

Escuela Superior Politécnica del Litoral

Facultad de Ingeniería en Electricidad y Computación

QuakePredictEC: Plataforma de análisis y predicción de sismos en Ecuador

asociados a variaciones climáticas extremas

TECH-450

Proyecto Integrador

Previo la obtención del Título de:

Ingeniero/a en Ciencias de la Computación

Presentado por:

Carlos Alexis Loor Hoyos

Yanaleen Britney Plúas Macías

Guayaquil - Ecuador

Año: 2025

Dedicatoria

Dedico este trabajo, en primer lugar, a Dios por permitirme culminar esta importante etapa de mi vida. A mis padres Iván y Silvia, por su amor y apoyo incondicional. A mis hermanos Eyleen, Elkyn, Arlet, por ser mis mejores amigos y una fuente constante de fortaleza. A mi compañero de vida, Benjamín, por su paciencia y creer en mí desde el día uno. Finalmente, dedico este trabajo a mi yo interior por nunca rendirse y mantenerse perseverante, aun cuando las cosas se ponen difíciles.

Yanaleen Britney Plúas Macías

Dedicatoria

El presente proyecto lo dedico a mis padres, quienes han sido mi motivación constante para superar todas las dificultades que se me presentaron en la carrera.

A mis hermanas, abuelos, tíos y primos, quienes siempre han estado pendientes de mí y alentándome a conseguir mis objetivos.

A las personas que me acompañaron en este camino y que hoy ya no están físicamente conmigo, pero cuyo recuerdo, apoyo y ejemplo seguirán presentes.

Carlos Alexis Loor Hoyos

Agradecimientos

En primera instancia quiero agradecer a Dios, por brindarme salud y acompañarme en este largo camino.

A mi papá Iván, por enseñarme a ser una mujer fuerte; y a mi mamá, Silvia, por inculcarme la perseverancia.

A mi otra mitad, mi hermana Eyleen, por ser mi mayor apoyo y mi ejemplo para seguir; sin ti no sería nada de lo que soy hoy en día.

A mis hermanos Elkyn y Arlet, por darme la fortaleza para no rendirme. A mi compañero de vida, Benjamín, por su amor y su paciencia infinita, por hacerme sentir acompañada desde el día uno, tu apoyo ha sido fundamental para mí.

Finalmente, agradezco a mis mentores y a esta prestigiosa universidad, no solo por brindarme formación académica, sino también enseñarme que, con esfuerzo y sacrificio, todo es posible, sin importar el tiempo que te tome alcanzar las metas.

Yanaleen Britney Plúas Macías

Agradecimientos

Mi más sincero agradecimiento a nuestra tutora y profesora, por su guía y acompañamiento que nos permitieron concluir este proyecto.

A los docentes que formaron parte de mi desarrollo académico, cuyas enseñanzas me permitirán alcanzar mis metas profesionales.

A la Escuela Superior Politécnica del Litoral, por brindarnos los recursos académicos, técnicos y humanos necesarios para nuestra formación.

A mi familia por creer siempre en mí y darme la confianza necesaria para avanzar en toda mi etapa académica a pesar de las dificultades.

Carlos Alexis Loor Hoyos

Declaración Expresa

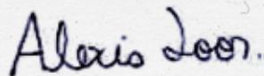
Nosotros Carlos Alexis Loor Hoyos y Yanaleen Britney Plúas Macías acordamos y reconocemos que:

La titularidad de los derechos patrimoniales de autor (derechos de autor) del proyecto de graduación corresponderá al autor o autores, sin perjuicio de lo cual la ESPOL recibe en este acto una licencia gratuita de plazo indefinido para el uso no comercial y comercial de la obra con facultad de sublicenciar, incluyendo la autorización para su divulgación, así como para la creación y uso de obras derivadas. En el caso de usos comerciales se respetará el porcentaje de participación en beneficios que corresponda a favor del autor o autores.

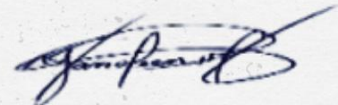
La titularidad total y exclusiva sobre los derechos patrimoniales de patente de invención, modelo de utilidad, diseño industrial, secreto industrial, software o información no divulgada que corresponda o pueda corresponder respecto de cualquier investigación, desarrollo tecnológico o invención realizada por mí/nosotros durante el desarrollo del proyecto de graduación, pertenecerán de forma total, exclusiva e indivisible a la ESPOL, sin perjuicio del porcentaje que me/nos corresponda de los beneficios económicos que la ESPOL reciba por la explotación de mi/nuestra innovación, de ser el caso.

En los casos donde la Oficina de Transferencia de Resultados de Investigación (OTRI) de la ESPOL comunique al/los autor/es que existe una innovación potencialmente patentable sobre los resultados del proyecto de graduación, no se realizará publicación o divulgación alguna, sin la autorización expresa y previa de la ESPOL.

Guayaquil, 17 de octubre del 2025.



Carlos Alexis Loor Hoyos



Yanaleen Britney Plúas Macías

Evaluadores

Lucia Marisol Villacrés Falconi

Profesor de materia

Adriana Elisa Collaguazo Jaramillo

Tutora de proyecto

Resumen

Ecuador es uno de los países con la mayor actividad sísmica del mundo, una situación que requiere que se busquen nuevas formas de gestionar el riesgo que implica un sismo y anticiparnos al peligro para proteger e informar a la población. Este proyecto, llamado QuakePredictEC, inicia con una hipótesis que se ha investigado en los últimos cinco años: “Aunque el origen de los sismos es fundamentalmente tectónico, ciertas variaciones meteorológicas pueden alterar el equilibrio de la corteza terrestre y favorecer su ocurrencia”.

