias cosechas.
Cuerpo de Agua	Natural	Superficie y volumen asociado de agua estática o en movimiento.
	Artificial	Superficie y volumen asociados con las actividades antrópicas y el manejo del recurso hídrico.
Zona Antrópica	Área Poblada	Áreas principalmente ocupadas por viviendas y edificios destinados a colectividades o servicios públicos.
	Infraestructura	Obra civil de transporte, comunicación, agroindustrial y social.
Otras Tierras	Área sin cobertura vegetal	Áreas generalmente desprovistas de vegetación, que por sus limitaciones edáficas, climáticas, topográficas o antrópicas, no son aprovechadas para uso agropecuario o forestal, sin embargo pueden tener otros usos.
	Glaciar	Nieve y hielo localizados en las cumbres de las elevaciones andinas.
Sin Información		Corresponde a áreas que no han podido ser mapeadas.

1.2 Carga de puntos en formato shapefile a GEE

En función de pruebas preliminares se vió la necesidad de modificar los puntos de la actividad **T.1.5 Selección y fotointerpretación de un conjunto de puntos que serán usados para validar la clasificación de US como “verdad terreno”**. En primer lugar se varió los porcentajes de datos para entrenamiento y validación, siendo usados el 70% (en lugar del 60%) para entrenamiento y el 30% (en lugar del 40%) para validación. El detalle de los cambios realizados en los puntos está disponible en la carpeta **“T.1.6.Anexos”**: <https://drive.google.com/drive/folders/10Ba7LJe9zPsumNKQO4Egu6bFjwgjQSiU>

Los shapes recibidos se cargaron en la base de datos (BD) PostgreSQL/Postgis **“asentamientos_humanos”**. Para ello en primer lugar el shapefile debió proyectarse al

datum EPSG:4326 - WGS 84 y luego evaluar las columnas (nombres y tipos de datos) descritas en la siguiente imagen:

	c2018	c2008	c2000	c1990	id	tipo	zona	codigo
1	TIERRA AGROP...	TIERRA AGROP...	TIERRA AGROP...	TIERRA AGROP...	1	TRAIN	Quito	1
2	TIERRA AGROP...	TIERRA AGROP...	TIERRA AGROP...	TIERRA AGROP...	2	TEST	Quito	2
3	BOSQUE	BOSQUE	BOSQUE	BOSQUE	3	TRAIN	Quito	3
4	BOSQUE	BOSQUE	BOSQUE	BOSQUE	4	TRAIN	Quito	4
5	TIERRA AGROP...	TIERRA AGROP...	TIERRA AGROP...	TIERRA AGROP...	5	TRAIN	Quito	5
6	VEGETACION A...	VEGETACION A...	VEGETACION A...	VEGETACION A...	6	TRAIN	Quito	6
7	VEGETACION A...	VEGETACION A...	VEGETACION A...	VEGETACION A...	7	TRAIN	Quito	7
8	ZONA ANTROP...	TIERRA AGROP...	TIERRA AGROP...	TIERRA AGROP...	8	TRAIN	Quito	8

Al tener la información en una BD se podrá actualizar de una forma más eficiente, en caso de detectar inconsistencias a través de la inspección visual realizada a partir de las imágenes de satélite Landsat/Sentinel.

Posterior a ello a través del software QGIS se exportó la información a un formato shapefile, con el cual podremos generar un archivo comprimido del mismo, el cual será cargado a GEE.

Name	Date modified	Type	Size
cedia_sample_points_gee.cpg	5/2/2024 10:02 PM	CPG File	1 KB
cedia_sample_points_gee.dbf	5/2/2024 10:02 PM	DBF File	2,366 KB
cedia_sample_points_gee.prj	5/2/2024 10:02 PM	PRJ File	1 KB
cedia_sample_points_gee.qmd	5/2/2024 10:02 PM	QMD File	2 KB
cedia_sample_points_gee.shp	5/2/2024 10:02 PM	SHP File	55 KB
cedia_sample_points_gee.shx	5/2/2024 10:02 PM	SHX File	16 KB
cedia_sample_points_gee.zip	5/2/2024 10:03 PM	Compressed (zipp...	68 KB

Para ello utilizaremos la herramienta **Assets** y será renombrado como **cedia_sample_points_gee**.

Nota: Luego de una evaluación inicial se decidió reformular el porcentaje de puntos usados para entrenamiento y validación, siendo usado el 70% para entrenamiento y el 30% para validación. A partir de este último se formulan las métricas y matrices de confusión.



1.3 Índices de vegetación y otros a evaluar (Landsat 5,7,8,9 y Sentinel 2)

Para la construcción de los índices propuestos se usó la plataforma Google Earth Engine donde la codificación de las bandas, según el satélite, es la siguiente:

Año	1982-1993	1984-2012	1999-2021	2013-Pres	2021-Pres L9	2017-Pres Sentinel 2-MSI
Nombre	L4 TM	L5 TM	L7 ETM+	L8 OLI/TIRS	OLI-2/TIRS-2	2-MSI
R (Red)	SR_B3	SR_B3	SR_B3	SR_B4	SR_B4	B4
G (Green)	SR_B2	SR_B2	SR_B2	SR_B3	SR_B3	B3
B (Blue)	SR_B1	SR_B1	SR_B1	SR_B2	SR_B2	B2
NIR	SR_B4	SR_B4	SR_B4	SR_B5	SR_B5	B8
SWIR1	SR_B5	SR_B5	SR_B5	SR_B6	SR_B6	B11
SWIR2	SR_B7	SR_B7	SR_B7	SR_B7	SR_B7	B12
RedEdge1						B5
RedEdge2						B6
RedEdge3						B7
RedEdge4						B8A

*L4,L5,L7 y L8 corresponden a la misión Landsat

* Pres: Presente

A partir de la bibliografía se recolectó un listado de índices de vegetación, que contribuirán a las tareas de clasificación de imágenes.

Nombre Índice	Fórmula	Fórmula Landsat 7	Especialización	Referencias
Visible-band Difference Vegetation Index (VDVI)	$(2 \cdot G - R - B) / (2 \cdot G + R + B)$	$(2 \cdot B2 - B3 - B1) / (2 \cdot B2 + B3 + B1)$	Vegetación (extrae vegetación verde)	Wang et al. (2015) Wan et al. (2018)
Visible Atmospherically Resistant Index (VARI)	$(G - R) / (G + R - B)$	$(B2 - B3) / (B2 + B3 - B1)$	Vegetación (estima fracción de vegetación; menos sensible a efectos atmosféricos)	Wan et al. (2018) Gitelson et al. (2002)
Normalized Green-Red Difference Index (NGRDI)	$(G - R) / (G + R)$	$(B2 - B3) / (B2 + B3)$	Vegetación (indicador de fenología; estimador de biomasa)	Wan et al. (2018) Gitelson et al. (2002)
Red-Green Ratio Index (RGRi)	R/G	B3/B2	Vegetación (analiza sensibilidad angular de índices de vegetación; tratar la estructura de dosel)	Wang et al. (2018) Verrelst et al. (2008)
Modified Green Red Vegetation Index (MGRVI)	$(G^2 - R^2) / (G^2 + R^2)$	$(B2^2 - B3^2) / (B2^2 + B3^2)$	Vegetación (indicador de fenología; estimador de biomasa)	Wan et al. (2018) Bendig et al. (2015)
Excess Green Index (ExG)	$2 \cdot G - R - B$	$2 \cdot B2 - B3 - B1$	Vegetación (identifica vegetación verde sin influencia de condiciones de sombra a luz solar)	Wan et al. (2018) Woebbecke et al. (1995)
Color Index of Vegetation (CIVE)	$0.441 \cdot R - 0.881 \cdot G + 0.385 \cdot B + 18.787$	$0.441 \cdot B3 - 0.881 \cdot B2 + 0.385 \cdot B1 + 18.787$	Vegetación (identifica vegetación verde sin influencia de condiciones de sombra a luz solar)	Wan et al. (2018) Kataoka et al. (2003)
Vegetativen (VEG)	$G/R^a \cdot B^{(1-a)}$ a = 0.667	$B2/B3^a \cdot B1^{(1-a)}$ a = 0.667	Vegetación (identifica vegetación verde sin influencia de condiciones de sombra a luz solar)	Wan et al. (2018) Hague et al. (2006)
Ratio Vegetation Index (RVI)	R/NIR	B3/B4	Vegetación (estima y la biomasa verde en coberturas vegetales de alta densidad)	Xue & Su (2017)
Normalized Difference Vegetation Index (NDVI)	$(NIR - R) / (NIR + R)$	$(B4 - B3) / (B4 + B3)$	Vegetación (es el índice más conocido y usado para evaluar vegetación: estructura y fotosíntesis del dosel)	Xue & Su (2017)

Atmospherically Resistant Vegetation Index (ARVI)	$(NIR - RB)/(NIR + RB)$	$(B4 - B3 \cdot B1) / (B4 + B3 \cdot B1)$	Vegetación y efectos atmosféricos (útil para eliminar efectos de aerosoles atmosféricos)	Xue & Su (2017)
Soil Adjusted Vegetation Index (SAVI)	$((NIR - R) / (NIR + R + L)) * (1+L)$ L = valor entre 0 y 1 (0 zonas con densidad vegetal alta; 1 para zonas con escasa densidad vegetal)	$((B4 - B3) / (B4 + B3 + L)) * (1+L)$	Vegetación y suelo (mejora la sensibilidad del NDVI en contextos de suelos)	Xue & Su (2017)
Normalized Difference Built-up Index (NDBI)	$(SWIR - NIR) / (SWIR + NIR)$	$(B6 - B5) / (B6 + B5)$ Landsat 8	Zonas construidas frente a zonas con vegetación	Prasomsup et al. (2020)
Built-up Index (BUI)	NDBI - NDVI		Útil para analizar patrones urbanos usando NDBI y NDVI	Prasomsup et al. (2020)
Modified Normalized Difference Water Index (MNDWI)	$(G - NIR) / (G + NIR)$	$(B3 - B5) / (B3 + B5)$ Landsat 8	Agua (delimita cuerpos de agua)	Prasomsup et al. (2020)
Índice de Vegetación RGB derivados de imágenes de vehículos aéreos no tripulados	(1) $I = B + G + R$ (2) Green to red ratio = G / R , (3) Red to blue ratio = R / B (4) Green normalized by blue (NGB) = $(G-B) / (G+B)$ (5) Red normalized by blue (NRB) = $(R-B) / (R+B)$.		Vegetación/ Clasificación de especies de árboles forestales de hoja ancha a través de imágenes adquiridas por un UAV/ Alto potencial para la identificación automática de la biomasa vegetal/ Monitorear el contenido de humedad.	Sadeghi & Sohrabi (2018)
Row image bands: raw bands / vegetation indices / raw bands + vegetation indices.			Vegetación/ Identificación de especies de árboles.	Sadeghi & Sohrabi (2018)

* Debido a la gran cantidad de índices se puede revisar la siguiente [tabla](#) *

1.4. Selección de variables con RFE (Recursive Feature Elimination) y Smile Random Forest; Random Forest

A partir del algoritmo smile RandomForest de GEE se pudo determinar la importancia de las variables en la clasificación. A partir de este valor se ejecutó un proceso recursivo denominado RFE (Recursive Feature Elimination), que consiste en eliminar las variables de menor importancia para llegar a determinar el mínimo número de variables útiles para la clasificación. Entre las bandas disponibles de Landsat¹ y los índices generados se obtuvo un total de 18 variables que son: SR_B1 (Blue), SR_B2 (Green), SR_B3 (Red), SR_B4 (NIR), SR_B5 (SWIR1), SR_B7 (SWIR2), NDVI, VDI, VARI, NGRDI, RGRI, MGRVI, ExG, CIVE, RVI, NDBI, BUI y MNDWI.

