

"Predicción de la Deforestación en Caldas, Colombia: Un Enfoque Sociodemográfico mediante el Uso de Modelos Machine Learning"

Alejandra Rivera Basto¹, Sebastian Pajoy Peña², y José John Fredy González Veloza³

^{1,2,3}Fundación Universitaria Los Libertadores

Resumen

Contexto. En el contexto de la creciente preocupación por la deforestación y sus impactos ambientales y sociales, esta investigación se centra en comprender las relaciones entre las características sociodemográficas y la deforestación en el departamento de Caldas, Colombia. **Propósito.** El propósito principal de este estudio es desarrollar un modelo de Machine Learning de regresión que permita predecir la deforestación en función de variables sociodemográficas, con el objetivo de facilitar la planificación territorial y la mitigación de la deforestación en la región. **Metodología.** La investigación se llevó a cabo mediante la recopilación de datos sociodemográficos del Departamento Nacional de Estadística (DANE) y registros satelitales de deforestación de la Universidad de Maryland. Estos datos se integraron para analizar la relación entre factores sociodemográficos y deforestación. Se utilizaron algoritmos de Machine Learning y herramientas computacionales para desarrollar y evaluar modelos de regresión. **Resultados.** Los resultados revelan relaciones significativas entre las variables sociodemográficas y la deforestación en Caldas. El modelo de Machine Learning logró prever la deforestación con precisión, proporcionando una valiosa herramienta para la planificación territorial basada en datos. **Conclusiones.** Según los resultados obtenidos, se resalta la importancia de implementar políticas de planificación territorial que tengan en cuenta ciertas características de la población. Estas características incluyen la edad, especialmente aquella superior a los 60 años, los niveles de formación educativa, considerando que las personas con un mayor grado de formación tienden a contribuir menos a la deforestación, y la ocupación de la población. Estos elementos se presentan como fundamentales en la gestión y mitigación de la deforestación en el departamento de Caldas.

Palabras clave— Deforestación, sociodemográfico, Machine Learning, Planificación territorial

Abstract

Context. In the context of growing concerns about deforestation and its environmental and social impacts, this research focuses on understanding the relationships between sociodemographic characteristics and deforestation in the department of Caldas, Colombia. **Purpose.** The main purpose of this study is to develop a regression-based Machine Learning model that predicts deforestation based on sociodemographic variables, aiming to facilitate territorial planning and deforestation mitigation in the region. **Methodology.** The research was conducted by collecting sociodemographic data from the National Department of Statistics (DANE) and satellite records of deforestation from the University of Maryland. These datasets were integrated to analyze the relationship between population and deforestation. Machine Learning algorithms and computational tools were used to develop and evaluate regression models. **Results.** The results reveal significant relationships between sociodemographic variables and deforestation in Caldas. The Machine Learning model accurately predicted deforestation, providing a valuable tool for data-driven territorial planning. **Conclusions.** According to these results, the need to implement territorial planning policies that take into account population characteristics becomes evident. These characteristics include age, especially those above 60 years old, educational levels, considering that individuals with higher education tend to contribute less to deforestation, and the occupation of the population. These elements are highlighted as key factors in the management and mitigation of deforestation in Caldas.

Keywords— Deforestation, Sociodemographic, Machine Learning, Territorial planning

1. Introducción

La deforestación es un fenómeno global que se refiere a la pérdida de bosques y su conversión en otros usos del suelo, como la agricultura, la ganadería o la minería. Este proceso tiene múltiples causas y efectos, y es una preocupación clave en la conservación ambiental y la sostenibilidad (Veldkamp y Lambin, 2001). La deforestación conlleva la pérdida de biodiversidad, la liberación de grandes cantidades de dióxido de carbono a la atmósfera, lo que contribuye al cambio climático, y

una serie de impactos sociales, incluyendo la afectación de comunidades locales que dependen de los bosques para su subsistencia (Burga Ríos, 2016). Se trata de un fenómeno históricamente atribuido a la actividad humana y representa un desafío ambiental crítico en diversas regiones del mundo.

En el contexto de Colombia, país de excepcional biodiversidad y ecosistemas únicos, la deforestación ha adquirido una dimensión preocupante (García, 2012). La transformación de tierras forestales en áreas agrícolas, la expansión de la frontera ganadera y la minería ilegal

han impulsado una pérdida acelerada de cobertura forestal en el país (González et al., 2011). Específicamente, el departamento de Caldas se ha visto afectado por este fenómeno, lo que amenaza no solo su entorno natural, sino también la calidad de vida de sus comunidades y su sostenibilidad a largo plazo.