Partiendo de esta premisa, analizamos la correlación estadística entre las variables climáticas y los registros históricos de sismos en el Ecuador, con el fin de llegar a entrenar un modelo predictivo de ocurrencia de sismos. A continuación, se unieron los registros históricos de sismos del Instituto Geofísico de la Escuela Politécnica Nacional con los datos meteorológicos de la plataforma Open Meteo. Se analizaron y procesaron las variaciones climáticas de los 30 días previos a cada sismo registrado.

Con estos datos, se entrenaron y compararon tres modelos de aprendizaje automático (Random Forest, Multi Layer Perceptron y XGBoost) que analizando las variaciones climáticas, como presión, temperatura y precipitación, de los últimos treinta días de cada cantón del país, pueda predecir si ocurrirá un sismo en los próximos días.

Los resultados indicaron que el modelo XGBoost fue el más efectivo. Obtuvo un Recall (sensibilidad) del 88%, lo que significa que es capaz de detectar la gran mayoría de los sismos. Más que una predicción, este proyecto ofrece una herramienta de apoyo para la gestión de riesgos y planificación, permitiendo a las autoridades pasar de un monitoreo general a una vigilancia focalizada en las zonas de mayor probabilidad de sismo.

Palabras claves: Sismos, Variables Climáticas, Aprendizaje Automático, Predicciones,

Abstract

Ecuador is one of the countries with the highest seismic activity in the world, a condition that demands the exploration of new ways to manage seismic risk and anticipate potential hazards in order to protect and inform the population. This project, called QuakePredictEC, is based on a hypothesis that has been investigated over the past five years: “Although the origin of earthquakes is fundamentally tectonic, certain meteorological variations may alter the equilibrium of the Earth’s crust and favor their occurrence.”

Based on this premise, we analyzed the statistical correlation between climatic variables and historical earthquake records in Ecuador, with the aim of training a predictive model for earthquake occurrence. To this end, historical seismic records from the Geophysical Institute of the National Polytechnic School were combined with meteorological data from the OpenMeteo platform. Climatic variations during the 30 days prior to each recorded earthquake were analyzed and processed.

Using these data, three machine learning models: Random Forest, Multi-Layer Perceptron, and XGBoost were trained and compared. By analyzing climatic variations such as pressure, temperature, and precipitation over the previous thirty days for each canton in the country, the models aim to predict whether an earthquake will occur in the following days.

The results indicated that the XGBoost model was the most effective, achieving a recall (sensitivity) of 88%, meaning it is capable of detecting the vast majority of earthquakes. Rather than providing a definitive prediction, this project offers a decision-support tool for risk management and planning, enabling authorities to move from generalized monitoring to focused surveillance in areas with a higher probability of seismic occurrence.

Keywords: Earthquake, Climate Indicators, Machine Learning, Prediction.

Índice General

Resumen.....	I
<i>Abstract</i>	II
Índice de figuras.....	V
Índice de tablas	VI
Capítulo 1.....	1
1. Introducción	2
1.1 Descripción del problema	2
1.2 Justificación del problema	4
1.3 Objetivos.....	5
1.3.1 <i>Objetivo general</i>	5
1.3.2 <i>Objetivos Específicos</i>	5
1.4 Marco teórico.....	5
Capítulo 2.....	11
2. Metodología	12
2.1 Diseño del modelo	12
2.1.1 <i>Obtención de datos</i>	12
2.1.2 <i>Procesamiento de variables</i>	13
2.1.3 <i>Descripción del conjunto de datos</i>	15
2.1.4 <i>Evaluación y selección de arquitecturas</i>	17
2.1.5 <i>Entrenar el modelo</i>	17
2.2 Requerimientos	19
2.2.1 <i>Requerimientos funcionales</i>	19
2.2.2 <i>Requerimientos no funcionales</i>	20
2.3 Tolerancia a fallos y recuperación:.....	20
2.4 Alcance de la solución.....	21
2.5 Limitaciones de la solución	21
2.6 Riesgos de la solución	22
2.7 Beneficios de la solución.....	23
2.8 Usuarios	23
2.8.1 <i>Historias de Usuario</i>	23
2.8.2 <i>Pruebas de usuario</i>	28
2.9 Diseño de la solución.....	29
2.9.1 <i>Arquitectura general del sistema (vista lógica)</i>	29

2.9.2 Estructura de procesamiento de datos (vista de desarrollo)	30
2.9.3 Diagrama de despliegue (vista física)	31
2.9.4 Diagrama de estados (vista de procesos)	33
2.9.5 Diagrama de casos de uso	34
Capítulo 3.....	36
3. Resultados y análisis	37
3.1 Resultados de la implementación del modelo	37
3.1.1 Análisis de la distribución de las variables climáticas respecto a la ocurrencia de sismos	38
Figura 8. Distribución de las variables climáticas respecto a la ocurrencia de sismos	38
3.1.2 Prueba de hipótesis para validación estadística de patrones entre las variables climáticas y la ocurrencia de sismos	40
3.1.3 Correlación entre las variables climáticas y la magnitud de los sismos.....	42
3.2 Resultados de la evaluación de los modelos.....	43
3.2.1 Modelo 1: XGBoost.....	43
3.2.2 Modelo 2: Random Forest	46
3.2.3 Modelo 3: Multi Layer Perceptron (MLP)	48
3.2.4 Tabla comparativa de los resultados de los modelos	51
3.3 Evaluación de usabilidad de la página web	53
3.3.1 Resultados de la evaluación del mapa interactivo	53
3.3.2 Resultados de la evaluación del módulo de alertas.....	54
Capítulo 4.....	56
4. Conclusiones	57
5. Recomendaciones.....	59
Referencias.....	60
ANEXOS	63