Finalmente RFE se ejecutó para cada una de las zonas de estudio y año de análisis. A partir del método no paramétrico Random Forest (**RF**) se pudo determinar en el estudio de Discriminación de especies de árboles mediante índices de vegetación RGB derivados de imágenes de vehículos aéreos no tripulados por Sima Sadeghi & Hormoz Sohrabi su precisión para la clasificación de especies de árboles, puesto que es menos sensible al número de variables de entrada y menos sensible al sobreajuste.(0)

El RF se calculó con 100 árboles en el bosque y 14 términos muestreados por división. LDA y RF se calcularon por separado para tres escenarios que incluyen bandas sin procesar, índices de vegetación; y bandas sin procesar + índices de vegetación como predictores para las clasificaciones.(0)

El porcentaje de clasificación errónea de RF fue del 10 % con una entropía R-cuadrada de 0,58. Los índices de vegetación, incluidos I, GtoR, así como la media azul y la varianza de la banda roja, desempeñaron un papel importante en las predicciones. Aquí también encontramos que el uso de bandas sin procesar junto con índices de vegetación da como resultado mejores predicciones. Además, según la matriz de confusión, Populus Caspica se identificó con mayor precisión que otras tres especies.(0)

De igual manera las referencias demostraron que (1) la reflectancia bidireccional, (2) las costuras y (3) las fluctuaciones de la luz ambiental debían considerarse para aplicaciones de imágenes de UAV. Las fluctuaciones de la luz ambiental fueron la fuente más eficaz que puede afectar el resultado de las aplicaciones de imágenes de UAV. La comparación de LDA y RF reveló que RF puede dar como resultado resultados de clasificación más precisos.(0)

¹ https://developers.google.com/earth-engine/datasets/catalog/LANDSAT_LT05_C02_T1_L2#bands

Referencias:

- ⁰ Sadeghi & Sohrabi (2018) Tree species discrimination using RGB vegetation indices derived from UAV images, 2-6.
- ¹ Fassnacht, F.E., Latifi, H., Stereńczak, K., Modzelewska, A, et al. (2016) Review of studies on tree species classification from remotely sensed data. Remote Sensing of Environment, 186, 64–87.
- ² Tuominen, S., Näsi, R., Honkavaara, E., Balazs, A., et al. (2017) Tree species recognition in species rich area using UAV-borne hyperspectral imagery and stereo-photogrammetric point cloud. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives, 42(3W3), 185–194.
- ³ Clark, M.L., Roberts, D.A. & Clark, D.B. (2005) Hyperspectral discrimination of tropical rain forest tree species at leaf to crown scales. Remote Sensing of Environment, 96(3–4), 375–398.
-

Se debe recalcar que en ciertos casos al no existir imágenes libres de nubosidad del año a evaluar, se seleccionó la imagen de fecha más cercana que pueda ser utilizada. A partir de las imágenes seleccionadas se desarrolló un microinforme bajo el siguiente formato: 1) la fecha de la imagen y sensor seleccionado para la clasificación. 2) La cantidad de puntos que pudieron obtener datos luego del filtro de nubes para entrenamiento y validación. 3) Las variables seleccionadas luego del proceso de RFE. 4) Una vista previa de la imagen seleccionada para evaluar las zonas de nubosidad. 5) El gráfico del proceso de RFE visto por la cantidad de variables eliminadas. 6) Un gráfico con la importancia de las variables seleccionadas con smileRandomForest.

ZONA DE ESTUDIO Y AÑO DE EVALUACIÓN

(1) Fecha: Sensor:	(4) Imagen xx usada	(5) RFE de variables
(2) Puntos usados: Train (556) 93% Test (361) 90%		(6) Importancia de variables
(3) variables seleccionadas:		

CUENCA 1990.

Fecha: 1989-12-28

Sensor: Landsat 5 TM

Puntos usados:

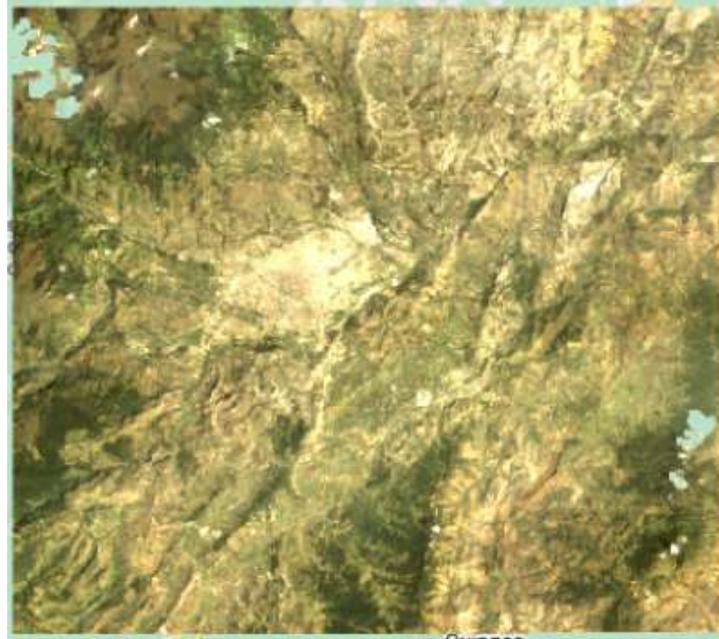
Train (692) 99%

Test (296) 99%

5 variables seleccionadas:

["SR_B1", "SR_B2", "SR_B4", "SR_B7", "NDBI"]

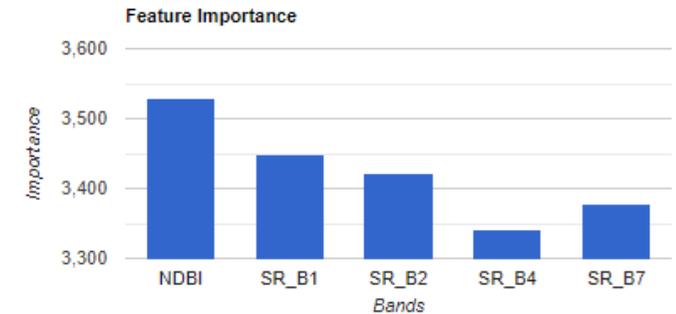
Imagen Landsat 5 usada



RFE de variables



Importancia de variables



Métricas de clasificación:

1. BOSQUE, 2. CUERPO DE AGUA, 3. TIERRA AGROPECUARIA, 4. VEGETACION ARBUSTIVA Y HERBACEA, 5. ZONA ANTROPICA

		Prediction				
		1	2	3	4	5
Real	1	47	0	4	2	1
	2	0	20	0	1	0
	3	14	0	89	11	7
	4	7	0	19	21	1
	5	0	0	16	0	36

	Acc	0.720	Usuario	Productor	FScore
Kappa	0.616	0.691	0.870	0.770	
		1.000	0.952	0.976	
		0.695	0.736	0.715	
		0.600	0.438	0.506	
		0.800	0.692	0.742	

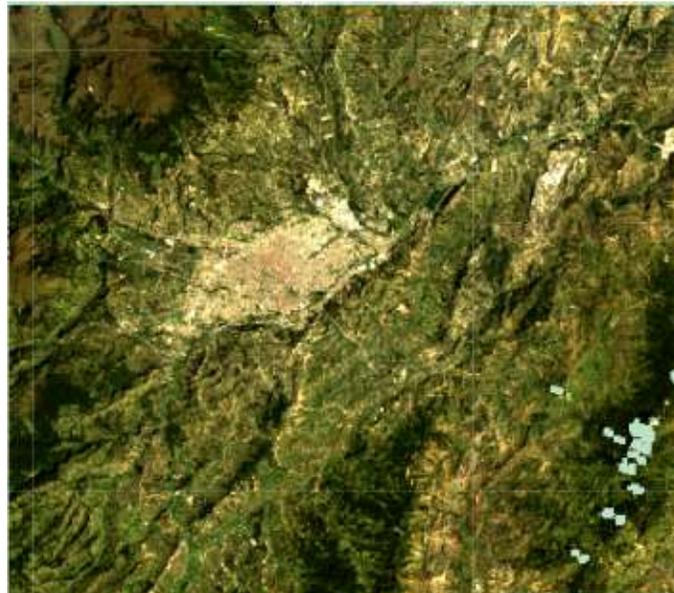
CUENCA 2000

Fecha: 2000-01-09
Sensor: Landsat 5 TM

Puntos usados:
 Train (694) 99%
 Test (296) 99%

9 variables seleccionadas:
 SR_B1,SR_B2,SR_B3,SR_B4,
 SR_B5,SR_B7,NDVI,RVI,BUI

Imagen Landsat 5 usada



RFE de variables



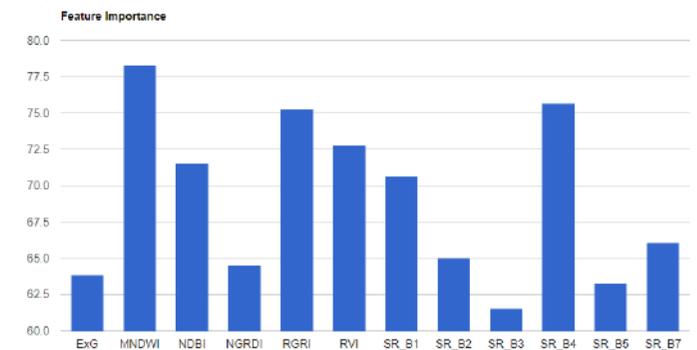
Métricas de clasificación:

1. BOSQUE, 2. CUERPO DE AGUA, 3. TIERRA AGROPECUARIA, 4. VEGETACION ARBUSTIVA Y HERBACEA, 5. ZONA ANTROPICA

		Prediction				
		1	2	3	4	5
Real	1	48	0	7	2	0
	2	0	19	0	2	0
	3	7	0	82	14	7
	4	5	1	22	25	1
	5	0	0	8	1	45

	Acc	0.740	Usuario	Productor	FScore
Kappa	0.652	0.800	0.842	0.821	
		0.950	0.905	0.927	
		0.689	0.745	0.716	
		0.568	0.463	0.510	
		0.849	0.833	0.841	

Importancia de variables



CUENCA 2008.