En este contexto, este trabajo se enfoca en analizar las complejas relaciones entre las características sociodemográficas y la deforestación en el departamento de Caldas, Colombia. La importancia de esta investigación radica en su enfoque preventivo y en la integración de factores demográficos en la comprensión y mitigación de la deforestación. A través de la combinación de datos censales recolectados por el Departamento Nacional de Estadística en los años 2005 (DANE, 2005) y 2018 (DANE, 2018), así como registros satelitales de deforestación de la Universidad de Maryland Hansen et al. (2013), la cual fue obtenida por los autores, a través del análisis multitemporal de imágenes satelitales Landsat; con esto se busca identificar patrones y tendencias que vinculen variables demográficas con la pérdida de vegetación.

La literatura científica ha documentado ampliamente la deforestación y sus consecuencias a nivel global y en diversas regiones (Green et al., 2013). No obstante, en el contexto de Caldas, no se encuentran estudios que exploren las relaciones específicas entre las características sociodemográficas y este fenómeno. Este estudio se distingue al enfocarse en esta área poco explorada y al emplear una metodología avanzada que combina análisis de correlación con modelos de Machine Learning para obtener resultados más precisos y aplicables.

La evaluación del riesgo de deforestación demanda una comprensión de fenómenos complejos que involucran variables diversas e interrelacionadas. Si bien las técnicas estadísticas han sido tradicionalmente preferidas, las técnicas de Machine Learning están ganando popularidad (Mayfield, Smith, Gallagher, y Hockings, 2020). La relevancia de esta investigación radica en la metodología propuesta, que abarca análisis de correlación y la implementación de modelos de Machine Learning. La aplicación de técnicas avanzadas de análisis de datos no solo permite establecer conexiones significativas entre factores demográficos y deforestación, sino también desarrollar un modelo predictivo que visualice cómo estas variables interactúan en el tiempo. Esta herramienta no solo informará la toma de decisiones en entes de control territorial, sino que también sentará las bases para la formulación de políticas efectivas centradas en la prevención de la deforestación y la promoción de un desarrollo sostenible.

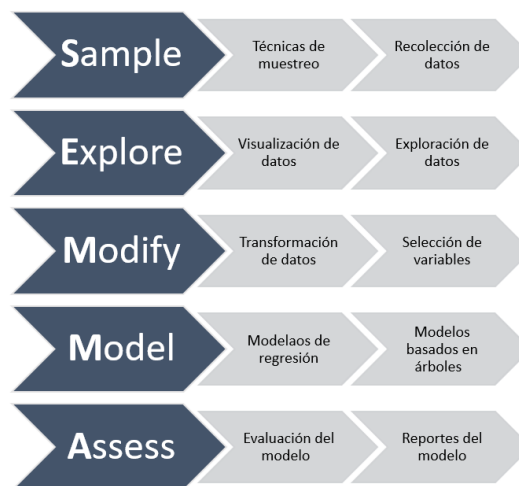


Figura 1: Esquema de la metodología SEMMA.

2. Metodología

En esta sección se presenta la metodología utilizada para llevar a cabo el análisis y la predicción de la deforestación en el departamento de Caldas, Colombia, mediante modelos de Machine Learning de regresión. Para alcanzar este objetivo, se emplea una metodología SEMMA que abarca desde la recolección y preparación de datos hasta la construcción y evaluación de los modelos de regresión. La figura 1 ilustra los pasos seguidos haciendo uso de esta metodología.

2.1. Recolección de datos

La primera etapa del proceso involucra la obtención de datos de múltiples fuentes. Se recopilieron datos sociodemográficos de la población de Caldas a partir de las fuentes proporcionadas por el DANE. Además, se utilizó información espacial obtenida de los datos de Cambio Global de Bosques 2000-2018, recopilados por Hansen (Hansen et al., 2013) a partir de imágenes multitemporales de Landsat.

2.2. Exploración y modificación de datos

Una vez recopilados, los datos fueron sometidos a un proceso integral de preparación y limpieza. Este proceso abarcó la eliminación de valores faltantes, la normalización de variables y la transformación de datos satelitales, asegurando así la idoneidad de los datos para su análisis y la construcción de modelos de regresión.

