



UNIVERSIDAD TECNOLÓGICA ECOTEC

TEMA:

COMPARAR ALGORITMOS DE MACHINE LEARNING PARA LA PREDICCIÓN DE ÁREAS DE MAYOR RIESGO DE DESLIZAMIENTOS DE TIERRA EN ZONAS DEFORESTADAS.

LÍNEA DE INVESTIGACIÓN: TECNOLOGÍAS DE LA INFORMACIÓN Y COMUNICACIÓN.

MODALIDAD DE TITULACIÓN: PROPUESTA TECNOLÓGICA

CARRERA

INGENIERÍA EN SISTEMAS

TÍTULO A OBTENER

INGENIERO EN SISTEMAS CON ÉNFASIS EN ADMINISTRACIÓN DE REDES

AUTOR

LUIS EFRÉN HINOSTROZA CABRERA

TUTOR

Mgtr. MANUEL RAMÍREZ PIREZ.

SAMBORONDÓN - ECUADOR

2023

DEDICATORIA

Dedico este trabajo a mis padres, migrantes quienes por los años 80 buscaron un nuevo futuro y que incansablemente, de domingo a domingo han trabajado para sacar adelante a tres hijos, padres maravillosos quienes hasta lo último siguieron impulsándome hacia adelante, viéndome cumplir mis sueños deportivos, vocacionales y en esta ocasión profesionales, los cuales a partir de hoy nos abren un nuevo mundo.

Padres pronto nos encontraremos allá en el norte y viviremos en ese mundo que si es para nosotros.

Luis Efrén Hinostroza Cabrera

AGRADECIMIENTO

Primeramente agradezco a Dios por la bendición de haber culminado esta etapa y por la lucidez con la cual he podido llevarla a un feliz término, agradezco a mis padres quienes me formaron y con su ejemplo gestaron en mí el carácter lo suficientemente fuerte para afrontar cualquier reto que me proponga, a mi hermano, hombre soñador y aventurero quien abrió las puertas de ese futuro soñado y que desde lejos no dejó nunca de pensar en mí y apoyarme en lo que fuera necesario, a mi hermana, quien siendo una niña muy frágil siempre pudo sostener ese buen ánimo con el cual me seguían demostrando que yo si podía vivir los mismo sueños que ellos, agradezco a mi mujer Fernanda Herrera quien ha tenido que sobrellevar muchas noches de traspasado, mientras yo procesaba algún dato o revisaba algún documento ella intentaba dormir pero siempre pendiente por si me faltaba algo, por ultimo y no menos importante a mi Hela quien moviendo la colita siempre me daba ánimos con su inagotable sonrisa, a Vito y Nilo quienes con sus travesuras animaban esas noches de estudio.

Luis Efrén Hinostroza Cabrera

**CERTIFICADO DE APROBACIÓN DEL TUTOR PARA LA PRESENTACIÓN A
REVISIÓN DEL TRABAJO DE TITULACIÓN**

ANEXO N° 14

**CERTIFICADO DE APROBACIÓN DEL TUTOR PARA LA PRESENTACIÓN A
REVISIÓN DEL TRABAJO DE TITULACIÓN**

Samborondón, 7 de julio de 2023

Magíster
Erika Ascencio Jordán
Decana de la Facultad
Ingeniería
Universidad Tecnológica ECOTEC

De mis consideraciones:

Por medio de la presente comunico a usted que el trabajo de titulación TITULADO: ALGORITMOS DE MACHINE LEARNING PARA LA PREDICCIÓN DE ÁREAS DE MAYOR RIESGO DE DESLIZAMIENTOS DE TIERRA EN ZONAS DEFORESTADAS según su modalidad, PROPUESTA TECNOLÓGICA ; fue revisado, siendo su contenido original en su totalidad, así como el cumplimiento de los requerimientos establecidos en la guía para la elaboración del trabajo de titulación, Por lo que se autoriza a: Hinostriza Cabrera Luis Efrén para que proceda a su presentación para la revisión de los miembros del tribunal de sustentación.

ATENTAMENTE,



Firmado electrónicamente por:
**MANUEL OSMANY
RAMIREZ PIREZ**

Ing. Manuel Ramírez Pírez, Msc

Tutor

CERTIFICADO DEL PORCENTAJE DE COINCIDENCIAS

ANEXO N°15

CERTIFICADO DEL PORCENTAJE DE COINCIDENCIAS

Habiendo sido nombrado Manuel Osmany Ramírez Pérez, tutor del trabajo de titulación ALGORITMOS DE MACHINE LEARNING PARA LA PREDICCIÓN DE ÁREAS DE MAYOR RIESGO DE DESLIZAMIENTOS DE TIERRA EN ZONAS DEFORESTADAS elaborado por: LUIS EFRÉN HINOSTROZA CABRERA, con mi respectiva supervisión como requerimiento parcial para la obtención del título de Ingeniero en Sistemas. Se informa que el mismo ha resultado tener un porcentaje de coincidencias 3 (%) mismo que se puede verificar en el siguiente link: (copiar y pegar el https del resultado). Adicional se adjunta print de pantalla de dicho resultado.



CERTIFICADO DE ANÁLISIS
magister

ALGORITMOS DE MACHINE LEARNING PARA LA PREDICCIÓN DE ÁREAS DE MAYOR RIESGO DE DESLIZAMIENTOS DE TIERRA EN ZONAS DEFORESTADAS.

3% Similitudes

< 1% Texto entre comillas
0% similitudes entre comillas

3% Idioma no reconocido

Nombre del documento: ALGORITMOS DE MACHINE LEARNING PARA LA PREDICCIÓN DE ÁREAS DE MAYOR RIESGO DE DESLIZAMIENTOS DE TIERRA EN ZONAS DEFORESTADAS.docx ID del documento: 18e4d39dc03b1b789a3da571829fdba2f60220df Tamaño del documento original: 786,81 kB	Depositante: MANUEL OSMANY RAMIREZ PIREZ Fecha de depósito: 7/7/2023 Tipo de carga: interface fecha de fin de análisis: 7/7/2023	Número de palabras: 9411 Número de caracteres: 67.469
---	---	--

Ubicación de las similitudes en el documento:





Firmado electrónicamente por:
**MANUEL OSMANY
RAMIREZ PIREZ**

FIRMA DEL TUTOR

NOMBRES Y APELLIDOS DEL TUTOR

CERTIFICADO DE APROBACIÓN DEL TUTOR

ANEXO N°16

CERTIFICADO DE APROBACIÓN DEL TUTOR PARA LA PRESENTACIÓN DEL TRABAJO DE TITULACIÓN CON INCORPORACIÓN DE LAS OBSERVACIONES DE LOS MIEMBROS DEL TRIBUNAL

Samborondón, 21 de agosto de 2023

Magíster

Erika Ascencio Jordan

Decana de la Facultad

Ingenierías.

Universidad Tecnológica ECOTEC

De mis consideraciones:

Por medio de la presente comunico a usted que el trabajo de titulación TITULADO: **COMPARAR ALGORITMOS DE MACHINE LEARNING PARA LA PREDICCIÓN DE ÁREAS DE MAYOR RIESGO DE DESLIZAMIENTOS DE TIERRA EN ZONAS DEFORESTADAS.** según su modalidad PROPUESTA TECNOLÓGICA; fue revisado y se deja constancia que el estudiante acogió e incorporó todas las observaciones realizadas por los miembros del tribunal de sustentación por lo que se autoriza a: **LUIS EFRÉN HINOSTROZA CABRERA,** para que proceda a la presentación del trabajo de titulación para la revisión de los miembros del tribunal de sustentación y posterior sustentación.

ATENTAMENTE,



Firmado electrónicamente por:
**MANUEL OSMANY
RAMIREZ PIREZ**

Mgtr/ Manuel Ramírez Pirez

Tutor

RESUMEN

La Organización de las Naciones Unidas (ONU) manifiesta que uno de los problemas sobre los diversos ecosistemas del planeta es la deforestación, a causa de una destrucción descontrolada de árboles sin gestiones de contingencia como la reforestación, destaca que son aproximadamente 10 millones de hectáreas anuales de bosques perdidos.

De acuerdo al Centro para el Control y Prevención de Enfermedades (CDC, 2018) describe a los deslizamientos de tierra como un conjunto de tierra, rocas y descomposición de masa sólida que caen desde una pendiente, adjunto a esto también perciben el nombre de aludes de barro que comúnmente forman canales debido a su gran descomposición de tierra y mayor parte de agua.

Machine Learning, traducido al español se define como aprendizaje automático, se concibe como parte esencial de la inteligencia artificial y como una técnica relacionada a la automatización y detección de patrones sobre una agrupación de datos cualitativos y cuantitativos mediante la integración de sistemas matemáticos, estadísticas e informáticos, se refiere al machine learning como un complemento o subcategoría de la inteligencia artificial que incluye herramientas estadísticas y algoritmos con la finalidad de proveer a las máquinas un nivel de capacidad de aprendizaje y garantizar resultados eficientes sobre una orden específica luego de un procesamiento de información sin previa aprobación e intervención externa por parte del programador.

Una red neuronal artificial (por sus siglas RNA) está basada en un mecanismo de aprendizaje automático cuya inspiración es el cerebro humano, así logran establecer relaciones no lineales entre las variables de entrada y las variables de salida. Una de las principales ventajas de esta técnica de machine Learning es que permite analizar, clasificar y procesar la información en patrones complejos.

Palabras clave: Deforestación, deslizamiento de tierra, inteligencia artificial, aprendizaje automático, redes neuronales, predicción

ABSTRACT

The United Nations (UN) states that deforestation is one of the problems affecting diverse ecosystems on the planet, caused by uncontrolled destruction of trees without contingency measures such as reforestation. Approximately 10 million hectares of forests are lost annually.

According to the Centers for Disease Control and Prevention (CDC, 2018), landslides are described as a combination of soil, rocks, and solid mass decomposition that fall from a slope. They are also commonly known as mudslides, which often form channels due to extensive soil decomposition and high water content.

Machine Learning, translated into Spanish as "aprendizaje automático," is an essential part of artificial intelligence and a technique related to automation and pattern detection in a collection of qualitative and quantitative data through the integration of mathematical, statistical, and computer systems. It refers to machine learning as a complement or subcategory of artificial intelligence that includes statistical tools and algorithms to provide machines with a level of learning capacity and ensure efficient results without prior approval or external intervention from the programmer.

An Artificial Neural Network (ANN) is based on a machine learning mechanism inspired by the human brain, allowing it to establish nonlinear relationships between input and output variables. One of the main advantages of this machine learning technique is its ability to analyze, classify, and process information in complex patterns.

Keywords: deforestation, landslide, artificial intelligence, machine learning, neural networks, prediction.

ÍNDICE GENERAL

DEDICATORIA.....	I
AGRADECIMIENTO	II
CERTIFICADO DE APROBACIÓN DEL TUTOR PARA LA PRESENTACIÓN A REVISIÓN DEL TRABAJO DE TITULACIÓN.....	III
CERTIFICADO DEL PORCENTAJE DE COINCIDENCIAS	IV
CERTIFICADO DE APROBACIÓN DEL TUTOR	V
RESUMEN.....	VI
ABSTRACT	VII
ÍNDICE GENERAL	8
ÍNDICE DE TABLAS.....	XI
ÍNDICE DE FIGURAS	XII
INTRODUCCIÓN	1
Antecedentes.....	1
Planteamiento del problema	3
Objetivo general.....	6
Objetivos específicos	6
CAPÍTULO 1: MARCO TEÓRICO.....	8
Deforestación.....	8
Deslizamiento de tierra	8
Machine learning	10
Fases de Machine learning.....	11
Tipos de Machine learning	12
1. Aprendizaje supervisado.....	12
2. Aprendizaje no supervisado.....	13
3. Aprendizaje reforzado.....	14

Algoritmos de Machine learning.....	15
1. Algoritmo de clasificación.....	15
2. Algoritmo de regresión.....	16
3. Árbol de decisiones.....	17
4. Redes neuronales.....	18
5. Bosques aleatorios.....	20
CAPÍTULO 2: METODOLOGÍA.....	24
Diseño de la investigación.....	24
Tipo de investigación.....	24
Enfoque de investigación.....	24
Operacionalización de variables.....	25
Población y muestra.....	25
Técnicas e instrumentos de investigación.....	25
CAPÍTULO 3: ANÁLISIS E INTERPRETACIÓN DE RESULTADOS.....	29
Análisis de las entrevistas.....	29
CAPÍTULO 4: IMPLEMENTACIÓN DE LA SOLUCIÓN TECNOLÓGICA.....	32
Objetivo.....	32
Desarrollo.....	32
1. Sectorización de áreas bajo estudio.....	32
2. Desarrollar la base de datos.....	33
3. Ejecutar un reprocesamiento de la información de la base de datos...34	34
4. Designación de las variables de entrada.....	35
5. Desarrollar de los algoritmos en Python.....	36
a. Modelo de Regresión Logística.....	36
b. Modelo de Bosques Aleatorios (random forest).....	42
c. Modelo de Redes Neuronales.....	45
d. Modelo de Aprendizaje Automático Supervisado XGBoost.....	50

CONCLUSIONES	53
RECOMENDACIONES	57
Bibliografía	58
ANEXOS	65

ÍNDICE DE TABLAS

Tabla 1 Matriz de operacionalización de variables	26
Tabla 2 Matriz de analisis de las entrevistas- Parte 1	29
Tabla 3 Matriz de analisis de las entrevistas-Parte 2	30
Tabla 4 Registros de zonas para clasificación de su nivel de severidad	32
Tabla 5 Base de datos.....	34
Tabla 6 Fases de reprocesamiento de datos	35
Tabla 7 Variables de entrada para la predicción de deslizamientos	35
Tabla 8 Comparación de resultados de las matrices de confusión	55

ÍNDICE DE FIGURAS

Figura 1 Diagrama de Machine Learning	11
Figura 2 Etapas principales en el machine learning.....	12
Figura 3 Proceso del aprendizaje supervisado	13
Figura 4 Proceso del aprendizaje no supervisado	14
Figura 5 Características generales y diferencias entre cada tipo de machine learning	15
Figura 6 Función matemática y grafica de la regresión logística	16
Figura 7 Ejemplo de árbol de decisión basado en el tipo y nombre de animal .	17
Figura 8 Ventajas y desventajas de los árboles de decisión	18
Figura 9 Características de un RNA	19
Figura 10 Capas o topología de un RNA	20
Figura 11 Representación gráfica de la topología de un RNA	20
Figura 12 Diferencia entre árbol de decisión y bosques aleatorios.....	21
Figura 13 Proceso de elaboración y estructura de un Random Forest	22
Figura 14 Importamos las librerías y llamamos las bases de datos.....	36
Figura 15 Presentación de datos en un datasheet.....	36
Figura 16 Separación de datos y segregación para el entrenamiento y prueba	37
Figura 17 Validación y remplazo de datos	37
Figura 18 Algoritmo de Regresión Logística	37
Figura 19 Importación de la matriz de confusión.	38
Figura 20 Matriz de Confusión de Regresión Logística 1.....	39
Figura 21 Se balancean los datos para obtener mejores resultados	39
Figura 22 Matriz de Confusión de Regresión Logística 2.....	40
Figura 23 Se balancean los datos una vez más para obtener mejores resultados	41
Figura 24 Matriz de Confusión de Regresión Logística 3.....	42
Figura 25 Importamos las librerías y llamamos las bases de datos.....	42
Figura 26 Presentación de datos en un datasheet.....	43
Figura 27 Separación de datos y segregación para el entrenamiento y prueba	43

Figura 28 Validación de datos y relleno de datos vacíos	43
Figura 29 Inicialización del entrenamiento y balanceo del peso de las variables	44
Figura 30 Importación de la Matriz de Confusión.....	44
Figura 31 Matriz de Confusión de Bosques Aleatorios (random forest)	45
Figura 32 Importación de librerías, inicialización de base de datos y validación de datos	45
Figura 33 Despliegue del Datasheet.....	46
Figura 34 Cambio de datos y balanceo para implementación de la red neuronal	46
Figura 35 Importación de librerías e inicio del entrenamiento.....	47
Figura 36 Entrenamiento de la red neuronal en 100 pasos	47
Figura 37 Reevaluación de datos y llamado a presentar la predicción	49
Figura 38 Importación de la Matriz de Confusión.....	49
Figura 39 Matriz de Confusión de Redes Neuronales	50
Figura 40 Importación de librerías, inicialización de base de datos y validación de datos	50
Figura 41 Balanceo de datos e inicio del proceso de aprendizaje	51
Figura 42 Despliegue de la Matriz de Confusión	51
Figura 43 Matriz de Confusión de XGBoost.....	52

INTRODUCCIÓN

Antecedentes

En los últimos años, la tendencia tecnológica se direcciona hacia el aprendizaje automático como solución a muchos problemas de predicciones y análisis de datos, según Pal (2023) el aprendizaje automático, referido comúnmente en sus términos en inglés “machine learning”, es una subdivisión de la inteligencia artificial que ayuda a una máquina a aprender automáticamente de datos anteriores con resultados de predicción más precisos, además que su uso e implementación se integra en soluciones enfocadas en diferentes áreas de la ciencia entre las que se destacan ciencias de la salud, ingeniería, economía, marketing, la gestión de riesgos y desastres entre otras. Mientras que Díaz (2021) bajo su revisión conceptual propuesta desde varias perspectivas de autores, describe al machine learning o aprendizaje automático como un área de estudio donde las computadoras programan o predicen desde datos históricos o tendencias.

En base al enfoque del presente estudio Pal (2023) menciona que describe que la aplicación de Machine Learning en la gestión de desastres naturales es un campo de creciente importancia y prometedor. Por ende, pueden ayudar en la predicción de eventos naturales, como terremotos, tsunamis, huracanes, inundaciones, entre otros. Además, también puede ayudar en la planificación y la respuesta a estos eventos, proporcionando información valiosa y en tiempo real.

Así como también, la predicción de eventualidades a causa de la deforestación, el cual se da por la tala de árboles en una zona boscosa, eliminando así la cubierta natural que cubre los suelos de la inclemencia climática, es decir, la cubierta que aísla y protege del calor irradiante del sol, además, de la energía y caudal con la que la gota de lluvia llega al suelo, provocando así, la erosión de las áreas. En sí, la deforestación se produce por la expansión de la agricultura, la ganadería, la minería, la construcción de infraestructuras, y la explotación de madera y otros recursos forestales (Criado, 2018).