Índice de figuras

Figura 1. Etapas de la metodología (Autoría propia).....	12
Figura 2. Modelo “4+1” de Kruchten [9].....	29
Figura 3. Arquitectura general del sistema	30
Figura 4. Diagrama de procesamientos de datos	31
Figura 5. Diagrama de Despliegue.....	32
Figura 6. Diagrama de estados.....	34
Figura 7. Diagrama de casos de uso.....	35
Figura 8. Distribución de las variables climáticas respecto a la ocurrencia de sismos.....	38
Figura 9. Gráfica de curva de aprendizaje del modelo XGBoost	44
Figura 10. Matriz de confusión del modelo XGBoost.....	45
Figura 11. Matriz de confusión del modelo Random Forest	47
Figura 12. Gráfica de curva de aprendizaje del modelo Multi Layer Perceptron.....	49
Figura 13. Matriz de confusión del modelo Multi Layer Perceptron	50

Índice de tablas

Tabla 1. Diccionario de variables	16
Tabla 2. Requerimientos funcionales.....	19
Tabla 3. Requerimientos no funcionales.....	20
Tabla 4. Historias de Usuario General	24
Tabla 5. Historias de usuario del cliente	25
Tabla 6. Resultados de las pruebas de hipótesis U de Mann-Whitney	41
Tabla 7. Coeficientes de correlación de las variables climáticas respecto a la magnitud.....	42
Tabla 8. Tabla de los resultados de los modelos.....	51

Capítulo 1

1. Introducción

1.1 Descripción del problema

Ecuador se encuentra en el Cinturón de Fuego del Pacífico, una zona donde se concentra la mayor parte de la actividad sísmica del planeta. En esta región, la placa oceánica de Nazca se mueve y se introduce lentamente bajo la placa Sudamericana. Como resultado de este proceso, se acumulan grandes cantidades de energía que al liberarse, provocan frecuentes sismos principalmente en la costa y en la cordillera de los Andes. [2]

Ecuador ha experimentado desde temblores leves hasta terremotos devastadores. Entre ellos destacan el de Esmeraldas en 1906, con magnitud 8.8, considerado uno de los más grandes del planeta; el de Bahía de Caráquez en 1998, de magnitud 7.1 y el de Pedernales en 2016, de magnitud 7.8, el cual causó más de 670 víctimas mortales, 28 000 heridos y pérdidas económicas estimadas en 3 300 millones de dólares. La recurrencia de estos eventos sísmicos evidencia la necesidad de contar con sistemas de monitoreo y prevención sísmica robustos, especialmente para las zonas más vulnerables del país. [2]

En este sentido, el país dispone de redes de monitoreo sísmico y volcánico operadas por el Instituto Geofísico de la Politécnica Nacional (IG-EPN), que incluyen una Red Nacional de Sismógrafos (RENSIG) de alrededor de 120 estaciones, una Red Nacional de Acelerógrafos (RENAC) y sistemas de transmisión de datos en tiempo real. Estos sistemas registran vibraciones del suelo y envían la información al IG-EPN para el análisis del movimiento sísmico. [1]

De igual manera, el Instituto Oceanográfico de la Armada (INOCAR), también mantiene estaciones meteorológicas y boyas oceanográficas para el monitoreo de las condiciones marítimas y atmosféricas de la costa. [1]

No obstante, aunque estas instituciones cumplen una labor importante en la observación, su función principal es detectar y registrar los fenómenos naturales una vez que el sismo ya ha empezado. En consecuencia, es imposible para la población y las autoridades tomar medidas preventivas antes de que ocurra el movimiento telúrico. [1]

A nivel mundial, la predicción sísmica ha sido un desafío porque los procesos que causan los terremotos son muy complejos, dado que, ocurren a gran profundidad e involucran múltiples fallas y condiciones geológicas a las que los instrumentos de medición no llegan. Para un país sísmico como Ecuador, la prevención de desastres depende de comprender a fondo qué ocasiona estos eventos, por eso es necesario comprender el fenómeno para reducir riesgos y fortalecer la prevención de eventos sísmicos. [1]

En este contexto, varios estudios científicos han explorado una relación entre la actividad sísmica y variaciones climáticas extremas. Se ha observado que fenómenos como intensas lluvias, sequías extremas o cambios bruscos de presión atmosférica pueden alterar la estabilidad del terreno y afectar el comportamiento de las fallas geológicas. Por ejemplo, investigaciones realizadas en California, Italia y el Himalaya sugieren que los cambios en la cantidad de agua subterránea, ya sea por acumulación durante periodos lluviosos o por pérdida en tiempos de sequía; pueden alterar la presión en el subsuelo y facilitar el deslizamiento de las fallas [3]. Es decir, el clima no causa directamente los sismos, pero puede influir en la liberación de la energía acumulada en la corteza terrestre.