El objetivo principal de este proceso fue generar una base de datos que integrara información espacial con datos censales proporcionados por el Departamento Administrativo Nacional de Estadística (DANE) correspondientes a los años 2005 y 2018. Esto permitió la conjunción de datos sociodemográficos con información espacial relativa a la deforestación. Las encuestas realizadas por el DANE proporcionaron datos con diferentes estructuras. El Censo de 2005 organizó la información espacial por veredas, mientras que el Censo de 2018 ofreció información a nivel de hogares.

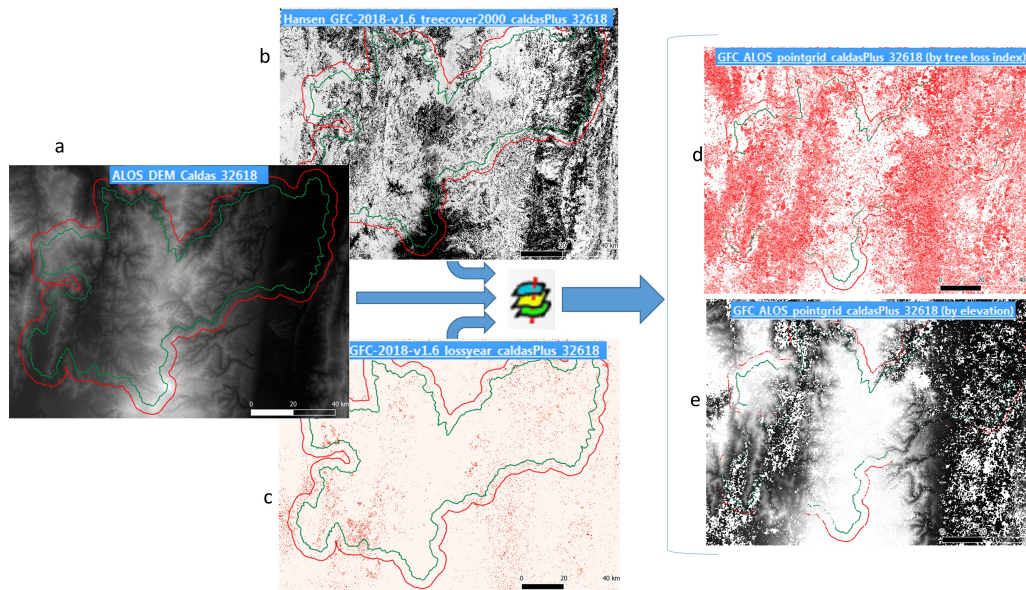


Figura 2: Integración de información Espacial de diferentes Fuentes. a. Información altura AW3D, JAXA (2022), b. Cobertura vegetal Hansen et al. (2013), c. pérdida de vegetación Hansen et al. (2013) d. grilla integrada por pérdida de vegetación, e. grilla integrada por elevación

Para superar esta disparidad y lograr una integración completa, se utilizaron los datos de la encuesta de 2005 como puntos de referencia o proyectores. Estos datos de 2005 se emplearon como base sobre la cual se proyectaron los cambios y actualizaciones que habían ocurrido en los datos de la encuesta de 2018.

En cuanto a los datos de deforestación, se obtuvieron a través del análisis multitemporal de imágenes satelitales proporcionadas por la Universidad de Maryland, desde el año 2005 hasta 2018 (disponibles hasta 2022). Estos datos presentaron una distribución espacial detallada, asignando un porcentaje de deforestación a cada píxel en la imagen satelital.

La integración de estos datos de deforestación con las encuestas se llevó a cabo mediante la generación de centroides y la creación de buffers (zonas circulares) en el municipio de Caldas, donde se desarrolló el estudio. Estos centroides y buffers sirvieron como puntos de referencia, estableciendo una relación espacial entre los datos de las encuestas y los niveles de deforestación en las áreas circundantes.