Por su parte, Gallo (2020) manifiesta que a la actualidad se están integrando estrategias para predecir e identificar áreas de mayor riesgo de afectaciones por deslizamientos de tierra mediante la combinación de técnicas de análisis de datos avanzadas como modelos de simulación y tecnologías innovadoras para la reducción de riesgos e impactos de los desastres naturales en la población, la infraestructura y la economía.

Tal como lo manifiesta Kuglitsch et. al. (2022) que el machine learning ha tomado fuerza en su implementación para la reducción y mitigación de riesgos de desastres naturales partiendo desde una predicción de situaciones extremas, elaboración de mapas de peligro, presentación de eventos en tiempo real para la aportación de información relevante y proponer una gestión efectiva de decisiones.

Por su parte, Ospina y Aristizábal (2021) explica en su estudio la factibilidad de integrar aplicaciones de inteligencia artificial mediante el uso de algoritmos de machine learning para la evaluación y determinación de movimientos en masa. Como resultados se evidenció que aquellos modelos de boosting manifiestan una mejor presentación de datos con respecto a desempeño y nivel de capacidad de predicción al ser contrastados con esquemas paramétricos lineales ya que su limitación trasciende en la presentación y análisis de problemas complejos.

Asimismo, Cepeda (2021) recalca que su estudio la importancia de conocer el origen de los deslizamientos ya que pueden ser translacionales o rotacionales con un estado sólido, líquido o semilíquido impulsados por la presencia de lluvias o vertientes de aguas cercanas, y así poder definir un sistema de alertas efectivo ya que se identificó la presencia de varias alertas con defectos, por ende, precisó la integración de susceptibilidad en relación a factores que condicionan un deslizamiento, adicionalmente, adjunto factores desencadenantes para la especificación de alertas.

En base a lo descrito, se estima un alto nivel de investigación sobre como prever deslizamientos de tierra ya que esto remite a la perdida de muchas vidas humanas o restricción de movilidad. Por lo tanto, se observa como expertos integran las nuevas tecnologías y sus componentes para lograr un sistema

efectivo de predicción sobre áreas susceptibles a deslizamientos y así prevenir el asentamiento de personas y viviendas.

Como parte de esta constante investigación se integra a Gallardo (2022) que bajo la implementación de algoritmos de machine learning como redes neuronales artificiales, bosque aleatorios y vectores de soporte, que contribuye a la identificación de zonas con alto nivel de probabilidad que ocurra una eventualidad geológica. Para lo cual integró 11 variables como: inventario de FRM, topografía, pendiente, aspecto, índice topográfico de humedad, índice de potencial de flujo, infiltración, precipitación, litología, cobertura vegetal y tipo de suelo que primero fueron gestionadas por el programa Rstudio, que como resultados de los tres algoritmos aplicados se determinó precisiones de 99.78%, 94.27% y 99.34%, donde el modelo con más precisión fue de redes neuronales artificiales.

En sí, trasciende la importancia de realizar un trabajo comparativo de algoritmos de machine learning para garantizar un mayor resultado de precisión sobre eventualidades como deslizamientos de tierra en zonas deforestadas para prevenir el asentamiento de poblaciones y viviendas, de tal manera que se erradique en un futuro casos de pérdidas humanas por deslizamientos.

Planteamiento del problema

A nivel global, se observa cambios inevitables en los integrantes del comercio o en las decisiones empresariales que a medida pasa el tiempo se evidencia que su crecimiento económico se debe al impulso de estar siempre a la vanguardia brindando oportunidades, además adjunto a esto se visualiza su interés por beneficiarse de las diversas tecnologías disponibles promoviendo a una constitución o creación de nuevos medios para evitar o prevenir cualquier inconveniente desde un contexto empresarial hasta social. Por ende, ante los cambios persistentes del mercado y entorno las empresas deben replantear su cadena de valor e implementar soluciones efectivas que incluyan las nuevas tecnologías para generar grandes resultados (Maisueche, 2019)

En definitiva, se convive en un entorno dinámico guiado por la tecnología y comunicación, lo cual ha impulsado a innovar y desarrollar grandes invenciones

como en un inicio fue el internet, que permitió la comunicación entre personas, familias u organizaciones a grandes distancias e impulso al desarrollo de perfiles empresariales hasta aplicaciones tecnológicas para presentar y dar a conocer estrategias como publicidad y promociones de bienes o servicios que brinda una entidad. Por lo tanto, extraer grandes soluciones y obtener efectivos resultados son una expectativa general y unánime a nivel mundial (Viteri et. al., 2018).

Como parte de esta innovación e integración de nuevas tecnologías se hace vital el conocimiento y manejo de capacidades y herramientas de inteligencia artificial, que un claro ejemplo es la industria automotriz, la cual integra la creación y fabricación de automóviles autónomos desde el año 2020. Históricamente, la presencia del concepto de inteligencia artificial fue promulgada por McCarthy en 1955, profesor de la Universidad de Standford, describiéndola como aquella ciencia que permite la creación de máquinas, productos o aplicaciones inteligentes. En sí, se describe que la evolución de esta inteligencia es motivada por factores característicos que impulsan a las personas a desarrollar sistemas que permitan resolver situaciones complejas (Villaseca & González, 2021).

Desde esta perspectiva, se ha desarrollado sistemas que han creado discusiones y campos de debates incluyendo implicaciones éticas, dicho avance parte desde la creación de la computación basado en códigos de programación concretas para lograr resultados deseados. Posteriormente, en los 70 un científico de la empresa IBM desarrolló una aplicación de autoaprendizaje, el cual consistía en potencializar los códigos del sistema para que ganen ante jugadores expertos de ajedrez de Estados Unidos, aunque esto no se evidencio hasta 1997 cuando por primera vez una maquina gana al campeón de ajedrez del mundo. Desde entonces, fue el punto de partida para la creación y evolución de la inteligencia artificial, dando origen a diversas áreas como de búsqueda y planificación, representación del conocimiento y razonamiento, machine learning, Deep learning, computer vision, natural language processing y robótica (Villaseca & González, 2021).

Por su parte, Florez & Pérez (2019) describe que el interés por predecir eventos naturales ha motivado a geocientíficos, ingenieros y ambientalistas buscar medios o sistemas predictivos que contribuyan a la mitigación de riesgos en la

población y su entorno. Se destaca que esta problemática toma fuerza desde los años 70 donde empezó la importancia y popularidad de desarrollar sistemas que aporten información y datos geográficos más certeros promoviendo que el mercado busque técnicas optimizadoras y precisas para predecir espacios vulnerables a producir deslizamientos.

Los deslizamientos de tierra son eventos en los cuales grandes cantidades de tierra se desplazan hacia abajo en una pendiente. Estos pueden ser causados por diferentes factores, incluyendo la lluvia intensa, la actividad sísmica, la erosión del suelo, la deforestación y la construcción inadecuada de infraestructura. Casamitjana & Carl, (2020) afirman que, los deslizamientos, desprendimientos rocosos y los movimientos en masa, en general, constituyen uno de los peligros naturales que más víctimas y daños materiales generan.

En Ecuador, los deslizamientos de tierra son un problema común debido a la topografía montañosa del país y la intensa actividad sísmica, en años recientes, el país ha experimentado varios eventos significativos de deslizamientos de tierra. Los deslizamientos (de tierra) hoy en día constituyen un grave problema que afectan a gran parte de la sociedad, estos problemas deben ser considerados a la hora de planificar, coordinar y ejecutar un ordenamiento del territorio (Medina & Astudillo, 2017).

En la actualidad el Ecuador no cuenta con un sistema de alerta temprana ni de análisis de un gran volumen de datos estadísticos y/o meteorológicos enfocados a la predicción y detección de zonas vulnerables a deslizamientos de tierra a fin de restringir las actividades económicas y sociales. El Ecuador históricamente ha crecido sin una planificación en lo que a soluciones habitacionales se refiere, al correcto uso del suelo y peor aún, el poco control de las invasiones y colonizaciones de terrenos que no están destinados para la habitabilidad, genera en la mayoría de los casos un aumento indiscriminado en la deforestación y erosión del suelo en esas zonas.

Esta propuesta tecnológica tiene como fin reconocer la utilidad del Machine Learning para ayudar en la predicción de desastres naturales, realizando una revisión, evaluación y monitoreo de indicadores históricos y actuales de las

zonas con mayor índice en deforestación, siendo esta una de las principales causales de los deslizamientos de tierra. Por lo tanto, la pregunta de investigación se centra: ¿Cuál será el algoritmo más efectivo de machine learning para garantizar una detección precisa de zonas vulnerables a deslizamientos de tierra?

Objetivo general

Comparar algoritmos de machine learning para la predicción de áreas de mayor riesgo de afectación por terremotos en el Ecuador.

Objetivos específicos

- Efectuar una revisión de la literatura científica acerca del uso de Machine Learning en la gestión de desastres naturales, enfocada en la predicción de deslizamientos de tierra.
- Identificar los factores de riesgo relacionados con los deslizamientos de tierra en el Ecuador relacionándolos con las técnicas de Machine Learning para la predicción de desastres naturales
- Diseñar una propuesta tecnológica que integre la aplicación de técnicas de Machine Learning para la predicción de deslizamientos de tierra en zonas vulnerables.

CAPÍTULO 1: MARCO TEÓRICO

Deforestación

De acuerdo a la Organización de las Naciones Unidas (ONU) manifiesta que uno de los problemas sobre los diversos ecosistemas del planeta es la deforestación, a causa de una destrucción descontrolada de árboles sin gestiones de contingencia como la reforestación, se destaca un aproximado de 10 millones de hectáreas anuales de bosques. Además, se describe como la pérdida y destrucción de vegetación (plantas o árboles) provocada por desastres naturales o por mano del hombre que degradan el proceso de regeneración de dichas áreas e incrementa el nivel de dióxido de carbono (Ambientum, 2023).

En base a perspectivas de entes gubernamentales internacionales se tenía como meta la reducción de deforestación para el año 2020 pero lamentablemente dicho objetivo no se cumplió y más bien se estima un recuperación y control de la deforestación en aproximadamente 25 años, además se ha diagnosticado una pérdida total de bosques de 178 millones hectáreas desde 1990. Pese a estos resultados, se observa un mayor control de bosques mediante la clasificación de áreas protegidas lo cual incremento en aproximadamente 535 millones hectáreas (ONU, 2020).

Entre los continentes que se destacan con la tasa más alta de pérdida de bosque se sitúa a África con una tasa anual en los últimos 10 años de 3.9 millones de hectáreas, seguida de Latinoamérica con un total de 2.6 millones de hectáreas. A nivel país, se destaca a Brasil, Bolivia, Paraguay, República Democrática del Congo, Indonesia, Angola, República Unida de Tanzania, Myanmar, Camboya, y Mozambique. Mientras que el continente que tuvo una mayor y eficiente gestión sobre la deforestación fue Asia, pero también se destaca países como Chile, Estados Unidos, China, Vietnam, Francia, Rumania, Italia, Turquía, Australia e India (ONU, 2020).

Deslizamiento de tierra

De acuerdo al Centro para el Control y Prevención de Enfermedades (CDC, 2018) describe a los deslizamientos como un conjunto de tierra, rocas y descomposición de masa solida que caen desde una pendiente, adjunto a esto

también perciben el nombre de aludes de barro que comúnmente forman canales debido a su gran descomposición de tierra y mayor parte de agua. Además, recalca que la causa principal de dichos eventos se debe por alteraciones naturales de pendientes, lomas, laderas, montañas o en zonas deforestadas provocadas por la erosión, presencia de lluvias, sismos y erupciones volcánicas.

Por su parte, Suarez (1998) describe a los deslizamientos desde distintas perspectivas como progresivo, de una sola masa, varias masas o masas semindependientes, provocados por desastres naturales o desestabilización de masas provenientes a cortes, erosiones o rellenos, deforestación, entre otros. En sí, se destacan oficialmente dos tipos de deslizamientos como rotacionales y translacionales, además conocer el tipo de dicho evento permite definir el sistema de análisis y estabilización idóneo.

Por ende, el deslizamiento rotacional implica la formación de una falla en forma de una U justo al centro de esta, encima del mismo centro de gravedad, que desde una vista lateral se observa grietas cóncavas y concéntricas en la dirección del deslizamiento. Mientras los de traslación se enfoca en agrietamientos con dirección hacia abajo o fuera de una superficie ligeramente plana, curvada o simplemente inclinada (Pilligua & Guaranda, 2021).

Está problemática natural y ambiental viene desde años atrás, donde se reconoce precedentes y casos fatídicos por la presencia de deslizamientos, en India durante el año 2013 presencié lluvias extremadamente fuertes que provocaron inundaciones y deslizamientos causando la muerte de aproximadamente 6000 personas. Uganda se presencié un deslizamiento de tierras cubriendo y arrastrando tres pueblos del monte Elgon con un aproximado de 350 muertes. En Afganistán, en el 2014 se suscita un deslizamiento de tierra y piedra provocando la reducción de un valle y sepultación de un pueblo con una pérdida de 350 personas. En el caso Favelas, Brasil en 2010 y 2011, a causa de lluvias se notó la presencia de deslizamientos y pérdidas humanas de aproximadamente 800 y 450, respectivamente (Díaz V. , 2017).

Machine learning

Machine Learning, traducido al español se define como aprendizaje automático, se concibe como parte esencial de la inteligencia artificial y como una técnica relacionada a la automatización y detección de patrones sobre una agrupación de datos cualitativos y cuantitativos mediante la integración de sistemas matemáticos, estadísticas e informáticos (Lee, 2019).

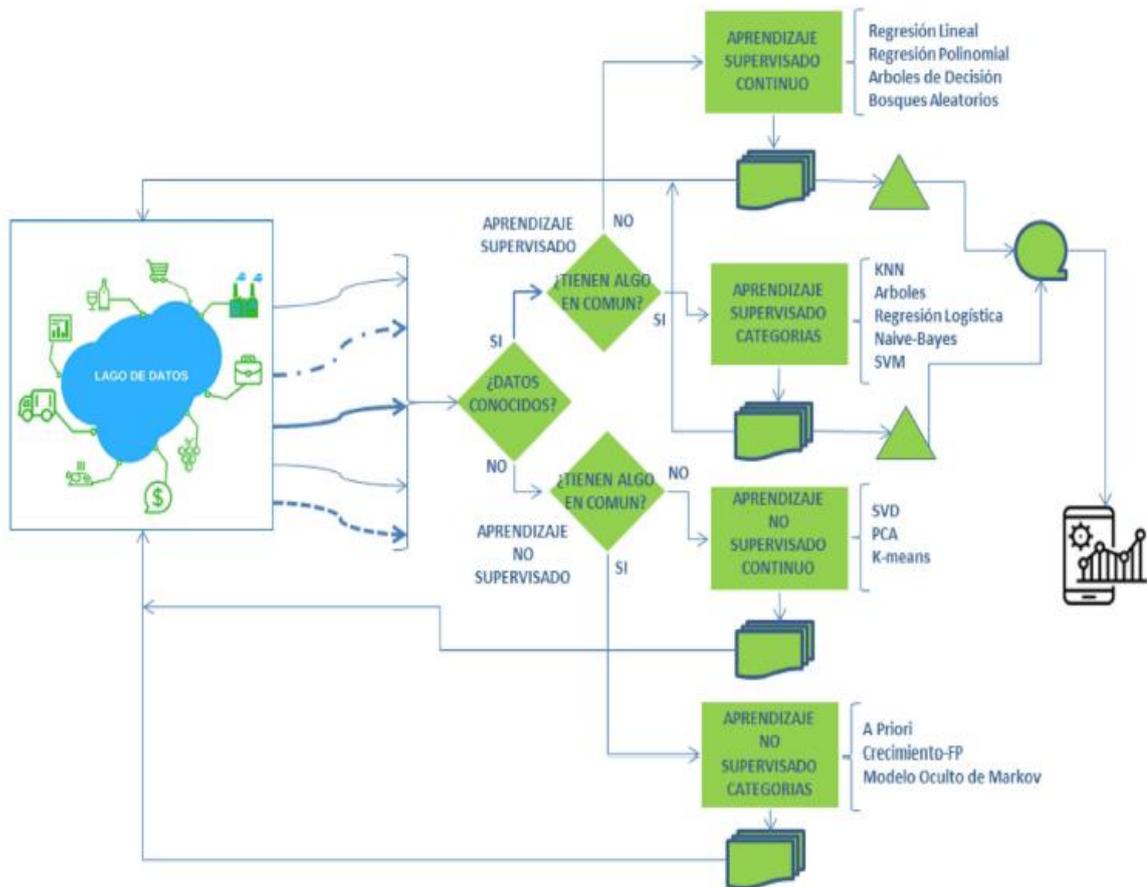
De igual manera, Álagos (2018) se refiere al machine learning como un conjunto de herramientas correspondientes a la inteligencia artificial que mediante la intervención de esquemas estadísticos se prevé búsquedas relevantes sobre patrones que permitan la fabricación de un artefacto capaz de aprender y ejecutar resultados basados en información empírica.

Por otra parte, el funcionamiento o proceso de machine learning parte de la existencia de un nivel idóneo y específico de conocimiento en máquinas, computadoras y dispositivos informáticos o digitales como si fuera un ser humano, adjunto a un sistema automático para el desarrollo de diversas actividades (Alba & Calle, 2020).

Por su parte, Núñez et. al. (2019) se refiere al machine learning como un complemento o subcategoría de la inteligencia artificial que incluye herramientas estadísticas y algoritmos con la finalidad de proveer a las máquinas un nivel de capacidad de aprendizaje y garantizar resultados eficientes sobre una orden específica luego de un procesamiento de información sin previa aprobación e intervención externa por parte del programador.

De tal manera, Alba y Calle (2020) manifiestan que la cantidad de información que se incluye o administra a dichas técnicas es demasiada y es vital contar con sistema que regula y verifique la seguridad y proveniencia de dicha información con la única intención de administrar información suficiente y necesaria para el mejoramiento e incremento del nivel de rendimiento de sistemas predictivos y obtener resultados efectivos para una consciente toma de decisiones bajo la ejecución de patrones y una gestión automática de datos.