Aunque estos estudios no se han realizado en Ecuador, hay mucha información que se encuentra distribuida entre distintas instituciones, por un lado, el Instituto Geofísico de la Escuela Politécnica Nacional (IG-EPN); por otro, el Instituto Nacional de Meteorología e Hidrología (INAMHI) y el Instituto Oceanográfico de la Armada (INOCAR). Sin embargo, no

existen plataformas que integren estos datos climáticos y sísmicos para identificar patrones o correlaciones que preceden la ocurrencia de un sismo. [1]

Esta falta de integración y análisis limita la capacidad de las instituciones para identificar patrones que anticipen desastres que amenacen la vida, la seguridad, la economía de los ciudadanos. En un país con alta actividad sísmica como Ecuador, esta brecha tecnológica representa un obstáculo para la implementación de estrategias preventivas efectivas. Por ende, al ejecutar las acciones de forma adecuada sería beneficioso para la población ecuatoriana estar alerta ante situaciones de riesgos.

1.2 Justificación del problema

En primer lugar, la importancia de este proyecto radica en su potencial para fortalecer los mecanismos de prevención ante eventos sísmicos en Ecuador. Actualmente, la mayoría de los sistemas se centran únicamente en el monitoreo sísmico, sin integrar otras variables que permitirían ampliar la comprensión del fenómeno para conseguir una predicción sísmica.

En este contexto, el proyecto propone un enfoque progresivo aplicando técnicas de Machine Learning para combinar los registros de variaciones climáticas asociadas a los registros de sismos, el cual es considerado una herramienta capaz de manejar relaciones no lineales. El objetivo es analizar la correlación entre ambas variables y a su vez abrir la posibilidad de descubrir patrones no identificados previamente. Al visualizar estos patrones, las instituciones municipales podrán gestionar el riesgo de forma preventiva mediante planificación territorial, simulacros, información a los ciudadanos, etc.

Finalmente, integrar este modelo en una interfaz gráfica interactiva y un mapa en tiempo real que proporcione un acceso directo a la información, permitiendo que tanto autoridades como ciudadanos comprendan los resultados de manera sencilla y práctica. Esto amplía la capacidad de respuesta y contribuye a reducir los impactos sociales, económicos y

humanos de posibles sismos. De esta manera, se visualizarán indicios sobre posibles actividades sísmicas próximas a ocurrir brindando información oportuna para la toma de decisiones.

1.3 Objetivos

1.3.1 Objetivo general

Crear una plataforma web que analice datos sísmicos y climáticos del Ecuador aplicando técnicas de inteligencia artificial para la identificación de patrones y predicciones sobre actividad sísmica asociada a variaciones climáticas extremas.

1.3.2 Objetivos Específicos

- Aplicar las técnicas de estadística y de minería de datos para la identificación de la relación entre variables climáticas y actividad sísmica.
- Elaborar un modelo de machine learning para el entrenamiento de patrones en variables sísmicas y climáticas.
- Desarrollar una interfaz gráfica interactiva que permita la visualización de los resultados del modelo en tiempo real.
- Integrar el modelo de machine learning y la plataforma web para el procesamiento de datos, predicción y visualización.

1.4 Marco teórico

Un sismo ocurre cuando la tierra libera energía que se ha ido acumulando en su interior de forma repentina. Esto es atribuible a la estructura de la corteza terrestre, que está formada por enormes fragmentos conocidos como placas tectónicas, los cuales se mueven lentamente unas con otras. Con el tiempo, ese movimiento genera presión en las zonas donde se encuentran

las placas. En consecuencia, cuando la presión es demasiado grande y las rocas no resisten más, se rompen y se desplazan bruscamente. Provocando una liberación repentina de energía que se propaga en forma de ondas sísmicas y viajan por el suelo produciendo el temblor que percibimos en la superficie durante un sismo.[2]

Estos eventos siguen un proceso llamado ‘ciclo sísmico’, que se divide en 3 etapas. Primero durante el período intersísmico, la energía se acumula lentamente en las fallas, que son fracturas profundas en la corteza terrestre. Luego, en la fase cosísmica, esa energía se libera repentinamente y ocurre el sismo. Finalmente, en el período postsísmico, algunas zonas continúan moviéndose de manera muy lenta, liberando parte de la tensión restante. [2]

En el caso de Ecuador y Colombia, la alta actividad sísmica que comparten ambos países está directamente relacionada con la intensa interacción tectónica que ocurre frente a sus costas. En esta región convergen la placa de Nazca y la placa sudamericana, generando un proceso de subducción responsable de la acumulación progresiva de tensiones en un tramo aproximado de 900 km de extensión. Como resultado, se han producido grandes terremotos en el último siglo, como el de 1906, con una magnitud entre 8.6 y 8.8 y otros eventos importantes en 1942, 1958, 1979 y 2016. [2]

La investigación geológica sugiere que la zona de subducción entre Ecuador y Colombia podría generar otro evento sísmico de gran magnitud similar al de 1906, ya que los registros históricos y las mediciones actuales indican que la energía se sigue acumulando en la profundidad. [2]

En los últimos años se han llevado a cabo distintas investigaciones con el objetivo de descubrir si ciertos cambios en las condiciones climáticas a un nivel extremo, como lluvias intensas o largos ciclos húmedos, pueden tener alguna influencia sobre la ocurrencia de terremotos. Aunque el movimiento de las placas tectónicas es y siempre será el principal

causante de los sismos, se ha observado que ciertas condiciones climáticas y cambios ambientales pueden alterar las condiciones de presión y tensión en el subsuelo. Lo cual favorece la liberación de energía en fallas que ya estaban cerca de su punto de quiebre.[2]