A continuación se detalla el procesos realizado:

Inicialmente, la información con la que se contaba presentaba formatos diferentes. Por un lado, las encuestas realizadas por el DANE proporcionaban datos geoespaciales, pero esta información estaba representada mediante centroides. Por otro lado, se disponía de imágenes satelitales procedentes de distintas fuentes, abarcando aspectos como la altitud, cuyos datos utilizados en este estudio fueron proporcionados por AW3D, una división de la Agencia de Exploración Aeroespacial de Japón (JAXA), la deforestación y la cobertura vegetal, proporcionadas por Hansen, (Universidad de Meryland. La tarea principal consistía en unificar esta información heterogénea para poder complementar los datos de las encuestas del DANE. Para llevar a cabo esta unificación, se tomaron todas las imágenes satelitales que se iban

a utilizar y se fusionaron utilizando la herramienta "point sampling tools" de QGIS. como se observa en la figura 2. Este proceso permitió vincular los datos geoespaciales de las encuestas con la valiosa información de las imágenes satelitales, creando así un conjunto de datos integrado y coherente que facilitaría el análisis y la comprensión de los patrones y tendencias relacionados con la deforestación y la cobertura vegetal en la región.

Los datos sociodemográficos proporcionados por el DANE, como se presenta en la Tabla1, se enriquecieron con información adicional sobre parques naturales y usos del suelo del departamento de Caldas, información que fue obtenida de la corporación autónoma regional de Caldas (Corporación Caldas) que se indica en la tabla2

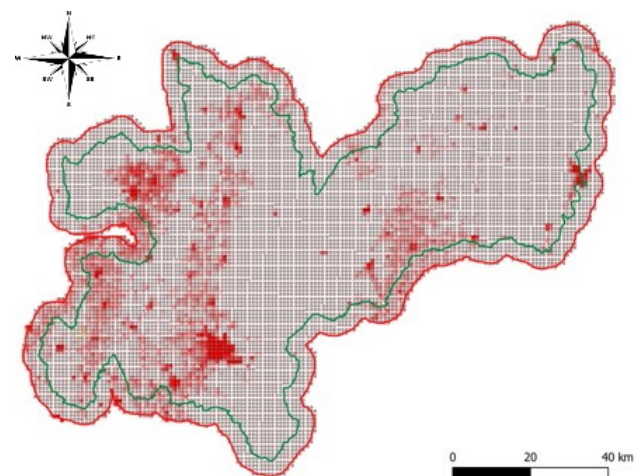


Figura 3: Integración Total de Datos Por Grilla en El Departamento del Caldas. Los puntos corresponden a los grillas base para construir los centroides km^2

Luego de combinar las imágenes satelitales, se procede a realizar un recorte de estas por el departamento de Caldas como se puede observar en la figura 3. Una

2005		2010	
Identificador Variable	Descripción de la variable	Identificador Variable	Descripción de la variable
p25b_sexo	sexo	VA1_ESTRATO	Estrato de la vivienda (según servicio de energía)
pc09b_edad	edad	VB_ACU	Cuenta con servicio de acueducto
p30b4_pais_5ano	país donde vivía hace 5 años	VC_ALC	Cuenta con servicio de alcantarillado
p30b5_ano_llego	en qué año llegó al país	VD_GAS	Cuenta con servicio de gas natural conectado a red pública
p31b_clase_5anos	clase de área donde vivía hace 5 años	VE_RECIBAS	Cuenta con servicio de recolección de basura
p41b1_alfabeta	sabe leer y escribir	VF_INTERNET	Cuenta con servicio de internet (fijo o móvil)
p44b1_tip_estud	nivel de estudios que aprobó	V_TIPO_SERSA	Tipo de servicio sanitario (inodoro)
p44b2_ult_ano	último año aprobado en el nivel de estudios		
p47b_ocupacion	condición de ocupación de la persona		

Tabla 1: Variable tomadas encuestas DANE 2005, 2018

cód. nivel 1	cód. nivel 2	Cobertura suelo 2010 nivel 1	Cobertura suelo 2010 nivel 2
1	11	Territorios Artificializados	Zonas Urbanizadas
1	12	Territorios Artificializados	Zonas Industriales-Comerciales-Redes Comunicación
1	13	Territorios Artificializados	Zonas de Extracción Mineras y Escombreras
1	14	Territorios Artificializados	Zonas Verdes Artificializadas No Agrícolas
2	21	Territorios Agrícolas	Cultivos Anuales y Transitorios
2	22	Territorios Agrícolas	Cultivos Permanentes
2	23	Territorios Agrícolas	Pastos
2	24	Territorios Agrícolas	Áreas Agrícolas Heterogéneas
3	31	Bosques y Áreas Seminaturales	Bosques
3	32	Bosques y Áreas Seminaturales	Áreas de Vegetación Herbácea y/o Arbustiva
3	33	Bosques y Áreas Seminaturales	Áreas Abiertas, sin o poca Vegetación
4	41	Áreas Húmedas	Áreas Húmedas Continentales
5	51	Superficies de Agua	Aguas Continentales

Tabla 2: Variables usos del suelo Tomados de la Corpocaldas

vez definidas estas divisiones, se crea una grilla espacial que servirá de base para la generación de centroides esto se puede observar en la figura 4. Estos centroides se utilizan para combinar la información sociodemográfica proporcionada por el DANE con los datos satelitales previamente obtenidos. El resultado es una base de datos integral que incluye información espacial, datos de encuestas y detalles sobre el uso del territorio.