Fecha: 2009-11-09

Sensor: Landsat7 ETM+

Puntos usados:

Train (700) 100%

Test (300) 100%

* El relleno de datos de Landsat puede alterar la radiometría de las coberturas vegetales.

7 variables seleccionadas:

SR_B1,SR_B3,SR_B4,SR_B7, RGRI,BUI,MNDWI

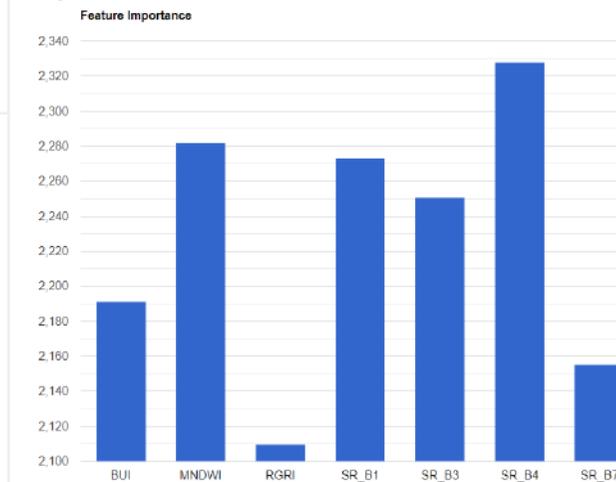
Imagen Landsat 7 usada



RFE de variables



Importancia de variables



Métricas de clasificación:

1. BOSQUE, 2. CUERPO DE AGUA, 3.TIERRA AGROPECUARIA, 4. VEGETACION ARBUSTIVA Y HERBACEA, 5.ZONA ANTROPICA

		Prediction				
		1	2	3	4	5
Real	1	43	1	12	6	0
	2	1	19	1	1	0
	3	13	1	99	8	5
	4	6	0	8	18	1
	5	0	0	9	0	48

	Acc	0.757	Usuario	Productor	FScore
Kappa	0.664	0.683	0.694	0.688	
		0.905	0.864	0.884	
		0.767	0.786	0.776	
		0.545	0.545	0.545	
		0.889	0.842	0.865	

CUENCA 2018

Fecha: 2018-03-02
Sensor: Sentinel 2 L1C

Puntos usados:
 Train 700 (100%)
 Test 300 (100%)

4 variables seleccionadas:
 B3,B9,NDVI,NDBI

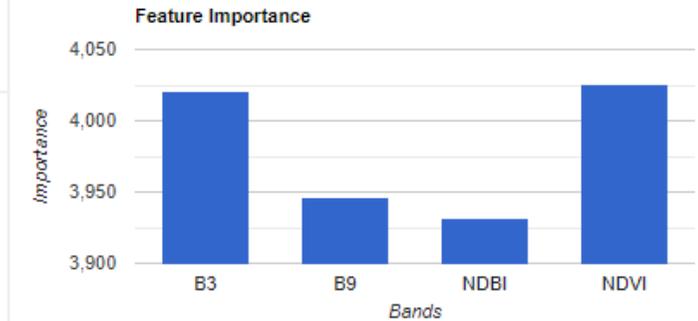
Imagen Sentinel 2 L1C usada



RFE de variables



Importancia de variables



Métricas de clasificación:

1. BOSQUE, 2. CUERPO DE AGUA, 3.TIERRA AGROPECUARIA, 4. VEGETACION ARBUSTIVA Y HERBACEA, 5.ZONA ANTROPICA

		Prediction				
		1	2	3	4	5
Real	1	31	2	19	1	0
	2	0	19	1	2	0
	3	6	0	111	5	3
	4	1	0	8	18	1
	5	0	0	18	1	53

	Acc	0.773	Usuario	Productor	FScore
Kappa	0.676	0.816	0.585	0.681	
		0.905	0.864	0.884	
		0.707	0.888	0.787	
		0.667	0.643	0.655	
		0.930	0.736	0.822	

QUITO 1990

Fecha: 1991-10-15

Sensor: Landsat 5 TM

Puntos usados:

Train (670) 96%

Test (292) 97%

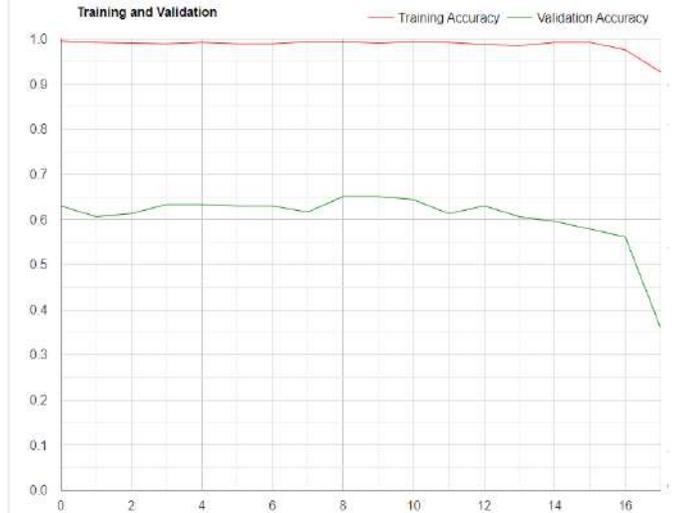
10 variables seleccionadas:

SR_B1,SR_B2,SR_B4,SR_B5,
SR_B7,NDVI,VARI,RGRI,ExG
,BUI

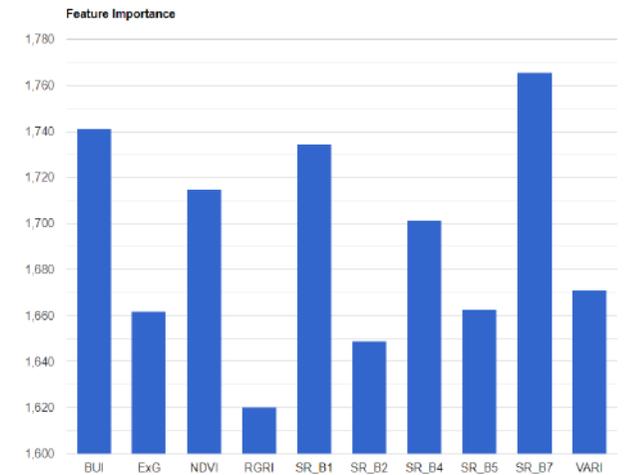
Imagen Landsat 5 usada



RFE de variables



Importancia de variables



Métricas de clasificación:

1. BOSQUE, 2. CUERPO DE AGUA, 3.TIERRA AGROPECUARIA, 4. VEGETACION ARBUSTIVA Y HERBACEA, 5.ZONA ANTROPICA

		Prediction					NA: No aplica				
		1	2	3	4	5	Acc	Usuario	Productor	FScore	
Real	1	61	0	20	13	0	Kappa	0.640	0.718	0.649	0.682
	2	NA	NA	NA	NA	NA		NA	NA	NA	
	3	22	0	96	24	2		0.738	0.667	0.701	
	4	1	0	9	20	0		0.308	0.667	0.421	
	5	1	0	5	8	10		0.833	0.417	0.556	

QUITO 2000

Fecha: 1999-08-18

Sensor: Landsat 5 TM

Puntos usados:

Train (694) 99%

Test (297) 99%

9 variables seleccionadas:

SR_B1,SR_B2,SR_B4,SR_B7,

NDVI,VARI,ExG,NDBI,BUI

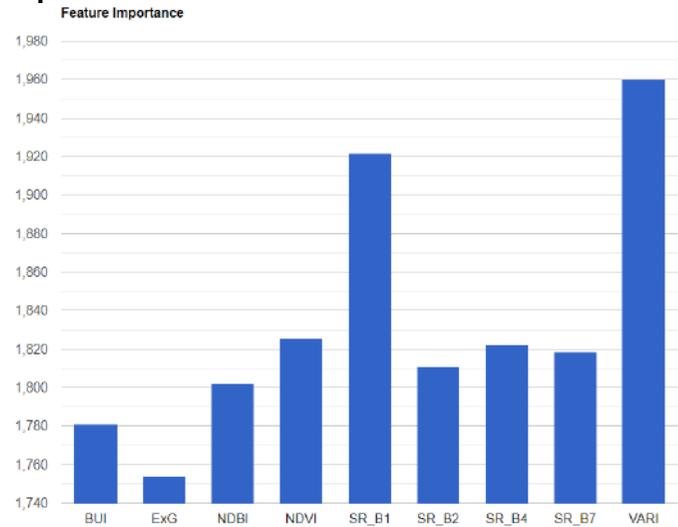
Imagen Landsat 5 usada



RFE de variables



Importancia de variables



Métricas de clasificación:

1. BOSQUE, 2. CUERPO DE AGUA, 3.TIERRA AGROPECUARIA, 4. VEGETACION ARBUSTIVA Y HERBACEA, 5.ZONA ANTROPICA

		Prediction					NA: No aplica				
		1	2	3	4	5	Acc	0.680	Usuario	Productor	FScore
Real	1	81	0	9	14	0	Kappa	0.547	0.827	0.779	0.802
	2	NA	NA	NA	NA	NA					
	3	15	0	76	24	10					
	4	1	0	5	25	3					
	5	1	0	8	5	20					
									0.776	0.608	0.682
									0.368	0.735	0.490
									0.606	0.588	0.597

