

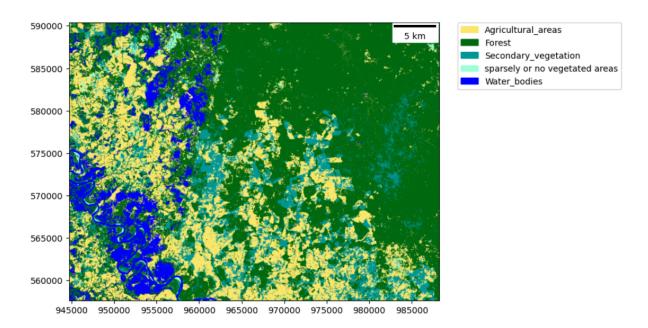
GEOPROCESAMIENTO - 2028638

Título: Workshop 3. Image Classification

Presentado por: María Fernanda López-Suárez¹, Megan García Antivar²

05/06/2025

Resultados obtenidos con Random forest



Visualmente se puede apreciar que hay claros errores en la categoría de agua. Sin embargo, a continuación se analizan las métricas obtenidas:

El clasificador Random Forest aplicado sobre la zona de estudio demostró un desempeño sólido, logrando una exactitud global del 89.77%. A continuación se analizan los principales resultados por clase, así como las métricas globales e implicaciones espaciales del modelo.

¹ malopezsu@unal.edu.co. Especialización en Análisis Espacial, Universidad Nacional de Colombia.

 $^{^2\} megarciaan @unal.edu.co.\ Especializaci\'on\ en\ An\'alisis\ Espacial, Universidad\ Nacional\ de\ Colombia.$

Desempeño por clase

Agricultural areas:

o Precision: 90.0%

o Recall (Sensibilidad): 94.7%

o F1-Score: 92.3%

Esta clase fue clasificada con alta precisión y mínima omisión, indicando que el modelo reconoce correctamente la mayoría de las zonas agrícolas y comete pocos falsos positivos.

• Forest:

Precision: 92.9% Recall: 72.2% F1-Score: 81.3%

Aunque la precisión es alta, el recall es más bajo, lo que sugiere que algunas zonas boscosas fueron clasificadas como otras coberturas. Esto puede deberse a la heterogeneidad espectral o a la confusión con vegetación secundaria.

Secondary vegetation:

Precision: 92.9% Recall: 86.7% F1-Score: 89.7%

Esta clase fue bien detectada, aunque existe una leve sobreestimación (falsos positivos). Su espectro similar al del bosque puede explicar las confusiones.

Water bodies:

Precision: 81.0% Recall: 94.4% F1-Score: 87.2%

El modelo es muy sensible a cuerpos de agua, detectándolos con facilidad, pero también muestra propensión a sobreclasificarlos (falsos positivos, especialmente en sombras o zonas húmedas).

Sparsely or no vegetated areas:

o Precision: 94.7%

Recall: 100%

o F1-Score: 97.3%

Esta clase muestra un excelente desempeño, sin omisiones y con mínima comisión, lo que indica una alta separación espectral respecto a las demás coberturas.

Análisis de la matriz de confusión

La matriz evidencia que:

- Hay leves confusiones entre Agricultural areas y Secondary vegetation.
- Se observa cierta sobreestimación de Water bodies, como lo indica la baja precisión relativa.
- La diagonal es dominante, reflejando un modelo robusto con mayoría de clasificaciones correctas.

Métricas globales

- Exactitud global: 89.77%
- Kappa de Cohen: 0.87 → Excelente acuerdo más allá del azar.
- Balanced Accuracy: 89.61% → Desempeño equitativo entre clases.
- MCC (Matthews Correlation Coefficient): 0.87 → Alta correlación entre clases reales y predichas.
- **Hamming Loss**: 0.10 → Solo el 10% de las etiquetas fueron incorrectas.
- Micro average: Calculado directamente como la exactitud global.
- Macro average:

Precision: 90.3% Recall: 89.6% F1-score: 89.5%

Promedio no ponderado que indica un rendimiento balanceado entre clases, incluso las menos representadas.

Weighted average:

Precision: 90.2% Recall: 89.8% F1-score: 89.6%

Refleja el rendimiento considerando el peso de cada clase en el total de muestras.

Área estimada por clase (en m²)

• Forest: 905,010,198.48

Agricultural areas: 348,875,167.60Secondary vegetation: 92,591,790.01

• Water bodies: 65,366,841.69

Sparsely/no vegetated areas: 14,621,699.27

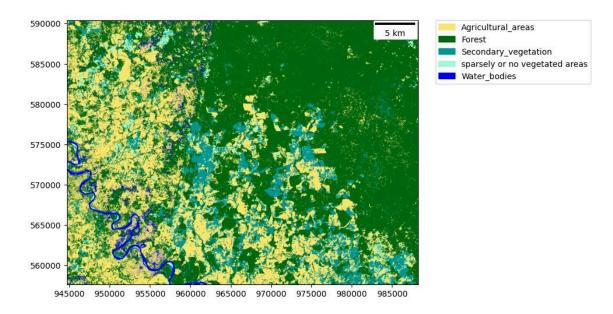
Esto muestra que el bosque domina el paisaje, seguido de áreas agrícolas. La vegetación secundaria y cuerpos de agua también tienen representatividad considerable.