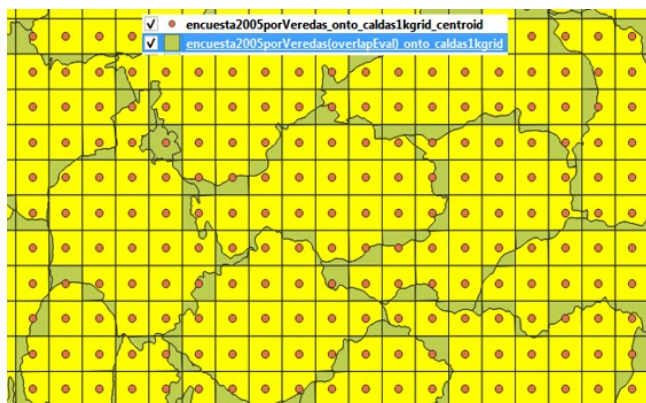


Figura 4: Construcción de Centroides

2.3. Modelo

En esta sección, se realizó la construcción y evaluación de modelos de Machine Learning con el objetivo de predecir la deforestación en el departamento de Caldas, Colombia, utilizando las características sociodemográficas

previamente seleccionadas. Se exploraron varios algoritmos de regresión, tales como el Random Forest Regressor y Light Gradient Boosting Machine, entre otros, con el propósito de identificar el modelo más eficaz para este estudio. La evaluación de los modelos se llevó a cabo de manera exhaustiva, empleando métricas pertinentes para la regresión, incluyendo el Error Absoluto Medio (MAE), el Error Cuadrado Medio (MSE), la Raíz Cuadrada del Error Cuadrado Medio (RMSE) y el Coeficiente de Determinación (R²). Los resultados de esta evaluación se presentan en la Tabla 3, que enumera los modelos probados junto con sus métricas correspondientes.

Como se puede observar en la tabla 3, Los resultados de esta evaluación se presentan en la Tabla 3, que enumera los modelos probados junto con sus métricas correspondientes. Como se puede observar en la tabla 3, se destaca el modelo Random Forest, el cual obtuvo las mejores métricas, con un MAE de 1.2760%, un MSE de 2.9540%², un RMSE de 1.7157% y un R² de 0.5156. Estos valores indican una alta precisión y capacidad de predicción del modelo en relación con la deforestación medida en porcentaje. En virtud de su excelente desempeño en las métricas evaluadas, se seleccionó el Random Forest Regressor como el modelo óptimo.

A continuación, se detalla el proceso de evaluación de este modelo con el fin de garantizar su robustez y capacidad de generalización

Modelo	MAE	MSE	RMSE	R2
Random Forest Regressor	1.2936 % ²	3.0651 %	1.7477 %	0.5156
Light Gradient Boosting Machine	1.3782 % ²	3.3755 %	1.8347 %	0.4659
Extra Trees Regressor	1.3453 % ²	3.4049 %	1.8418 %	0.4617
Gradient Boosting Regressor	1.6247 % ²	4.3165 %	2.0763 %	0.3172
K Neighbors Regressor	1.8094 % ²	5.4791 %	2.3385 %	0.1315
AdaBoost Regressor	1.8879 % ²	5.5718 %	2.3597 %	0.1180
Decision Tree Regressor	1.7034 % ²	5.7501 %	2.3943 %	0.0909

