

## **Informe de Proyecto de Investigación**

**Evaluación del impacto climático en la erosión del suelo del departamento del Tolima usando modelamiento *Machine Learning***

Elaborado por:  
Sergio Herrera Quijano  
Juan Camilo Hernandez Sanchez  
Oscar Ivan Bautista Santana

Universidad Ean  
Escuela de Formación en Investigación  
Seminario de Investigación de Pregrado  
Bogotá  
13/10/2024

# Contenido

Resumen.....	4
Problema de Investigación.....	4
Objetivos .....	5
Objetivo general.....	5
Objetivos específicos .....	5
Justificación.....	5
Marco Teórico .....	5
La erosión del suelo .....	6
Variabilidad Climática.....	6
Modelos de Machine Learning.....	7
Aprendizaje supervisado y no supervisado .....	7
Variables Geológicas y Climáticas que Afectan la Erosión del Suelo .....	10
Datos históricos .....	13
Estudios de caso específicos en el Tolima .....	14
Metodología .....	15
Enfoque, alcance y diseño de la investigación.....	15
Definición de Variables.....	16
Población y Muestra.....	16
Selección de métodos o instrumentos para recolección de información .....	16
Técnicas de análisis de datos .....	17
Herramientas para la recolección y procesamiento de datos .....	17
Modelos predictivos y algoritmos de machine learning .....	18
Validación del modelo.....	19
Resultados .....	19
Distribución de la Variable Objetivo (Factor R) .....	20
Preprocesamiento y Transformación de Datos .....	21
Optimización y Evaluación del Modelo .....	21
Conclusiones.....	22
Lista de referencias .....	23

## Lista de figuras

Figura 1. Mapa de precipitación del departamento del Tolima. Tomado de <a href="http://www.tolima.gov.co">www.tolima.gov.co</a> .....	11
Figura 2. Mapa de temperatura en el departamento del Tolima. Tomado de: <a href="https://www.todacolombia.com/departamentos-de-colombia/tolima/clima.html">https://www.todacolombia.com/departamentos-de-colombia/tolima/clima.html</a> .....	12
Figura 3. Mapa de radiación solar promedio. Tomado y modificado de: (( <i>UPME</i> ), 2005) .....	13
Figura 4. Matriz de actividades y objetivos específicos.....	16
Figura 5. Mapa de calor de correlación de las variables de la base de datos.....	20
Figura 6. Regresión lineal de los datos de entrenamiento en comparación con los datos de validación de predicción.....	21

## Resumen

La erosión del suelo es un fenómeno geológico que afecta gravemente a la estabilidad del terreno, la productividad agrícola y la calidad de los ecosistemas. A nivel global, la erosión ha sido exacerbada por el cambio climático que altera los patrones de precipitación y temperatura, provocando una mayor vulnerabilidad en las regiones más afectadas. El departamento del Tolima, ubicado en una región montañosa de Colombia, enfrenta graves desafíos relacionados con la erosión del suelo, un fenómeno que amenaza la estabilidad de los ecosistemas locales y la productividad agrícola. Este proyecto propone desarrollar un modelo predictivo basado en machine learning que utiliza datos climáticos y geológicos del departamento del Tolima para anticipar las tasas de erosión, permitiendo identificar áreas de alto riesgo y desarrollar estrategias de mitigación.

## Problema de Investigación

La erosión del suelo es un fenómeno geológico que afecta gravemente a la estabilidad del terreno, la productividad agrícola y la calidad de los ecosistemas. A nivel global, la erosión ha sido exacerbada por el cambio climático, que altera los patrones de precipitación y temperatura, provocando una mayor vulnerabilidad en las regiones más afectadas. El departamento del Tolima, ubicado en una región montañosa de Colombia, enfrenta graves desafíos relacionados con la erosión del suelo, un fenómeno que amenaza la estabilidad de los ecosistemas locales y la productividad agrícola. La erosión es particularmente preocupante en la Cuenca Mayor del Río Lagunilla, donde se ha observado que aproximadamente el 63% del área total presenta una erosión actual ligera, y un 5% enfrenta erosión severa, con pérdidas superiores a las 100 toneladas por hectárea anualmente (Cortolima, 2009). La variabilidad climática, caracterizada por cambios en los patrones de precipitación, juega un papel crucial en la intensificación de este fenómeno, exacerbando los problemas de pérdida de suelo en la región. Estudios previos sugieren que bajo condiciones de manejo inapropiado y sin la adecuada cobertura vegetal, la erosión potencial podría afectar hasta el 51% de la cuenca. Sin embargo, la cuantificación precisa de la relación entre la variabilidad climática y la erosión del suelo utilizando técnicas avanzadas como machine learning es todavía limitada.

La falta de herramientas predictivas efectivas que permitan anticipar las tasas de erosión del suelo bajo diferentes escenarios climáticos genera incertidumbre en la planificación territorial y la gestión de recursos naturales en el Tolima. La erosión no solo afecta la productividad agrícola, sino que también compromete la infraestructura y los ecosistemas locales (SIMMA - SGC, n.d.). Este proyecto propone desarrollar un modelo predictivo basado en machine learning que utiliza datos climáticos y geológicos del departamento del Tolima para anticipar las tasas de erosión. Este enfoque permitirá identificar áreas de alto riesgo y desarrollar estrategias de mitigación más efectivas y orientadas a la preservación del suelo.

### Pregunta de investigación.

¿Cómo puede un modelo basado en machine learning mejorar la predicción de la erosión del suelo en función de la variabilidad climática en el departamento del Tolima?

# Objetivos

## Objetivo general

Desarrollar un modelo predictivo basado en machine learning para evaluar y predecir la erosión del suelo en función de la variabilidad climática en el departamento del Tolima

## Objetivos específicos

1. Analizar la relación entre la variabilidad climática (precipitación y temperatura) y la erosión del suelo en el departamento del Tolima utilizando técnicas de machine learning.
2. Identificar las variables climáticas y geológicas más influyentes en la predicción de la erosión del suelo en el Tolima.
3. Desarrollar y entrenar un modelo de machine learning para predecir la erosión del suelo en el Tolima bajo diferentes escenarios climáticos.
4. Validar y evaluar la precisión del modelo desarrollado mediante la comparación con datos históricos y estudios de caso específicos del Tolima.
5. Proponer recomendaciones basadas en los resultados del modelo para la gestión de riesgos de erosión en el departamento del Tolima.

## Justificación

Este proyecto es fundamental para mejorar la gestión del territorio en el departamento del Tolima, donde la erosión del suelo representa una amenaza significativa para la agricultura, la infraestructura y los ecosistemas locales. La implementación de un modelo predictivo basado en machine learning permitirá a las autoridades locales y agricultores anticipar y mitigar los efectos de la erosión, contribuyendo a la sostenibilidad y conservación del suelo. Además, la investigación aporta un enfoque innovador que combina geociencias y ciencia de datos, alineándose con las necesidades de la población vulnerable y así mismo, servir de referente para futuras aplicaciones en otras localidades.

## Marco Teórico

La erosión del suelo es un fenómeno natural que se ve afectado por varios factores, entre los que destacan el clima y la actividad humana. Los estudios han demostrado que la variabilidad climática, expresada en términos de precipitación y temperatura, ha presentado cambios abruptos en los últimos años. Estos cambios han alterado la estabilidad del ecosistema, influyendo directamente en los procesos erosivos y la dinámica del suelo y su cobertura vegetal. La erosión del suelo no solo tiene implicaciones ecológicas, sino que afecta gravemente la productividad agrícola, la disponibilidad de recursos hídricos y la sostenibilidad de los ecosistemas. Herramientas tecnológicas como el uso de modelos predictivos y algoritmos de aprendizaje automático pueden jugar un papel decisivo en la gestión y mitigación de estos problemas.

## **La erosión del suelo**

Los suelos, según Brady y Weil (2008), se forman mediante un proceso continuo y complejo que involucra la interacción dinámica entre cinco factores clave: material parental (roca madre), clima, organismos vivos, relieve y tiempo. Estos factores determinan las características específicas del suelo en cualquier región, incluida la del departamento del Tolima, que es objeto de esta investigación sobre erosión y cambio climático.

En la región de Tolima, la génesis del suelo está marcada por su variado relieve y condiciones climáticas extremas. Los procesos erosivos se ven acelerados por fenómenos meteorológicos como “El Niño” y “La Niña”, que provocan alteraciones en los patrones de precipitación, con lluvias más intensas y sequías prolongadas. La erosión del suelo es particularmente crítica en estas condiciones, ya que implica la pérdida de la capa superficial, rica en nutrientes y materia orgánica, fundamental para el crecimiento de la vegetación y la productividad agrícola.