El estudio realizado por Pier Luigi Bragato (2022) en los Apeninos italianos resaltó una coincidencia significativa entre los periodos de clima húmedo y la ocurrencia de terremotos fuertes. Al analizar más de un siglo de datos climáticos y sísmicos, se observó que la mayoría de los grandes sismos ocurrieron en los mismos años o muy poco después de los tiempos de máxima humedad del suelo. Este patrón no fue solo una coincidencia, los resultados mostraron una clara relación estadística entre ambas variables. Esto sugiere que, los aumentos en la recarga de agua subterránea contribuirían a la desestabilización de fallas geológicas que ya se encontraban cerca de su límite de resistencia. [3]

La explicación física señala que, cuando llueve mucho durante años seguidos o se acumula agua por deshielos y suelos saturados, el peso del agua y la presión ejercida sobre las rocas se pueden transmitir hacia capas más profundas de la corteza. De esta manera, esa presión adicional, aunque sea muy pequeña comparada con las fuerzas tectónicas, puede alterar el equilibrio interno y facilitar la ruptura de la falla. Asimismo, en zonas con suelos porosos, rocas con grietas y/o cavidades por donde el agua circula con facilidad, este efecto puede ser aún más fuerte.[3]

El trabajo de Bragato también revisó registros históricos que abarcaban más de 800 años y encontró que los grandes terremotos en Italia tendían a concentrarse en épocas prolongadas de clima frío y húmedo, como durante la llamada “Pequeña Edad de Hielo” en Europa. En cambio, durante los periodos más secos o cálidos, la actividad sísmica era notablemente baja. [3]

Estos resultados evidencian la necesidad de abordar la relación entre factores climáticos y actividad sísmica mediante enfoques cuantitativos y basados en datos, capaces de procesar fenómenos no lineales y dependencias multivariadas. Dado que las interacciones entre el clima, la hidrosfera y la corteza terrestre son complejas y dinámicas, los métodos estadísticos tradicionales presentan limitaciones para capturar dichas relaciones. [3]

En este escenario, el uso de técnicas de Machine Learning ha ganado relevancia como una alternativa para el estudio de fenómenos complejos, especialmente cuando intervienen múltiples variables y relaciones no lineales difíciles de modelar mediante métodos tradicionales. Estas técnicas permiten analizar grandes volúmenes de información y extraer patrones que no son evidentes a simple vista. En la actualidad, los algoritmos de Machine Learning se emplean en tareas como la detección temprana de eventos, el análisis de relaciones entre factores geofísicos y ambientales, así como en la estimación espacial del riesgo sísmico [4].

Durante la última década, diversos estudios han incorporado algoritmos de Machine Learning en el ámbito de la sismología, abordando problemas como la clasificación automática de señales sísmicas, la estimación de magnitudes, la detección de eventos en tiempo real y el análisis conjunto de variables tectónicas, meteorológicas y ambientales [5]. Entre los enfoques más utilizados se encuentran las redes neuronales artificiales (ANN), los modelos de Random Forest y las Máquinas de Vectores de Soporte (SVM), los cuales presentan diferentes ventajas en función del tipo de datos y del objetivo del análisis, particularmente en contextos donde convergen información sísmica y climática [5].

Por ejemplo, el estudio realizado por Hajikhodaverdikan et al.(2018) [4] integra variables meteorológicas como precipitación, temperatura y velocidad del viento, con registros sísmicos usando un modelo híbrido *Particle Filter-Support Vector Regression (PR-SVR)* [4],

logrando un coeficiente de correlación de $R = 0.96$ en la predicción de magnitudes promedio. Sus resultados evidencian que las variaciones climáticas locales pueden modular la actividad sísmica superficial, especialmente cuando las fallas se encuentran próximas a su ruptura.

Simultáneamente, estudios como el de Li et al.(2018) [6] han utilizado machine learning para la discriminación de ondas sísmicas en sistema de alerta temprana, mejorando la detección y reduciendo el tiempo de respuesta ante eventos reales. Por su parte, Moghadam Nejad et al. (2026) [7] realizó una revisión exhaustiva de los principales algoritmos aplicados a la predicción y el análisis sísmico. A partir de estos hallazgos, se ha observado que los modelos híbridos y las arquitecturas de Deep Learning tienden a ofrecer mejores resultados cuando los conjuntos de datos incorporan información climática o ambiental adicional.

Posteriormente, Omni y Hashemi (2024) [5] aplicaron modelos ANN, RF Y SVM en la región de Zagros (Irán) para identificar patrones previos a sismos de magnitud mayor a 5.5. Su estudio alcanzó una precisión del 96.2% con ANN, demostrando que este tipo de redes es capaz de reconocer relaciones no lineales entre múltiples indicadores sísmicos.

En conjunto estas investigaciones demuestran que el ML no solo puede emplearse para predecir sismos, sino también sirven para analizar correlaciones entre el clima y la actividad tectónica. De este modo, se surgen modelos interpretativos que aporten a la comprensión del riesgo sísmico desde una perspectiva multidimensional.

En el caso de Ecuador, el monitoreo sísmico nacional está a cargo principalmente del instituto geofísico de la escuela politécnica nacional (IG - EPN) que opera la Red Nacional de Sismógrafos (RENSING) y la Red Nacional de Acelerógrafos (RENAC). Estas redes, con más de 120 estaciones distribuida a lo largo del país, registran y transmiten datos de vibración del suelo en tiempo real, los cuales son procesados en el centro nacional de monitoreo sísmico [1]

El Instituto Oceanográfico de la Armada (INOCAR) complementa este sistema mediante estaciones meteorológicas y boyas oceanográficas que monitorean las condiciones atmosféricas y oceánicas. Del mismo modo, el INAMHI proporciona los registros climáticos terrestres (precipitación, temperatura, humedad y presión atmosférica). Sin embargo, estas bases de datos funcionan de forma independiente, sin una integración estructurada entre variables sísmicas y meteorológicas. Por ende, esto limita el análisis conjunto de fenómenos naturales. [1]

Tras el terremoto de Pedernales (2016), estudios como el de Beauval et al. (2016) [8] y Quinde Martínez & Reinoso Angulo (2016) recalcaron la importancia de ajustar los modelos de atenuación del movimiento del suelo (GMPE) al contexto ecuatoriano, ya que los modelos extranjeros sobreestiman o subestiman las intensidades observadas. Por ello, estos trabajos evidencian la necesidad de contar con modelos locales basados en datos integrados y actualizados.