Tabla 3: Métricas de los Modelos de Regresión

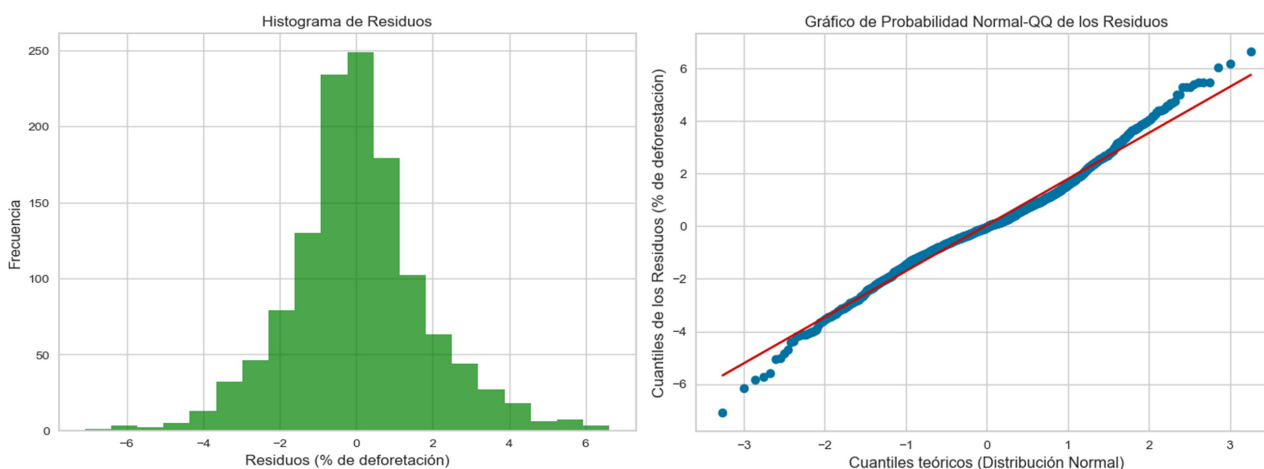


Figura 5: Resultados residuales Modelos Random Forest

2.4. Evaluación

En la etapa final del proceso de modelado, se realizó una evaluación exhaustiva del rendimiento del modelo seleccionado, que en este caso es el Random Forest Regressor. Esta evaluación se llevó a cabo utilizando un conjunto de datos de prueba que no se había utilizado durante el proceso de entrenamiento. El principal objetivo de esta evaluación es garantizar la confiabilidad y utilidad del modelo en la predicción de la deforestación en el departamento de Caldas, Colombia.

En el estudio realizado, se llevó a cabo la evaluación del rendimiento del modelo mediante la comparación de las predicciones generadas con los valores reales de deforestación en el conjunto de datos de prueba. Esta evaluación se basó en el empleo del Coeficiente de Determinación (R^2), métrica que permite medir la proporción de la variabilidad en los datos de prueba que puede ser explicada por las predicciones del modelo. Una vez seleccionado el modelo, se procedió a realizar pruebas adicionales enfocadas en los residuos generados por el mismo. Los residuos, que son las diferencias entre los valores reales y las predicciones del modelo, fueron sometidos a análisis para evaluar la idoneidad del modelo y verificar si cumplía con las suposiciones necesarias para realizar inferencias válidas. Estas pruebas incluyeron la evaluación de la normalidad de los residuos y otros aspectos relacionados con su comportamiento.

La combinación de la comparación entre las predicciones y los valores reales a través del uso del R^2 , junto con el análisis de los residuos, contribuyó a una evaluación exhaustiva del rendimiento y la confiabilidad del modelo en la predicción de la deforestación en el departamento de Caldas, Colombia. En la figura 5 se puede apreciar la distribución normal de los mismo lo que es un indicador de que la regresión lineal, y proporciona una base sólida para la interpretación de los resultados del modelo.

3. Resultados y discusión

Al examinar detenidamente Al realizar un examen detenido de la Figura 6, que ilustra la importancia de las variables en el modelo, se revela una conexión significativa entre las características demográficas y la deforestación en el departamento de Caldas, Colombia. En particular, se ha observado una relación notable entre las personas con educación secundaria completa y aquellas mayores de 60 años en el departamento de Caldas. Sin embargo, hasta este punto, no se ha definido si esta relación influye positivamente o negativamente en el aumento o disminución de la deforestación en la región. En este contexto, se procedió a realizar un análisis exhaustivo mediante la construcción y evaluación de dos modelos: el

modelo Random Forest y el modelo Ridge a través de un modelo de ensamble.