Daniels y Buol (2019) señalan que los suelos que se forman bajo condiciones de alta precipitación son especialmente vulnerables a la erosión hídrica. El agua actúa como el principal agente erosivo, transportando partículas del suelo y causando su desgaste. Además, Brady y Weil (2008) destacan que la estructura del suelo y su contenido de materia orgánica son factores determinantes en su susceptibilidad a la erosión. Suelos con escasa cobertura vegetal, o aquellos manejados inapropiadamente, son más propensos a sufrir erosión. En el contexto del Tolima, los cambios en los patrones climáticos asociados al calentamiento global pueden exacerbar la frecuencia y la intensidad de estos eventos, haciendo necesario el desarrollo de estrategias de manejo del suelo más efectivas (Blanco & Lal, 2010).

## **Variabilidad Climática**

El cambio climático ha tenido un impacto directo en la estabilidad de los suelos y los ecosistemas. Según Reddy y Khanna (2019), las alteraciones en los regímenes de lluvia, la mayor intensidad de las tormentas y el aumento global de las temperaturas incrementan el riesgo de erosión en áreas vulnerables, como el Tolima. En regiones montañosas, las lluvias intensas provocan una mayor escorrentía, lo que aumenta el transporte de sedimentos y la pérdida de suelo. A su vez, los periodos prolongados de sequía compactan el suelo, reduciendo su capacidad de regeneración y aumentando la susceptibilidad a la erosión cuando llegan las lluvias.

La erosión hídrica es especialmente severa en áreas con insuficiente cobertura vegetal, lo que puede ser resultado de la deforestación o de prácticas agrícolas no sostenibles. En el caso de Tolima, donde gran parte del terreno se utiliza para la agricultura, es crucial identificar las áreas de mayor riesgo y desarrollar estrategias de manejo sostenible que minimicen el impacto erosivo del agua. Schaetzl y Anderson (2005) subrayan la importancia de la cobertura vegetal en la protección del suelo frente a los procesos erosivos. Las plantas actúan como una barrera natural, disminuyendo la velocidad del agua y reduciendo la erosión. En su ausencia, las lluvias intensas causan una mayor pérdida de suelo (Blanco & Lal, 2010).

Por otro lado, la Agencia Europea de Medio Ambiente (2019) señala que los cambios en los patrones de precipitación debido al cambio climático intensifican la vulnerabilidad del suelo, promoviendo la pérdida de su capa fértil a causa de la escorrentía superficial. Las sequías prolongadas también reducen la capacidad del suelo para retener nutrientes y regenerarse, lo que afecta negativamente la productividad agrícola y los ecosistemas locales. Esto es

particularmente relevante en regiones montañosas como Tolima, donde las variaciones altitudinales crean microclimas que afectan directamente los procesos erosivos (Blanco & Lal, 2010).

## **Modelos de Machine Learning**

El Machine Learning es una herramienta que ha mostrado un gran potencial para predecir la erosión del suelo. Según Bobadilla (2020), los algoritmos de aprendizaje automático permiten extraer patrones complejos a partir de datos geospaciales y climatológicos, ofreciendo soluciones eficientes a problemas que antes requerían métodos manuales o empíricos. Estos modelos son capaces de identificar correlaciones entre variables como precipitación, topografía, tipo de suelo y cobertura vegetal, permitiendo prever futuros escenarios de erosión bajo diferentes condiciones climáticas.

Se han realizado varios estudios que demuestran la eficacia de los modelos de Machine Learning en la predicción de la erosión del suelo. Por ejemplo, Liu et al. (2018) utilizaron redes neuronales artificiales para predecir la erosión en la cuenca del río Yangtsé, logrando una alta precisión en sus resultados. Otro estudio de Youssef et al. (2016) aplicó algoritmos de bosque aleatorio y máquinas de soporte vectorial en Egipto, mostrando que estos modelos superan a los métodos tradicionales en términos de exactitud y eficiencia.

Forharini et al. (2023) realizaron un estudio sobre la erosión del suelo en subcuencas del norte de Portugal, utilizando modelos de soporte vectorial y bosque aleatorio. El análisis se centró en identificar los factores predictivos de la erosión, evaluando 761 subcuencas, y demostraron que las técnicas de Machine Learning tienen ventajas sobre los métodos tradicionales al manejar múltiples variables con alta complejidad. Los resultados subrayan la importancia de incorporar modelos predictivos en la gestión sostenible de los suelos y recursos hídricos. Utilizando datos de lluvia, morfometría y cobertura del suelo, el estudio mostró la relación directa entre estos parámetros y el aumento de la erosión en áreas afectadas por incendios forestales (Blanco & Lal, 2010).

Además, Chen et al. (2019) implementaron modelos de aprendizaje profundo para mapear la susceptibilidad a la erosión en áreas montañosas de China. Sus resultados indican que los modelos de deep learning pueden capturar relaciones no lineales complejas entre las variables geográficas y climáticas, mejorando la precisión de las predicciones en comparación con los modelos de Machine Learning tradicionales.

En el contexto del departamento del Tolima, la aplicación de modelos de Machine Learning es especialmente relevante debido a la complejidad geográfica y climática de la región. Estos modelos permiten integrar una gran cantidad de variables y manejar la heterogeneidad de los datos, proporcionando herramientas valiosas para la predicción y gestión de la erosión del suelo.

## **Aprendizaje supervisado y no supervisado**

El aprendizaje supervisado es una de las técnicas más utilizadas en Machine Learning. Este enfoque utiliza conjuntos de datos etiquetados para entrenar modelos predictivos, que luego pueden generalizar y predecir resultados para nuevos datos no etiquetados. En el caso de la erosión del suelo, se puede entrenar un modelo con datos históricos de precipitación, cobertura del suelo y características del relieve para predecir la erosión futura. Por otro lado, el aprendizaje

no supervisado, en el que los datos no están etiquetados, también puede ser útil en la clasificación de diferentes tipos de suelos o en la identificación de patrones ocultos en los datos geoespaciales. En la investigación sobre erosión, los algoritmos de clustering se utilizan para agrupar áreas con características similares de erosión, permitiendo así una gestión diferenciada y más precisa del territorio (Blanco & Lal, 2010). Investigaciones recientes han explorado el uso de algoritmos de clustering para identificar zonas con patrones similares de erosión. Por ejemplo, Borrelli et al. (2020) aplicaron técnicas de aprendizaje no supervisado para clasificar áreas en Europa según su susceptibilidad a la erosión, lo que facilitó la implementación de políticas de conservación del suelo adaptadas a cada región. Además, las Redes Neuronales Artificiales (RNA) han demostrado ser herramientas eficaces en este contexto; estas redes, inspiradas en las conexiones neuronales biológicas, permiten a las máquinas aprender y reconocer patrones complejos a partir de datos. Liu et al. (2018) emplearon RNA para predecir la erosión en la cuenca del río Yangtsé, obteniendo resultados precisos, mientras que Meena et al. (2019) las utilizaron junto con datos de sensores remotos para mapear áreas propensas a la erosión en el Himalaya. Las Máquinas de Soporte Vectorial (SVM), por su parte, son algoritmos de aprendizaje supervisado que buscan encontrar el hiperplano que mejor separa los datos de diferentes clases, maximizando el margen entre ellos. En estudios de erosión del suelo, las SVM han sido útiles para modelar la relación entre variables como la precipitación y la pendiente, permitiendo identificar áreas de alto riesgo de erosión en el Tolima (Huang et al., 2019).

## **Predicción de Erosión del suelo**

La erosión del suelo es un problema ambiental importante en muchas regiones, y su medición se ha mejorado mediante el uso de técnicas avanzadas, como el aprendizaje automático. Este enfoque ha demostrado ser eficaz para predecir la erosión del suelo y optimizar la gestión de los recursos hídricos y la conservación de la tierra. Forharini et al. (2023) realizaron un estudio en el que analizaron la erosión del suelo en subcuencas del norte de Portugal utilizando modelos de aprendizaje automático basados en soporte vectorial y bosque aleatorio. El estudio se centró en identificar los factores que predicen la erosión, examinando 761 subcuencas y estableciendo relaciones significativas entre parámetros como la lluvia, la morfometría y la cobertura del suelo. Los autores enfatizan que el aprendizaje automático combina lo mejor de los métodos tradicionales, ya que puede manejar múltiples variables y su complejidad.

Otro estudio significativo es el de Rahmati et al. (2017), quienes utilizaron modelos de regresión logística y aprendizaje automático para mapear la susceptibilidad a la erosión en Irán. Sus hallazgos destacan que los modelos basados en Machine Learning ofrecen mayor precisión y pueden integrarse eficazmente con Sistemas de Información Geográfica (SIG) para mejorar la gestión del suelo.

Además, el estudio destaca que la erosión del suelo no solo afecta la calidad de los recursos hídricos, sino que también reduce la productividad agrícola y daña los servicios ecosistémicos. El uso de herramientas analíticas modernas es esencial para mejorar la conservación y desarrollar políticas que reduzcan el impacto de la erosión. Estas herramientas permiten predecir las condiciones futuras de erosión en diferentes climas y usos del suelo, lo que es crucial para planificar la gestión sostenible de los ríos.