Figura 1 Diagrama de Machine Learning



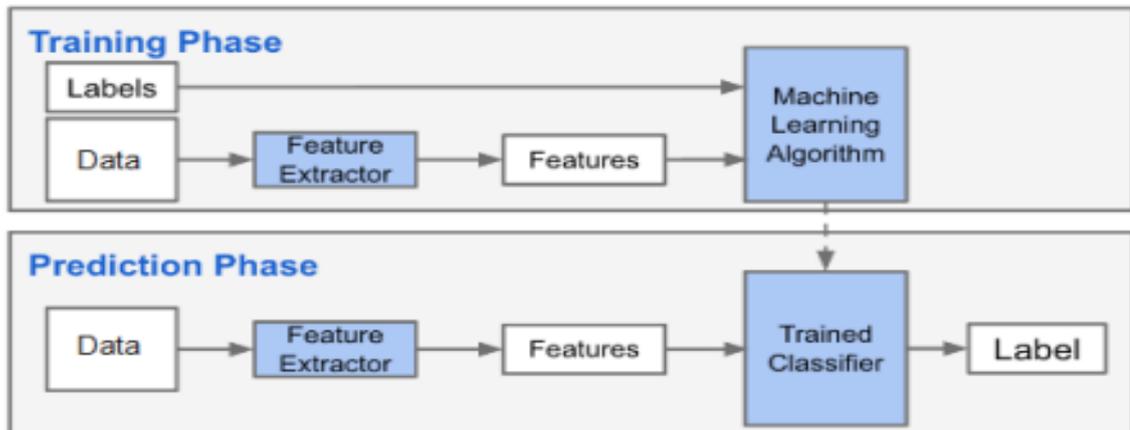
Nota. Tomado de Cepeda (2020, p. 28).

Fases de Machine learning

El machine learning se da por la presencia de dos etapas, la primera denominada de entrenamiento, donde puede poseer una gran cantidad de datos, por lo tanto, una parte podría ser dividida para entrenar el algoritmo y así ir recolectando toda la información necesaria que aprenden ellos, permitiendo encontrar comportamientos y patrones con resultados esperados (Sandoval, 2018).

Mientras que la segunda etapa se proceda a la gestión de prueba, la cual consiste en la validación de resultados con los datos que quedaron del entrenamiento para poder interrogar al algoritmo si es correcto o incorrecto lo que ha aprendido, en caso de que los datos no coincidan se tendrá que agregar datos o hacer uso de un método diferente, si coinciden entre el 80% a 90%, se podrá decir que ha aprendido y poder aplicar el algoritmo (Sandoval, 2018).

Figura 2 Etapas principales en el machine learning



Nota. Tomado de Alejandro y Figueroa (2020).

Tipos de Machine learning

Para la aplicación del machine learning es fundamental precisar de ciertas herramientas como algoritmos o codificación que permitan la administración eficiente de información, por ende, se describe que dicha técnica se utiliza para el entrenamiento de información y así, el modelo aplicado en el computador o máquina aprende y posteriormente, definir parámetros específicos para futuros análisis.

Por consiguiente, Klaine et. al. (2017) describen que el machine learning está integrado por tres tipos de algoritmos denominados como supervisado, no supervisado y reforzado o también denominado reinforcement learning. Por su parte, Bobadilla (2020) expresa la necesidad de herramientas de calidad para la identificación de tareas mediante la implementación de diversos algoritmos o tipos de machine learning, quién, además, agrupa dichos algoritmos en cuatro categorías y cuatro subcategorías: aprendizaje supervisado de regresión y clasificación, no supervisado por clúster o agrupamiento y por reducción de dimensiones, semisupervisado y por refuerzo.

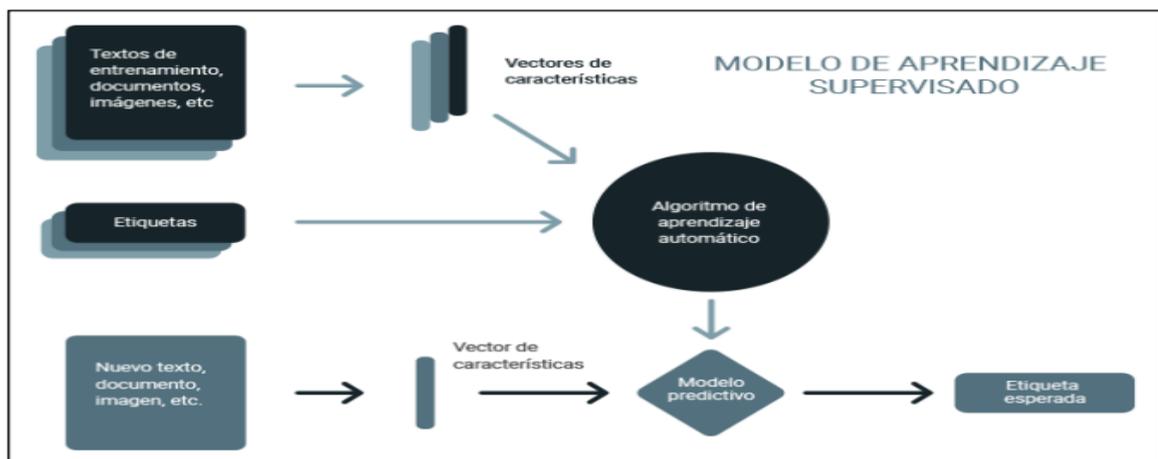
1. Aprendizaje supervisado

Dicho aprendizaje se centra en la gestión de información o datos con previa etiqueta sobre respuestas correctas, es decir que se registra información bajo

cierta etiqueta conforme se va ingresando con la finalidad de que los resultados obtenidos sean los más precisos posible (Ramírez et. al., 2020).

Mientras que Trejo y Miramá (2018) manifiestan la intervención de una persona como instructor para fomentar un proceso de entrenamiento en el sistema a utilizar mediante la inserción de acontecimientos que preserven información relacionada para evidenciar resultados precisos. Al igual que Maisueche (2019) y Gallardo (2022), describen que dicho algoritmo se origina con ayuda del programador, quién ingresará información para el proceso de entrada y salida como parte del entrenamiento del algoritmo y así definir resultados en base a la información de entrada.

Figura 3 Proceso del aprendizaje supervisado



Nota. Tomado de Ruiz (2019, p. 18).

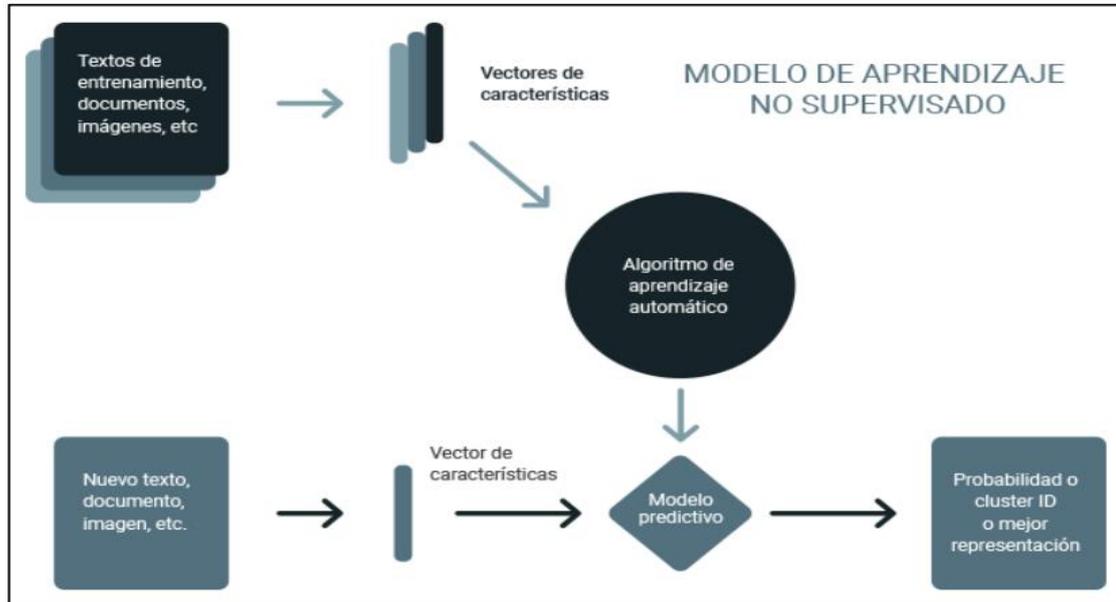
2. Aprendizaje no supervisado

En este caso, no se necesario la intervención de un desarrollador e instructor para la gestión de información de entrada o salida, solo se centra en características y no en etiquetas, por ende, al momento del análisis se ejecutará una agrupación por características (Sandoval, 2018).

Por su parte, Ruiz (2019) manifiesta que los ajustes del modelo predictivo se centran en el ingreso de información sin etiquetas, clasificaciones o distintivos específicos, sino que solo basta que analice la información mediante dos esquemas de entrenamiento que son el de imágenes y el algoritmo K-medias o

en inglés K-means. En sí, este algoritmo implica resultados eficientes basados en la identificación de características.

Figura 4 Proceso del aprendizaje no supervisado



Nota. Tomado de Ruiz (2019, p. 18).

Asimismo, Wenjian et. al. (2017) se enfocan sobre este tipo de machine learning como aquel modelo sin supervisión ni integración de distintivos o etiquetas, sino que posee la capacidad de arrojar y agrupar datos de salidas específicos basado en patrones similares.

Por ende, se entiende que el machine learning no supervisado tiene la capacidad de analizar y clasificar información conforme a características similares sin previa etiquetación al momento de ingresar dicha información.

3. Aprendizaje reforzado

Finalmente, Trejo y Miramá, (2018) manifiestan que el aprendizaje reforzado es muy parecido al de no supervisado, solo que es agregado un mecanismo de recompensa que se encarga de premiar o penalizar al sistema en base a la calidad de las decisiones ya sean estas buenas o malas, produciendo que el sistema se actualice continuamente por sí mismo.

Es un algoritmo que le permite a la máquina aprender en base a resultados con las acciones que le proporciona el supervisor, es decir, podrá interactuar

automáticamente con los datos aprendidos y decidir si la acción que eligió fue correcta o incorrecta (Alejandro & Figueroa, 2020).

Figura 5 Características generales y diferencias entre cada tipo de machine learning



Nota. Adaptado de Aprendizaje automático con Python de Raschka y Mirjalili (2019)

Algoritmos de Machine learning

1. Algoritmo de clasificación

Torres (2018) menciona que dicho aprendizaje se atribuye a una subcategoría de enfoque de clasificación, originada por el ingreso previo de datos que ya cuentan con la debida agrupación o clasificación para proveer de un modelo específico y eficiente bajo un esquema previo de procesamiento, características y diversos algoritmos, entre los que se destaca: redes neuronales artificiales, inferencia bayesiana, sistemas de soporte de vectores o árbol de decisión.

En efecto, Gallardo (2022) afirma que el aprendizaje supervisado por clasificación comprende modelos o grupos específicos acorde al tema u objeto en estudio dando como resultado una variable categórica o discreta, evidenciada como binaria, solo atribuye a dos valores o clasificaciones; múltiple, refleja entre dos o más agrupaciones; ordenada, en base a esquema de valor o de importancia.

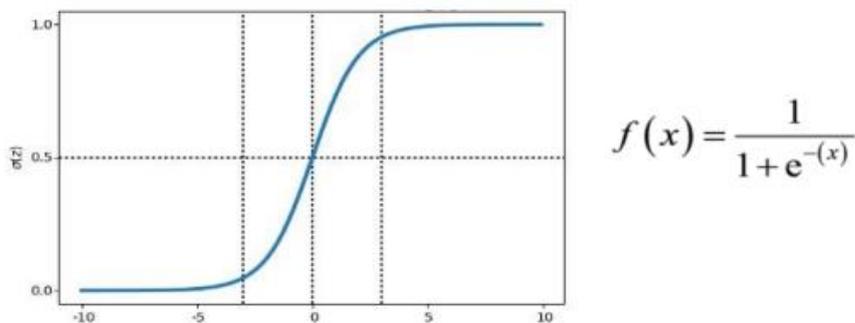
2. Algoritmo de regresión

Mientras que el aprendizaje supervisado por regresión no implica a una variable descrita dentro de una agrupación sino más bien que da como resultado un dato específico y numérico (Sandoval, 2018).

De acuerdo a Aimacaña y Columba (2021) se manifiesta dos tipos de algoritmos uno lineal y otro logístico, éste último es implementado en problemas de clasificación binomial y multinomial, para la predicción de una sola variable cualitativa y dependiente, es decir, “casa”, “gato”, “si o no”, de tal manera que se van ajustando los datos bajo una función logística.

Por lo tanto, esta gráfica designa dos valores comprendidos entre 0 y 1, donde si el resultado es 0.70 se interpreta como un nivel de probabilidad del 70% de que sea uno y el restante 30% se considera como las probabilidades de que el resultado sea 0. Por su parte, Lee (2019) también respalda que la regresión logística no consiste en la predicción numérica de un conjunto de entradas y sus resultados se centra entre 0 o 1.

Figura 6 Función matemática y grafica de la regresión logística



Nota. Tomado de Análisis comparativo de algoritmos de Machine Learning para la detección de plagas en los cultivos representativos de la sierra ecuatoriana por Aimacaña y Columba (2021).

Mientras que en el caso de la regresión lineal se centra en un análisis de predicción de variables cuantitativas entre una variable independiente y dependiente, que a su vez, identifica una regresión múltiple, la cual analiza una relación lineal entre dos o más variables independientes y una variable dependiente; regresión polinomial, donde existe una relación entre una variable

independiente y una dependiente bajo el uso de una función polinomial de n grado y por último, regresión múltiple polinomial, se manifiesta una relación y análisis de predicción entre dos o más variables independientes y una dependiente con la implementación de una función polinomial de n grado. Sin embargo, el algoritmo de regresión lineal es el más simple utilizado dentro del esquema del aprendizaje automatizado para las relaciones entre etiquetas y entidades (Lee, 2019).

3. Árbol de decisiones

De acuerdo a Sandoval (2018), dichos algoritmos resultan ser de alta precisión, estabilidad y sencillez para la realización de interpretaciones ya que se definen por construir reglas y normas de decisión como una visualización de forma de árbol, así mismo manifiesta el autor que pese a su alto nivel de capacidad predictiva pierde rendimiento.

Asimismo, Aimacaña y Columba (2021) describen que el algoritmo de árbol de decisiones se usa para temas de clasificación debido a que resulta ser el más comprensible por parte de un individuo. Identificándose nodos en su raíz como aquellos diversos atributos a ejecutar, las hojas reflejadas por los resultados mientras que las ramas son las normas de decisión, y por tal razón se deriva dicho nombre.

Figura 7 Ejemplo de árbol de decisión basado en el tipo y nombre de animal



Nota. Tomado de Análisis comparativo de algoritmos de Machine Learning para la detección de plagas en los cultivos representativos de la sierra ecuatoriana por Aimacaña y Columba (2021).

Figura 8 Ventajas y desventajas de los árboles de decisión

Ventajas	Desventajas
<ul style="list-style-type: none">• Fácil interpretación y visualización• Tratamiento de información heterogénea con atributos numéricos y categóricos.• No requieren de supuestos• Buena capacidad para la gestión de datos perdidos.	<ul style="list-style-type: none">• Tendencia de sobreajuste• Sensible a los cambios• Pequeñas variaciones pueden crear árboles distintos.• Dar resultados sesgados.• No siempre son óptimos

Nota. Adaptado de Árboles de decisión en R con Random Forest por Alaminos (2023).

De acuerdo a Alaminos (2023), el proceso de estructurar y crear arboles de decisiones integra una serie de fases que incluyen: seleccionar atributos, dividir nodos, podar el árbol. En relación a la selección de atributos se refiere a la identificación de características sobre un conjunto de datos, la cual proporcionará información para ejecutar la división respectiva del árbol, mediante la aplicación de entropía, ganancia de información, razón de ganancia, índice Gini y reducción del error.

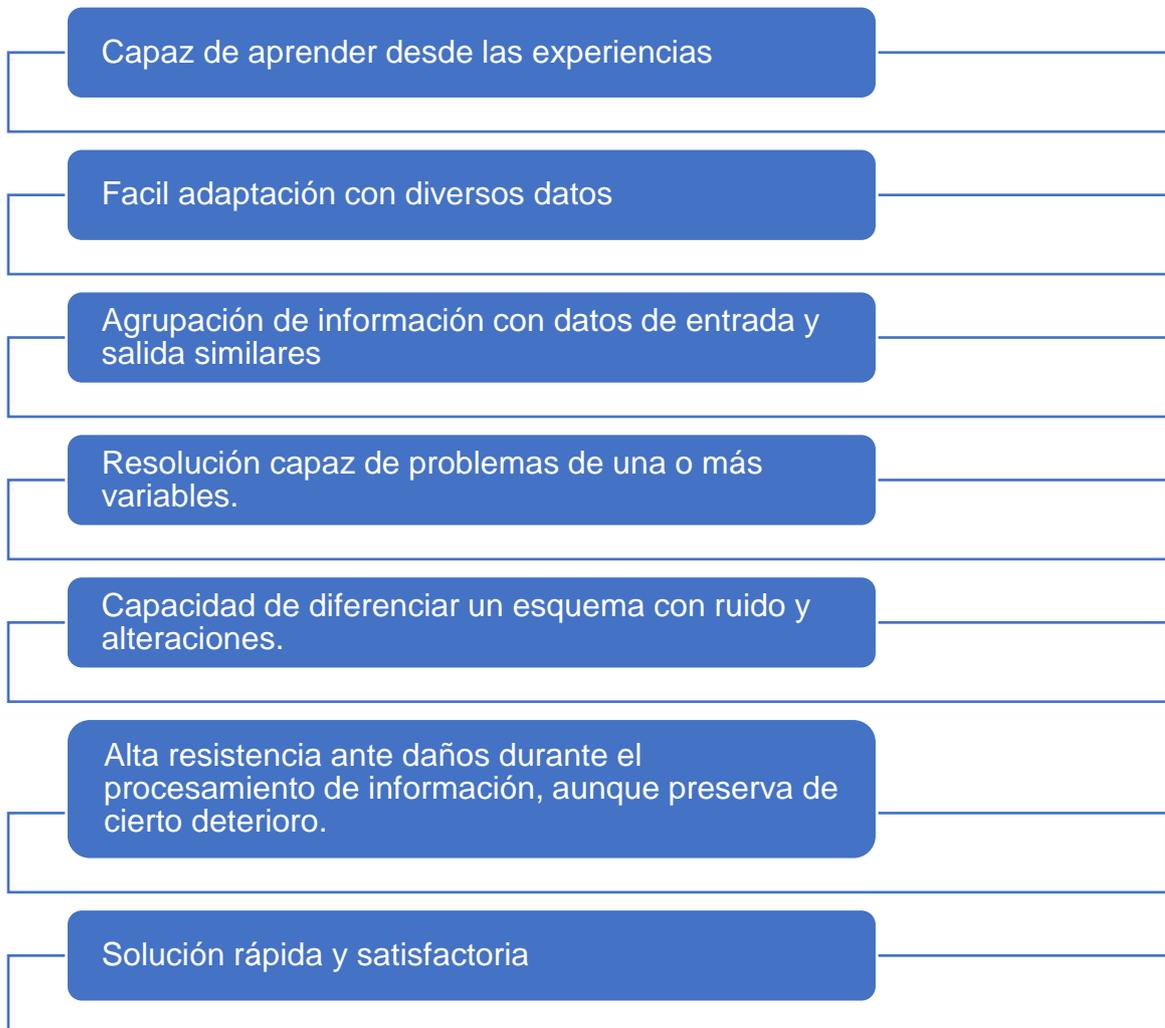
4. Redes neuronales

Una red neuronal artificial (por sus siglas RNA) está basada en un mecanismo de aprendizaje automático cuya inspiración es el cerebro humano, así logran establecer relaciones no lineales entre las variables de entrada y las variables de salida. Una de las principales ventajas de esta técnica de machine Learning es que permite analizar, clasificar y procesar la información en patrones complejos. La conexión entre los diferentes nodos que conforman una red neuronal se realiza a través de sinapsis, y a cada nodo se le da diferente peso dependiendo del impacto que tenga en la predicción que se desea realizar (Alba & Calle, 2020).