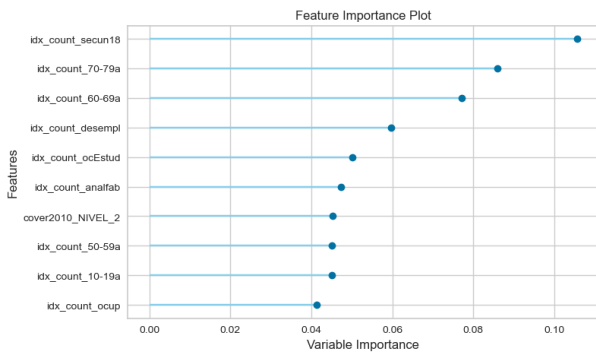


Figura 6: Variables que explican mejor la pérdida de vegetación.

Calcular la importancia promedio de las variables predictoras en un ensamble de modelos que combina un modelo Ridge y un modelo Random Forest, puede reducir la dependencia de un solo modelo y mejorar la robustez de las predicciones. Tanto el modelo Ridge como el modelo Random Forest emplean enfoques distintos para realizar predicciones. El modelo Ridge se basa en una regresión lineal regularizada, mientras que Random Forest utiliza árboles de decisión. La combinación de ambos modelos aprovecha sus respectivas fortalezas y compensa sus debilidades, lo que conduce a la reducción del sesgo y la varianza. El modelo Ridge, al aplicar penalizaciones a los coeficientes, tiende a generar predicciones menos variables pero puede tener un sesgo más pronunciado. Por otro lado, Random Forest puede capturar relaciones no lineales, aunque podría ser susceptible al sobreajuste. La conjunción de ambos modelos contribuye a lograr un equilibrio entre sesgo y varianza.

Calcular la importancia promedio de las variables predictoras en este ensamble se presenta como una técnica valiosa para identificar cuáles características ejercen un impacto más significativo en las predicciones finales. Esto proporciona una comprensión más profunda de las variables que influyen en el fenómeno de la deforestación. Como resultado de este ensamble, se obtiene una figura 7 que muestra qué variables estudiadas tienen un impacto positivo (aumentando la deforestación) y cuáles tienen un impacto negativo (disminuyendo la deforestación) en el Departamento de Caldas

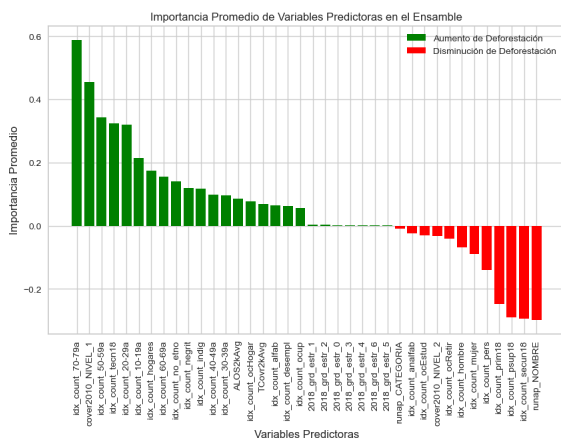


Figura 7: Importancia de variables en modelo.

Según la gráfica de ensamble, es evidente que los factores que más contribuyen al aumento de la deforestación son la población de 60 años o más, en particular, los grupos de edad de 70 a 79 años, seguidos por los de 60 a 69 años. Además, la presencia de usos del suelo clasificados como agrícolas también desempeña un papel significativo en el aumento de la deforestación.

Por otro lado, se observa que variables como la educación superior o posgradual y la presencia de reservas naturales están asociadas con una reducción en la deforestación en el departamento. Esta asociación tiene sentido, ya que las áreas protegidas suelen experimentar menos actividad de deforestación. Además, se demuestra que a medida que aumenta el nivel de formación de la población, se reduce el impacto en la deforestación. Esto es coherente con estudios realizados en latinoamérica, en donde se ha evidenciado que la educación del hogar reduce la probabilidad de tala en aproximadamente un 3% (Hübler, 2017). Estos hallazgos sugieren la importancia de implementar programas de educación por parte de las autoridades, ya que podrían tener un impacto positivo significativo en la reducción de la deforestación en el departamento de Caldas.

En lo que respecta al uso de modelos de machine learning, es fundamental destacar que el modelo generado en este estudio tiene la capacidad de realizar predicciones de probabilidad de deforestación en diferentes regiones. Esta funcionalidad representa una notable ventaja, ya que elimina la necesidad de depender exclusivamente de registros satelitales costosos o estudios extensos. En lugar de eso, se puede utilizar información demográfica fácilmente accesible para predecir el riesgo de deforestación en diversas áreas. Esto simplifica la recopilación de datos y reduce los costos asociados con la vigilancia de la deforestación.