Uno de los enfoques más comunes para medir y predecir la erosión del suelo es la Ecuación Universal de Pérdida de Suelo (USLE), que se ha utilizado ampliamente desde su desarrollo por

Wischmeier y Smith en 1965. La USLE estima la pérdida de suelo en función de varios factores, incluyendo:

- **R:** el **factor de erosividad de la lluvia**, que cuantifica el efecto de la precipitación sobre la erosión. Este factor se basa en la intensidad y la cantidad de la lluvia y se calcula mediante un índice que considera la energía de la lluvia en un área específica (Wischmeier & Smith, 1978).
- **K:** el **factor de erodibilidad del suelo**, que refleja la susceptibilidad del suelo a la erosión según su textura, estructura y contenido de materia orgánica.
- **LS:** el **factor topográfico**, que considera la longitud y la pendiente de la ladera.
- **C:** el **factor de cobertura**, que evalúa el efecto de la vegetación y las prácticas de manejo del suelo sobre la erosión.
- **P:** el **factor de prácticas de conservación**, que incluye prácticas de manejo que pueden reducir la erosión, como la labranza de contornos o la siembra de cultivos de cobertura.

La integración de modelos de Machine Learning con la USLE ha demostrado mejorar la precisión en la predicción de la erosión del suelo. Por ejemplo, Behera et al. (2018) combinaron la USLE con algoritmos de Machine Learning para mapear la erosión del suelo en India, logrando una mejor representación de la complejidad espacial y temporal de la erosión. Esta integración permite aprovechar las fortalezas de ambos enfoques: la solidez teórica de la USLE y la capacidad de los algoritmos de aprendizaje automático para manejar grandes volúmenes de datos y capturar patrones complejos.

La USLE ha demostrado ser útil en diversas investigaciones, incluyendo aquellas en el departamento del Tolima, donde se ha adaptado para considerar las variaciones locales en clima y topografía (Blanco & Lal, 2010). A medida que se avanza hacia una mayor comprensión de la erosión del suelo, los modelos de Machine Learning se están integrando con enfoques tradicionales como la USLE para mejorar la precisión de las predicciones. Estos modelos pueden procesar grandes volúmenes de datos geoespaciales y climáticos, permitiendo identificar patrones de erosión y proyecciones futuras bajo diferentes escenarios climáticos (Ramos Castillo & Orozco Medina, 2020).

El uso de modelos climáticos globales ha sido fundamental para predecir los impactos del cambio climático, particularmente en relación con las variaciones de temperatura y precipitación. Esta investigación emplea modelos del CMIP5 y escenarios de radiación RCP4.5, RCP6.0 y RCP8.5 del IPCC, los cuales reflejan condiciones extremas de forzamiento radiactivo, proporcionando un amplio rango de posibles futuros climáticos. Para obtener resultados más detallados a nivel local, se implementa la técnica de *downscaling*, que permite ajustar los datos globales a escalas regionales y de cuenca. Además, el software Terrset® se utiliza para proyectar cambios en la cobertura vegetal y el uso del suelo, basándose en datos históricos de 1997 y 2016, y generando proyecciones hasta el año 2035. Sin embargo, las proyecciones para fechas posteriores revelaron inconsistencias, como una inesperada disminución de las áreas urbanas, lo que llevó a limitar el análisis al horizonte temporal de 2035 (Ramos Castillo & Orozco Medina, 2020).

Por ejemplo, Behera et al. (2018) combinaron la USLE con algoritmos de Machine Learning para mapear la erosión del suelo en India. Esta integración permitió una mejor representación de la complejidad espacial y temporal de la erosión, proporcionando información más detallada para la gestión del suelo.

Adicionalmente, se han realizado estudios que combinan Machine Learning con datos de sensores remotos para mejorar la predicción de la erosión del suelo. Meena et al. (2019) emplearon imágenes satelitales junto con algoritmos de aprendizaje automático para mapear áreas propensas a la erosión en el Himalaya occidental. Sus resultados facilitaron la identificación de zonas críticas y contribuyeron al desarrollo de estrategias de conservación del suelo.

Otro ejemplo es el trabajo de Tien Bui et al. (2020), quienes aplicaron técnicas de aprendizaje profundo para evaluar la susceptibilidad a la erosión en regiones costeras de Vietnam. Este estudio demostró que los modelos de deep learning pueden manejar eficientemente grandes conjuntos de datos y ofrecer predicciones más precisas que los métodos tradicionales.

Además, Abba et al. (2017) investigaron el uso de algoritmos de clasificación, como el árbol de decisión y el k-NN, para predecir la erosión del suelo en Nigeria. Encontraron que estos modelos pueden ser herramientas útiles en regiones con datos limitados, proporcionando información valiosa para la gestión del suelo y la planificación agrícola.

## **Variables Geológicas y Climáticas que Afectan la Erosión del Suelo**

La erosión del suelo es un proceso complejo influenciado por una serie de variables geológicas y climáticas. Comprender estas variables es crucial para el manejo sostenible del suelo y la mitigación de la erosión, especialmente en regiones vulnerables como el departamento del Tolima en Colombia.

### Variables Geológicas

Material Parental: El tipo de roca del que se deriva el suelo, conocido como material parental, es fundamental en la formación y resistencia del suelo a la erosión. Los suelos formados sobre rocas sedimentarias tienden a ser más susceptibles a la erosión en comparación con aquellos desarrollados sobre rocas ígneas o metamórficas, debido a su menor cohesión y estabilidad estructural (Blanco & Lal, 2010; Smith et al., 2011). No obstante, esto también depende de la estabilidad química y mecánica de las rocas y los minerales que contiene; minerales ferromagnesianos en rocas ígneas básicas (máficas) tales como los piroxenos o anfíboles, se meteorizan (degradación en otros minerales más estables) más fácilmente ante agentes ambientales como el agua y el oxígeno, que las rocas ricas en cuarzo como una cuarzoarenita o un granito.

Textura del Suelo: La textura del suelo, que se refiere a la proporción de partículas de arena, limo y arcilla, afecta su capacidad para retener agua y nutrientes. Suelos arenosos, con grandes partículas y baja cohesión, son particularmente vulnerables a la erosión, ya que tienen una menor capacidad para mantener la humedad y la estructura (Miller et al., 2012). En contraste, los suelos arcillosos, aunque retienen más agua, pueden sufrir erosión por compactación y escorrentía.

Estructura del suelo: La estructura del suelo se refiere a la forma en que las partículas de suelo están agrupadas. Suelos con buena estructura, que presentan agregados estables, son más resistentes a la erosión. Una estructura deficiente, que a menudo resulta de prácticas de manejo inadecuadas, puede incrementar la susceptibilidad a la erosión, especialmente en condiciones de lluvia intensa (Brady & Weil, 2008; Borrelli et al., 2017).

**Topografía:** La inclinación y el relieve del terreno son factores críticos que determinan la velocidad de escorrentía del agua. Las pendientes pronunciadas facilitan la escorrentía rápida, lo que a su vez puede incrementar la erosión del suelo. Estudios han demostrado que las laderas con pendientes superiores al 5% son significativamente más propensas a la erosión hídrica (Ramos et al., 2014).

## Variables Climáticas

**Precipitación:** La cantidad, intensidad y frecuencia de las precipitaciones son factores determinantes en la erosión del suelo. Lluvias intensas pueden causar escorrentía rápida, lo que resulta en la pérdida de suelo y nutrientes. Según la Agencia Europea de Medio Ambiente (2019), los cambios en los patrones de precipitación debido al cambio climático están aumentando la vulnerabilidad del suelo a la erosión.

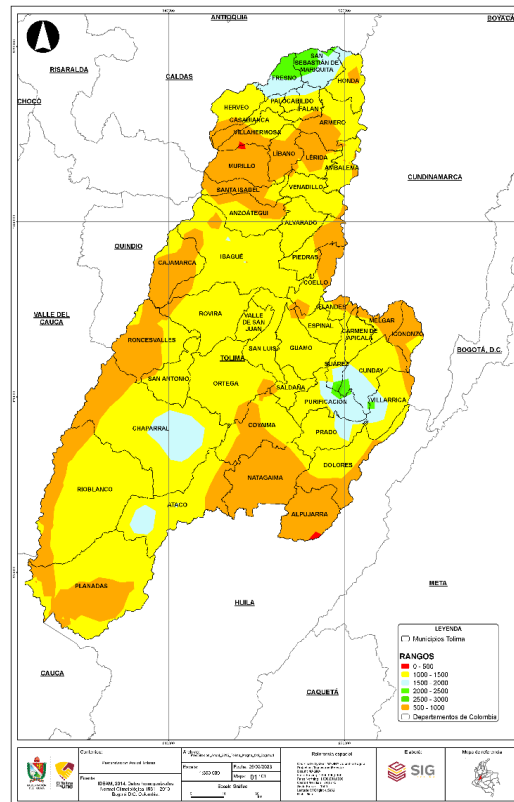


Figura 1. Mapa de precipitación del departamento del Tolima. Tomado de [www.tolima.gov.co](http://www.tolima.gov.co)

**Temperatura:** La temperatura del aire influye en la tasa de descomposición de la materia orgánica y en la actividad biológica del suelo. Un aumento en las temperaturas, asociado con el cambio climático, puede acelerar estos procesos, afectando la estructura del suelo y su capacidad para resistir la erosión (Huang et al., 2019).