QUITO 2008

Fecha: 2008-11-06
Sensor: Landsat 7 ETM+

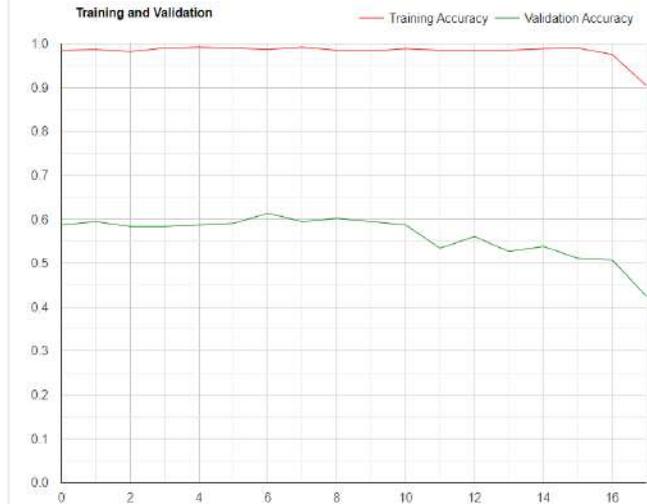
Puntos usados:
 Train (615) 88%
 Test (264) 88%

12 variables seleccionadas:
 SR_B1,SR_B2,SR_B5,SR_B7,
 VDMI,VARI,RGRI,MGRVI,
 CIVE,RVI,BUI,MNDWI

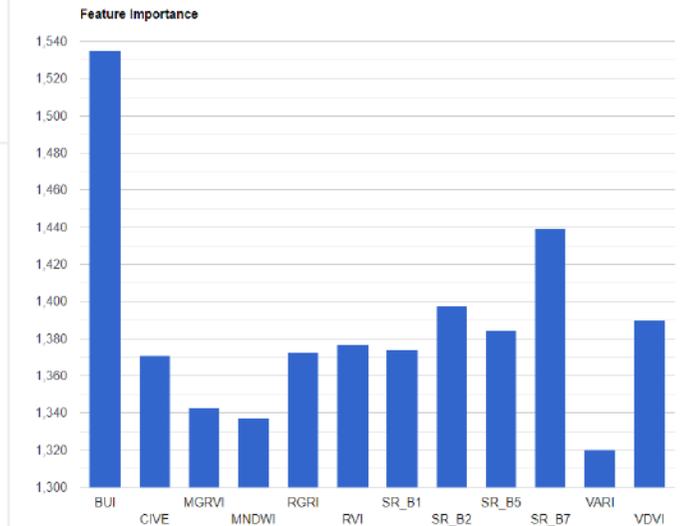
Imagen Landsat 7



RFE de variables



Importancia de variables



Métricas de clasificación:

1. BOSQUE, 2. CUERPO DE AGUA, 3.TIERRA AGROPECUARIA, 4. VEGETACION ARBUSTIVA Y HERBACEA, 5.ZONA ANTROPICA

		Prediction					NA: No aplica				
		1	2	3	4	5	Acc	0.617	Usuario	Productor	FScore
Real	1	48	0	12	3	0	Kappa	0.463	0.706	0.762	0.733
	2	NA	NA	NA	NA	NA			NA	NA	NA
	3	10	0	69	15	7			0.595	0.683	0.636
	4	9	0	19	18	5			0.450	0.353	0.396
	5	1	0	16	4	28			0.700	0.571	0.629

QUITO 2018

Fecha: 2018-10-01
Sensor: Landsat 8 ETM +

Puntos usados:
 Train(684) 98%
 Test (298) 99%

7 variables seleccionadas:
 SR_B1,SR_B2,SR_B3,SR_B5,
 NGRDI,ExG,RVI

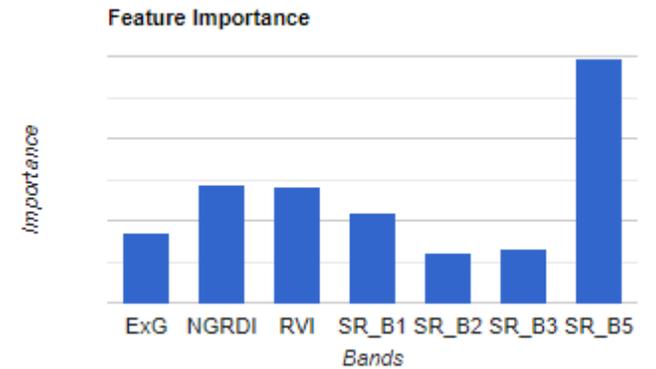
Imagen Landsat 8



RFE de variables



Importancia de variables



Métricas de clasificación:

1. BOSQUE, 2. CUERPO DE AGUA, 3.TIERRA AGROPECUARIA, 4. VEGETACION ARBUSTIVA Y HERBACEA, 5.ZONA ANTROPICA

		Prediction					NA: No aplica				
		1	2	3	4	5	Acc	0.738	Usuario	Productor	FScore
Real	1	68	0	6	5	4	Kappa	0.646	0.861	0.819	0.840
	2	NA	NA	NA	NA	NA			NA	NA	NA
	3	8	0	61	7	4			0.693	0.763	0.726
	4	2	0	11	26	11			0.553	0.520	0.536
	5	1	0	10	9	65			0.774	0.765	0.769

LOJA 1990

Fecha: 1989-11-10

Sensor: Landsat 5 TM

Puntos usados:

Train (699) ~100%

Test (300) ~100%

11 variables seleccionadas:

SR_B1,SR_B2,SR_B3,SR_B4,S

R_B7,NDVI,VARI,RGRI,CIVE,

NDBI, BUI

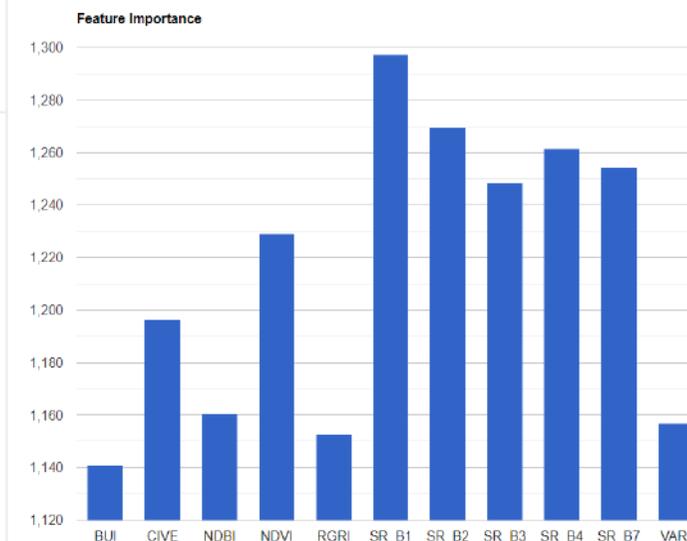
Imagen Landsat 5



RFE de variables



Importancia de variables



Métricas de clasificación:

1. BOSQUE, 2. CUERPO DE AGUA, 3.TIERRA AGROPECUARIA, 4. VEGETACION ARBUSTIVA Y HERBACEA, 5.ZONA ANTROPICA

		Prediction				
		1	2	3	4	5
Real	1	110	0	11	8	0
	2	1	2	2	1	0
	3	4	0	58	9	1
	4	9	0	13	19	0
	5	0	1	1	0	50

	Acc	Usuario	Productor	FScore
Acc	0.797			
Kappa	0.713	0.887	0.853	0.870
		0.667	0.333	0.444
		0.682	0.806	0.739
		0.514	0.463	0.487
		0.980	0.962	0.971

LOJA 2000

Fecha: 2000-09-05
Sensor: Landsat 5 TM

Puntos usados:
 Train (687) 98%
 Test (292) 97%

7 variables seleccionadas:
 SR_B1,SR_B4,SR_B7,VDVI,M
 GRVI,ExG,BUI

Imagen Landsat 5 usada

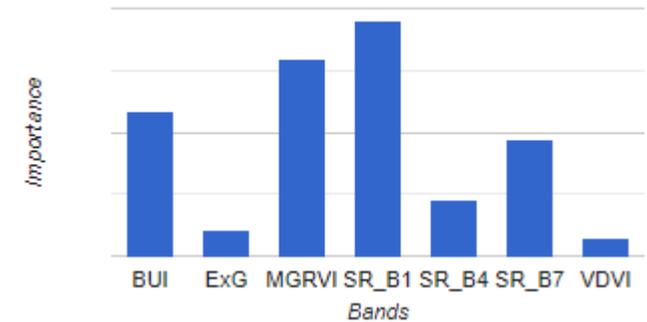


RFE de variables



Importancia de variables

Feature Importance



Métricas de clasificación:

1. BOSQUE, 2. CUERPO DE AGUA, 3.TIERRA AGROPECUARIA, 4. VEGETACION ARBUSTIVA Y HERBACEA, 5.ZONA ANTROPICA

		Prediction					Acc 0.736	Usuario	Productor	FScore			
		1	2	3	4	5							
Real	1	94	2	11	3	0	Kappa 0.633	0.746	0.855	0.797			
	2	1	1	1	4	1					0.333	0.125	0.182
	3	8	0	57	7	0					0.671	0.792	0.726
	4	22	0	12	16	0					0.533	0.320	0.400
	5	1	0	4	0	47					0.979	0.904	0.940

LOJA 2008

Fecha: 2008-09-19

Sensor: Landsat 7 ETM +

Puntos usados:

Train (669) 96%

Test (295) 98%

6 variables seleccionadas:

SR_B7,NDVI,NGRDI,MGRVI,CIVE,BUI

Imagen Landsat 7 usada

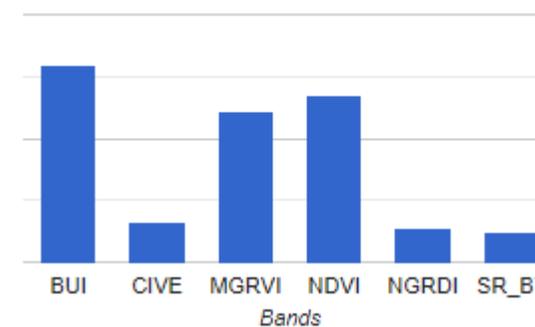


RFE de variables



Importancia de variables

Feature Importance



Métricas de clasificación:

1. BOSQUE, 2. CUERPO DE AGUA, 3.TIERRA AGROPECUARIA, 4. VEGETACION ARBUSTIVA Y HERBACEA, 5.ZONA ANTROPICA

		Prediction				
		1	2	3	4	5
Real	1	80	2	4	11	0
	2	6	0	2	2	1
	3	6	0	51	15	1
	4	10	0	10	38	0
	5	1	0	3	4	48

	Acc	0.736	Usuario	Productor	FScore
Kappa	0.646	0.777	0.825	0.800	
		0.000	0.000	NaN	
		0.729	0.699	0.713	
		0.543	0.655	0.594	
		0.960	0.857	0.906	

LOJA 2018

Fecha: 2018-10-24
Sensor: Sentinel 2 L1C (17MPR)