Conclusión

El modelo Random Forest ofrece una clasificación precisa y confiable para el área de estudio, con excelente desempeño en la mayoría de las clases. Las diferencias entre precisión y recall, especialmente en Forest y Water bodies, pueden guiar mejoras mediante:

- Posclasificación para corregir confusiones comunes.
- Refinamiento del conjunto de entrenamiento.
- Evaluación de modelos complementarios.

Resultados obtenidos con Gradient Boosting



Visualmente se puede apreciar un mejor resultado en la clasificación de todas las categorías. Sin embargo, a continuación se analizan las métricas obtenidas:

El modelo Gradient Boosting aplicado sobre el conjunto de validación ha mostrado un desempeño robusto, alcanzando una exactitud global del 84.99% y un coeficiente Kappa de 0.8158, lo que indica un acuerdo sustancial entre predicción y referencia.

Desempeño por clase

• Clase 1: Agricultural_areas

Precision: 75% Recall: 98% F1-score: 85%

Esta clase es identificada con gran sensibilidad (recall), aunque su precisión es menor, lo que implica sobreclasificación o falsos positivos.

• Clase 2: Forest

Precision: 96% Recall: 81% F1-score: 88%

Muestra buen equilibrio, aunque con algunos casos no detectados (falsos negativos).

• Clase 3: Secondary_vegetation

Precision: 98% Recall: 90% F1-score: 94%

Excelente desempeño general; una de las clases mejor modeladas.

• Clase 4: sparsely or no vegetated areas

Precision: 92% Recall: 91% F1-score: 91%

Muy buen equilibrio entre precisión y sensibilidad.

• Clase 5: Water_bodies

Precision: 85% Recall: 84% F1-score: 84%

Rendimiento sólido y balanceado.

• Clase 6

o Precision: 76%

o Recall: 79%

F1-score: 78%

Clase con desempeño aceptable, aunque con margen de mejora respecto a la precisión.

Análisis de la matriz de confusión

- Se observan **ligeras confusiones** entre las clases 2, 3 y 4, lo que puede deberse a similitudes espectrales o zonas de transición.
- La diagonal está bien definida, lo que indica que la mayoría de los puntos fueron correctamente clasificados.
- La clase 0 no tiene datos (support = 0), por lo que no debe considerarse en el análisis.

Métricas globales

• Exactitud global: 84.99%

• Kappa de Cohen: 0.8158 → indica un alto grado de concordancia.

• Macro promedio:

Precision: 78% Recall: 74% F1-score: 75%

Promedio sin ponderar que sugiere variabilidad en el rendimiento entre clases.

Promedio ponderado (weighted avg):

Precision: 93% Recall: 85% F1-score: 89%

Al considerar la distribución real de clases, el modelo mantiene un rendimiento alto y consistente.

Conclusión

El clasificador **Gradient Boosting** se comporta de forma **altamente efectiva**, siendo especialmente fuerte en clases dominantes como las clases 3 y 4. El rendimiento es ligeramente inferior al observado con Random Forest (89.77% vs. 84.99%), aunque comparable.

Este modelo:

- Podría ser una alternativa adecuada si se busca mayor precisión en ciertas clases.
- Requiere una evaluación más profunda del tiempo de entrenamiento y recursos si se desea implementar operativamente.

Selección del mejor modelo para la clasificación final

Con base en las métricas de precisión obtenidas durante la evaluación, el modelo **Random Forest** es el que presenta el mejor desempeño general y, por tanto, se selecciona como el clasificador final para el área de estudio.

Resumen comparativo:

Métrica	Random Forest	Gradient Boosting
Exactitud global	89.77%	84.99%
Kappa de Cohen	0.87	0.82
F1 macro promedio	89.5%	75%
F1 ponderado promedio	89.6%	89%
Exactitud balanceada	89.6%	74%

Justificación:

- Random Forest logró una mayor exactitud global y una mejor distribución en el desempeño por clase.
- Mostró excelente precisión y recall en clases clave como Forest, Agricultural areas y Water bodies, con valores superiores al 90%.
- El valor de Kappa de 0.87 refleja un alto nivel de acuerdo entre los valores clasificados y los valores de referencia.
- Aunque Gradient Boosting también obtuvo buenos resultados, mostró menor recall en varias clases y una mayor disparidad entre clases, lo que afectó su balance general.

Conclusión:

Random Forest es el modelo más adecuado debido a su mayor rendimiento en las métricas clave, su solidez ante posibles sobreajustes y su consistencia en la clasificación de todas las coberturas. Por ello, se elige como clasificador final del proyecto.