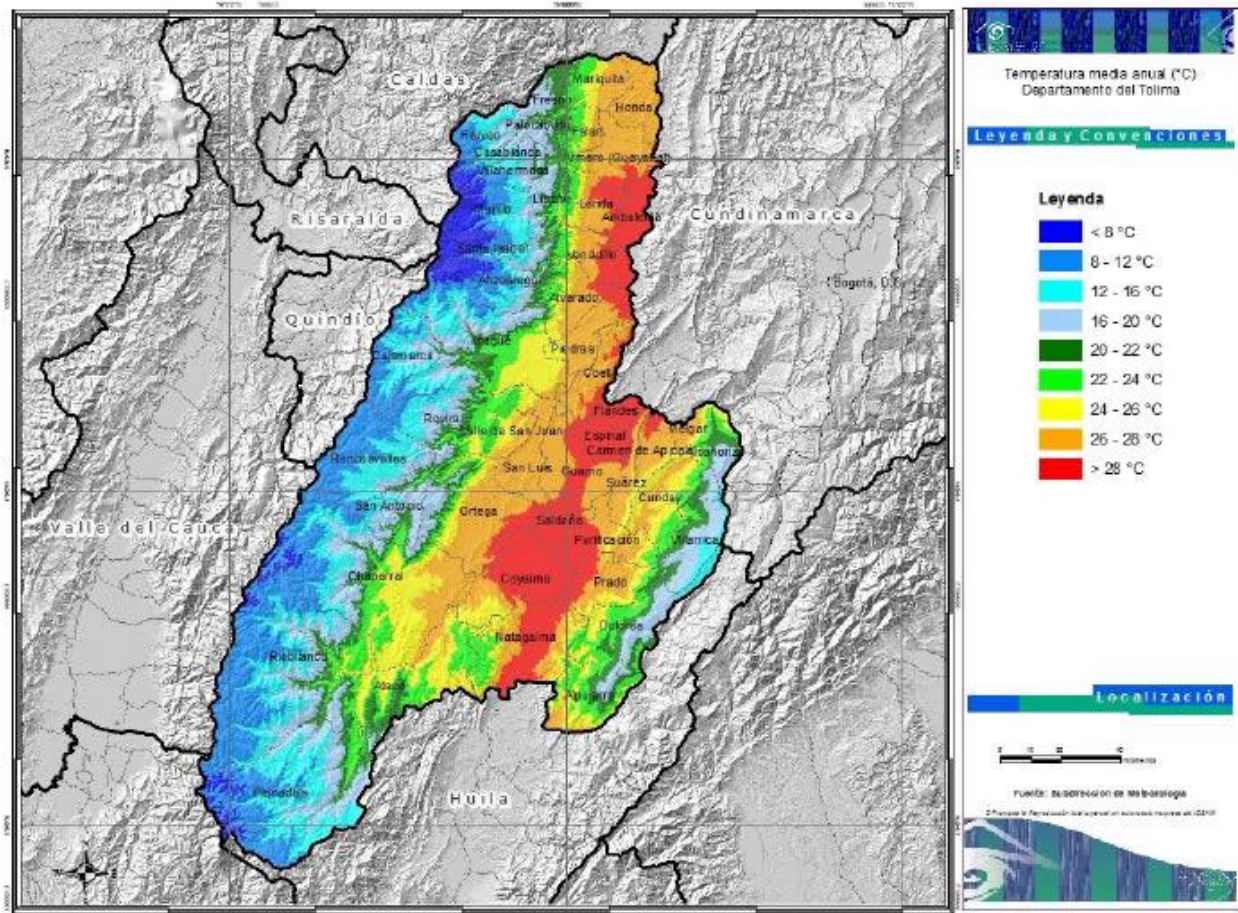


Figura 2. Mapa de temperatura en el departamento del Tolima. Tomado de: <https://www.todacolombia.com/departamentos-de-colombia/tolima/clima.html>

**Humedad Relativa:** La humedad del suelo es un factor importante que afecta su estabilidad y resistencia a la erosión. Un suelo excesivamente húmedo puede volverse inestable y más susceptible a la erosión, mientras que suelos secos pueden no ser capaces de retener adecuadamente los nutrientes y el agua (Dumanski et al., 2014).

**Radiación Solar:** La radiación solar afecta la temperatura y la evaporación en el suelo. Una mayor radiación puede incrementar la temperatura del suelo y la evaporación, lo que a su vez afecta la humedad del suelo y su resistencia a la erosión (Khan et al., 2020).

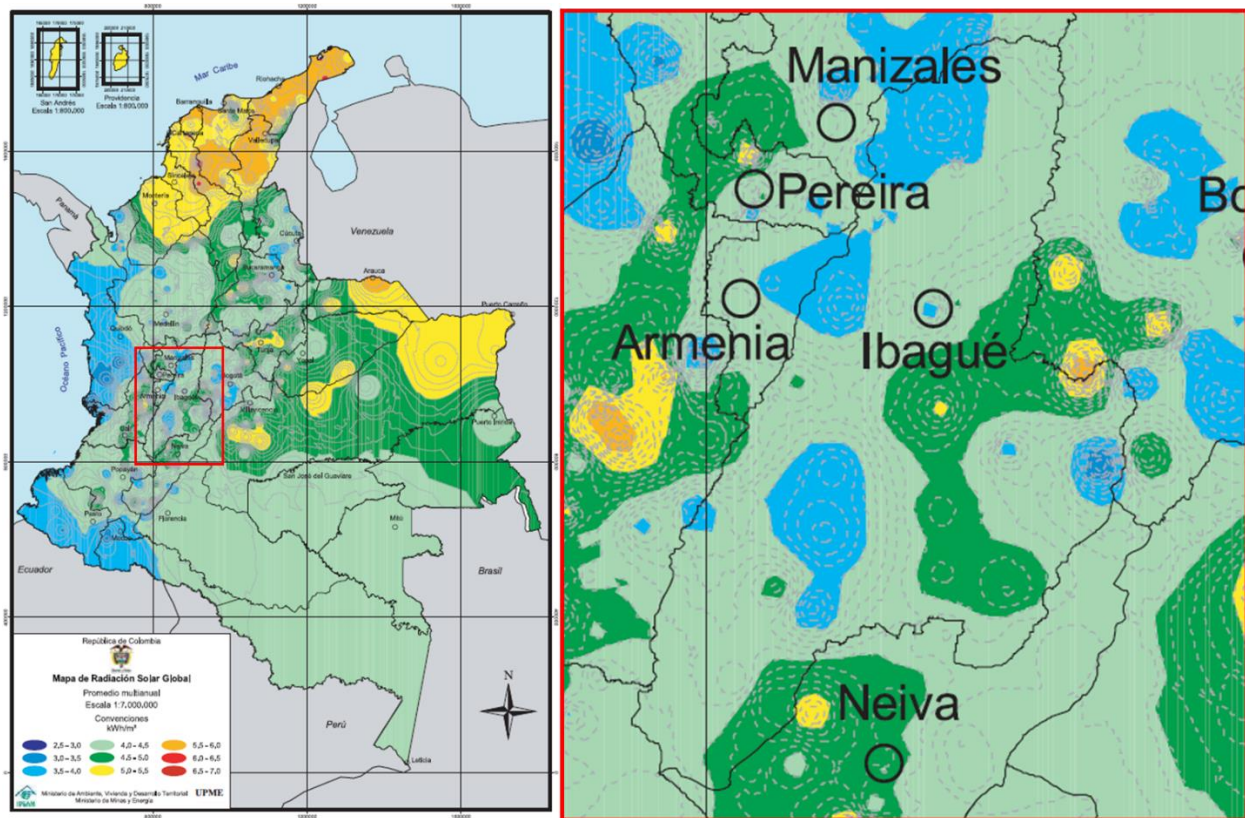


Figura 3. Mapa de radiación solar promedio. Tomado y modificado de: ((UPME), 2005)

Las variables geológicas y climáticas interactúan de manera compleja, influenciando los procesos de erosión del suelo en múltiples niveles. La integración de esta comprensión en la planificación y gestión del uso del suelo es fundamental para desarrollar estrategias efectivas que mitiguen la erosión, preserven la calidad del suelo y promuevan la sostenibilidad agrícola en el contexto del cambio climático.

## Datos históricos

La comprensión de la erosión del suelo en el departamento del Tolima requiere un análisis exhaustivo de datos históricos climáticos y geográficos, los cuales son proporcionados por instituciones clave como el Instituto de Hidrología Meteorología y Estudios Ambientales (IDEAM) y el Instituto Geográfico Agustín Codazzi (IGAC). Estas entidades desempeñan un papel crucial en la generación y divulgación de información sobre el estado de los recursos naturales y el medio ambiente en Colombia. El IDEAM ofrece datos históricos de variables climáticas fundamentales, como la precipitación, temperatura, humedad relativa y vientos, esenciales para analizar los patrones de erosión del suelo. Por su parte, el IGAC se encarga de producir y actualizar la cartografía oficial del país, proporcionando información sobre suelos, cobertura y uso del suelo. Además, los estudios de caso específicos en Tolima, como el análisis de los suelos fluvio-volcánicos de Armero, permiten comprender la capacidad de recuperación de estos suelos tras eventos climáticos extremos, como la erupción del volcán Nevado del Ruiz. La erosión del suelo, uno de los problemas ambientales más críticos en Colombia, se agrava por el cambio

climático, siendo un fenómeno que se manifiesta de manera alarmante en varias regiones, incluido el departamento del Tolima.