Puntos usados:

Train (700) 100%
 Test (300) 100%

6 variables seleccionadas:

B2,B3,B5,NGRDI,RVI,NDBI

Imagen Sentinel 2 usada

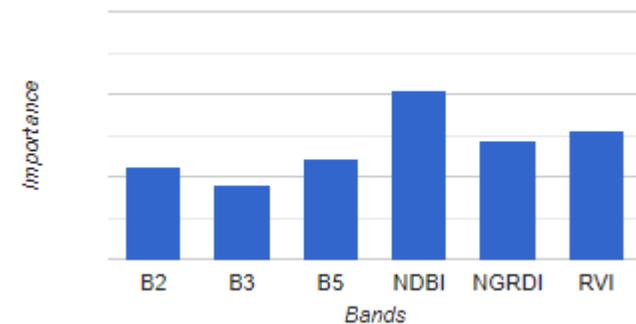


RFE de variables



Importancia de variables

Feature Importance



Métricas de clasificación

1. BOSQUE, 2. CUERPO DE AGUA, 3.TIERRA AGROPECUARIA, 4. VEGETACION ARBUSTIVA Y HERBACEA, 5.ZONA ANTROPICA

		Prediction					Acc	Usuario	Productor	FScore
		1	2	3	4	5				
Real	1	96	0	3	14	0	0.783	0.842	0.850	0.846
	2	1	0	1	2	1	0.704	0.000	0.000	NaN
	3	2	0	51	9	4		0.750	0.773	0.761
	4	15	0	8	34	0		0.576	0.596	0.586
	5	0	0	5	0	54		0.915	0.915	0.915

GUAYAQUIL 1990

Fecha: 1991-04-29

Sensor: Landsat 5 TM

Puntos usados:

Train (656) 94%

Test (295) 98%

9 variables seleccionadas:

SR_B1,SR_B2,SR_B4,SR_B5,SR_B7,NDVI,RVI,BUI,MNDWI

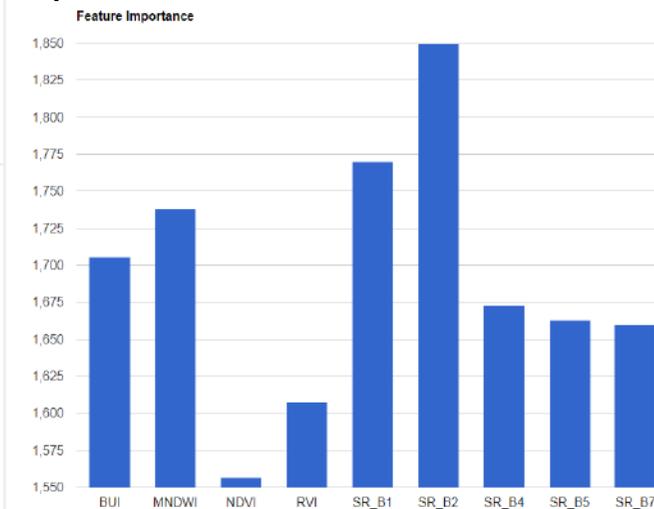
Landsat 5 usada



RFE de variables



Importancia de variables



Métricas de clasificación

1. BOSQUE, 2. CUERPO DE AGUA, 3. TIERRA AGROPECUARIA, 4. VEGETACION ARBUSTIVA Y HERBACEA, 5. ZONA ANTROPICA

		Prediction				
		1	2	3	4	5
Real	1	2	0	15	0	0
	2	1	24	9	0	3
	3	25	3	121	0	4
	4	0	0	0	0	0
	5	0	0	4	0	9

	Acc	0.692	Usuario	Productor	FScore
Kappa	0.499		0.641	0.667	0.654
			0.828	0.649	0.727
			0.703	0.791	0.745
			0.000	0.000	NaN
			0.563	0.692	0.621

GUAYAQUIL 2000

Fecha: 2000-06-08

Sensor: Landsat 5 TM

Puntos usados:

Train (587) 84%

Test (255) 85%

12 variables seleccionadas:

SR_B1,SR_B2,SR_B3,SR_B4,S
R_B5,SR_B7,NDVI,MGRVI,CIVE
,NDBI,BUI,MNDWI

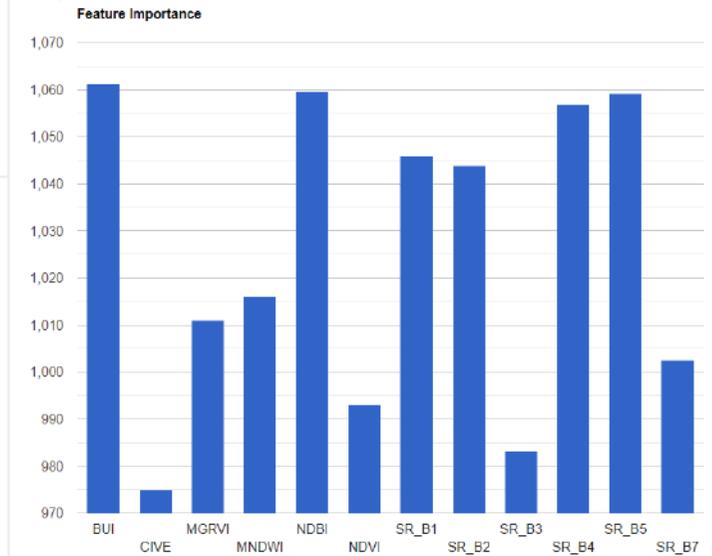
Landsat 5 usada



RFE de variables



Importancia de variables



Métricas de clasificación

1. BOSQUE, 2. CUERPO DE AGUA, 3.TIERRA AGROPECUARIA, 4. VEGETACION ARBUSTIVA Y HERBACEA, 5.ZONA ANTROPICA

		Prediction				
		1	2	3	4	5
Real	1	51	0	21	0	1
	2	1	19	6	0	1
	3	10	2	115	0	3
	4	1	0	1	0	0
	5	0	1	11	0	11

	Acc	0.769	Usuario	Productor	FScore
Kappa	0.619	0.810	0.699	0.750	
		0.864	0.704	0.776	
		0.747	0.885	0.810	
		0.000	0.000	NaN	
		0.688	0.478	0.564	

GUAYAQUIL 2008

Fecha: 2008-11-29

Sensor: Landsat 7 TM

Puntos usados:

Train (565) 94%

Test (372) 93%

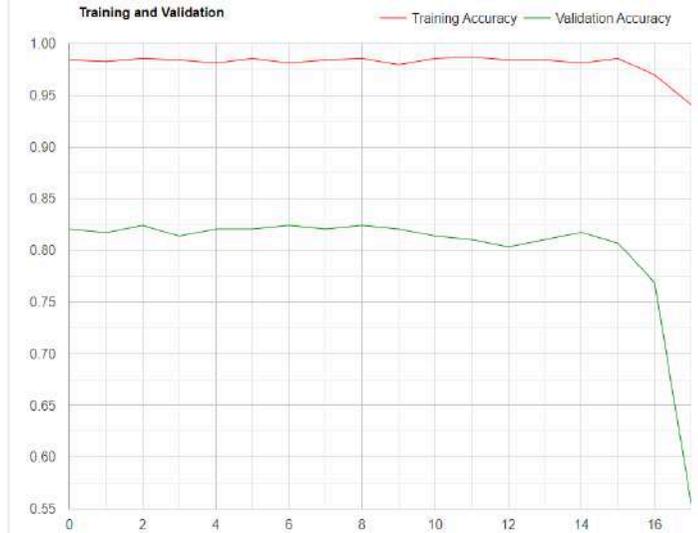
4 variables seleccionadas:

SR_B1,SR_B4,NDBI,MNDWI

Landsat 7 usada (relleno bandeado)



RFE de variables

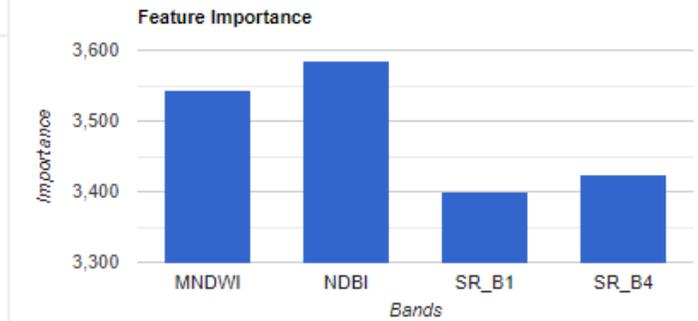


Métricas de clasificación

1. BOSQUE, 2. CUERPO DE AGUA, 3.TIERRA AGROPECUARIA, 4. VEGETACION ARBUSTIVA Y HERBACEA, 5.ZONA ANTROPICA

		Prediction					Acc	Kappa	Usuario	Productor	FScore
		1	2	3	4	5					
Real	1	41	1	20	0	0	0.803	0.649	0.759	0.661	0.707
	2	0	25	8	0	2			0.962	0.714	0.820
	3	13	0	153	0	5			0.810	0.895	0.850
	4	0	0	0	0	0			0.000	0.000	NaN
	5	0	0	8	0	14			0.667	0.636	0.651

Importancia de variables



GUAYAQUIL 2018

Fecha: 2019-05-07

Sensor: Sentinel 2 MSI

Puntos usados:

Train (625) 89%

Test (278) 93%

7 variables seleccionadas:

B5,B7,B11,B12,NDVI,NDBI,MNDWI

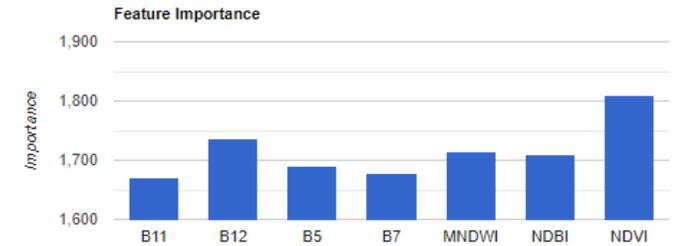
Imagen Sentinel 2 usada



RFE de variables



Importancia de variables



Métricas de clasificación

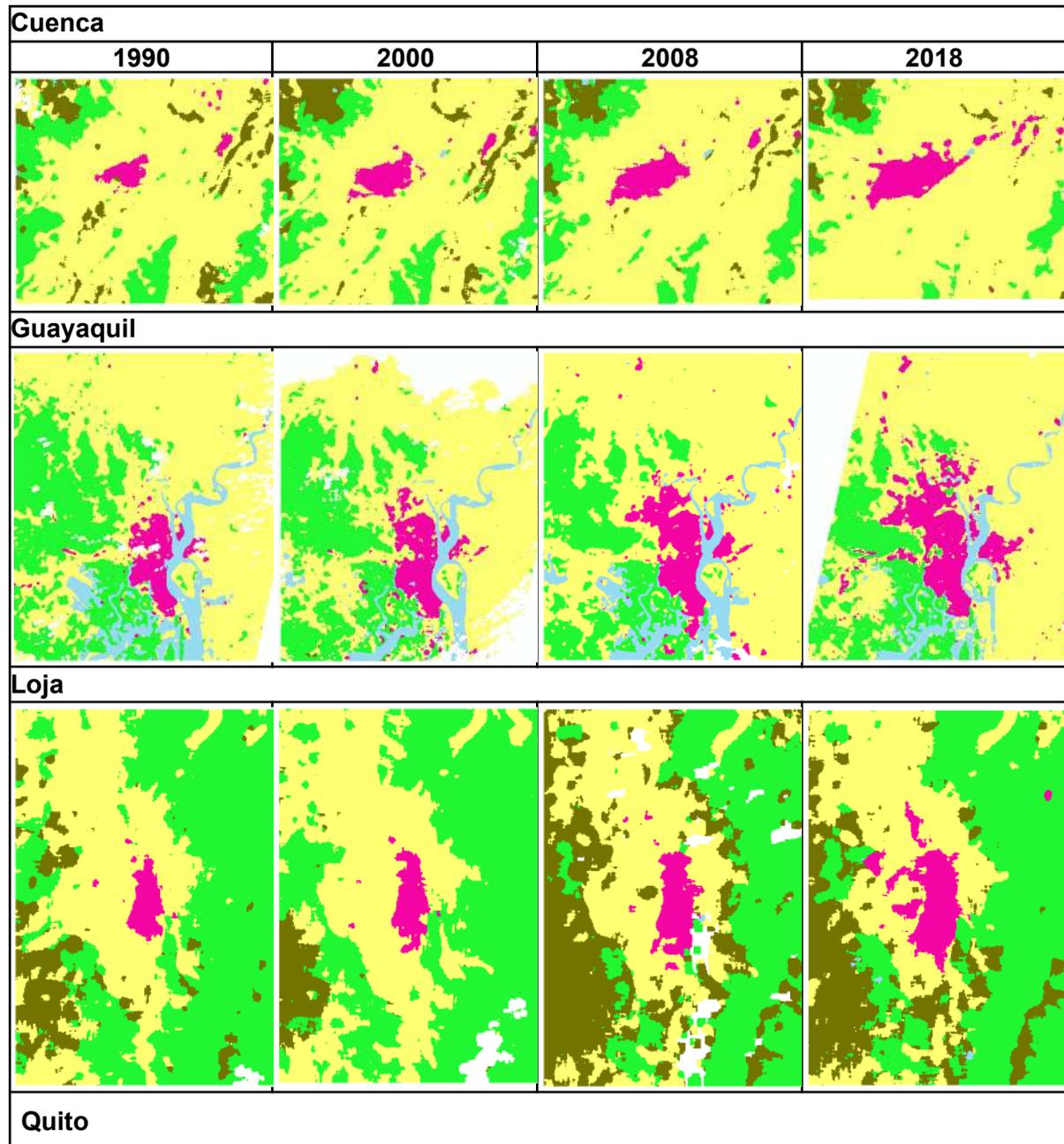
1. BOSQUE, 2. CUERPO DE AGUA, 3.TIERRA AGROPECUARIA, 4. VEGETACION ARBUSTIVA Y HERBACEA, 5.ZONA ANTROPICA

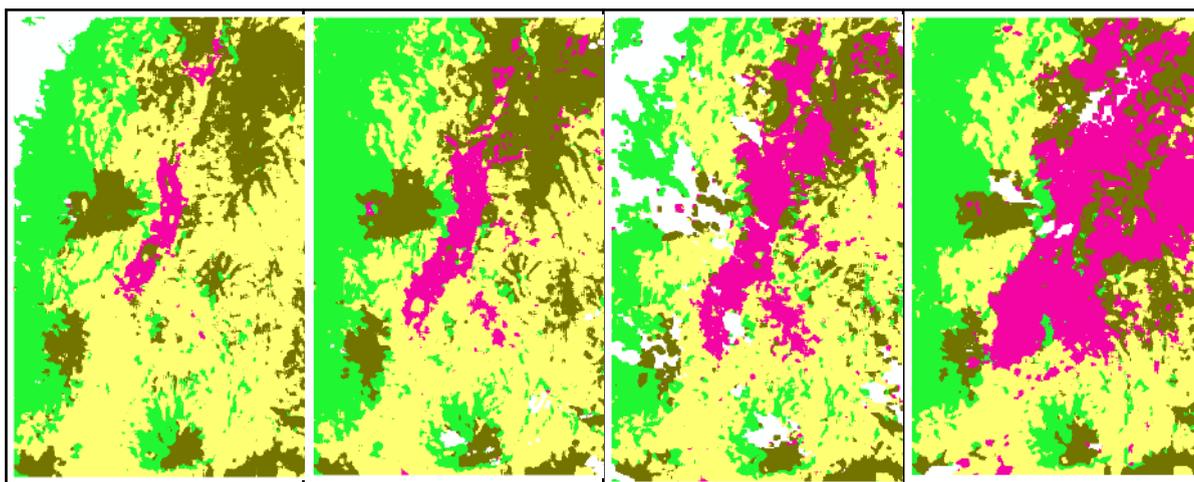
		Prediction					Acc 0.781	Usuario	Productor	FScore			
		1	2	3	4	5							
Real	1	37	0	6	0	0	Kappa 0.622	0.740	0.860	0.796			
	2	0	14	18	0	0					0.875	0.438	0.583
	3	8	2	141	0	11					0.801	0.870	0.834
	4	2	0	3	0	0					0.000	0.000	NaN
	5	3	0	8	0	25					0.694	0.694	0.694

1.5 Clasificación de imágenes con Smile Random Forest

Los resultados de la clasificación fueron colocados en la carpeta de Drive https://drive.google.com/drive/folders/1jFC7cqcxObLTe_87hifJrk4ifiJjifxb?usp=sharing. Las zonas que fueron recortadas por la nubosidad se encuentran vacías de información. Una vista previa de los mismos se presenta a continuación:

■ 1 BOSQUE
 ■ 2 AGUA
 ■ 3 TIERRA AGROPECUARIA
 ■ 4 VEGETACION ARBUSTIVA
 ■ 5 ZONA ANTROPICA/OTRAS TIERRAS





1.6 Métricas evaluadas (Accuracy, Usuario, Productor, F score)

Overall Accuracy: Es el porcentaje de píxeles clasificados correctamente en el conjunto de datos de validación (Deus, 2016). Es un único valor que representa la precisión de la clasificación. El cálculo de este valor se realiza comparando una muestra de ubicaciones en el mapa con las mismas ubicaciones en datos de referencia y verificando la cantidad de veces que existe concordancia (Congalton & Green, 2009). Se calcula simplemente sumando la diagonal principal (las unidades clasificadas correctamente) dividida por el número total de unidades de muestra de toda la matriz de confusión.

De acuerdo a Anderson et al. (1976) (citado en Congalton & Green, 2009), un límite de resultados aceptables e inaceptables en cuanto a la precisión es de 85 %, aunque se debe considerar que la precisión del mapa depende de muchos factores (por ej. nivel de detalle, variabilidad de categorías a cartografiar, entre otros).

Error (Confusion) Matrix: Según Congalton & Green (2009) esta matriz describe claramente las precisiones individuales de cada categoría junto con los errores de comisión y los de omisión presentes en la clasificación. El **error de comisión** es la inclusión de un área en una categoría cuando no pertenece a esta; mientras que el **error de omisión** excluye un área de la categoría a la que pertenece. Todos los errores son una omisión de la categoría correcta y una comisión de la categoría equivocada.

Junto con mostrar los errores antes descritos, la matriz de confusión permite calcular otras medidas (además de overall accuracy, descrita previamente) que se detallan a continuación:

Producer Accuracy: Se utiliza para estimar el error de omisión a una clase determinada y es la probabilidad de que un sitio de referencia se haya clasificado correctamente (Deus, 2016).

User Accuracy: Se utiliza para estimar el error de comisión y es la probabilidad de que un píxel clasificado en la imagen signifique la clase real sobre el terreno (Deus, 2016).

Ambas precisiones (producer & user) representan las precisiones de las categorías individuales en lugar de sólo la precisión de la clasificación global.

F-Score: También conocido como F1, combina la precisión de productor y usuario. Fusiona las métricas antes mencionadas mediante el cálculo de la media armónica entre precisión y recall (Schuster, et al. 2012), con lo cual se obtiene una mejor evaluación de las diferentes coberturas de suelo. Varía entre 0 (el peor resultado) y 1 (la mejor precisión que se puede alcanzar) (Deus, 2016).

$$F_1 \text{ score} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} = 2 \times \frac{\text{user's accuracy} \times \text{producer's accuracy}}{\text{user's accuracy} + \text{producer's accuracy}}$$

[*Enlace de acceso al libro de Congalton & Green](#)

1.7 Análisis de las métricas obtenidas

Para una mejor comprensión del análisis, es necesario revisar los aspectos teóricos detallados previamente. Además se adjunta la denominación de cada clase:

1 Bosque **2** Cuerpo de agua **3** Tierra agropecuaria **4** Vegetación arbustiva y herbácea
5 Zona antrópica.

CUENCA

Cuenca 1990: El valor de accuracy global (0.731) se ubica debajo del valor óptimo definido por Congalton & Green (2009), lo cual indica una clasificación poco precisa.

User accuracy: No se obtuvieron buenos valores de precisión para las clases 1 y 4 ya que ambos son menores a 0.65. Esto quiere decir que la clasificación de las muestras no refleja con exactitud la realidad en terreno. Las clases 3 y 5 presentan mejores valores de precisión (mayores a 0.75) aunque son menores al valor óptimo. La única clase que muestra una precisión óptima es la 2 con una precisión de 0.952.

Producer accuracy: La clase 4 no se está clasificando correctamente pues tiene un valor muy bajo de esta métrica (0.321). De manera similar, aunque con un poco más de precisión (con valores mayores a 0.62), las clases 1 y 5 presentan una clasificación poco precisa. Las únicas clases que cumplen con valores óptimos de esta métrica son la 2 (0.87) y 3 (0.866).

F-Score: Las clases 2 (0.909) y 3 (0.806) muestran un valor aceptable (mayor a 0.8) al analizar esta métrica; mientras que la clase 4 presenta un valor bajo de clasificación (0.429).