A nivel institucional, Ecuador carece de plataformas que combinen el análisis sísmico y climático para detectar correlaciones o patrones asociados a la ocurrencia de sismos. Esto representa una brecha tecnológica y científica que impide aprovechar de manera conjunta la información disponible y restringe la capacidad de generar predicciones más completas y contextualizadas basadas en múltiples variables.

Por lo expuesto, resulta pertinente desarrollar una plataforma como QuakePredictEC, que busca integrar datos de distintas fuentes oficiales como, IG-EPN, INAMHI, INOCAR en un solo entorno analítico. De tal manera, que sea posible visualizar predicciones sísmicas que consideren simultáneamente variables climáticas. Por consiguiente, generando información útil y oportuna para que las autoridades puedan tomar decisiones preventivas y alertar a la ciudadanía ante la posible ocurrencia de un evento sísmico

Capítulo 2

2. Metodología

2.1 Diseño del modelo

En el siguiente apartado se describe la metodología empleada para el desarrollo de la plataforma web QuakePredictEC, así como el proceso de entrenamiento del modelo predictivo orientado al análisis integrado de datos sísmicos y climáticos del Ecuador. Por consiguiente, como se observa en el Diagrama 1, se establecieron cuatro etapas que organizan el flujo de trabajo que van desde la obtención de la información, hasta la validación del modelo analítico.



Figura 1. Etapas de la metodología (Autoría propia)

2.1.1 Obtención de datos

Los datos sísmicos utilizados en este estudio fueron obtenidos a través del servicio de catálogos sísmicos del Instituto Geofísico (IG) del Ecuador, disponible públicamente en su sitio web institucional. Este catálogo es el principal repositorio nacional de información sobre actividad sísmica y se alimenta de la Red Sísmica Nacional, la cual ha evolucionado de manera significativa desde su creación.

El IG inició en 1987 la operación continua de la Red Nacional de Sismógrafos (RENSIG), conformada inicialmente por 11 estaciones de período corto. Con el paso del tiempo, esta red fue ampliándose, modernizándose e incorporando sensores de banda ancha y sistemas digitales de mayor resolución. En 2012, gracias a un proyecto conjunto con la SENESCYT, se completó la modernización de la red, lo que permitió mejorar la precisión de

las localizaciones y reducir los tiempos de procesamiento, obteniendo resultados confiables en un lapso de 2 a 3 minutos después de ocurrido un evento. [10]

El catálogo sísmico nacional incluye registros históricos desde 1901, complementado con registros de sismos relevantes ocurridos entre los siglos XVI y XIX, localizados mediante métodos indirectos [10]. Aunque el catálogo completo abarca más de un siglo de actividad sísmica, para este trabajo se seleccionaron únicamente los registros comprendidos desde el año 1940 hasta julio de 2025. Esta decisión se debe a que la API de Open-Meteo, utilizada para descargar las variables climáticas, solo cuenta con datos disponibles a partir del año 1940. El conjunto de datos sísmicos seleccionado incluye un total de 14031 registros, los cuales fueron utilizados para el entrenamiento y validación de modelos de aprendizaje automático. Los campos tomados fueron: fecha del evento, latitud y longitud.

En cuanto a la información climática, esta fue obtenida a través de la API de Open-Meteo, un servicio de acceso abierto que proporciona datos meteorológicos tanto históricos como actualizados, con una cobertura de hasta 80 años. Las variables consideradas en este estudio corresponden a factores potencialmente asociados a variaciones ambientales previas a la ocurrencia de sismos como son: Temperatura, precipitación y presión superficial. Estos datos fueron extraídos para las ubicaciones geográficas de cada evento sísmico registrado.

2.1.2 Procesamiento de variables

El procesamiento de las variables consistió en preparar los datos sísmicos y climáticos antes de utilizarlos en el modelo de aprendizaje automático. Este procedimiento siguió la lógica de un flujo ETL, cuyo propósito fue integrar la información proveniente de las diferentes fuentes y garantizar que ingrese al modelo en un formato uniforme y limpio. Tanto los registros sísmicos del Instituto Geofísico como los datos meteorológicos pasaron por un conjunto de transformaciones para asegurar su consistencia.

En esta etapa se realizaron tareas como la depuración de valores faltantes o no válidos, la corrección de inconsistencias, la estandarización de unidades y formatos de fecha para permitir la correlación precisa entre variables.

Una parte fundamental para construir el conjunto de datos de entrenamiento fue capturar las variaciones climáticas previas a cada evento sísmico. Para ello, se estableció una ventana temporal de 30 días anteriores a la fecha de cada sismo, los datos climáticos diarios dentro de esta ventana fueron procesados mediante medidas de tendencia central y dispersión (media, desviación estándar y rangos). Esta técnica permitió resumir los datos de 30 días en un conjunto de indicadores estadísticos que incluyen: Presión Atmosférica Media, Diferencial de Presión Atmosférica, Precipitación Acumulada, Variabilidad Térmica, Temperatura media.