## **Estudios de caso específicos en el Tolima**

En el marco del estudio de los suelos fluvio-volcánicos de Armero, Tolima, la composición física y química del suelo es clave para entender su capacidad de recuperación tras la erupción del volcán Nevado del Ruiz. Según Torres-Benítez et al. (2017), los suelos afectados presentan una variabilidad significativa en nutrientes como fósforo, calcio y magnesio, lo que influye en la regeneración de la vegetación y la productividad agrícola. Este análisis es fundamental para implementar estrategias de manejo sostenible del suelo en la región.

Las condiciones edáficas tras la deposición de los lodos volcánicos favorecieron la aparición inicial de 294 especies vegetales, con *Sarcostemma clausum*, *Melinis repens* y *Cyperus ligularis* como especies pioneras. A medida que avanzaba la sucesión vegetal, la influencia de las leguminosas mejoró la fertilidad del suelo, principalmente debido a la fijación de nitrógeno por sus nódulos. Este proceso facilitó el establecimiento de especies arbóreas y arbustivas, reduciendo la acidez y mejorando la calidad del suelo a lo largo del tiempo (Torres-Benítez et al., 2017).

La erosión del suelo, uno de los problemas ambientales más críticos de Colombia, si no se atiende debidamente, ve agravados sus impactos por el cambio climático. La Fundación Alejandro Ángel Escobar (2015) señala que la degradación del suelo por erosión se está produciendo a un ritmo alarmante en diversas zonas, incluido el departamento del Tolima, en el país. El fenómeno está relacionado con alteraciones en el uso del suelo que también han sido influenciadas por la variabilidad climática reciente.

El deterioro del régimen de lluvias, tanto en su cantidad como en su distribución, junto con el alargamiento de los períodos secos, constituyen factores erosivos que transforman suelos anteriormente fértiles en tierras secas e improductivas. Específicamente en las tierras agrícolas del Tolima, esto adquiere una dimensión muy preocupante, pues las prácticas no sostenibles encuentran un terreno fértil en las áreas deforestadas.

En la región de Santa Isabel, Tolima, la relación entre el uso del suelo y el medio ambiente es un aspecto fundamental que se ha estudiado ampliamente. De acuerdo con las investigaciones de la Universidad del Tolima: “el uso intensivo del suelo para actividades agrícolas sin prácticas sostenibles ha conducido a una significativa degradación del recurso suelo, lo que se traduce en un aumento de la erosión”. A su vez, utilizando el mismo criterio, es posible inferir que este evento testimonia una condición sumamente necesaria a las regiones andinas en general, a saber, “una condición natural crítica en la que la topografía, el clima y la intervención humana proporcionan las condiciones ideales para la pérdida de suelo”.

Adicionalmente, el cambio climático representa una amenaza, ya que cambia los patrones de lluvia y permite que los eventos climáticos extremos dañen el suelo a través de la erosión. Mientras que, en ciertas temporadas del año, el municipio de Santa Isabel en el Tolima es afectado por intensas lluvias que resultan en escorrentía superficial, a su vez, este evento conduce a los procesos de transformación y pérdida de suelo cultivable por descomposición de material y lavado.

Los estudios han sugerido que la erosión, además de dañar la calidad superficial del suelo, afecta la calidad del agua y la salud de los ecosistemas vecinos. La sedimentación de cuerpos de agua y la reducción en la capacidad de los suelos para capturar y almacenar agua son efectos colaterales no insignificantes. Por lo tanto, es una prioridad urgente desarrollar e implementar estrategias de manejo sostenible que incorporen prácticas agrícolas que minimicen la erosión y promuevan la conservación de suelos a largo plazo. La investigación local señala la necesidad de sensibilizar a los agricultores y educarlos sobre cómo sus prácticas de conservación, como la rotación de cultivos y la reforestación, son efectivas contra el impacto de la erosión en esta región frágil.

El informe también sostiene que la erosión hídrica es más dañina en áreas con baja cobertura vegetal. Esta erosión resulta en una pérdida de recursos hídricos y una disminución en la retención de agua del suelo, lo que afecta directamente la producción agrícola y la biodiversidad nativa de los sitios. Por lo tanto, se deben adoptar acciones encaminadas a garantizar la conservación del suelo y la restauración del entorno dañado para reducir estas amenazas erosivas de alto riesgo y garantizar la continuidad de estas áreas productivas (Fundación Alejandro Ángel Escobar, 2015).

## **Metodología**

El presente proyecto seguirá un diseño metodológico estructurado, orientado a desarrollar un modelo predictivo de erosión del suelo en el departamento del Tolima, Colombia, utilizando técnicas de Machine Learning y datos climáticos y geológicos. Para lograr esto, se adoptará el estándar CRISP-DM (Cross-Industry Standard Process for Data Mining) como guía para el proceso de minería de datos y modelado, adaptándolo al contexto de la investigación académica.

El enfoque metodológico se dividirá en las siguientes etapas:

- Comprensión del Negocio (Investigación)
- Comprensión de los Datos
- Preparación de los Datos
- Modelado
- Evaluación
- Implementación de Recomendaciones

## **Enfoque, alcance y diseño de la investigación**

El presente estudio adoptará un enfoque cuantitativo utilizando un diseño no experimental de tipo descriptivo-correlacional. El objetivo es analizar la relación entre la erosión del suelo en el departamento del Tolima y las variaciones climáticas, empleando técnicas de machine learning para predecir la tasa de erosión. Este tipo de estudio es adecuado ya que se busca describir el fenómeno sin manipular las variables, sino observando y correlacionando los datos de erosión y clima previamente recopilados.

## Definición de Variables

Objetivo específico	Actividades	Resultado esperado
1. Analizar la relación entre la variabilidad climática (precipitación y temperatura) y la erosión del suelo en el departamento del Tolima utilizando técnicas de machine learning.	Descripción estadística de datos históricos	Encontrar una distribución normal de los datos
	Consultar bases de datos gubernamentales	Extraer datos para hacer modelamiento
2. Identificar las variables climáticas y geológicas más influyentes en la predicción de la erosión del suelo en el Tolima.	Usar correlación de Pearson del total de las variables con una variable objetivo que describa la erosión	Obtener 10 variables más correlacionadas
3. Desarrollar y entrenar un modelo de machine learning para predecir la erosión del suelo en el Tolima bajo diferentes escenarios climáticos.	Hacer limpieza de la base de datos	Preparar base de datos para el modelado
	Escoger variable objetivo y dependientes	
	Explorar 5 modelos predictivos para evaluar su desempeño en la base de datos	Modelar y escoger el modelo con el mejor rendimiento
4. Validar y evaluar la precisión del modelo desarrollado mediante la comparación con datos históricos y estudios de caso específicos del Tolima.	Desplegar predicciones con una subdivisión de los datos en set de validación	Probar la exactitud de las predicciones
5. Proponer recomendaciones basadas en los resultados del modelo para la gestión de riesgos de erosión en el departamento del Tolima	Concluir la utilidad y rendimiento del modelo	Analizar la utilidad del método para la aplicación en estamentos nacionales.

Figura 4. Matriz de actividades y objetivos específicos.

## Población y Muestra

La población objeto del estudio comprende las áreas geográficas del departamento del Tolima el cual tiene 47 municipios y limita con los departamentos Caldas, Quindío, Risaralda y Valle del Cauca, El río Magdalena atraviesa por todo su territorio y lleva otras cuencas a su vez como Gualí, Sabandija, Recio, Lagunillas, Opía, Anchique, Chenche y Atá, su clima es semihúmedo por sus diversas precipitaciones anuales. Debido a las limitaciones en tiempo y recursos, se utilizará un muestreo por conveniencia, tomando datos históricos.

## Selección de métodos o instrumentos para recolección de información

La recolección de datos se realizará a través de la integración de bases de datos climatológicas históricas, mapas geospaciales y estudios sobre erosión del suelo en la región. Para ello, se emplearán técnicas de extracción y preprocesamiento de datos para garantizar la homogeneidad de las variables. Los datos serán limpiados y normalizados antes de ser ingresados al modelo predictivo.

## Técnicas de análisis de datos

El presente estudio sobre la erosión del suelo en el departamento del Tolima y su relación con las variables climáticas requiere una serie de instrumentos y técnicas que permitan procesar, analizar y modelar los datos de manera efectiva. Para lograr este objetivo, se emplearán diversas herramientas tecnológicas y metodológicas que optimicen la recolección de datos, el procesamiento y la generación de modelos predictivos utilizando técnicas de *machine learning*. A continuación, se describen los instrumentos, herramientas y enfoques metodológicos que se emplearán en cada fase del análisis.