Cuenca 2000: El valor de accuracy global (0.688) se ubica debajo del valor óptimo definido por Congalton & Green (2009), lo cual indica una clasificación poco precisa.

User accuracy: No se obtuvieron buenos valores de precisión para las clases 1 y 4 ya que ambos son menores a 0.6. Esto quiere decir que la clasificación de las muestras no refleja con exactitud la realidad en terreno. La clase 3 presenta un mejor valor de precisión (0.722) aunque es menor al valor óptimo. Las clases que muestra una precisión aceptable por su cercanía al valor óptimo son la 2 (0.84) y la 5 (0.83)

Producer accuracy: La clase 4 no se está clasificando correctamente pues tiene un valor muy bajo de esta métrica (0.359). De manera similar, aunque con un poco más de precisión (con valores mayores a 0.6), las clases 1 y 5 presentan una clasificación poco precisa. Las únicas clases que cumplen con valores aceptables para esta métrica son la 2 (clasificación óptima con 0.875) y 3 (0.837).

F-Score: La clase 2 muestra un buen valor de clasificación al analizar esta métrica (0.857); mientras que la clase 4 presenta un valor bajo de clasificación (0.434).

Cuenca 2008: El valor de accuracy global (0.748), aunque se ubica debajo del valor óptimo definido por Congalton & Green (2009), presenta cierta mejora con respecto a la precisión de los periodos anteriores.

User accuracy: No se obtuvieron buenos valores de precisión para las clases 1 y 4 ya que ambos son menores a 0.62. Esto quiere decir que la clasificación de las muestras no refleja con exactitud la realidad en terreno. Las clases 3 y 5 presentan un mejor valor de precisión (0.781 y 0.73 respectivamente) aunque estos son menores al valor óptimo. La única clase que muestra una precisión idónea es la 2 (0.955).

Producer accuracy: La clase 4 no se está clasificando correctamente pues tiene un valor muy bajo de esta métrica (0.302). De manera similar, aunque con un poco más de precisión, la clase 1 (0.656) presenta una clasificación poco precisa. Las únicas clases que cumplen con valores óptimos para esta métrica son la 2 (0.875) y 3 (0.865).

F-Score: La clase 2 muestra un valor idóneo de clasificación al analizar esta métrica (0.913); mientras que la clase 4 presenta un valor bajo de clasificación (0.405).

Cuenca 2018: El valor de accuracy global (0.776), aunque se ubica debajo del valor óptimo definido por Congalton & Green (2009), es superior al de las clasificaciones de periodos de tiempo anteriormente analizados para esta zona.

User accuracy: No se obtuvieron buenos valores de precisión para las clase 4 (0.55) ni para la clase 1, aunque en esta, la precisión ha mejorado (0.622). Estos valores indican que la clasificación de las muestras no refleja con exactitud la realidad en terreno. Las clases 3 y 5 presentan un mejor valor de precisión (0.799 y 0.829) los cuales están próximos al valor óptimo. La única clase que muestra una precisión ideal es la 2 (0.95).

Producer accuracy: La clase 4 no se está clasificando correctamente pues tiene un valor extremadamente bajo de esta métrica (0.268). De manera similar, aunque con un poco más de precisión, la clase 1 (0.583) presenta una clasificación poco precisa. Las únicas clases que cumplen con valores idóneos para esta métrica son la 2 (0.864) y 3 (0.899). La clase 5 presenta un valor aceptable de precisión (0.81), aunque es menor al valor ideal establecido.

F-Score: La clase 2 muestra un valor idóneo de clasificación (0.905) al analizar esta métrica; mientras que la clase 4 presenta un valor bajo (0.361) de clasificación.

Conclusiones (i): En el análisis de la zona de estudio correspondiente a Cuenca, existe cierta variabilidad en los niveles de precisión analizados; no obstante se ha observado un patrón establecido que muestra que la clase 4 se clasifica incorrectamente en todos los periodos; al contrario de la clase 2, que presenta una clasificación óptima en cada periodo analizado.

QUITO

Quito 1990: El valor de accuracy global (0.64), se ubica debajo del valor óptimo definido por Congalton & Green (2009) e indica falencias en la clasificación de coberturas.

User accuracy: No se obtuvieron buenos valores de precisión para la clase 4 (0.30) Para las clases 1, 3 y 5 la precisión mejoró (0.71; 0.73 y 0.83) valores cercanos al óptimo. Los valores antes descritos indican que la clasificación de las muestras refleja baja exactitud la realidad en terreno. La clase 2 presenta un valor de 0, con lo cual se infiere la falta de muestras de cuerpos de agua que impiden su clasificación.

Producer accuracy: La clase 5 muestra una clasificación de un valor bajo (0.41). La clase 4 tampoco se está clasificando correctamente pues tiene un valor (0.66) al igual que la clase 1 (0.64) y la clase 3 presenta un valor (0.66).; mientras que la clase 2, tal como en el análisis anterior, presenta un valor de 0.

F-Score: Analizando esta métrica, se puede decir que ninguna de las clases muestra un valor idóneo de clasificación; no obstante, la clase 3 presenta el valor (0.70) y la clase 5 presenta el valor bajo (0.55) de clasificación. No se ha considerado a la clase 2, ya que no ha producido resultados para este periodo.

Quito 2000: El valor de accuracy global (0.68), se ubica debajo del valor óptimo definido por Congalton & Green (2009) e indica falencias en la clasificación de coberturas.

User accuracy: No se ha obtenido un valor adecuado de precisión para la clase 4 (0.36), pues refleja que la clasificación de las muestras no representa con exactitud la realidad en terreno. Las clases 3 y 5 presentan mejores valores de precisión (0.077 y 0.60 respectivamente) aunque son inferiores al óptimo. En este periodo de tiempo la clase que muestra una buena precisión es la 1, con una precisión de 0.82; mientras que la clase 2, al igual que en el periodo anterior, presenta un valor de 0, con lo cual se infiere la falta de muestras de cuerpos de agua que impiden su clasificación.

Producer accuracy: Las clases 3; 4 y 5 muestran una clasificación de precisión bajo (0,60; 0.73 y 0.58). Las clases 1 presentan un valor aceptable de clasificación (0.77); mientras que la clase 2, tal como en el análisis anterior, presenta un valor de 0.

F-Score: Esta métrica refleja que la clase 1 se clasifica aceptablemente, ya que presenta el valor más alto (0.80), al contrario de la clase 4 que presenta un valor bajo (0.49) de clasificación para esta métrica. No se ha considerado a la clase 2, pues no ha producido resultados para este periodo.

Quito 2008: El valor de accuracy global (0.61), se ubica debajo del valor óptimo definido por Congalton & Green (2009) e indica falencias en la clasificación de coberturas, además de ser el valor de precisión global más bajo de todos los periodos analizados en esta zona.

User accuracy: No se ha obtenido un valor adecuado de precisión para la clase 4 (0.45), pues su valor bajo refleja que la clasificación de las muestras no representa con exactitud la realidad en terreno. La clase 3 también presenta un valor bajo de clasificación, aunque un tanto superior al anterior (0.59). Las clases 1 y 5 presentan mejores valores de precisión (0.706 y 0.70) aunque son inferiores al óptimo. La clase 2, al igual que en los periodos anteriores, presenta un valor de 0, con lo cual se infiere la falta de muestras de cuerpos de agua que impiden su clasificación.

Producer accuracy: Las clases 4 y 5 muestran una clasificación defectuosa reflejada en valores de precisión bajos (0.352 y 0.571 respectivamente); la clase 1 un valor de clasificación no óptima, aunque algo superior a los de las clases antes descritas (0.762). La clase 3 presenta un valor bajo clasificación (0.68). La clase 2, tal como en el análisis anterior, presenta un valor de 0.

F-Score: Esta métrica refleja que la clase 1 se clasifica aceptablemente, ya que presentan un valor mayor a 0.7 y la clase 3 (0.636) un valor de clasificación baja. La clase 4 presenta el valor más bajo de clasificación para esta métrica (0.353) . No se ha considerado a la clase 2, pues no ha producido resultados para este periodo.

Quito 2018: El valor de accuracy global (0.738), se ubica debajo del valor óptimo definido por Congalton & Green (2009) e indica falencias en la clasificación de coberturas. **User accuracy:** No se ha obtenido un valor adecuado de precisión para la clase 4 (0.553), el cual es bajo y refleja que la clasificación de las muestras no representa con exactitud la realidad en terreno. Las clases 3 y 5 presentan valores bajos de precisión (0.69 y 0.774) aunque son inferiores al óptimo. En este periodo de tiempo la clase que muestra una mejor precisión es la 1, con un valor 0.86 (valor ideal); mientras que la clase 2, al igual que en los periodos anteriores, presenta un valor de 0, con lo cual se infiere la falta de muestras de cuerpos de agua que impiden su clasificación.

Producer accuracy: Las clases 4 y 5 muestran una clasificación defectuosa reflejada en valores de precisión bajos (0.520 y 0.765). La clase 3 presenta un valor aceptable de clasificación (0.76) y la clase 1 un valor casi óptimo de clasificación (0.819). La clase 2, tal como en el análisis anterior, presenta un valor de 0.

F-Score: Esta métrica refleja que la clase 1 se clasifica aceptablemente, ya que presenta el valor más alto (0.84). La clase 4 presenta un valor bajo (0.536) de clasificación para esta métrica. No se ha considerado a la clase 2, pues no ha producido resultados para este periodo.

Conclusiones (ii): En el análisis de la zona de estudio correspondiente a Quito y con base en la mayoría de periodos analizados, se observa que la clase 4 presenta una clasificación defectuosa; mientras que la clase 1 presenta aceptables valores de clasificación.

LOJA

Loja 1990: El valor de accuracy global (0.797), se ubica debajo del valor óptimo definido por Congalton & Green (2009) e indica una clasificación aceptable aunque con imprecisiones.

User accuracy: No se obtuvieron buenos valores de precisión para las clases 3 (0.682) y 4 (0.514). Los valores antes descritos indican que la clasificación de las muestras no refleja con exactitud la realidad en terreno. Para las clases 1 y 5 la precisión muestra valores óptimos para la clasificación (0.887 y 0.98). La clase 2 presenta un valor de 0.667, con lo cual se infiere la falta de muestras de cuerpos de agua que impiden su clasificación.

Producer accuracy: La clase 4 muestra una clasificación muy defectuosa reflejada en un valor extremadamente bajo (0.514). Las clases 1 y 5 muestran un valor de clasificación aceptable (0.853 y 0.962 respectivamente). La única clase que muestra una clasificación óptima es la 3 (0.806); mientras que la clase 2, presenta un valor de 0.33.