De acuerdo a Arias (2018) posee dos redes de conexión, la primera hacia delante, también conocida como Feed-forward networks son redes que ocurren en una sola dirección, las conexiones en los nodos no tienen ciclos o bucles. Mientras que la segunda se enfoca hacia atrás: También conocida como Feed-

back networks son redes que realizan conexiones de retroalimentación entre nodos anteriores, por lo tanto, producen ciclos o bucles.

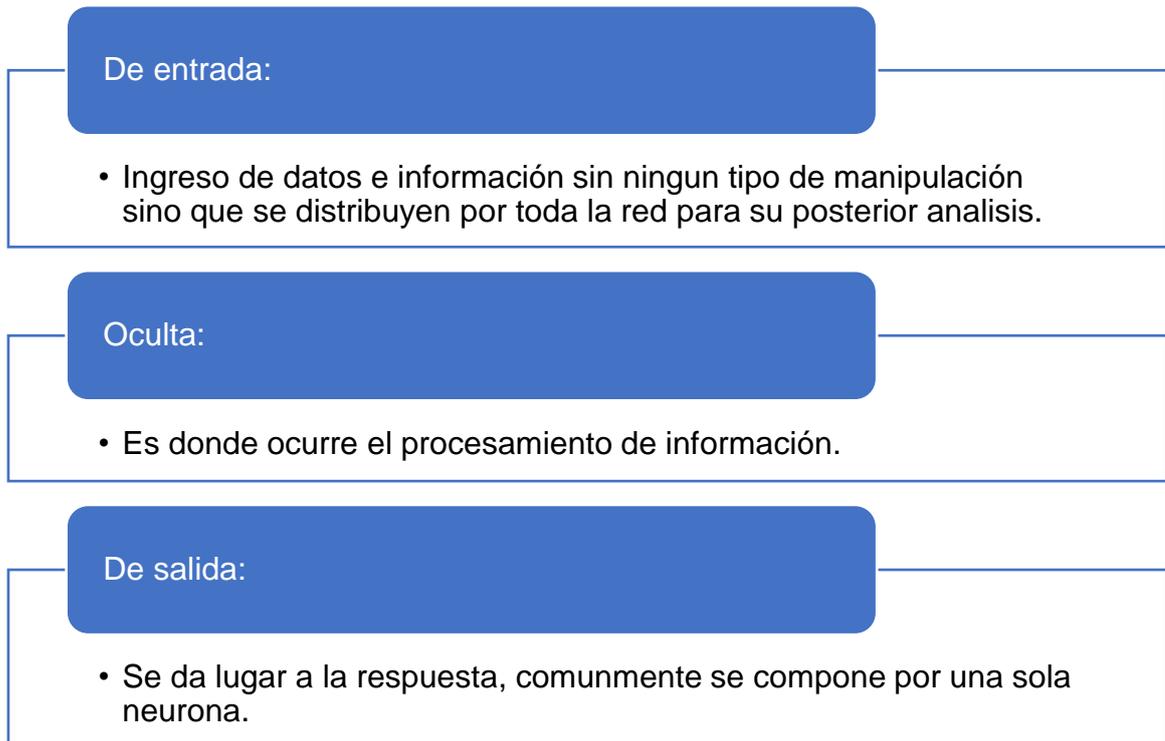
Figura 9 Características de un RNA



Nota. Adaptado de Arias (2018, págs. 29 - 30)

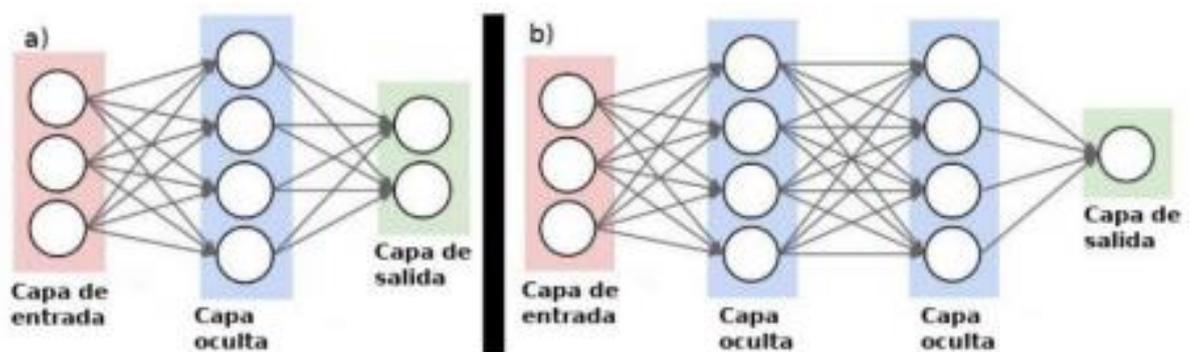
Por su parte, Arias (2018) son redes interconectadas que colaboran entre sí para producir una salida son aplicadas en diagnósticos médicos, reconocimiento de imágenes, crisis empresariales, análisis del clima, educación y economía.

Figura 10 Capas o topología de un RNA



Nota. Adaptado de Arias (2018), Alba y Calle (2020)

Figura 11 Representación gráfica de la topología de un RNA



Nota. Tomado de Alejandro y Figueroa (2020).

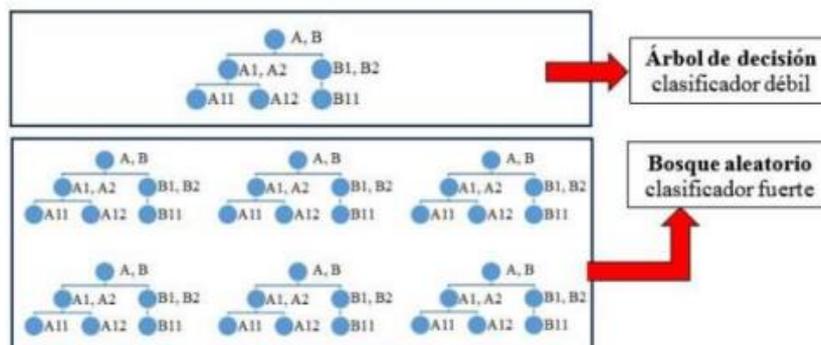
5. Bosques aleatorios

También conocido como Random Forest es un algoritmo de clasificación del machine learning que se basa de la construcción de árboles de decisiones, la cual fue iniciada en Breiman en 2001. De acuerdo a Fernández (2022), previo a la descripción de bosques aleatorios se presenta el bagging como un referente del presente algoritmo, el mismo que fue desarrollado en Breiman en 1996 desde

la introducción y aplicación de árboles de decisiones complejas con un alto nivel de variabilidad. En sí, esta primera aparición de bosque aleatorios pretendía reducir dicha variabilidad para no elevar el sesgo mediante la implementación algoritmos simples y en paralelo. Por lo tanto, para crear resultados más precisos y atributos más definidos se integra random forest, el cual agrandará el tamaño de árbol y nodos sobre un estimado d variables garantizando un nivel de independencia. Es decir, que los bosques aleatorios contribuyen a la reducción de correlación entre los diversos árboles para evitar la similitud.

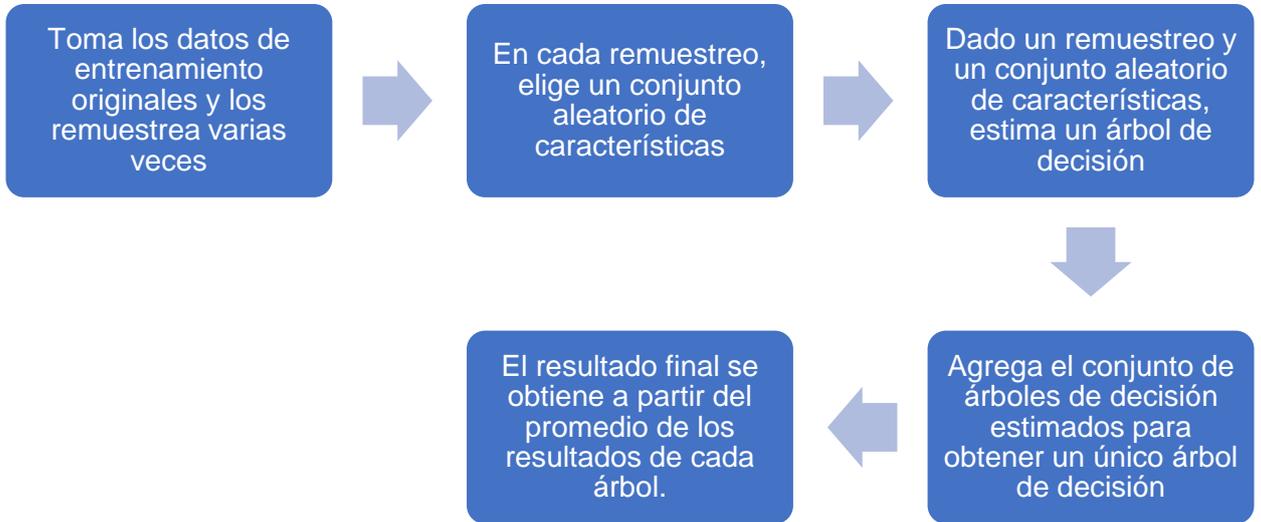
Por su parte, Carmano (2018) menciona que los bosques aleatorios están compuestos por la combinación de dos o más árboles de decisión, donde cada árbol depende de los valores de un vector aleatorio creado independientemente y con la misma distribución de todos los árboles que conforman el bosque, el mismo que integra el modelo bagging y aleatorio para construir una estructura extensa de árboles predictores con variación controlada.

Figura 12 Diferencia entre árbol de decisión y bosques aleatorios



Nota. Tomado de Bosques aleatorios como extensión de los árboles de clasificación con los programas R y Python por Medina y Ñique (2017)

Figura 13 Proceso de elaboración y estructura de un Random Forest



Nota. Adaptado de Evaluación de susceptibilidad por fenómenos de remoción en masa mediante machine learning. Caso de estudio: vía Papallacta-Baeza, Gallardo (2022)

CAPÍTULO 2: METODOLOGÍA

Diseño de la investigación

Para el presente estudio, se predispone de un diseño no experimental ya que no era necesario recurrir a la manipulación o alteración de las variables sino a la obtención de información directa sin previa experimentación. tal como lo describe Hernández et. al., (2014) que son aquellas investigaciones que no se apegan a una manipulación deliberada de los elementos que son examinados en la problemática de investigación, meditando el fenómeno en su contexto natural.

Tipo de investigación

Para el desarrollo del presente trabajo se utilizará un tipo de investigación descriptiva, de acuerdo con Castillero (2019), este es usado con frecuencia ya que su propósito es determinar una descripción detallada del fenómeno estudiado, midiendo sus características y los procesos que lo conforman.

En efecto, la investigación exploratoria contribuye a un análisis de profundidad sobre el objeto de estudio y así conocer como los factores tanto internos como externos que influyen en las variables de estudio. Así como lo determina Hernández & Mendoza (2018) que busca “identificar conceptos o variables promisorias a indagar, establecer prioridades para futuros estudios o sugerir afirmaciones, hipótesis y postulados, con el propósito de familiarizarnos con un planteamiento desconocido, poco estudiado o novedoso” (p.116).

Enfoque de investigación

Será de enfoque cuantitativo, a diferencia del cualitativo, es objetivo, secuencial y probatorio. Realiza la recolección de datos e información con el propósito de probar hipótesis a través de la medición numérica y análisis estadísticos. Los fenómenos observados no tienen influencia por el investigador, por lo que debe evitar la subjetividad para que esta no repercuta en los resultados del estudio o afecte los procesos (Hernández et. al., 2014).

El presente proyecto de investigación propone elaborar un análisis comparativo con respecto a los algoritmos empleados en machine learning para la predicción y prevención de deslizamientos de tierra.

Operacionalización de variables

Variable dependiente: Predicción de áreas de mayor riesgo de deslizamientos de tierra en zonas deforestadas

Variable independiente: Algoritmos de machine learning

Población y muestra

Para el presente estudio se tomará en consideración como población la Secretaría Nacional de Gestión de Riesgos y Emergencias, donde se considerará una muestra bajo criterio del investigador y disponibilidad de atención por parte de los entrevistados, se estima cubrir 3 funcionarios de dicha entidad, el Mayr. (B) Reinaldo Carbo Rodríguez, Asesor del Ministro de Riesgos, el Cap. (B) Ing. Mario Ordeñana Carmigniani Director de Operaciones de la Secretaria Nacional de Gestión de Riesgos y la Ing. Mariana Quispillo Motoya Directora de Monitoreo de Eventos Adversos.

Técnicas e instrumentos de investigación

Se procederá aplicar la entrevista como técnica cualitativa bajo la implementación de la herramienta “guía de preguntas”, las cuales estarán estructuradas de acuerdo a la necesidad de la investigación, es decir, conocer qué tipo de algoritmo de machine learning están empleado y en caso de no hacerlo que algoritmo recomendaría.

Tabla 1 Matriz de operacionalización de variables

Variables	Definición conceptual	Definición Operacional	Dimensiones/Indicadores	Escala de medición
Independiente: Algoritmos de machine learning	Facilitan que las computadoras construyan modelos, identificando patrones en los datos disponibles, y que automaticen el proceso de toma de decisiones basándose en la entrada de datos y en la experiencia (Calderon et. al., 2021)	Permite conocer el nivel de precisión, sensibilidad y especificidad que posee los algoritmos empleados de acuerdo a los entrevistados y cuales consideran que debe emplearse en la predicción de áreas de mayor riesgo en deslizamientos de tierra	Precisión Sensibilidad Especificidad	Entrevista Análisis cualitativo

<p>Dependiente:</p> <p>Predicción de áreas de mayor riesgo de deslizamientos de tierra en zonas deforestadas</p>	<p>Los deslizamientos de tierra son un fenómeno natural que se presentan en zonas montañosas (Flórez & Flórez, 2021)</p>	<p>Mediante el desarrollo de entrevistas se pretende conocer la estabilidad del suelo ecuatoriano y que zonas están propensas a sufrir un deslizamiento.</p>	<p>Identificación del tipo de deslizamiento.</p> <p>Estudio de suelos</p> <p>Estudio de amenazas</p>	<p>Entrevista</p> <p>Análisis cualitativo</p>
--	--	--	--	---

CAPÍTULO 3: ANÁLISIS E INTERPRETACIÓN DE RESULTADOS

Analisis de las entrevistas

Tabla 2 Matriz de analisis de las entrevistas- Parte 1

Preguntas	Analisis de los tres entrevistados
1. Bajo su criterio, ¿se evalúan constantemente las zonas deforestadas para conocer si el lugar está propenso a un deslizamiento?	<p>De acuerdo a los entrevistados, manifiestan que a la actualidad se gestiona un control sobre las zonas deforestadas de manera satelital y estudio de campo para conocer las condiciones del lugar y evitar futuros deslizamientos de tierras y prever consecuencias catastróficas.</p> <p>Los entrevistados coinciden de que la innovación tecnológica es un recurso aun limitado en esta área, lo cual se restringe la posibilidad de conocer a tiempo zonas propensas de deslizamiento, lo que ha conllevado perdidas humanas, materiales y económicas, y a su vez, manifiesta que sería esencial renovar e integrar el conocido aprendizaje automático e inteligencia artificial que se utilizan para medir o conocer las perspectivas de personas para empresas dedicadas a las ventas.</p>
2. En base a su nivel de experiencia laboral, ¿se ha implementado un buen sistema para la predicción de deslizamientos de tierra en zonas deforestadas?	<p>Conforme a las opiniones receptadas no posee ningún sistema, solo el de control satelital y mediante planificadores o personal de campo que se dirigen a las áreas vulnerables para hacer las respectivas pruebas y determinar si a corto o largo plazo existirá un deslizamiento.</p>
3. ¿Qué tipo y cuántos sistemas han empleado para la predicción de deslizamientos de tierra en zonas deforestada?	

Tabla 3 Matriz de analisis de las entrevistas-Parte 2

Preguntas	Análisis de los tres entrevistados
<p>4. ¿Cree usted que el actual sistema tiene un alto nivel de precisión, sensibilidad y especificidad, y de qué manera ha contribuido en la predicción de deslizamientos de tierra en zonas deforestada?</p>	<p>Por lo general, el sistema tradicional que se plantea para el análisis de deslizamientos no trae resultados exactos ni específicos que impliquen una predicción adecuada.</p>
<p>5. ¿Cree usted que al implementar un algoritmo de machine learning beneficiaría en la predicción de deslizamientos de tierra en zonas deforestada?</p>	<p>En base a la opinión de los entrevistados, se observa que poseen conocimientos sobre el machine learning en asuntos de predecir deslizamientos de tierra, por lo que, si tienen conocimiento para fortalecer el sistema de control de deslizamientos, lo cual resulta importante para que el estado y ente competente asigne recursos y tiempo para capacitar a su personal y emplear un sistema basado en machine learning.</p>
<p>6. ¿Qué tipo de algoritmo de machine learning recomendaría para la predicción de deslizamientos de tierra en zonas deforestada y por qué?</p>	<p>Asimismo, se denoto gran aprecio por algoritmos de predicción sustentados en funciones de librería Python y de red neuronal, ya que bajo estudios previos que han revisado presenta un gran aporte y nivel de exactitud y calidad de información al emplear dicho algoritmo.</p>

CAPÍTULO 4: IMPLEMENTACIÓN DE LA SOLUCIÓN TECNOLÓGICA

Conforme a la presente investigación se determina la necesidad de estructurar una propuesta metodológica y técnica que detalle la intervención de un algoritmo de machine learning en la predicción de deslizamientos de tierra en zonas deforestadas y cada uno de los códigos requeridos partiendo de información histórica para la posterior comparación y combinación de datos y condiciones actuales para conocer el nivel de probabilidad de que se produzca un deslizamiento de tierra, de tal manera que se pueda tomar acciones preventivas.