### Herramientas para la recolección y procesamiento de datos

#### Fuentes de datos

El análisis de la erosión del suelo exige la recopilación de datos históricos y actuales de diversas fuentes confiables. Las cuales principales se basan en bases de datos climatológicas, que idealmente contienen datos sobre temperatura, precipitación, velocidad y dirección del viento, humedad relativa y radiación solar. Estos datos serán obtenidos de fuentes gubernamentales como el IDEAM (Instituto de Hidrología, Meteorología y Estudios Ambientales)

#### Limpieza y preprocesamiento de datos

Antes de poder llevar a cabo un análisis riguroso, es necesario realizar una limpieza y preprocesamiento de los datos. En esta fase, se eliminarán datos incompletos o inconsistentes, y se manejarán los valores faltantes. Se aplicarán técnicas de imputación de datos, donde los valores faltantes serán estimados basados en datos similares o patrones identificados dentro del conjunto de datos. Las etapas del preprocesamiento incluirán:

- Transformación de datos: Los datos se transformarán en formatos estandarizados para que puedan ser utilizados en los algoritmos de machine learning. Esto incluye la conversión de formatos geoespaciales, la estandarización de unidades de medida y la agregación de datos a intervalos temporales consistentes (por ejemplo, por mes o año).
- Normalización de datos: Algunos de los modelos de machine learning, como la regresión logística y las redes neuronales, requieren que los datos estén en un rango específico. Por lo tanto, se aplicarán técnicas de normalización para escalar los datos y garantizar que no existan grandes disparidades entre las variables.
- Manejo de datos categóricos: Las variables categóricas (por ejemplo, tipos de suelo, clasificación del uso del terreno) se codificarán utilizando técnicas como la codificación one-hot para que puedan ser tratadas por los algoritmos de machine learning.
- Reducción de dimensionalidad: Para optimizar el rendimiento de los modelos y reducir la complejidad computacional, se utilizarán técnicas de reducción de dimensionalidad como el análisis de componentes principales (PCA, por sus siglas en inglés). Esta técnica permite identificar las variables más relevantes para predecir la erosión, eliminando redundancias y datos no significativos.

## **Modelos predictivos y algoritmos de machine learning**

Una vez recopilados y preprocesados los datos, el siguiente paso es la construcción de modelos predictivos que permitan estimar la tasa de erosión del suelo en función de las variables climáticas y geoespaciales. En este estudio, se aplicarán diversos algoritmos de machine learning que se seleccionarán en función de la naturaleza de los datos y el tipo de predicción deseada.

### Algoritmos de regresión

La erosión del suelo es un fenómeno continuo que puede ser modelado utilizando técnicas de regresión. Los modelos de regresión proporcionan una forma de estimar la relación entre las variables independientes (factores climáticos y geoespaciales) y la variable dependiente (tasa de erosión). Entre los algoritmos de regresión que se utilizarán se encuentran:

- **Regresión lineal:** Se implementará un modelo de regresión lineal para obtener una relación simple entre la erosión del suelo y cada uno de los factores climáticos. Si bien este enfoque asume una relación lineal entre las variables, será útil como punto de partida para identificar tendencias básicas.
- **Regresión polinómica:** Dado que las relaciones entre los factores climáticos y la erosión pueden no ser lineales, se utilizará un modelo de regresión polinómica que permita capturar relaciones más complejas.
- **Regresión Ridge y Lasso:** Para evitar el sobreajuste del modelo y manejar los datos con alta dimensionalidad, se aplicarán técnicas de regularización como la regresión Ridge y Lasso. Estas técnicas permiten controlar la complejidad del modelo al penalizar las características menos significativas, mejorando así la precisión de las predicciones.

### Algoritmos de clasificación

Además de estimar las tasas de erosión, es importante clasificar las zonas en diferentes categorías de riesgo de erosión (bajo, medio, alto). Para ello, se utilizarán algoritmos de clasificación como:

- **Bosques aleatorios (Random Forest):** Este algoritmo se compone de múltiples árboles de decisión y es particularmente útil para manejar grandes conjuntos de datos con múltiples variables. Random Forest genera varios modelos y combina sus resultados para mejorar la precisión de las predicciones, además de ofrecer información sobre la importancia de cada variable en la clasificación de las zonas de riesgo de erosión.
- **Máquinas de soporte vectorial (SVM):** Se aplicará este algoritmo para clasificar las zonas de estudio en categorías de riesgo, maximizando el margen entre las distintas clases. SVM es eficaz en problemas de clasificación binaria y multiclase, y será útil para determinar áreas críticas de erosión.
- **Redes neuronales:** Para capturar patrones complejos y no lineales en los datos, se emplearán redes neuronales artificiales. Estas redes son capaces de aprender de los

datos y mejorar su precisión a través de la retropropagación, siendo especialmente útiles para identificar relaciones no triviales entre las variables climáticas y la erosión.

## **Validación del modelo**

Una vez entrenados los modelos de regresión y clasificación, se procederá a su validación para garantizar que las predicciones sean precisas y robustas. La validación se llevará a cabo mediante las siguientes técnicas:

### Validación cruzada

Para evitar que el modelo se ajuste excesivamente a los datos de entrenamiento, se utilizará la técnica de validación cruzada k-fold. En este método, el conjunto de datos se divide en k partes iguales, y el modelo se entrena k veces, utilizando una parte como conjunto de prueba y las restantes como conjunto de entrenamiento. Este enfoque permite evaluar el desempeño del modelo en diferentes subconjuntos de datos, reduciendo la posibilidad de sobreajuste.

### Métricas de evaluación

Se emplearán diversas métricas para evaluar la precisión y eficacia de los modelos predictivos:

- **Error cuadrático medio (MSE):** Esta métrica será utilizada para los modelos de regresión, evaluando la diferencia promedio entre los valores predichos y los valores reales. Un MSE bajo indicará que el modelo tiene un buen rendimiento.
- **Precisión y F1-Score:** Para los modelos de clasificación, se evaluará el rendimiento utilizando métricas como la precisión (proporción de predicciones correctas) y el F1-Score (la media armónica de la precisión y la sensibilidad). Estas métricas proporcionarán una evaluación integral del modelo en términos de su capacidad para clasificar correctamente las zonas de riesgo de erosión.
- **Curva ROC y AUC:** En los modelos de clasificación, se empleará la curva ROC para medir la capacidad del modelo de distinguir entre clases. El área bajo la curva (AUC) será utilizada como métrica de desempeño global, donde un valor más cercano a 1 indica un modelo más preciso.

## **Resultados**

La presente investigación tuvo como objetivo principal desarrollar un modelo predictivo para evaluar la erosión del suelo en el departamento del Tolima, empleando técnicas de *machine learning*. Este enfoque innovador permitió analizar de manera detallada los factores climáticos, geográficos y geológicos que influyen en el fenómeno de la erosión. A continuación, se presentan los hallazgos más relevantes derivados del análisis de datos y el desarrollo del modelo.

La información utilizada provino de un archivo Excel cargado en Google Colab, conteniendo datos de estaciones meteorológicas, precipitaciones, altitud y categorías climáticas, entre otras variables. Tras una revisión preliminar, se identificaron las columnas clave, se evaluó la calidad del dataset y se seleccionaron las variables más relevantes para el análisis.

## Distribución de la Variable Objetivo (Factor R)

La variable objetivo, representada por el promedio de *grid\_code* (lluvia acumulada), fue analizada junto con otras variables para comprender su distribución y posibles relaciones con factores geográficos. Este proceso inicial permitió establecer las bases para el análisis exploratorio y la preparación de los datos.

El análisis exploratorio de datos (EDA) se centró en identificar patrones y relaciones entre las variables mediante herramientas como mapas de correlación (*heatmaps*). Se observó que la altitud tenía una fuerte correlación negativa (-0.63) con las precipitaciones, lo que refleja que en zonas más altas las lluvias son menores. Por el contrario, la latitud presentó una correlación positiva (0.54), indicando un incremento en las precipitaciones en ciertas regiones. Aunque variables como el mes y el valor mostraron correlaciones más bajas, se incluyeron en el modelo debido a su posible impacto acumulativo en el comportamiento del sistema.