Con las variables ya procesadas, se procedió a la construcción del conjunto de datos de entrenamiento mediante la integración de la información sísmica y climática. Cada registro quedó compuesto por un conjunto de variables de entrada, conformado por los indicadores climáticos calculados y la ubicación geográfica del evento, representada por la latitud y la longitud, así como por una variable de salida que indica la ocurrencia o no de un evento sísmico. Adicionalmente, se incluyeron registros correspondientes a días sin actividad sísmica, con el objetivo de que el modelo pudiera aprender a diferenciar entre condiciones asociadas a eventos sísmicos y periodos de calma.

Una vez definido el conjunto de datos, se llevó a cabo un análisis estadístico comparativo orientado a evaluar el potencial predictivo de las variables climáticas. Este análisis se enfocó en identificar posibles diferencias en la distribución de los indicadores atmosféricos entre días con ocurrencia de sismos y días sin actividad sísmica, permitiendo establecer de manera preliminar el grado de información aportado por dichas variables antes de su utilización en los algoritmos de clasificación.

Para finalizar, el conjunto de datos resultante fue almacenado en Google Drive, lo que permitió su acceso directo desde el entorno de Google Colab durante las etapas de desarrollo, entrenamiento y evaluación de los modelos de aprendizaje automático.

2.1.3 Descripción del conjunto de datos

Fuente de Datos Sísmicos: Catálogo del Instituto Geofísico de la Escuela Politécnica Nacional (IGEPN).

Fuente de Datos Climáticos: API de archivos históricos de Open-Meteo

Cobertura Temporal: Datos comprendidos entre el 1 de enero de 1940 hasta julio de 2025.

Volumen de Datos: El conjunto de datos de entrenamiento cuenta con un total de 27165 registros.

- **Eventos Sísmicos (Clase 1):** 14031 registros.
- **No Sismo (Clase 0):** 13134 registros.

Resolución e Intervalo:

- **Datos Sísmicos:** Registro por evento (instante exacto de ocurrencia con fecha y hora).
- **Datos Climáticos:** Intervalo diario (24 horas).

Datos de Entrenamiento: El dataset de entrenamiento no es una serie temporal continua, sino un conjunto de ventanas de observación. Cada fila del dataset representa un resumen estadístico de una ventana de 30 días consecutivos previos a un evento sísmico o no-sísmico.

Tabla 1. Diccionario de variables

Nombre de la Variable	Tipo	Unidad	Descripción
latitud	Entrada	Grados (°)	Coordenada geográfica (latitud) del epicentro del sismo o del punto de control aleatorio.
longitud	Entrada	Grados (°)	Coordenada geográfica (longitud) del epicentro del sismo o del punto de control aleatorio.
temp_mean	Entrada	Celsius (°C)	Temperatura Media: Promedio aritmético de la temperatura del aire a 2 metros sobre el suelo durante la ventana de 30 días previos al evento.
temp_std	Entrada	Celsius (°C)	Desviación Estándar de la Temperatura: Medida de la variabilidad térmica durante los 30 días previos. Indica qué tan fluctuante fue la temperatura en ese periodo.
pres_mean	Entrada	Hectopascales (hPa)	Presión Atmosférica Media: Promedio de la presión superficial en la ubicación dada durante la ventana de 30 días.
pres_delta	Entrada	Hectopascales (hPa)	Diferencial de Presión: Cambio neto en la presión atmosférica dentro de la ventana de tiempo (Presión del día final - Presión del día inicial). Indica si hubo una caída o aumento brusco antes del evento.
precip_sum	Entrada	Milímetros (mm)	Precipitación Acumulada: Suma total de lluvia caída en la ubicación durante los 30 días previos. Representa la carga hidrológica potencial sobre el suelo.
target_ocurrencia	Salida	Binario (no sismo/sismo)	Variable objetivo para el modelo. 0: No Sismo 1: Sismo

2.1.4 Evaluación y selección de arquitecturas

La selección de la arquitectura del modelo se basó en la naturaleza de los datos procesados y en la necesidad de gestionar el desbalance de clases asociado a los catálogos sísmicos. Aunque las investigaciones destacan la eficacia de arquitecturas de aprendizaje profundo, como las redes neuronales convolucionales (CNN) o redes de memoria a largo plazo (LSTM), para el análisis de formas de onda o series temporales directas [11, 12], nuestro enfoque transformó la dimensión temporal en un conjunto tabular de indicadores estadísticos (medias, desviaciones estándar y diferenciales).

Dado que el conjunto de datos obtenido presenta una estructura tabular y describe un fenómeno de comportamiento complejo, se consideró inicialmente el uso de modelos simples, como la regresión lineal, con el fin de evaluar posibles relaciones directas entre las variables. No obstante, este enfoque resultó insuficiente, ya que la interacción entre las condiciones climáticas y la ocurrencia de sismos no responde necesariamente a relaciones lineales o proporcionales. En estos casos, pueden existir patrones y combinaciones de variables cuya relación no se puede capturar correctamente mediante modelos lineales.

Para determinar la arquitectura apropiada, se realizó un análisis comparativo entre tres algoritmos de aprendizaje automático: Random Forest, XGBoost y Multilayer Perceptron (MLP). Estas arquitecturas fueron preseleccionadas debido a su capacidad probada para modelar relaciones no lineales y encontrar conexiones complejas entre las variables climáticas y los eventos sísmicos. [13]

2.1.5 Entrenar el modelo

El proceso de entrenamiento se ejecutó siguiendo una metodología experimental para identificar la arquitectura de aprendizaje automático más eficiente. Esta etapa no consistió

únicamente en el ajuste de un algoritmo, sino en la comparación y evaluación de tres modelos candidatos: Random Forest, Perceptrón Multicapa (MLP) y XGBoost.