Objetivo

Desarrollar un sistema de alerta temprana mediante la descripción de especificaciones que contribuyen a la predicción de deslizamientos de tierra en zonas deforestadas.

Desarrollo

1. Sectorización de áreas bajo estudio

En esta fase se establece un análisis por zonas o sectores bajo sus características geomorfológicas, tales como estabilidad, relieve y drenaje, de tal manera que se pueda identificar previamente las zonas más inestables y más adelante confirmar bajo el algoritmo de machine learning. Por lo cual, se describe el siguiente formato de registro:

Tabla 4 Registros de zonas para clasificación de su nivel de severidad

Código de Zona	Descripción	ABS. Inicial	ABS. Final	Susceptibilidad	Sitios de severidad
CZ-1					
CZ-2					
CZ-n					

2. Desarrollar la base de datos

Una vez definidas las áreas susceptibles a deslizamientos se requiere conocer ciertos parámetros para el desarrollo de la base de datos que se aplicará en el algoritmo de machine learning.

- a. Precipitación media
- b. Ciclo anual de precipitaciones
- c. Temperatura media valor máximo
- d. Temperatura media valor mínimo
- e. Nivel de humedad media
- f. Nivel de evaporación media

Tabla 5 Base de datos

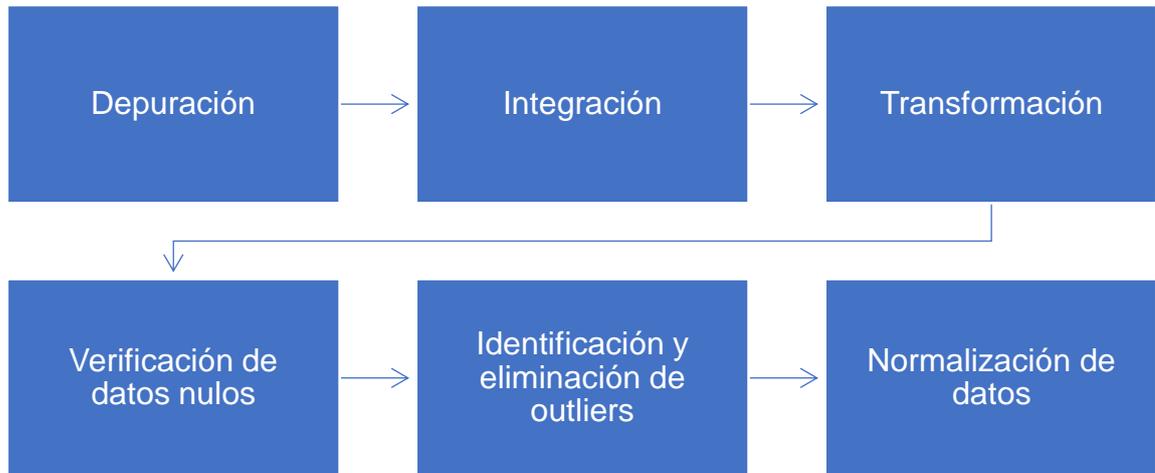
Año	mes	dia	Precipitación	Temperatura máxima	Temperatura mínima	Humedad	Evaporación
1964	1	1	3.6	28.5	25	86	1002.3
1964	1	2	0.4	20.7	15.2	92	1109.3
1964	1	3	0	28	21	76	196.8
1964	1	4	0.3	21	16	75	192
1964	1	5	0.4	27	18	67	1007.4
1964	1	6	0.4	27	20	74	190.5
1964	1	7	2.3	28.5	25	86	1002.3
1964	1	8	12.7	20.7	15.2	92	1109.3
1964	1	9	2	28	21	76	196.8
1964	1	10	47.2	21	16	75	192
1964	1	11	7.4	28	21	76	196.8
1964	1	12	3.8	21	16	75	192
1964	1	13	11.6	27	18	67	1007.4
1964	1	14	28.5	27	20	74	190.5
1964	1	15	25.3	28.5	25	86	1002.3
1964	1	16	23.1	27	18	67	1007.4
1964	1	17	0.8	27	20	74	190.5
1964	1	18	17.4	28.5	25	86	1002.3
1964	1	19	18.2	20.7	15.2	92	1109.3
1964	1	20	15.7	28	21	76	196.8
1964	1	21	119.3	21	16	75	192
1964	1	22	49.1	28.5	25	86	1002.3
1964	1	23	23.8	20.7	15.2	92	1109.3
1964	1	24	46	28	21	76	196.8
1964	1	25	67.8	21	16	75	192
1964	1	26	25.4	28	21	76	196.8
1964	1	27	48.3	21	16	75	192
1964	1	28	17.6	27	18	67	1007.4
1964	1	29	3.7	28.5	25	86	1002.3
1964	1	30	4.7	20.7	15.2	92	1109.3
1964	1	31	1.2	28	21	76	196.8
1964	2	1	11.1	21	16	75	192
1964	2	2	10	27	18	67	1007.4
1964	2	3	4.6	27	20	74	190.5
1964	2	4	3.6	28.5	25	86	1002.3
1964	2	5	32	20.7	15.2	92	1109.3
1964	2	6	5.9	28	21	76	196.8
1964	2	7	0.6	21	16	75	192
1964	2	8	5.2	28	21	76	196.8
1964	2	9	0.8	21	16	75	192
1964	2	10	45.5	27	18	67	1007.4
1964	2	11	9.8	27	20	74	190.5
1964	2	12	18.6	28.5	25	86	1002.3
1964	2	13	13.4	27	18	67	1007.4
1964	2	14	5	27	20	74	190.5
1964	2	15	15.3	28.5	25	86	1002.3
1964	2	16	1.3	20.7	15.2	92	1109.3
1964	2	17	0.1	28	21	76	196.8
1964	2	18	0.8	21	16	75	192
1964	2	19	2.5	28.5	25	86	1002.3
1964	2	20	2.5	20.7	15.2	92	1109.3
1964	2	21	1.4	28	21	76	196.8
1964	2	22	57.7	21	16	75	192
1964	2	23	41.4	28	21	76	196.8
1964	2	24	20.2	21	16	75	192
1964	2	25	14.2	27	18	67	1007.4
1964	2	26	9.3	28.5	25	86	1002.3
1964	2	27	15.5	20.7	15.2	92	1109.3
1964	2	28	0.3	28	21	76	196.8
1964	2	29	1.4	21	16	75	192
1964	3	1	10	27	18	67	1007.4
1964	3	2	5.5	27	20	74	190.5
1964	3	3	14.6	28.5	25	86	1002.3
1964	3	4	4.3	20.7	15.2	92	1109.3
1964	3	5	12.7	28	21	76	196.8
1964	3	6	0.1	21	16	75	192
1964	3	7	40.4	28	21	76	196.8
1964	3	8	36.8	21	16	75	192
1964	3	9	12.6	27	18	67	1007.4
1964	3	10	0.8	27	20	74	190.5
1964	3	11	0	28.5	25	86	1002.3
1964	3	12	0	27	18	67	1007.4
1964	3	13	1	27	20	74	190.5
1964	3	14	3.4	28.5	25	86	1002.3
1964	3	15	2.6	20.7	15.2	92	1109.3
1964	3	16	7.2	28	21	76	196.8

3. Ejecutar un reprocesamiento de la información de la base de datos

Esta fase es importante cuando se desea interactuar con aplicación o algoritmos de machine learning, ya que se debe preservar datos de calidad para retribuir excelentes resultados. En sí, esta fase se convierte en esencial debido a que, al ingresar, transcribir o detallar información de un archivo a otro está expuesto a ciertas deficiencias causadas por errores en tipeo, desactualización de datos o envío erróneo de datos, entre otros.

Por ende, para el presente caso que se ha designado el algoritmo de red neuronal requiere de una efectiva preparación de datos caso contrario el algoritmo y sistema no funcionara, pero principalmente, los resultados serán inaceptables. Para lo cual esta etapa está formada por los siguientes pasos o subetapas:

Tabla 6 Fases de reprocesamiento de datos



4. Designación de las variables de entrada

Posterior a la reprocesamiento de datos, se predispone de las variables específicas de entrada para la ejecución de algoritmo.

Tabla 7 Variables de entrada para la predicción de deslizamientos

Variables	Descripción	Unidad de medición
Fecha (año, mes, día)	Tiempo que ocurrió las llluvias	Tiempo
Precipitación	Medir la lluvia	mm
Temperatura Máxima	Nivel de temperatura alto en el día	Centígrados
Temperatura Mínima	Nivel de temperatura más bajo en el día	Centígrados
Humedad	Vapor de agua en el aire	g/m3
Evaporación	Cantidad de agua que se convierte en vapor	mm

5. Desarrollar de los algoritmos en Python

Para la ejecución del algoritmo de predicción se toma en consideración Pandas, una librería Python, donde se pretende crea una simulación, manipulación y análisis de datos y así estimará cualquier error comparable previa la implementación total.

Por otra parte, para el diseño y funcionamiento del algoritmo de predicción se considerará la interfaz TensorFlow, la cual está disponible en Python y permite el desarrollo efectivo del algoritmo de redes neuronales, para lo cual se requerirá del comando Sequential permitiendo el ingreso de una capa tras otra.

a. Modelo de Regresión Logística

Figura 14 Importamos las librerías y llamamos las bases de datos.

```
import pandas as pd
import numpy as np

data_final_path = 'dataFinal.csv'
final_data = pd.read_csv(data_final_path)

final_data.head()
```

Figura 15 Presentación de datos en un datasheet.

Unnamed: 0.1	Unnamed: 0	FECHA	PROVINCIA	CANTON	TARGET	PRECIPITACION	PRECIPITACION_LAG1	PRECIPITACION_LAG2	PRECIPITACION_LAG3	...	high_precipitation_f
0	6	6	2016-01-07	17	1701	0	0.0	0.1	0.1	0.0	...
1	20	20	2016-01-21	17	1701	0	1.0	0.0	0.0	0.0	...
2	21	21	2016-01-22	17	1701	0	0.0	1.0	0.0	0.0	...
3	22	22	2016-01-23	17	1701	0	0.0	0.0	1.0	0.0	...
4	23	23	2016-01-24	17	1701	0	21.0	0.0	0.0	1.0	...

5 rows × 28 columns

Figura 16 Separación de datos y segregación para el entrenamiento y prueba

```
from sklearn.model_selection import train_test_split

final_data = final_data.drop(columns=["Unnamed: 0", "Unnamed: 0.1",
"FECHA"])

X = final_data.drop(columns="TARGET")
y = final_data["TARGET"]

# (80% training, 20% testing)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
random_state=42, stratify=y)

X_train.shape, X_test.shape, y_train.shape, y_test.shape
```

Figura 17 Validación y remplazo de datos

```
X_train.replace([np.inf, -np.inf], np.nan, inplace=True)
X_test.replace([np.inf, -np.inf], np.nan, inplace=True)

X_train.fillna(X_train.median(), inplace=True)
X_test.fillna(X_train.median(), inplace=True)

np.any(np.isinf(X_train)), np.any(np.isinf(X_test)),
X_train.isna().sum().sum(), X_test.isna().sum().sum()
```

Figura 18 Algoritmo de Regresión Logística

```
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score, precision_score,
recall_score, f1_score, roc_auc_score

logistic_model = LogisticRegression(random_state=42)
logistic_model.fit(X_train, y_train)

y_pred = logistic_model.predict(X_test)

accuracy = accuracy_score(y_test, y_pred)
precision = precision_score(y_test, y_pred)
recall = recall_score(y_test, y_pred)
f1 = f1_score(y_test, y_pred)
roc_auc = roc_auc_score(y_test, logistic_model.predict_proba(X_test)[:,
1])

logistic_metrics = {
```

```

    "Accuracy": accuracy,
    "Precision": precision,
    "Recall": recall,
    "F1 Score": f1,
    "AUC-ROC": roc_auc,
}

logistic_metrics

{
'Accuracy': 0.8797814207650273,
'Precision': 0.5652173913043478,
'Recall': 0.14285714285714285,
'F1 Score': 0.22807017543859645,
'AUC-ROC': 0.7569045618967616}

```

Figura 19 Importación de la matriz de confusión.

```

import matplotlib.pyplot as plt
from sklearn.metrics import confusion_matrix, ConfusionMatrixDisplay

cm = confusion_matrix(y_test, y_pred)

disp = ConfusionMatrixDisplay(confusion_matrix=cm, display_labels=["No
Landslide", "Landslide"])

disp.plot(cmap=plt.cm.Blues)
plt.title("Confusion Matrix")
plt.show()

```

Figura 20 Matriz de Confusión de Regresión Logística 1

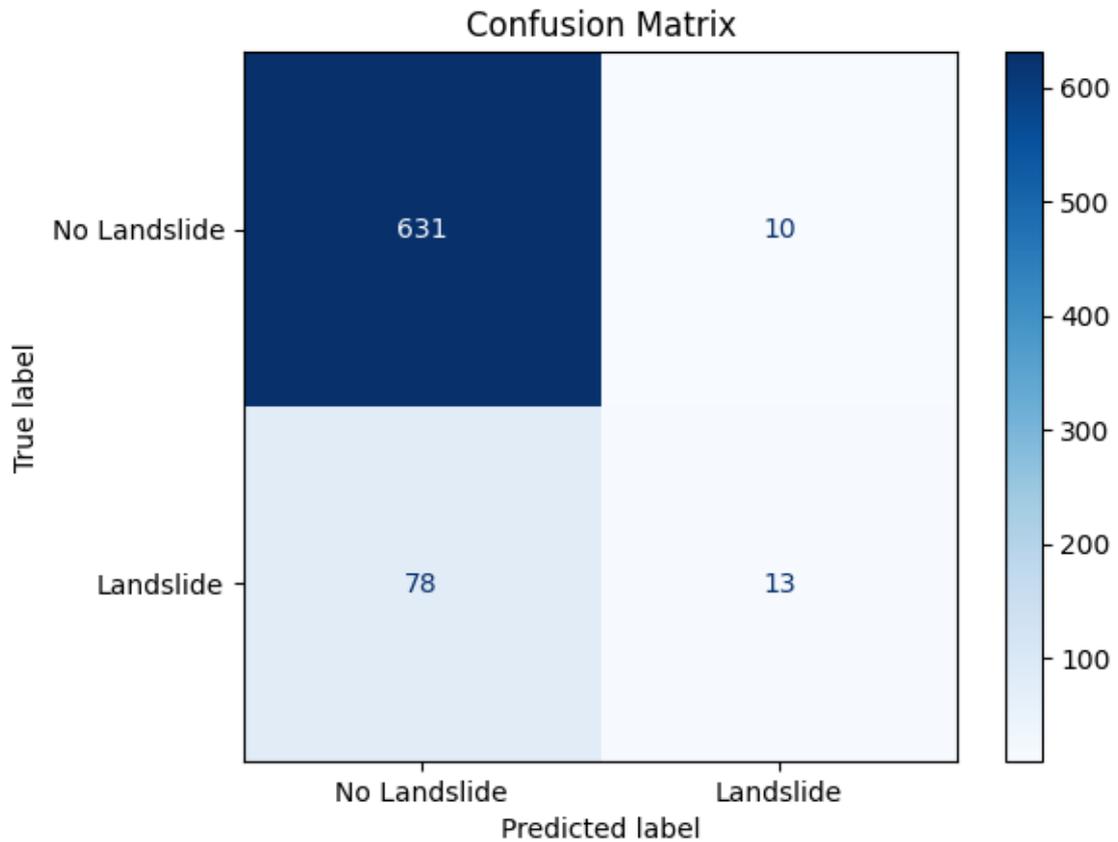


Figura 21 Se balancean los datos para obtener mejores resultados

```
from sklearn.linear_model import LogisticRegression

log_reg_weighted = LogisticRegression(class_weight='balanced')

log_reg_weighted.fit(X_train, y_train)

y_pred_weighted = log_reg_weighted.predict(X_test)

accuracy_weighted = accuracy_score(y_test, y_pred_weighted)
precision_weighted = precision_score(y_test, y_pred_weighted)
recall_weighted = recall_score(y_test, y_pred_weighted)
f1_weighted = f1_score(y_test, y_pred_weighted)

print("Accuracy:", accuracy_weighted)
print("Precision:", precision_weighted)
print("Recall:", recall_weighted)
print("F1 Score:", f1_weighted)

disp_weighted = ConfusionMatrixDisplay.from_estimator(
    log_reg_weighted,
```

```
X_test,  
y_test,  
display_labels=["No Landslide", "Landslide"],  
cmap=plt.cm.Blues  
)  
disp_weighted.ax_.set_title("Confusion Matrix with Weighted Loss")  
plt.show()  
  
{  
'Accuracy': 0.7459016393442623  
'Precision': 0.28110599078341014  
'Recall': 0.6703296703296703  
'F1 Score': 0.39610389610389607  
}
```