La capacidad predictiva de este modelo es de gran alcance y puede tener un impacto significativo en la gestión de la deforestación. Por ejemplo, las organizaciones de conservación pueden utilizar esta herramienta para identificar áreas de alto riesgo y dirigir recursos de manera eficiente hacia la prevención de la deforestación. Además, el modelo respalda la planificación territorial al proporcionar información crucial sobre cómo las políticas y las intervenciones pueden influir en la conservación de los bosques.

Finalmente, un aspecto relevante del estudio es la transformación de datos satelitales en grillas, lo cual presenta ventajas significativas. Esta técnica permite la integración de datos satelitales con información demográfica y sociodemográfica, facilitando así diversos análisis y sentando un precedente en la región.

4. Conclusiones

Este estudio ha revelado conexiones intrigantes entre las variables sociodemográficas y la deforestación en el departamento de Caldas, Colombia. En particular, se destacan dos factores clave: la edad, con un énfasis en

aquellos mayores de 70 años, y los niveles de educación, especialmente en personas con educación máxima hasta secundaria. Estos resultados subrayan la importancia de comprender cómo estas características influyen en las actividades de deforestación.

Las personas de 70 a 79 años y aquellas con educación máxima hasta secundaria han demostrado tener una mayor propensión a contribuir a la deforestación. Este hallazgo plantea la necesidad de abordar estas poblaciones de manera efectiva para prevenir futuras actividades de deforestación. Al considerar estrategias de ocupación significativas y programas educativos adaptados a estas poblaciones, se puede no solo preservar los recursos naturales, sino también mejorar las condiciones económicas y sociales locales.

Además, el modelo de machine learning desarrollado en este estudio tiene la capacidad de predecir la deforestación en diferentes regiones, eliminando la dependencia de costosos registros satelitales. Esta herramienta es esencial para identificar áreas de alto riesgo y dirigir recursos de manera eficiente hacia la prevención de la deforestación y respalda la planificación territorial.

Finalmente, la transformación de datos satelitales en grillas ha demostrado ser una técnica valiosa para integrar datos y realizar análisis precisos. En conjunto, estos resultados ofrecen una perspectiva integral para abordar la deforestación y tienen aplicaciones prometedoras en la conservación y la gestión de recursos naturales.

Referencias

- AW3D, JAXA. (2022). *Aloa3d* (Inf. Téc.). Agencia de Exploración Aeroespacial de Japón (JAXA). Descargado de https://www.eorc.jaxa.jp/ALOS/en/dataset/aw3d30/aw3d30_e.htm
- Burga Ríos, M. (2016). Incremento de la deforestación y sus consecuencias en la pérdida de biomasa en los bosques de la provincia alto amazonas del departamento de Loreto, 2000-2014.
- DANE. (2005). General, censo and de convivencia, encuesta and ciudadana, seguridad. *Departamento Administrativo Nacional de Estadística, Colombia*.
- DANE. (2018). General, censo and de convivencia, encuesta and ciudadana, seguridad. *Departamento Administrativo Nacional de Estadística, Colombia*.
- García, H. (2012). Deforestación en Colombia: Retos y perspectivas.
- González, J. J., Etter, A., Sarmiento, A., Orrego, S., Ramírez, C., Cabrera, E., ... Ordoñez, M. (2011). Análisis de tendencias y patrones espaciales de deforestación en Colombia. *Instituto de Hidrología, Meteorología y Estudios Ambientales-IDEAM. Bogotá DC, Colombia*.
- Green, J. M., Larrosa, C., Burgess, N. D., Balmford, A., Johnston, A., Mbilinyi, B. P., ... Coad, L. (2013). Deforestation in an African biodiversity hotspot: Extent, variation and the effectiveness of protected areas. *Biological Conservation*, 164, 62–72.
- Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., ... others (2013). High-resolution global maps of 21st-century forest cover change. *science*, 342(6160), 850–853.
- Hübner, M. (2017). How to curb poverty-related deforestation? *Applied Economics Letters*, 24(6), 374–380.
- Mayfield, H. J., Smith, C., Gallagher, M., y Hockings, M. (2020). Considerations for selecting a machine learning technique for predicting deforestation. *Environmental Modelling & Software*, 131, 104741.
- Veldkamp, A., y Lambin, E. F. (2001). *Predicting land-use change* (Vol. 85) (n.º 1-3). Elsevier.