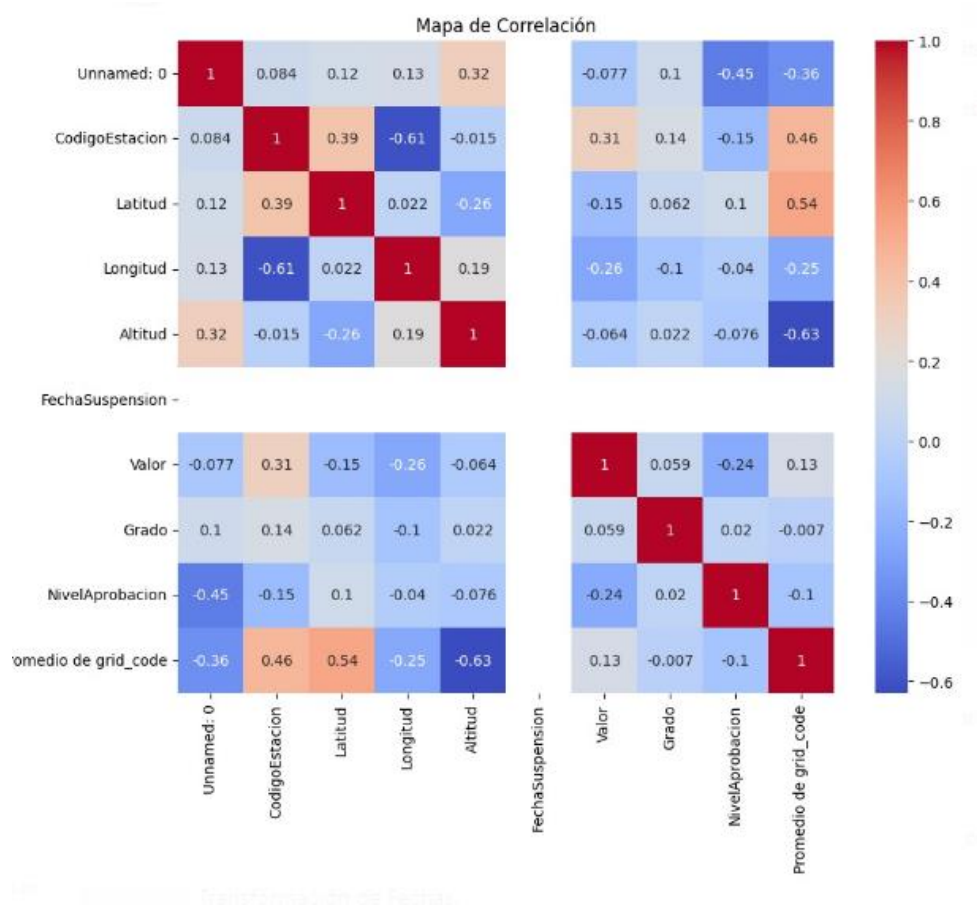


Figura 5. Mapa de calor de correlación de las variables de la base de datos

## Preprocesamiento y Transformación de Datos

Para garantizar que los datos estuvieran listos para el modelo, se realizaron transformaciones clave. Esto incluyó la extracción de información temporal como el mes, la codificación de variables categóricas mediante *One-Hot Encoding* y la división del dataset en conjuntos de entrenamiento (80%) y prueba (20%). Estas etapas aseguraron un formato adecuado y la calidad de los datos utilizados.

El modelo predictivo elegido fue un *Random Forest Regressor*, conocido por su capacidad para manejar datos mixtos y relaciones no lineales. En su configuración inicial, con 100 árboles, el modelo alcanzó un Error Cuadrático Medio (MSE) de 3642.32, indicando un buen ajuste inicial. Posteriormente, se aplicaron técnicas de optimización como la validación cruzada y el ajuste de hiperparámetros mediante *GridSearchCV*, logrando un  $R^2$  de 0.9975 tras ajustar parámetros como la profundidad máxima y la cantidad de árboles.

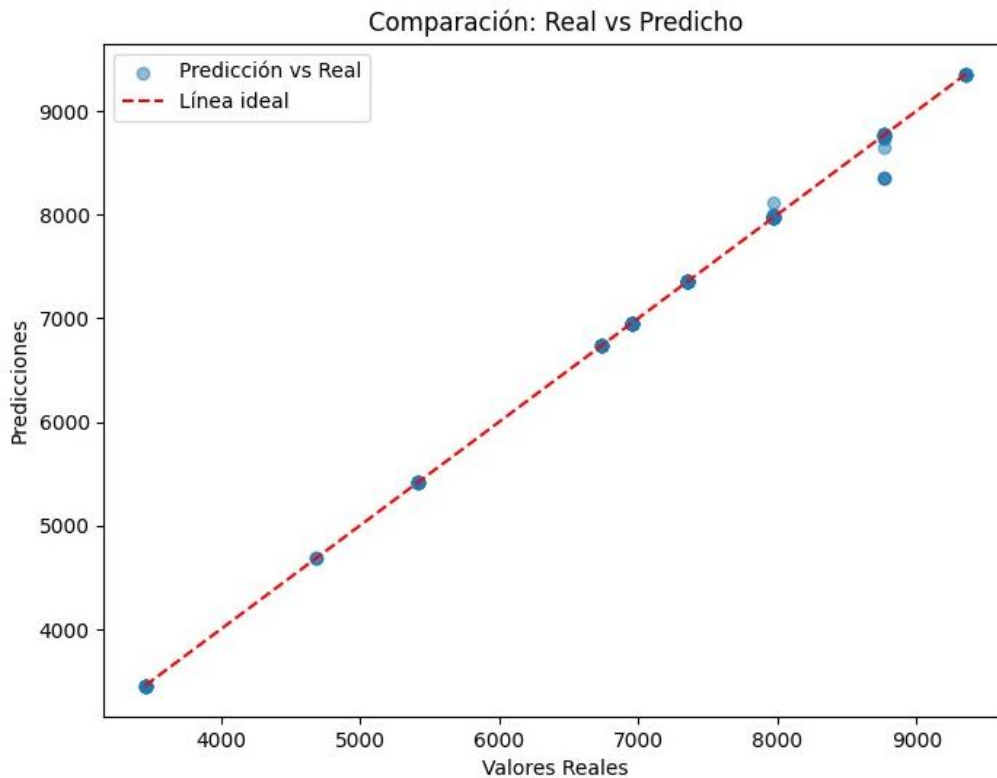


Figura 6. Regresión lineal de los datos de entrenamiento en comparación con los datos de validación de predicción.

## Optimización y Evaluación del Modelo

El análisis de importancia de variables reveló que la altitud fue el factor más influyente, representando un 72% del peso total del modelo, seguida por las categorías climáticas con un 17.8%. Estas variables demostraron ser críticas para la predicción precisa de la erosión del suelo.

El modelo fue evaluado mediante gráficos comparativos entre valores reales y predichos, mostrando una alta alineación con la línea ideal. Este resultado confirmó la precisión del modelo y su utilidad como herramienta para la planificación y mitigación de riesgos de erosión.

## Conclusiones

El proyecto cumplió satisfactoriamente con los objetivos específicos establecidos, utilizando técnicas avanzadas de análisis de datos y *machine learning* para abordar la erosión del suelo en el departamento del Tolima. Los resultados destacan tanto por su precisión como por su aplicabilidad en la gestión de riesgos.

- 1. Análisis de la relación entre variables climáticas y erosión**  
Se identificaron relaciones clave entre variables climáticas y la erosión del suelo. La altitud mostró una correlación negativa significativa (-0.63) con las precipitaciones, indicando que las zonas más altas son menos propensas a recibir lluvias abundantes. Por otro lado, la latitud presentó una correlación positiva (0.54), sugiriendo que ciertas áreas presentan mayores acumulaciones de lluvia. Estas relaciones evidencian cómo factores geográficos y climáticos influyen directamente en los patrones de erosión.
- 2. Identificación de variables más influyentes**  
El análisis de importancia de variables determinó que la altitud es el factor más relevante, representando un 72% del peso del modelo predictivo. La categoría climática principal, con un peso del 17.8%, también demostró ser un indicador clave. Este enfoque permitió priorizar las variables que tienen un mayor impacto en la erosión, facilitando una interpretación clara y centrada en los factores más críticos.
- 3. Entrenamiento y validación del modelo**  
Se desarrolló y validó un modelo basado en *Random Forest Regressor*, optimizado mediante técnicas como validación cruzada y ajuste de hiperparámetros. El modelo alcanzó un  $R^2$  de 0.9975, lo que refleja su alta precisión para predecir tasas de erosión. A pesar de su rendimiento sobresaliente, se observaron variaciones en subconjuntos específicos durante la validación cruzada, lo que sugiere la necesidad de manejar mejor los *outliers* y aumentar la representatividad del conjunto de datos.
- 4. Propuesta de recomendaciones para la gestión de riesgos**  
Basados en los resultados del modelo, se proponen recomendaciones prácticas para la gestión de riesgos de erosión. Estas incluyen:
  - Priorizar acciones de mitigación en zonas de alta altitud con baja acumulación de lluvia.
  - Implementar medidas preventivas en áreas categorizadas como críticas según el análisis climático.
  - Usar el modelo predictivo para evaluar escenarios futuros bajo diferentes condiciones climáticas, apoyando la planificación territorial y la toma de decisiones estratégicas.

Además, se recomienda continuar recopilando datos históricos y monitorear de manera continua los cambios en las variables climáticas y geográficas para mantener la precisión del modelo y ampliar su aplicabilidad.

En conclusión, este proyecto representa un avance significativo en la comprensión de los factores que influyen en la erosión del suelo. La combinación de técnicas de *machine learning* y análisis de datos geoespaciales no solo permitió alcanzar los objetivos específicos, sino que también sentó una base sólida para futuras investigaciones y aplicaciones en regiones con problemas similares.