Figura 22 Matriz de Confusión de Regresión Logística 2

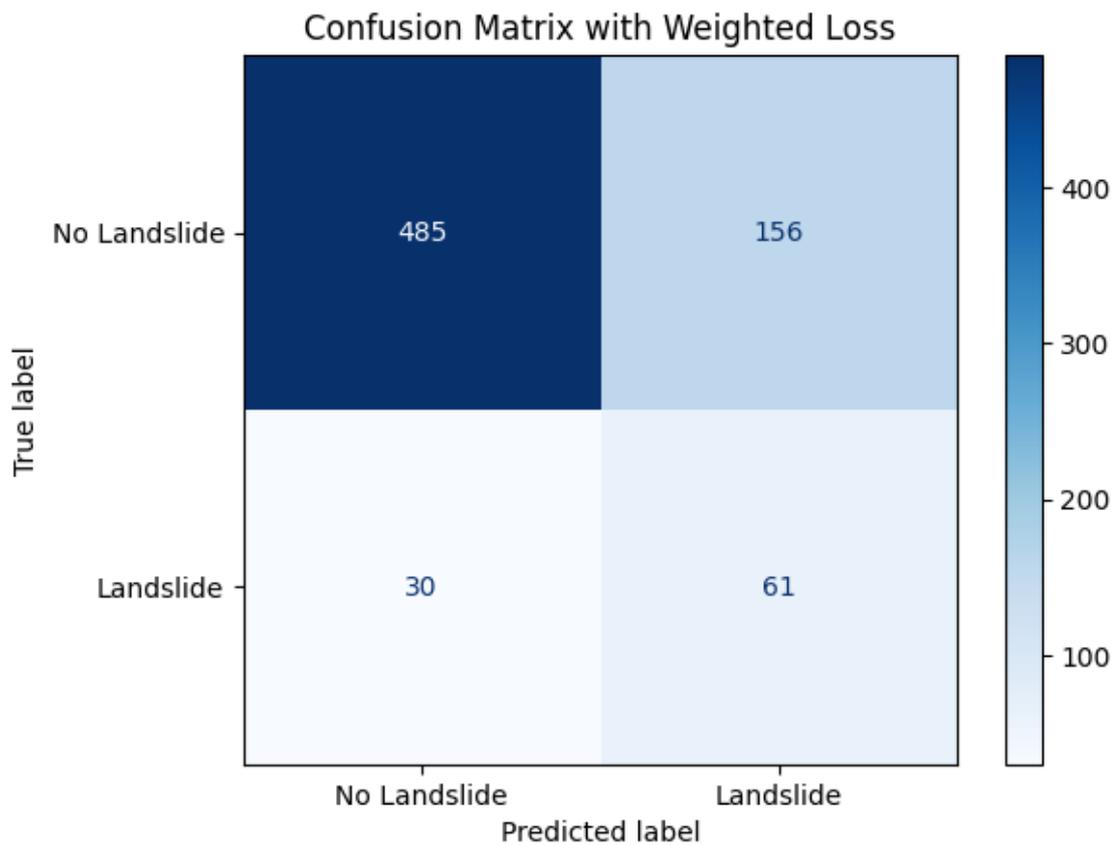


Figura 23 Se balancean los datos una vez más para obtener mejores resultados

```
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score, precision_score,
recall_score, f1_score
from sklearn.metrics import confusion_matrix, ConfusionMatrixDisplay
import matplotlib.pyplot as plt

log_reg_weighted = LogisticRegression(class_weight='balanced',
max_iter=10000)

log_reg_weighted.fit(X_train, y_train)

y_pred_weighted = log_reg_weighted.predict(X_test)

accuracy_weighted = accuracy_score(y_test, y_pred_weighted)
precision_weighted = precision_score(y_test, y_pred_weighted)
recall_weighted = recall_score(y_test, y_pred_weighted)
f1_weighted = f1_score(y_test, y_pred_weighted)

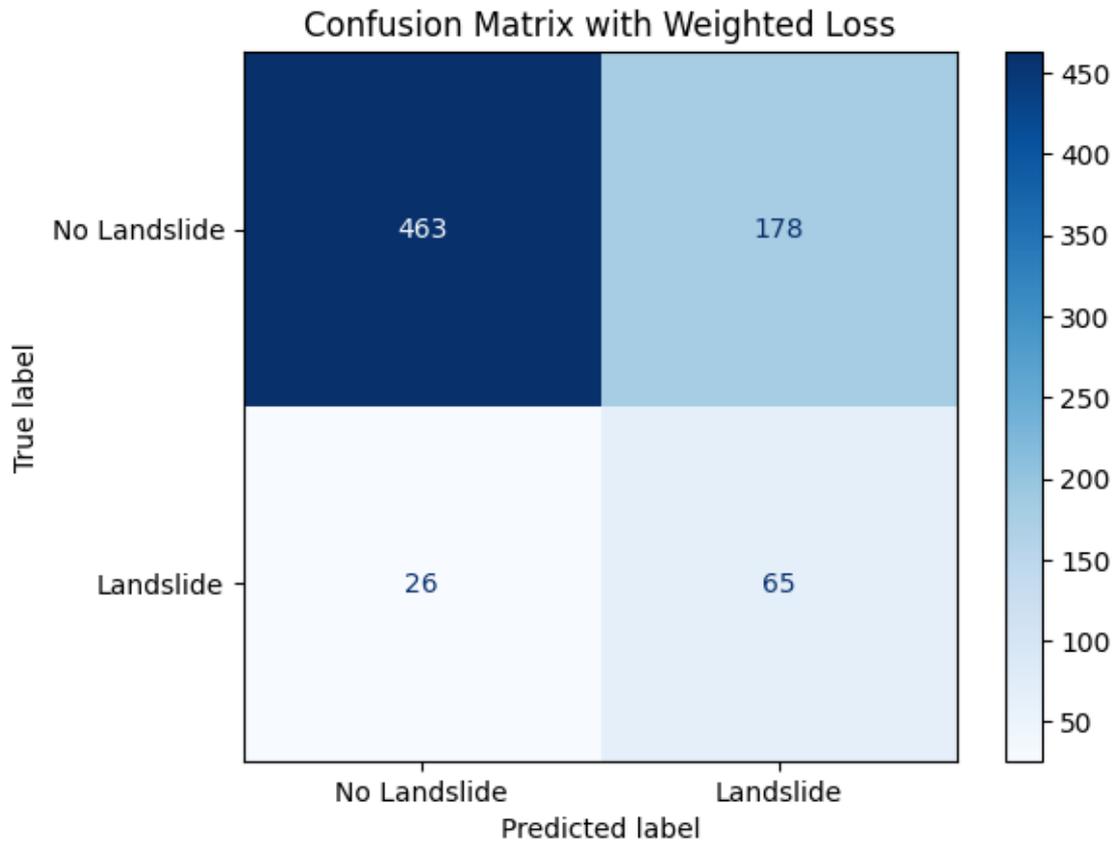
weighted_metrics = {
    "Accuracy": accuracy_weighted,
    "Precision": precision_weighted,
    "Recall": recall_weighted,
    "F1 Score": f1_weighted
}

cm_weighted = confusion_matrix(y_test, y_pred_weighted)
disp_weighted = ConfusionMatrixDisplay(confusion_matrix=cm_weighted,
display_labels=["No Landslide",
"Landslide"])
disp_weighted.plot(cmap=plt.cm.Blues)
plt.title("Confusion Matrix with Weighted Loss")
plt.show()

weighted_metrics

{
'Accuracy': 0.7213114754098361,
'Precision': 0.2674897119341564,
'Recall': 0.7142857142857143,
'F1 Score': 0.38922155688622756
}
```

Figura 24 Matriz de Confusión de Regresión Logística 3



b. Modelo de Bosques Aleatorios (random forest)

Figura 25 Importamos las librerías y llamamos las bases de datos

```
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score, precision_score,
recall_score, f1_score
from sklearn.model_selection import train_test_split
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

data_final_path = 'dataFinal.csv'
data_final = pd.read_csv(data_final_path)

data_final.head()
```

Figura 26 Presentación de datos en un datasheet

Unnamed: 0.1	Unnamed: 0	FECHA	PROVINCIA	CANTON	TARGET	PRECIPITACION	PRECIPITACION_LAG1	PRECIPITACION_LAG2	PRECIPITACION_LAG3	...	high_precipitation_f
0	6	6	2016-01-07	17	1701	0	0.0	0.1	0.1	0.0	...
1	20	20	2016-01-21	17	1701	0	1.0	0.0	0.0	0.0	...
2	21	21	2016-01-22	17	1701	0	0.0	1.0	0.0	0.0	...
3	22	22	2016-01-23	17	1701	0	0.0	0.0	1.0	0.0	...
4	23	23	2016-01-24	17	1701	0	21.0	0.0	0.0	1.0	...

5 rows x 28 columns

Figura 27 Separación de datos y segregación para el entrenamiento y prueba

```
# Dropping unnecessary columns
data_final = data_final.drop(columns=["FECHA"])

# Splitting features and target
X = data_final.drop(columns="TARGET")
y = data_final["TARGET"]

# Splitting the dataset into training and testing sets (80% training, 20% testing)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42, stratify=y)
```

Figura 28 Validación de datos y relleno de datos vacíos

```
# Replacing infinite values with NaN
X_train.replace([np.inf, -np.inf], np.nan, inplace=True)
X_test.replace([np.inf, -np.inf], np.nan, inplace=True)

# Filling NaN values with the median of the respective columns
X_train.fillna(X_train.median(), inplace=True)
X_test.fillna(X_train.median(), inplace=True)
```

Figura 29 Inicialización del entrenamiento y balanceo del peso de las variables

```
# Initializing and training the Random Forest Classifier with
class_weight='balanced'
rf_clf = RandomForestClassifier(class_weight="balanced", random_state=42)
rf_clf.fit(X_train, y_train)

# Making predictions
y_pred_rf = rf_clf.predict(X_test)

# Calculating performance metrics
rf_accuracy = accuracy_score(y_test, y_pred_rf)
rf_precision = precision_score(y_test, y_pred_rf)
rf_recall = recall_score(y_test, y_pred_rf)
rf_f1_score = f1_score(y_test, y_pred_rf)

rf_accuracy, rf_precision, rf_recall, rf_f1_score

(0.9316939890710383, 0.9019607843137255, 0.5054945054945055,
0.647887323943662)
```

Figura 30 Importación de la Matriz de Confusión

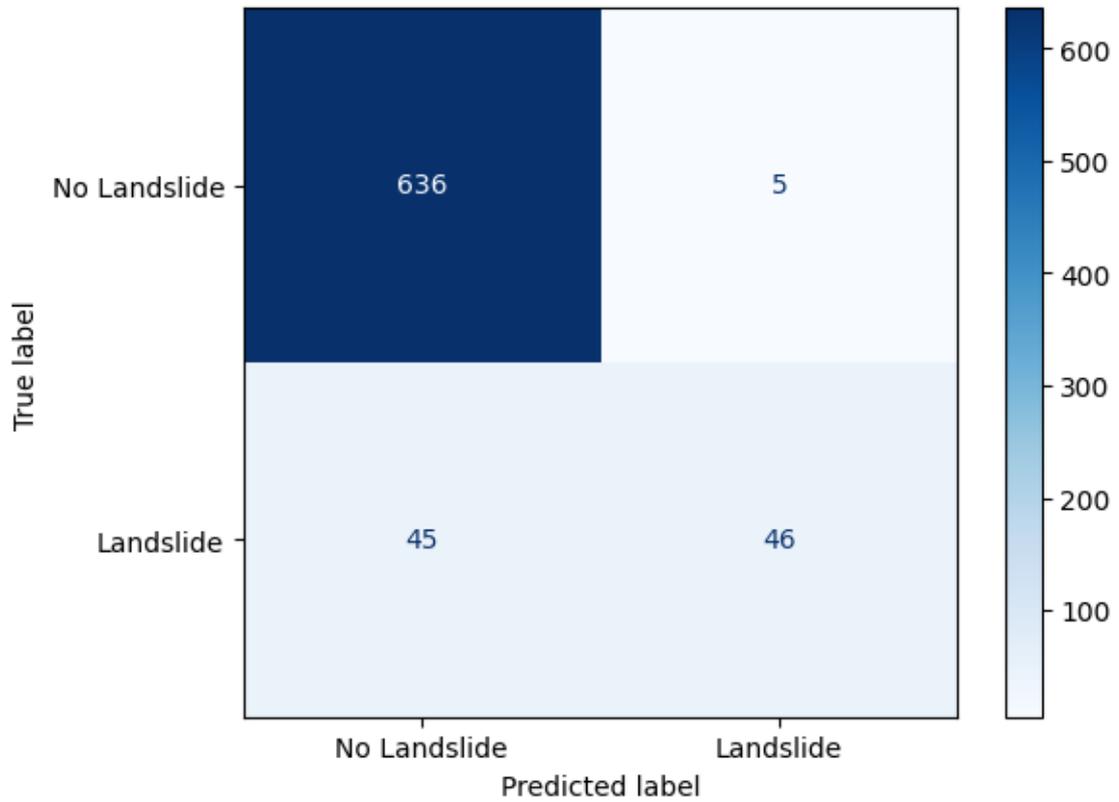
```
# Plotting the confusion matrix for the Random Forest model

from sklearn.metrics import ConfusionMatrixDisplay
from sklearn.metrics import confusion_matrix

# Calculating the confusion matrix
rf_cm = confusion_matrix(y_test, y_pred_rf)

# Displaying the confusion matrix
disp = ConfusionMatrixDisplay(confusion_matrix=rf_cm, display_labels=["No
Landslide", "Landslide"])
disp.plot(cmap=plt.cm.Blues)
plt.show()
```

Figura 31 Matriz de Confusión de Bosques Aleatorios (random forest)



c. Modelo de Redes Neuronales

Figura 32 Importación de librerías, inicialización de base de datos y validación de datos

```
import pandas as pd
import numpy as np

file_path = 'dataFinal.csv'

final_data = pd.read_csv(file_path)

final_data = final_data.drop(columns=["Unnamed: 0", "FECHA"])

final_data.replace([np.inf, -np.inf], np.nan, inplace=True)

final_data.fillna(final_data.median(), inplace=True)

final_data.isna().sum().sum(), np.any(np.isinf(final_data))

(0, False)
```

Figura 33 Despliegue del Datasheet

Unnamed: 0.1	PROVINCIA	CANTON	TARGET	PRECIPITACION	PRECIPITACION_LAG1	PRECIPITACION_LAG2	PRECIPITACION_LAG3	PRECIPITACION_ROLLING3_MEAN	PRECI
0	6	17	1701	0	0.0	0.1	0.1	0.0	0.066667
1	20	17	1701	0	1.0	0.0	0.0	0.0	0.333333
2	21	17	1701	0	0.0	1.0	0.0	0.0	0.333333
3	22	17	1701	0	0.0	0.0	1.0	0.0	0.333333
4	23	17	1701	0	21.0	0.0	0.0	1.0	7.000000
...
3651	11630	8	801	0	0.8	0.0	0.0	0.0	0.266667
3652	11631	8	801	0	0.0	0.8	0.0	0.0	0.266667
3653	11632	8	801	0	2.9	0.0	0.8	0.0	1.233333
3654	11633	8	801	0	0.3	2.9	0.0	0.8	1.066667
3655	11634	8	801	0	0.4	0.3	2.9	0.0	1.200000

3656 rows x 26 columns

Figura 34 Cambio de datos y balanceo para implementación de la red neuronal

```
array([[ 0.11814904,  0.94072254,  0.94072254, -0.48639744, -0.4779008 ,
        -0.48797098, -0.46486242, -0.68166511, -0.67478329, -0.84039127,
        -0.87374183,  0.05300982,  0.52430133, -0.47160156, -0.47322624,
        -0.48402073, -0.48939582, -0.18306078, -0.19333983, -0.1956718 ,
        -0.27014285, -0.18505699, -0.27959094, -0.26063921, -0.25763328],
       [ 0.3763396 ,  0.94072254,  0.94072254, -0.48639744, -0.45799186,
        -0.4673062 , -0.07900879, -0.66280201, -0.66196419, -0.69781585,
        -0.65735517,  0.65848138, -1.51095535, -0.47160156, -0.47322624,
        -0.48402073, -0.48939582, -0.18280572, -0.18993402, -0.19075336,
        -0.00528918, -0.18954958, -0.27959094, -0.26063921, -0.25763328],
       [-0.30656852,  0.94072254,  0.94072254, -0.16419689, -0.43144661,
        0.18707843,  4.28963912, -0.1975121 , -0.24728411,  0.66419871,
        1.79898716,  0.05300982,  1.54192967, -0.47160156, -0.47322624,
        -0.48402073,  2.04333581, -0.17115801, -0.09977165,  0.98530244,
        0.02596561, -0.25035367, -0.27959094, -0.26063921, -0.25763328],
       [-1.22025655, -0.11988001, -0.11988001,  0.631444121,  0.65027246,
        -0.5017475 , -0.48448209,  0.38095644,  0.23818906,  0.07712347,
        -0.10067189, -0.85519751,  1.54192967,  2.12043406,  2.11315417,
        -0.48402073, -0.48939582, -0.18306078, -0.19333983, -0.19576887,
        0.04692343,  0.13972181,  3.57665386, -0.26063921, -0.25763328],
       [ 1.07133197,  0.94072254,  0.94072254,  0.57226152, -0.41153767,
        -0.5017475 , -0.47140231, -0.15035434,  0.16341336, -0.55859515,
        -0.28186451,  1.26395294, -0.49332701,  2.12043406, -0.47322624,
        -0.48402073, -0.48939582, -0.18306078, -0.19297361, -0.19576887,
        -0.24019003, -0.30321854, -0.27959094, -0.26063921, -0.25763328]])
```

Figura 35 Importación de librerías e inicio del entrenamiento

```
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense
from tensorflow.keras.metrics import Precision, Recall, Accuracy

model = Sequential([
    Dense(64, activation='relu', input_shape=(X_train.shape[1],)),
    Dense(32, activation='relu'),
    Dense(1, activation='sigmoid')
])

model.compile(
    optimizer='adam',
    loss='binary_crossentropy',
    metrics=[Recall(name='recall'), Precision(name='precision'),
    Accuracy(name='accuracy')]
)

history = model.fit(X_train, y_train, epochs=100, batch_size=32,
validation_data=(X_test, y_test))
```