F-Score: Las clases 1 y 5 muestran valores aceptables de clasificación con un valor muy cercano al óptimo (0.87 y 0.97); la clase 3 tiene un valor ligeramente más bajo (0.739), mientras que las clases 2 y 4 presentan los valores más bajos de clasificación (0.444 y 0.487).

Loja 2000: El valor de accuracy global (0.736), se ubica debajo del valor óptimo definido por Congalton & Green (2009) e indica una clasificación aceptable aunque con imprecisiones.

User accuracy: No se obtuvieron buenos valores de precisión para las clases 2 (0.333), 3 (0.671) y 4 (0.533). Los valores antes descritos indican que la clasificación de las muestras no refleja con exactitud la realidad en terreno. La clase 1 presenta un valor aceptable de precisión (0.746). La clase 5 muestra valores óptimos de clasificación (0.979).

Producer accuracy: Las clases 2 y 4 muestran una clasificación muy defectuosa reflejada en un valor extremadamente bajo (0.125 y 0.32). La clase 3 muestra un valor de clasificación aceptable (0.79). Las clases 1 y 5 muestran una clasificación óptima (0.855 y 0.904).

F-Score: La clase 5 muestra un valor de clasificación muy cercano al óptimo (0.94); las clases 1 y 3 presentan valores aceptables (0.797 y 0.726), mientras que las clases 2 y 4 presentan los valores más bajos de clasificación (0.182 y 0.4).

Loja 2008: El valor de accuracy global (0.736), se ubica debajo del valor óptimo definido por Congalton & Green (2009) e indica una clasificación aceptable aunque con imprecisiones.

User accuracy: No se ha obtenido una buena precisión para la clase 4 (0.543). La clase 3 mejora el nivel de precisión (0.729); mientras que la clase 1 mejora aún más el nivel de precisión pero con un valor que no llega al óptimo (0.777). Los valores antes descritos, sobre todo de las clases 3 y 4 indican que la clasificación de las muestras no refleja con exactitud la realidad en terreno. Para la clase 5 la precisión muestra un valor óptimo de clasificación (0.960). La clase 2 presenta un valor de 0, con lo cual se infiere la falta de muestras de cuerpos de agua que impiden su clasificación.

Producer accuracy: Las clases 3 y 4 muestran una clasificación con valores bajos (0.699 y 0.655). La clase 1 presenta un valor aceptable (0.825). La clase 5 muestra un valor óptimo (0.857). La clase 2, tal como en el análisis anterior, presenta un valor de 0.

F-Score: La clase 5 presenta un óptimo valor de clasificación (0.906); mientras que la clase 4 presenta el valor más bajo de clasificación (0.594) que, sin embargo, presenta una mejora ostensible con respecto a los periodos anteriores analizados para esta zona. Las clases 1 y 3 tienen valores aceptables (0.80 y 0.713). No se ha considerado a la clase 2, ya que no ha producido resultados para este periodo.

Loja 2018: El valor de accuracy global (0.783), el cual es aceptable, pero se ubica debajo del valor óptimo definido por Congalton & Green (2009). En este caso, muestra la precisión más alta de todos los periodos analizados en esta zona de estudio.

User accuracy: La clase 4 muestra bajos niveles de precisión (0.576), lo que indica que la clasificación de las muestras no refleja con exactitud la realidad en terreno. La clase 3 tiene un valor aceptable (0.75). En las clases 1 y 5 la precisión muestra valores óptimos de clasificación (0.842 y 0.915). La clase 2 presenta un valor de 0, con lo cual se infiere la falta de muestras de cuerpos de agua que impiden su clasificación.

Producer accuracy: La clase 4 muestra una clasificación defectuosa reflejada en un valor bajo (0.596). Las clases 1 y 5 muestran valores de clasificación óptimos (0.85 y 0.915), mientras que la clase 3 muestra un valor aceptable (0.773). La clase 2, tal como en el análisis anterior, presenta un valor de 0.

F-Score: Las clases 1 y 5 presentan un óptimo valor de clasificación (0.846 y 0.915); la clase 3 muestra un valor aceptable (0.761), mientras que la clase 4 presenta el valor más bajo (0.586). No se ha considerado a la clase 2, ya que no ha producido resultados para este periodo.

Conclusiones (iii): En el análisis de la zona de estudio correspondiente a Loja, se observa que, en todos los periodos, la clase 4 presenta una clasificación defectuosa; mientras que las clases 1 y 5 presentan buenos valores de clasificación.

GUAYAQUIL

Guayaquil 1990: El valor de accuracy global (0.692), se ubica debajo del valor óptimo definido por Congalton & Green (2009) e indica una clasificación poco precisa.

User accuracy: No se ha obtenido una buena precisión para las clases 1 y 5 (0.641 y 0.563). La clase 3 mejora el nivel de precisión, con un valor aceptable (0.703); mientras que la clase 2 presenta valor próximo al nivel óptimo (0.828). Los valores antes descritos, específicamente para las clases 1 y 5, indican que la clasificación de las muestras no refleja con exactitud la realidad en terreno. La clase 4 presenta un valor de 0, con lo cual se infiere la falta de muestras de vegetación arbustiva y herbácea que impiden su clasificación.

Producer accuracy: La clase 2 muestra una clasificación poco precisa reflejada en un valor bajo (0.649). Las clases 1 y 5 muestran valores ligeramente superiores a la clase 2, reflejando imprecisión en la clasificación (0.667 y 0.692); mientras que la clase 3 muestra un valor aceptable (0.791). La clase 4, como en el análisis anterior, presenta un valor de 0.

F-Score: La clase 3 presenta el valor de clasificación más alto (0.745); mientras que la clase 5 presenta el valor más bajo (0.621). No se ha considerado a la clase 4, ya que no ha producido resultados para este periodo.

Guayaquil 2000: El valor de accuracy global (0.769), se ubica debajo del valor óptimo definido por Congalton & Green (2009) e indica una clasificación aceptable pero con imprecisiones.

User accuracy: Las clases 1 y 2 presentan niveles de precisión con valores próximos al nivel óptimo (0.810 y 0.864); mientras que la clase 3 presenta un nivel aceptable de precisión (0.747). El valor obtenido para la clase 5 refleja imprecisión en la clasificación (0.688). La clase 4 presenta un valor de 0, con lo cual se infiere la falta de muestras de vegetación arbustiva y herbácea que impiden su clasificación.

Producer accuracy: La clase 3 muestra un valor muy cercano al óptimo (0.885). Los valores obtenidos para las clases 1 y 2 reflejan ciertos niveles de imprecisión en la clasificación. El valor más bajo se obtuvo para la clase 5, lo que es indicativo de una clasificación defectuosa (0.478). La clase 4, como en el análisis anterior, presenta un valor de 0.

F-Score: La clase 3 presenta el valor de clasificación más alto (0.810); mientras que la clase 5 presenta el valor más bajo (0.564). No se ha considerado a la clase 4, ya que no ha producido resultados para este periodo.

Guayaquil 2008: El valor de accuracy global (0.803) es bastante aceptable, aunque se ubica debajo del valor óptimo definido por Congalton & Green (2009). Este es el valor más alto de precisión global de todos los periodos analizados en esta zona.

User accuracy: Las clases 2 y 3 presentan los mayores valores, reflejando niveles de precisión casi óptimos (0.962 y 0.810). La clase 1 presenta un nivel bastante aceptable de precisión en la clasificación (0.759); el valor más bajo se obtuvo para la clase 5, reflejando imprecisión en la clasificación (0.667). La clase 4 presenta un valor de 0, con lo cual se infiere la falta de muestras de vegetación arbustiva y herbácea que impiden su clasificación.

Producer accuracy: Las clases 1 y 5 muestran una clasificación muy imprecisa reflejada en sus valores bajos (0.661 y 0.636). La clasificación mejora para la clase 2 con un nivel aceptable (0.714), mientras que la clase 3 muestra un valor casi óptimo (0.895). La clase 4, como en el análisis anterior, presenta un valor de 0.

F-Score: La clase 3 presenta el valor de clasificación más alto (0.850); mientras que la clase 5 presenta el valor más bajo (0.651) de clasificación. No se ha considerado a la clase 4, ya que no ha producido resultados para este periodo.

Guayaquil 2018: El valor de accuracy global (0.781), se ubica debajo del valor óptimo definido por Congalton & Green (2009). Si bien este valor podría considerarse como aceptable, refleja imprecisiones en la clasificación.

User accuracy: Se ha obtenido una buena precisión para la clase 1 (0.740); mientras que para las clases 2 y 3, la precisión mejora aún más alcanzando niveles casi óptimos (0.875 y 0.801). El valor más bajo se obtuvo para la clase 5 (0.694). La clase 4 presenta un valor de 0, con lo cual se infiere la falta de muestras de vegetación arbustiva y herbácea que impiden su clasificación.

Producer accuracy: La clase 2 muestra una clasificación defectuosa reflejada en su valor bajo (0.438). La clasificación mejora para las clases 1 y 3 con valores casi óptimos (0.860 y 0.870), mientras que la clase 5 muestra un valor bajo, aunque cercano a términos aceptables (0.694). La clase 4, como en el análisis anterior, presenta un valor de 0.

F-Score: La clase 3 presenta el valor de clasificación más alto (0.834); mientras que la clase 2 presenta el valor más bajo (0.583) de clasificación. No se ha considerado a la clase 4, ya que no ha producido resultados para este periodo.

Conclusiones (iv): En el análisis de la zona de estudio correspondiente a Guayaquil, se aprecia variabilidad en los niveles de precisión para los diferentes periodos considerados, principalmente en las clases 2 y 3. Los valores más bajos de clasificación se encuentran en la clase 5.

Medidas posibles a tomar para mejorar la precisión de la clasificación: Con base en la literatura revisada, para mejorar la precisión de la clasificación se podría considerar, de forma general: (i) Aumentar la cantidad de puntos en las muestras; (ii) Mejorar la resolución de las imágenes (difícil por falta de disponibilidad); (iii) Probar diferentes porcentajes de equilibrio para datos de entrenamiento y validación (70% - 30% u 80% - 20%)

Elaborado por:

Ing.(MSc) Diego Pacheco P

Ing.(MSc) Paul Bravo López

María Eugenia González Calle.

Arq. Ma. Eugenia González

Aprobado por:

Ing.(MSc) Ivonne González
Loja

Dra. Teresa Pérez de Murzi
Guayaquil

Dra. Monserrath Mejía S
Quito

Ing.(MSc) Omar Delgado I
**Cuenca - Director del
proyecto**