## Lista de referencias

- Bobadilla, J. (2020). *Machine learning y deep learning: usando python, scikit y keras*. Ediciones de la U. <https://www-ebooks7-24-com.bdbiblioteca.universidadean.edu.co/?il=10277>
- Borrelli, P., et al. (2017). Soil erosion risk assessment: A review. *Environmental Science & Policy*, 68, 31-39.
- Blanco-Canqui, H., & Lal, R. (2010). *Principles of soil conservation and management*. Springer.
- Brady, N. C., & Weil, R. R. (2008). *The nature and properties of soils* (14th ed.). Pearson Prentice Hall.
- Buol, S. W., Southard, R. J., Graham, R. C., & McDaniel, P. A. (2011). *Soil genesis and classification* (6th ed.). Wiley-Blackwell.
- Daniels, R. B., & Buol, J. A. (2019). *Principles of soil genesis and classification*. Oxford University Press.
- Dumanski, J., et al. (2014). Soil conservation in the context of climate change: A review. *Sustainable Agriculture Research*, 3(2), 38-49.
- Eslava, J. (1992). El perfil altitudinal de la temperatura en Colombia y sus implicaciones en la estabilidad del suelo. Universidad Nacional de Colombia.
- Folharini, S., Vieira, A., Bento-Gonçalves, A., Silva, S., Marques, T., & Novais, J. (2023). Soil erosion quantification using machine learning in sub-watersheds of Northern Portugal. *Hydrology*, 10(1), 7. <https://doi.org/10.3390/hydrology10010007>
- Fundación Alejandro Ángel Escobar. (2015). *Estudio nacional de la degradación de suelos por erosión en Colombia*. Recuperado de <https://www.faae.org.co/premios-nacionales/ciencias-2017-estudio-nacional-de-la-degradacion-de-suelos-por-erosion-2015/>
- Khan, M. A., et al. (2020). Impact of climate on soil erosion and conservation. *Soil & Tillage Research*, 203, 104635.

- Miller, J. R., et al. (2012). Soil texture and erosion: A review. *Soil Science Society of America Journal*, 76(1), 1-12.
- Ramos Castillo, L. A., & Orozco Medina, I. (2020). Modelado de la producción de sedimentos en una cuenca con poca información incluyendo los potenciales efectos del cambio climático y el cambio de uso de suelo. *Acta Universitaria*, 30, 1-19. <https://doi.org/10.15174/au.2020.2901>
- Reddy, K. R., & Khanna, S. M. (2019). *Soil science: Agricultural and environmental perspectives*. Springer.
- Schaetzl, R. J., & Anderson, S. (2005). *Soils: Genesis and geomorphology*. Cambridge University Press.
- Smith, J. A., et al. (2011). The effects of geology on soil formation and erosion. *Geoderma*, 163(3-4), 255-267.
- Torres-Benítez, A. J., Esquivel, H. E., & Tinoco, F. (2017). Composición física y química de los suelos fluviovolcánicos de Armero, Tolima, Colombia. *Revista de la Academia Colombiana de Ciencias Exactas, Físicas y Naturales*, 41(158), 119-126.
- Universidad del Tolima. (2021). Investigaciones sobre el uso del suelo y su impacto en el medio ambiente en Santa Isabel, Tolima. Recuperado de <https://www.ut.edu.co/>
- Wischmeier, W. H., & Smith, D. D. (1978). *Predicting rainfall erosion losses: A guide to conservation planning*. USDA.
- Agencia Europea de Medio Ambiente. (2019). Impactos del cambio climático en la erosión del suelo en Europa.
- Huang, Y., et al. (2019). Impact of climate change on soil erosion: A review. *Land Degradation & Development*, 30(7), 745-762.
- Abba, S. I., Ibrahim, Y. O., & Edicha, J. (2017). Application of data-driven techniques in predicting soil erosion vulnerability in a semiarid watershed. *Modeling Earth Systems and Environment*, 3(4), 1585-1599.
- Behera, M. D., Borah, B., Yadav, V. P., & Ahmed, R. (2018). Soil erosion-susceptibility mapping and prioritization of sub-watersheds using machine learning algorithms. *Journal of Earth System Science*, 127(7), 95.
- Borrelli, P., Panagos, P., Märker, M., Modugno, S., Schütt, B., & Montanarella, L. (2020). Assessment of the impacts of clear-cutting on soil loss by water erosion in Italian forests: First comprehensive monitoring and modelling approach. *Catena*, 187, 104396.
- Chen, W., Zhao, X., Wang, S., & Xie, X. (2019). Analyzing urban flood vulnerability using remote sensing and GIS: A case study of Shanghai, China. *Sustainability*, 11(22), 6263.

- Liu, B., Zhang, K., Zhang, G., & Liu, G. (2018). Soil erosion prediction using Rusle in the GIS environment: A case study in the Danjiangkou Reservoir area of China. *Environmental Earth Sciences*, 59(3), 679-687.
- Meena, S. R., Kumar, A., & Tripathi, N. K. (2019). Soil erosion assessment using remote sensing data and artificial neural network modeling. *Environmental Earth Sciences*, 78(15), 432.
- Rahmati, O., Samadi, M., Keshavarzi, A., & Sadeghi, S. H. (2017). Application of GIS-based data driven random forest and maximum entropy models for groundwater potential mapping: A case study at Mehran Region, Iran. *Catena*, 155, 45-56.
- Tien Bui, D., Khosravi, K., Panahi, M., Chen, W., Shirzadi, A., Shahabi, H., ... & Li, S. (2020). Evaluation of the applicability of deep learning neural networks to spatial prediction of landslide susceptibility. *Landslides*, 17(1), 227-245.
- Youssef, A. M., Pradhan, B., & Hassan, A. M. (2016). Flash flood susceptibility assessment in Wadi Sudr Basin, Egypt using an integrated GIS-based model. *Natural Hazards*, 84(1), 389-410.
- Restrepo Leal, D. A., Vilorio Porto, J. P., & Robles Algarín, C. A. (2021). El camino a las redes neuronales artificiales. Editorial Unimagdalena.
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273-297.
- Haykin, S. (2009). *Neural Networks and Learning Machines* (3ra ed.). Prentice Hall.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer.
- Melesse, A. M., Wang, X., & McClain, M. E. (2011). Modeling soil erosion and sediment yield using the WEPP model: Case study of the Mara River Basin, Kenya and Tanzania. *International Journal of Sediment Research*, 26(3), 248-257.
- Mousavi, S. R., Singh, V. P., & Seo, Y. (2019). Modeling soil erosion using machine learning techniques: Comparative study between ANN, ANFIS, SVM, and GEP. *Environmental Monitoring and Assessment*, 191(9), 554.
- Murphy, K. P. (2012). *Machine Learning: A Probabilistic Perspective*. MIT Press.
- Pérez, D., González, E., & Rodríguez, M. (2021). Modelado de la erosión del suelo en el Tolima utilizando máquinas de soporte vectorial. *Ingeniería y Desarrollo*, 39(3), 65-80.
- Sarma, K., Saikia, R., & Sarma, B. (2018). Soil erosion susceptibility mapping using remote sensing and GIS techniques: A case study of Dihing River Basin, Assam, India. *Environmental Earth Sciences*, 77(14), 513.
- Schölkopf, B., & Smola, A. J. (2002). *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press.

- Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing*, 14(3), 199-222.
- Vapnik, V. N. (1998). *Statistical Learning Theory*. Wiley.
- Abba, S. I., Ibrahim, Y. O., & Edicha, J. (2017). Application of data-driven techniques in predicting soil erosion vulnerability in a semiarid watershed. *Modeling Earth Systems and Environment*, 3(4), 1585-1599.
- Akgun, A., Erkan, O., & Türk, N. (2012). Application of GIS-based fuzzy logic and analytical hierarchy process (AHP) approaches for landslide susceptibility mapping: A case study of Sultan Mountains (Akşehir, Turkey). *Environmental Earth Sciences*, 66(2), 319-333.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32.
- Chen, W., Li, W., Chai, H., Zhao, Z., & Chen, G. (2017). Mapping the susceptibility of rainfall-induced landslides using a novel ensemble model: A case study in Sichuan Province, China. *Environmental Earth Sciences*, 76(16), 567.
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273-297.
- Haykin, S. (2009). *Neural Networks and Learning Machines* (3rd ed.). Prentice Hall.
- Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1), 55-67.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning*. Springer.
- Mountrakis, G., Im, J., & Ogole, C. (2011). Support vector machines in remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(3), 247-259.
- Montgomery, D. C., Peck, E. A., & Vining, G. G. (2008). *Introduction to Linear Regression Analysis* (4th ed.). Wiley.
- Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267-288.
- Tibshirani, R. (2014). *Introduction to Statistical Learning with Applications in R*. Springer.
- Zhu, L., Lei, T., Wang, Q., Liu, H., & Yang, X. (2018). Modeling soil erosion and its response to the check dams in the Loess Plateau of China. *Environmental Earth Sciences*, 77(13), 471.