Figura 36 Entrenamiento de la red neuronal en 100 pasos

```
Epoch 1/100
92/92 [=====] - 2s 6ms/step - loss: 10.5491 -
recall: 0.2240 - precision: 0.2336 - accuracy: 0.2962 - val_loss: 3.2980
- val_recall: 0.0659 - val_precision: 0.3333 - val_accuracy: 0.0068
Epoch 2/100
92/92 [=====] - 0s 3ms/step - loss: 2.8300 -
recall: 0.2049 - precision: 0.2078 - accuracy: 0.0783 - val_loss: 3.6886
- val_recall: 0.0769 - val_precision: 0.3684 - val_accuracy: 0.0109
Epoch 3/100
92/92 [=====] - 0s 3ms/step - loss: 1.9436 -
recall: 0.2869 - precision: 0.3052 - accuracy: 0.0243 - val_loss: 2.3293
- val_recall: 0.1868 - val_precision: 0.4048 - val_accuracy: 0.0041
Epoch 4/100
92/92 [=====] - 0s 3ms/step - loss: 2.2373 -
recall: 0.2514 - precision: 0.2722 - accuracy: 0.0711 - val_loss: 2.1636
- val_recall: 0.6044 - val_precision: 0.3416 - val_accuracy: 0.0055
Epoch 5/100
92/92 [=====] - 0s 3ms/step - loss: 1.8313 -
recall: 0.3361 - precision: 0.3555 - accuracy: 0.0202 - val_loss: 2.9377
- val_recall: 0.8242 - val_precision: 0.2000 - val_accuracy: 0.0109
Epoch 6/100
92/92 [=====] - 0s 3ms/step - loss: 2.4303 -
recall: 0.2787 - precision: 0.2873 - accuracy: 0.1077 - val_loss: 2.3070
- val_recall: 0.0879 - val_precision: 0.4000 - val_accuracy: 0.0055
```

```
Epoch 7/100
92/92 [=====] - 0s 3ms/step - loss: 1.6340 -
recall: 0.3224 - precision: 0.3315 - accuracy: 0.0243 - val_loss: 4.1461
- val_recall: 0.1538 - val_precision: 0.4828 - val_accuracy: 0.0587
Epoch 8/100
92/92 [=====] - 0s 3ms/step - loss: 1.5198 -
recall: 0.3251 - precision: 0.3439 - accuracy: 0.0243 - val_loss: 2.5311
- val_recall: 0.0659 - val_precision: 0.2727 - val_accuracy: 0.0014
Epoch 9/100
92/92 [=====] - 0s 3ms/step - loss: 1.9478 -
recall: 0.3033 - precision: 0.3127 - accuracy: 0.0889 - val_loss: 3.6237
.
.
.
.
92/92 [=====] - 0s 3ms/step - loss: 1.0303 -
recall: 0.3798 - precision: 0.4371 - accuracy: 0.0694 - val_loss: 1.2176
- val_recall: 0.1758 - val_precision: 0.5926 - val_accuracy: 0.0014
Epoch 97/100
92/92 [=====] - 0s 3ms/step - loss: 0.4917 -
recall: 0.4290 - precision: 0.5286 - accuracy: 0.0068 - val_loss: 1.2534
- val_recall: 0.5604 - val_precision: 0.3893 - val_accuracy: 0.0041
Epoch 98/100
92/92 [=====] - 0s 3ms/step - loss: 0.5465 -
recall: 0.4262 - precision: 0.4875 - accuracy: 0.0044 - val_loss: 1.2907
- val_recall: 0.0659 - val_precision: 0.5455 - val_accuracy: 0.0014
Epoch 99/100
92/92 [=====] - 0s 3ms/step - loss: 0.6876 -
recall: 0.3989 - precision: 0.4679 - accuracy: 0.0041 - val_loss: 1.5241
- val_recall: 0.4505 - val_precision: 0.3832 - val_accuracy: 0.0068
Epoch 100/100
92/92 [=====] - 0s 3ms/step - loss: 0.5417 -
recall: 0.4454 - precision: 0.4939 - accuracy: 0.0055 - val_loss: 1.0824
- val_recall: 0.3077 - val_precision: 0.5600 - val_accuracy: 0.0027
```

Figura 37 Reevaluación de datos y llamado a presentar la predicción

```
loss, recall_score, precision_score, accuracy_score =
model.evaluate(X_test, y_test)
f1 = f1_score = 2 * (precision_score * recall_score) / (precision_score +
recall_score)

print(f"Recall: {recall_score}")
print(f"Precision: {precision_score}")
print(f"Accuracy: {accuracy_score}")
print(f"F1 SCORE: {f1}")

23/23 [=====] - 0s 2ms/step - loss: 1.0824 -
recall: 0.3077 - precision: 0.5600 - accuracy: 0.0027
Recall: 0.3076923191547394
Precision: 0.5600000023841858
Accuracy: 0.0027322403620928526
F1 SCORE: 0.3971631307158453

y_pred = model.predict(X_test)
y_pred_classes = (y_pred > 0.5).astype(int) # 0.5 para arriba es
verdadero
```

Figura 38 Importación de la Matriz de Confusión

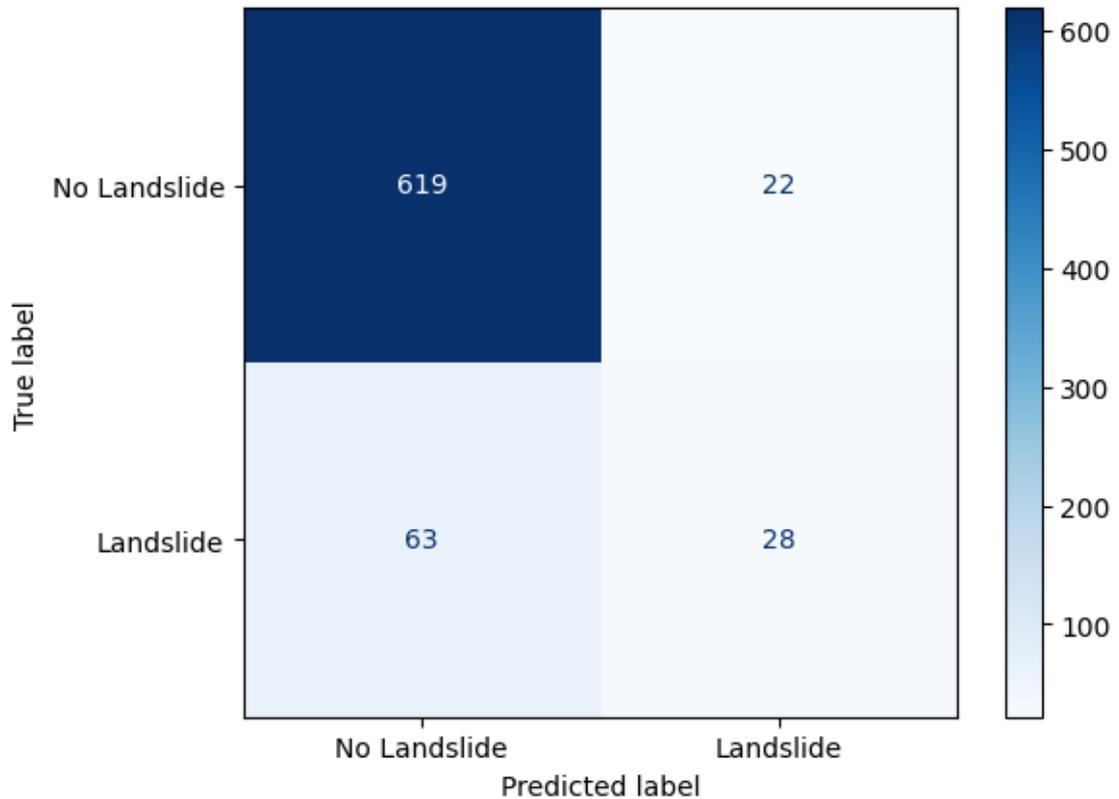
```
from sklearn.metrics import confusion_matrix

cm = confusion_matrix(y_test, y_pred_classes)

from sklearn.metrics import ConfusionMatrixDisplay
import matplotlib.pyplot as plt

disp = ConfusionMatrixDisplay(confusion_matrix=cm, display_labels=["No
Landslide", "Landslide"])
disp.plot(cmap=plt.cm.Blues)
plt.show()
```

Figura 39 Matriz de Confusión de Redes Neuronales



d. Modelo de Aprendizaje Automático Supervisado XGBoost

Figura 40 Importación de librerías, inicialización de base de datos y validación de datos

```
from xgboost import XGBClassifier
from sklearn.metrics import accuracy_score, precision_score,
recall_score, f1_score
from sklearn.model_selection import train_test_split
import pandas as pd
import numpy as np

final_data_xgb = pd.read_csv('dataFinal.csv')

final_data_xgb = final_data_xgb.drop(columns=["Unnamed: 0", "FECHA"])

X = final_data_xgb.drop(columns="TARGET")
y = final_data_xgb["TARGET"]

X.replace([np.inf, -np.inf], np.nan, inplace=True)
```

```

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
random_state=42, stratify=y)

X_train.fillna(X_train.median(), inplace=True)
X_test.fillna(X_train.median(), inplace=True)

np.any(np.isinf(X_train)), np.any(np.isinf(X_test)),
X_train.isna().sum().sum(), X_test.isna().sum().sum()

(False, False, 0, 0)

```

Figura 41 Balanceo de datos e inicio del proceso de aprendizaje

```

from xgboost import XGBClassifier
from sklearn.metrics import accuracy_score, precision_score,
recall_score, f1_score, confusion_matrix

xgb_model = XGBClassifier(scale_pos_weight=(1 - y_train.mean()) /
y_train.mean(), use_label_encoder=False)

xgb_model.fit(X_train, y_train)

y_pred_xgb = xgb_model.predict(X_test)

xgb_accuracy = accuracy_score(y_test, y_pred_xgb)
xgb_precision = precision_score(y_test, y_pred_xgb)
xgb_recall = recall_score(y_test, y_pred_xgb)
xgb_f1 = f1_score(y_test, y_pred_xgb)
xgb_confusion_matrix = confusion_matrix(y_test, y_pred_xgb)

xgb_accuracy, xgb_precision, xgb_recall, xgb_f1, xgb_confusion_matrix

```

Figura 42 Despliegue de la Matriz de Confusión

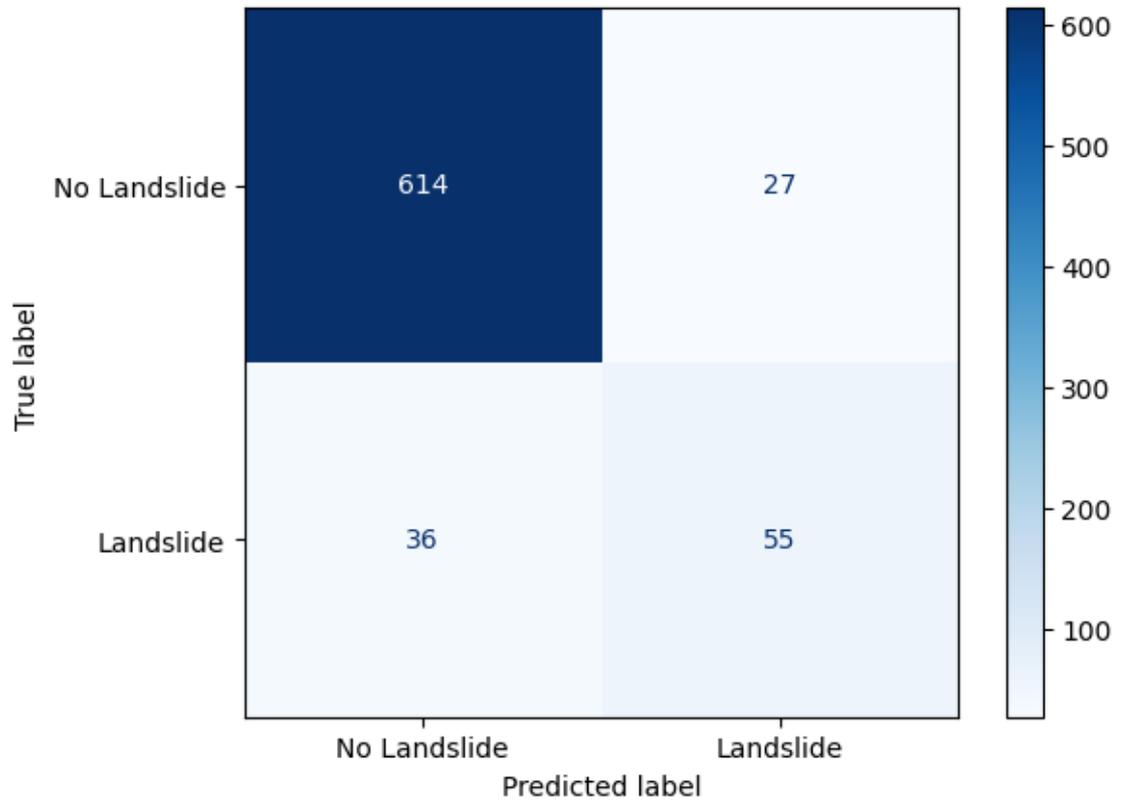
```

from sklearn.metrics import ConfusionMatrixDisplay
import matplotlib.pyplot as plt

# Displaying the confusion matrix
disp = ConfusionMatrixDisplay(confusion_matrix=xgb_confusion_matrix,
display_labels=["No Landslide", "Landslide"])
disp.plot(cmap=plt.cm.Blues)
plt.show()

```

Figura 43 Matriz de Confusión de XGBoost



CONCLUSIONES

De acuerdo a la presente investigación se analizó todo lo referente a machine learning con el objetivo principal de comparar cuatro algoritmos disponibles y definir el más idóneo para la predicción de áreas de mayor riesgo de afectación por terremotos en el Ecuador, lo cual conllevó a una especificación con el algoritmo de red neuronal.

Conforme a la revisión bibliográfica se pudo identificar estudios que buscan integrar la innovación tecnológica del machine learning para la gestión de riesgos relacionados a los diversos desastres naturales, pero sobre todo para detectar zonas que quedan vulnerables luego de la manifestación de un terremoto, temblor, inundación, lluvias hasta sequías. En sí, se concluye que el machine learning no solo está implícito en áreas de ventas o marketing como se dio a conocer, dado su gran aporte y fiabilidad en los niveles de predicción ha causada intriga en varias áreas de las ciencias.

Además, se destaca que los factores riesgos en el deslizamiento se revela pérdidas materiales, económicas y humanas, así como también una zona vulnerable que impide la agricultura y hábitat para cualquier ser vivo.

Con base a los datos obtenidos en las distintas bases de datos se puede concluir que el factor fundamental para que haya deslizamientos de tierra va mas allá del suelo deforestado, realmente el dato importante es la cantidad de precipitaciones que han acontecido en el territorio el día o en días anteriores al evento.

Se puede concluir que basar el análisis predictivo en un solo modelo no es adecuado ya que en base al análisis de las matrices de confusión se aprecia que hay modelos que son mejores para predecir verdaderos positivos, pero totalmente lo contrario para predecir los verdaderos negativos, además de dar un valor muy alto de falsos negativo o falsos positivo, generado incertidumbre, perdiendo así el objetivo de ser un medio de alerta temprana.

Se identifica que el algoritmo como mayor eficiencia en descartar deslizamientos de tierra es el de Bosques Aleatorios (random forest) conforme a las investigaciones previas ya que posee un menor índice de falsos positivos y el mejor índice en verdaderos negativos.

Sin embargo, el algoritmo que mejor predice verdaderos positivos es la Regresión Logística posterior a balancear sus variables, en contra punto de su capacidad de emitir muchos falsos positivos, los cuales generarían alertar y zozobra.

Por su parte, las Redes Neuronales son un muy mal algoritmo para predecir estos deslizamientos de tierra, debido a que posterior al análisis y balanceo de sus bases de datos de entrenamientos es la peor en predecir verdaderos positivos, la peor en emitir alertas que son falsos negativos, en contra punto es la segunda mejor en predecir verdaderos negativos, traduciéndose en un algoritmo que más se basa en verificar las NO actividades.

Por ultimo el algoritmo de aprendizaje supervisado XGBoost es el segundo mejor en predecir verdaderos positivos siendo un 4% menos eficiente que el primero en este ítem, el tercero mejor en predecir verdaderos negativos solo con 1% por detrás del primero, es el segundo mejor en identificar los falsos negativos con un 6 % por detrás del primero y el tercer mejor en los falsos positivos son un 3% por detrás del segundo.

Es necesario implementar más datos a la base, contrayendo un mayor problema que es el equipo físico que va a procesar los datos debe tener una robustez considerable en memoria RAM para poder llevar a cabo el procesamiento de los mismo.

Tabla 8 Comparación de resultados de las matrices de confusión

Índice	Regresión logística	Bosques aleatorios	Red neuronal	XGBoost	Calificación
Verdaderos Positivos	65	46	28	55	<ol style="list-style-type: none"> 1. Regresión Logística 34% 2. XGBoost 28% 3. Bosques aleatorios 24% 4. Redes Neuronales 14%
Verdaderos Negativos	463	636	619	614	<ol style="list-style-type: none"> 1. Bosques aleatorios 27% 2. Redes Neuronales 27% 3. XGBoost 26% 4. Regresión Logística 20%
Falsos Negativo	26	45	63	36	<ol style="list-style-type: none"> 1. Regresión Logística 15% 2. XGBoost 21% 3. Bosques aleatorios 27% 4. Redes Neuronales 37%
Falsos Positivos	178	5	22	27	<ol style="list-style-type: none"> 1. Bosques Aleatorios 2% 2. Redes Neuronales 9% 3. XGBoost 12% 4. Regresión Logística 77%

Ilustración 2 Gráfico de pastel de los Verdades Negativas

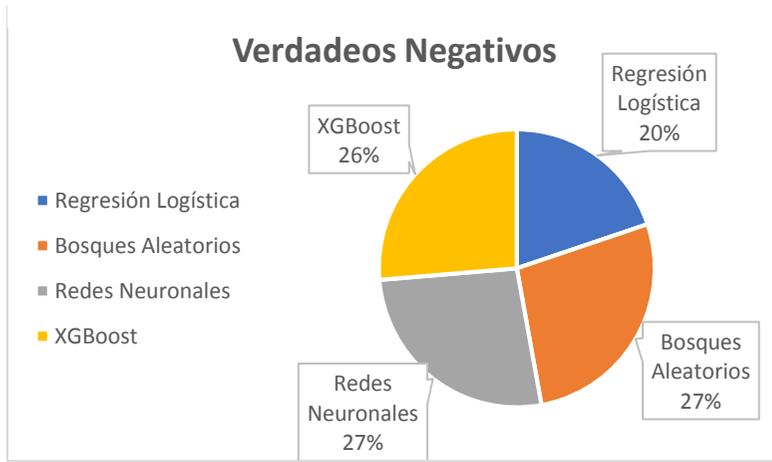


Ilustración 1 Gráfico de pastel de los Verdades Positivos

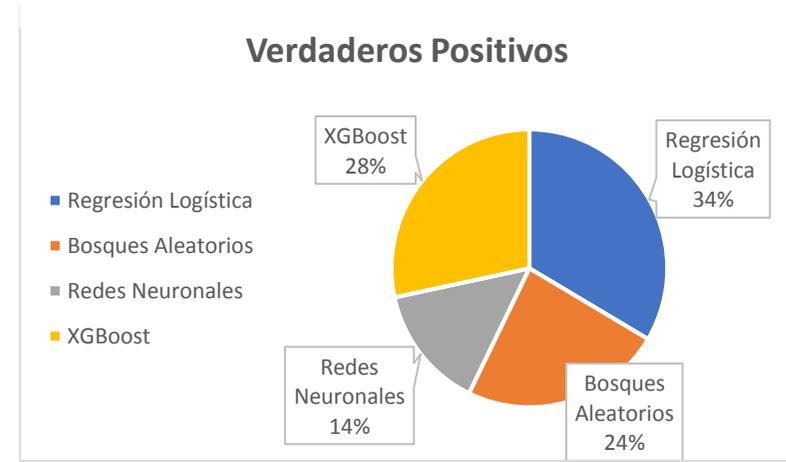


Ilustración 6 Gráfico de pastel de los Falsos Positivos

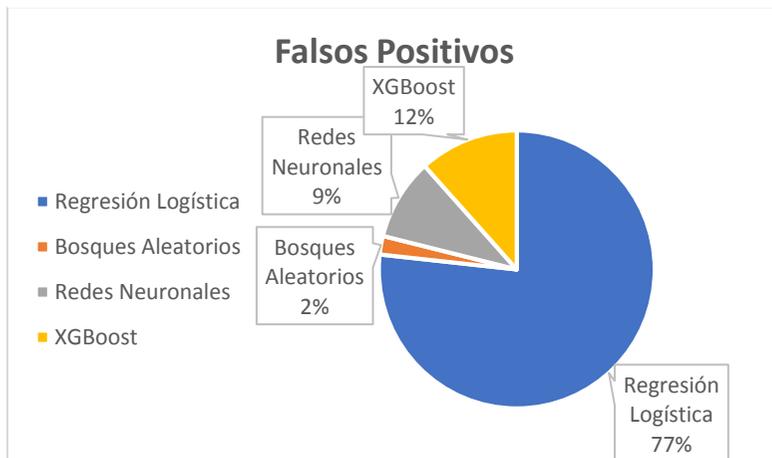
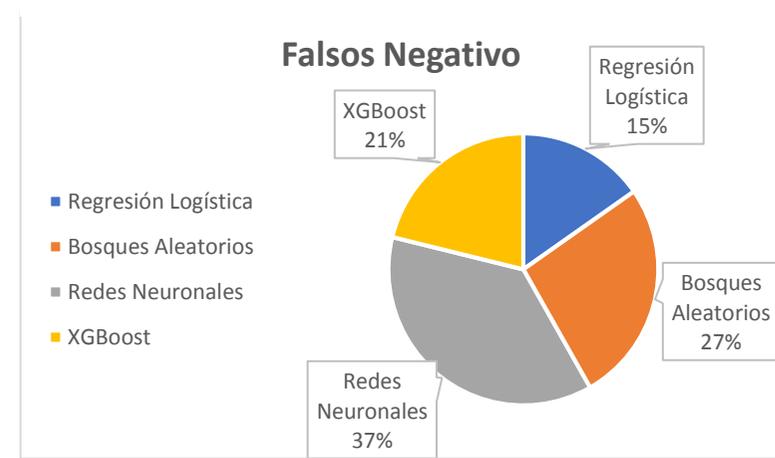


Ilustración 11 Gráfico de pastel de los Falsos Negativos



RECOMENDACIONES

Debido a la relevancia del algoritmo de predicción por red neuronal se recomienda integrar este modelo de solución en áreas de gestión de riesgos, con la finalidad de prever a tiempo cualquier deslizamiento de tierra y evitar fatídicas consecuencias. Por ende, se recomienda estructurar un algoritmo mixto bajo el uso de librería Python, donde primero se ejecutará un algoritmo de prueba previo al análisis de predicción con el pseudocódigo de la red neuronal.

Además, se recomienda que para el entrenamiento del algoritmo se integre datos escalados para minimizar y evitar una alteración en los datos. Se propone utilizar el RMSprop como el principal optimizador de entrenamiento, a su vez, utilizar el 80% de información recopilada en entrenamiento y el 20% de información validada con la intención de que se reciba datos pronosticados y así, tener resultados más precisos y reales.

Bibliografía

- Aimacaña, J., & Columba, A. (2021). *Análisis comparativo de algoritmos de Machine Learning para la detección de plagas en los cultivos representativos de la sierra ecuatoriana. [Tesis de grado. Universidad Central del Ecuador]*. Obtenido de <http://www.dspace.uce.edu.ec/bitstream/25000/24228/1/UCE-FING-ISI-AIMACA%c3%91A%20JEFFERSON-COLUMBA%20ALEXANDER.pdf>
- Álagos, F. (2018). *Introducción al Machine Learning*. Álagos. Obtenido de <http://alogos.es/introduccion-machine-learning/>
- Alaminos, A. (2023). *Árboles de decisión en R con Random Forest*. Alicante: Universidad de Alicante. Obets Ciencia Abierta. Obtenido de https://rua.ua.es/dspace/bitstream/10045/133067/1/Random_Forest_en_la_Investigacion_Social.pdf
- Alba, D., & Calle, J. (2020). *Aplicación de técnicas de Machine Learning basado en información sísmica para profundizar la probabilidad de terremotos mediante el uso de regresión logística y redes neuronales.[Tesis de grado. Universidad de Guayaquil]*. Obtenido de <http://repositorio.ug.edu.ec/bitstream/redug/48862/1/B-CISC-PTG-1717%20Alba%20Vega%20Daniel%20Alejandro%20-%20Calle%20Jara%20Javier%20Fernando.pdf>
- Alejandro, M., & Figueroa, A. (2020). *Diseño de un modelo de aprendizaje automático para la predicción de la precipitación de lluvias en la ciudad de Guayaquil.[Tesis de grado. Universidad de Guayaquil]*. Obtenido de <http://repositorio.ug.edu.ec/bitstream/redug/52744/1/B-CISC-PTG-1945-2021%20Alejandro%20Segura%20Michael%20Gilson%20-%20Figueroa%20Panchana%20Ana%20Jennifer.pdf>
- Ambientum. (10 de Enero de 2023). *La deforestación, un grave problema a nivel mundial*. Obtenido de <https://www.ambientum.com/ambientum/medio-natural/la-deforestacion-un-grave-problema-a-nivel-mundial.asp>

- Arias, E. (2018). *Control de la gesticulación de un robot social interactivo con aspecto humanoide*. Obtenido de <https://repositorio.uta.edu.ec:8443/jspui/handle/123456789/28115>
- Bobadilla, J. (2020). *Machine Learning y Deep Learning: Usando Python, Scikit y Keras*. Madrid : RAMA Editorial . Obtenido de <https://bibliotecas.ups.edu.ec:3488/es/ereader/bibliotecaups/222698?page=11>.
- Calderon, J., Morales, L., Roncal, M., & Solórzano, W. (Febrero de 2021). *Uso de algoritmos de machine learning para el diagnóstico de melanomas*. [Tesis de Grado. Universidad de Piura]. Obtenido de https://pirhua.udep.edu.pe/bitstream/handle/11042/4949/T_IME_2105.pdf?sequence=1&isAllowed=y
- Carmano, P. (2018). *Study on viability, strengths and weaknesses of bag of little bootstraps (BLB)*. [Tesis de Maestría. Universidad Autónoma de Madrid]. Obtenido de https://repositorio.uam.es/bitstream/handle/10486/682687/vina_carmona_pablode_tfm.pdf;jsessionid=82DEE60FC8F05E4D11435794362687FE?sequence=1
- Casamitjana, M., & Carl, R. (2020). *Movimientos en masa*. Fondo Editorial EIA.
- CDC. (12 de Enero de 2018). *Deslizamientos de tierra y aludes de barro*. Obtenido de <https://www.cdc.gov/es/disasters/landslides.html>
- Cepeda, J. (2021). Deslizamientos y Aprendizaje Automático. *Revista de Investigaciones UCA*, 27-33. Obtenido de <https://revistas.uca.edu.sv/index.php/iuca/article/view/6812/6804>
- Criado, M. (23 de Febrero de 2018). *La deforestación aumenta el impacto de la radiación solar en la superficie terrestre*. Obtenido de Diario El País: https://elpais.com/elpais/2018/02/22/ciencia/1519277366_574125.html
- Díaz, J. (2021). Aprendizaje Automático y Aprendizaje Profundo. *Ingeniare. Revista chilena de ingeniería*, 29(2), 182-183. Obtenido de

<https://www.scielo.cl/pdf/ingeniare/v29n2/0718-3305-ingeniare-29-02-180.pdf>

Díaz, V. (01 de Abril de 2017). *Los deslizamientos de tierra más mortales en el mundo desde 2010*. Obtenido de Diario El Comercio:
<https://www.elcomercio.com/actualidad/mundo/deslizamientos-muertes-avalancha-colombia-putumayo.html>

Estrada, E. (2020). Inteligencia emocional y resiliencia en adolescentes de una institución educativa pública de Puerto Maldonado. *Ciencia y Desarrollo*, 23(3), 27-35. doi:<http://dx.doi.org/10.21503/cyd.v23i3.2139>

Fernández, S. (Junio de 2022). *Métodos de regresión y clasificación basados en árboles*. [Tesis de grado. Universidad de Valladolid]. Obtenido de <https://uvadoc.uva.es/bitstream/handle/10324/53822/TFG-I-2181.pdf?sequence=1>

Florez, A., & Pérez, J. (2019). Técnicas para la predicción espacial de zonas susceptibles a deslizamientos. *Avances: Investigación en Ingeniería*, 16(1), 20-49. doi:<https://doi.org/10.18041/1794-4953/avances.1.5188>

Flórez, J., & Flórez, L. (2021). *Detección de deslizamientos de tierra utilizando técnicas de inteligencia artificial IA (2022)* [Tesis de grado. Pontificia Universidad Javeriana]. Obtenido de <https://repository.javeriana.edu.co/bitstream/handle/10554/61566/516-attachment-1657291268.pdf?sequence=1>

Gallardo, O. (2022). *Evaluación de susceptibilidad por fenómenos de remoción en masa mediante machine learning. Caso de estudio: vía Papallacta-Baeza*. [Tesis de grado. Universidad Central del Ecuador]. Obtenido de <http://www.dspace.uce.edu.ec/bitstream/25000/28627/1/UCE-FIGEMPA-CIG-GALLARDO%20OSWALDO.pdf>

Gallo, K. (10 de eNERO de 2020). *Uso de la tecnología en la prevención de desastres*. Obtenido de UTPL: <https://noticias.utpl.edu.ec/uso-de-la-tecnologia-en-la-prevencion-de-desastres>

- Hernández, R., & Mendoza, C. (2018). *Metodología de la investigación: Las rutas cuantitativa, cualitativa y mixta*. México: McGraw-Hill Interamericana.
- Hernández, R., Fernández, C., & Baptista, P. (2014). *Metodología de la investigación*. México : Mc Graw Hill.
- Klaine, P., Imra, M., Onireti, O., & Souza, R. (2017). A survey of machine learning techniques applied to selforganizing cellular networks. *IEEE Communications Surveys & Tutorials*, 19(4), 99.
doi:10.1109/COMST.2017.2727878
- Kuglitsch, M., Albayrak, A., Aquino, R., Craddock, A., Edward-Gill, J., Kanwar, R., . . . Luterbacher, J. (2022). La inteligencia artificial aplicada a la reducción de riesgos de desastre: oportunidades, retos y perspectivas. *Reducción de riesgos de desastre*, 71(1). Obtenido de <https://public.wmo.int/es/recursos/bolet%C3%ADn/inteligencia-artificial-y-la-reducci%C3%B3n-de-riesgos>
- Lee, W. (2019). *Python Machine Learning*. Wiley. Obtenido de <https://bibliotecas.ups.edu.ec:3488/es/ereader/bibliotecaups/186090?page=27>.
- Maisueche, A. (Septiembre de 2019). *Utilización del machine learning en I industria 4.0. [Tesis de Maestría. Universidad de Valladolid.]*. Obtenido de <https://uvadoc.uva.es/bitstream/handle/10324/37908/TFM-I-1372.pdf?sequence=1&isAllowed=y>
- Medina, C., & Astudillo, J. (2017). *Ubicación e identificación de los principales deslizamientos de la vía Macará-Sabiango de la provincia de Loja*. Obtenido de Universidad Nacional de Loja.
- Medina, R., & Ñique, C. (2017). Bosques aleatorios como extensión de los árboles de clasificación con los programas R y Python. *Interfases*, 165-189. doi:<http://dx.doi.org/10.26439/interfases2017.n10.1775>

- Núñez, A., Armengol, M., & Sánchez, M. (2019). Big Data Analysis y Machine Learning en medicina intensiva. *Medicina Intensiva*, 43(7), 416-426. Obtenido de <https://doi.org/10.1016/j.medin.2018.10.007>
- ONU. (21 de Julio de 2020). *La deforestación disminuye, pero no al ritmo suficiente para proteger al planeta*. Obtenido de <https://news.un.org/es/story/2020/07/1477741>
- Ospina, J., & Aristizábal, E. (2021). Aplicación de inteligencia artificial y técnicas de aprendizaje automático para la evaluación de la susceptibilidad por movimientos en masa. *Revista Mexicana de Ciencias Geológicas*, 38(1), 43-54. Obtenido de <https://geohazards.com.co/documentos/publicacion1.pdf>
- Pal, T. (2023). The Exploratory Study of Machine Learning on Applications, Challenges, and Uses in the Financial Sector. En *Advanced machine learning algorithms for complex financial applications* (págs. 156-165). doi:10.4018/978-1-6684-4483-2.ch010
- Pilligua, E., & Guaranda, B. (2021). *Análisis geotécnico del retroceso del acantilado, zona de Santa rosa – Manabí y su incidencia en la infraestructura vial. [Tesis de maestría. Escuela Superior Politécnica del Litoral]*. Obtenido de <https://www.dspace.espol.edu.ec/bitstream/123456789/53840/1/T-70620%20Guaranda-Pilligua.pdf>
- Ramírez, A., Sánchez, L., Anzures, M., Sánchez, S., & Larios, M. (2020). SISTEMA DE RECONOCIMIENTO DE VOZ BASADO EN UN MÉTODO DE APRENDIZAJE SUPERVISADO Y LA CORRELACIÓN DE PEARSON. *Pistas Educativas*, 42(136), 743-764. Obtenido de <https://pistaseducativas.celaya.tecnm.mx/index.php/pistas/article/view/2442/1961>
- Raschka, S., & Mirjalili, V. (2019). *Aprendizaje automático con Python*. Marcombo. Obtenido de https://books.google.es/books?id=5EtOEAAAQBAJ&printsec=frontcover&hl=es&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false

- Ruiz, G. (30 de Junio de 2019). *Modelo de análisis de datos utilizando técnicas de aprendizaje supervisado y no supervisado, para identificar patrones en la información generada por los pacientes, sometidos a juegos diseñados como un instrumento de apoyo terapéutico*. . Obtenido de [Tesis de grado. Universidad Jorge Tadeo Lozano]:
<https://expeditiorepositorio.utadeo.edu.co/bitstream/handle/20.500.12010/8502/Trabajo%20de%20grado.pdf?sequence=1&isAllowed=y>
- Sandoval, L. (2018). Algoritmos de aprendizaje automático para análisis y predicción de datos. *ITCA-FEDAPE*, 36-40. Obtenido de http://www.redicces.org.sv/jspui/bitstream/10972/3626/1/Art6_RT2018.pdf
- Sotomayor, N. (2020). *Planificación curricular docente y la calidad educativa de la Unidad Educativa Salitre, Guayaquil, 2020*. [Tesis de Maestría. Universidad Claudio Vallejo]. Obtenido de https://repositorio.ucv.edu.pe/bitstream/handle/20.500.12692/47413/Sotomayor_RNE-SD.pdf?sequence=1&isAllowed=y
- Suárez, J. (1998). *Deslizamientos y Estabilidad de Taludes en Zonas Tropicales*. Colombia: Instituto de Investigaciones sobre Erosión y Deslizamientos. Obtenido de <https://desastres.medicina.usac.edu.gt/documentos/docgt/pdf/spa/doc0101/doc0101.pdf>
- Torres, C. (Enero de 2018). *Introducción a Machine Learning Cuántico*. [Tesis de grado. Escuela Colombiana de Ingeniería Julio Garavito]. Obtenido de <https://repositorio.escuelaing.edu.co/bitstream/handle/001/900/Torres%20Torres%2c%20Camilo%20Andr%c3%a9s%20-%202018.pdf?sequence=1&isAllowed=y>
- Trejo, O., & Miramá, V. (2018). Machine learning algorithms for intercell interference coordination. *Sistemas & Telemática*, 16(46), 37-57. doi:doi:10.18046/syt.v16i46.3034

- Villaseca, D., & González, S. (2021). *De Silicon Valley a tu negocio: Innovación, data e inteligencia artificial*. Madrid: ESIC Editorial. Obtenido de https://books.google.es/books?hl=es&lr=&id=7MVAEAAAQBAJ&oi=fnd&pg=PA7&dq=El+machine+learning+como+una+evoluci%C3%B3n+tecnol%C3%B3gica+de+la+inteligencia+artificial:+ventajas+y+su+aplicaci%C3%B3n+pr%C3%A1ctica+en+el+%C3%A1rea+de+marketing.&ots=yA2M_47k-u&s
- Viteri, F., Herrera, L., & Bazurto, A. (2018). Importancia de las Técnicas del Marketing Digital. *Revista Científica Mundo de la Investigación y el Conocimiento*. 2(1)., 2(1), 764-783.
doi:10.26820/recimundo/2.1.2018.764-783
- Wenjian, H., Singh, R., & Scalettar, R. (2017). Discovering phases, phase transitions, and crossovers through unsupervised machine learning: A critical examination. *Phys. Rev*, 114.
doi:doi:10.1103/PhysRevE.95.062122.

ANEXOS

Anexo 1 Formato de entrevista

1. Bajo su criterio, ¿se evalúan constantemente las zonas deforestadas para conocer si el lugar está propenso a un deslizamiento?

2. En base a su nivel de experiencia laboral, ¿se ha implementado un buen sistema para la predicción de deslizamientos de tierra en zonas deforestadas?

3. ¿Qué tipo y cuántos sistemas han empleado para la predicción de deslizamientos de tierra en zonas deforestada?

4. ¿Cree usted que el actual sistema tiene un alto nivel de precisión, sensibilidad y especificidad, y de qué manera ha contribuido en la predicción de deslizamientos de tierra en zonas deforestada?

5. ¿Cree usted que al implementar un algoritmo de machine learning beneficiaría en la predicción de deslizamientos de tierra en zonas deforestada?

6. ¿Qué tipo de algoritmo de machine learning recomendaría para la predicción de deslizamientos de tierra en zonas deforestada y por qué?