El conjunto de datos se dividió en dos subconjuntos utilizando la técnica de muestreo estratificado (stratified split). Se asignó un 70% de los datos para la fase de entrenamiento y el 30% restante para la fase de pruebas. La estratificación fue importante para garantizar que la proporción de clases (Sismo vs. No Sismo) se mantuviera idéntica en ambos grupos, para evitar que las distintas magnitudes físicas sesguen el aprendizaje del modelo.

Antes de ingresar a los algoritmos, las variables numéricas (presión, precipitación, temperatura y coordenadas) fueron sometidas a un proceso de estandarización utilizando Standard Scaler. Este procedimiento transformó los datos para que tuvieran una media de 0 y una desviación estándar de 1, eliminando sesgos derivados de las diferentes escalas de medida (por ejemplo, evitar que la presión atmosférica en hPa domine numéricamente sobre la precipitación en mm).

Durante la fase de configuración y ajuste de los algoritmos, se realizaron pruebas iterativas modificando sus hiperparámetros clave para optimizar su desempeño:

- Para Random Forest, se exploró la cantidad de estimadores y la profundidad máxima de los árboles.
- Para MLP, se probaron distintas configuraciones de capas ocultas y funciones de activación.
- Para XGBoost, se ajustaron parámetros críticos como la tasa de aprendizaje (learning rate), el peso de las clases (scale_pos_weight) y la profundidad de los árboles.

2.2 Requerimientos

2.2.1 Requerimientos funcionales

Tabla 2. Requerimientos funcionales

ID	Nombre del Requerimiento	Descripción	Usuario
RF-1	Autenticación y cierre de sesión	La página web permitirá al administrador iniciar sesión con credenciales seguras y cerrar sesión para acceder al panel de gestión y funciones restringidas	Cliente
RF-2	Visualización del mapa de riesgo sísmico	La página web mostrará el mapa interactivo de Ecuador con zonas coloreadas según el nivel de riesgo sísmico calculado por el modelo predictivo.	Usuario general / Cliente
RF-3	Interacción con regiones del mapa	Al hacer clic en una ubicación del mapa, la página web mostrará una ventana emergente con la información: nivel de riesgo y variables climáticas asociadas que explican la predicción.	Usuario general / Cliente
RF-4	Explorador de reportes	Se visualizará todos los reportes generados del modelo en un formato de tabla y aplicar filtros de fecha y/o ubicación	Cliente
RF-5	Descarga de reportes	Se exportarán los reportes en formato separado por comas (csv) la tabla completa, que será mostrada en el explorador de reportes.	Cliente
RF-6	Configuración y generación de alertas automáticas	Cuando el riesgo sísmico en una zona supere el umbral configurado, la página web generará una alerta visible en la interfaz del panel de administración que incluye la predicción del sismo y la posible ubicación.	Cliente
RF-7	Registro Histórico de alertas	Todas las alertas generadas se almacenarán en la base de datos con su fecha, hora y ubicación, para futuras consultas.	Cliente
RF-8	Explorador de alertas	El consultarán todas las alertas emitidas por el sistema.	Cliente

RF-9	Descarga de alertas	Se exportarán en formato csv una tabla que contiene la información de todas las alertas emitidas por el sistema.	Cliente
RF-10	Actualización automática del mapa	La página web actualizará automáticamente el mapa cuando se generen nuevas predicciones del modelo, reflejando los niveles de riesgo más recientes	Usuario General / Cliente

2.2.2 Requerimientos no funcionales

Tabla 3. Requerimientos no funcionales

ID	Categoría	Descripción
RNF - 1	Seguridad	La página web debe garantizar la protección de los datos mediante comunicación cifrada (HTTPS) y controles de acceso por sesión.
RNF - 2	Usabilidad	La interfaz debe ser intuitiva, con navegación clara, iconografía descriptiva y colores coherentes con el mapa de riesgo, permitiendo su uso sin capacitación previa.
RNF - 3	Disponibilidad	El sistema debe estar disponible al menos el 95%, asegurando el acceso a la información y alertas en cualquier momento.
RNF - 4	Rendimiento	La carga inicial del mapa no debe superar los 5 segundos, y la respuesta al hacer clic en una región no debe superar los 2 segundos.
RNF - 6	Compatibilidad	La página web debe ser accesible desde navegadores modernos (Chrome, Edge, Firefox, Safari) y adaptarse a pantallas de escritorio, tablet y móvil.

2.3 Tolerancia a fallos y recuperación:

La página web detectará errores en los procesos de predicción o visualización que se originan principalmente por fallos en el acceso a internet o interrupciones en la conexión al servicio web. Ante estos eventos, la plataforma web intentará restablecer automáticamente la comunicación mediante reintentos controlados, si el problema persiste se utilizará copias

temporales de los datos más recientes o se notificará al usuario sobre la interrupción del servicio.

2.4 Alcance de la solución

Nuestro proyecto tiene como finalidad el desarrollo de una plataforma web interactiva que integre datos sísmicos y climáticos históricos del Ecuador. Estos datos se procesan mediante un modelo de aprendizaje automático (machine learning) entrenado para identificar patrones de comportamiento que permitan estimar el nivel de riesgo sísmico en distintas zonas del país.

El alcance de la plataforma comprende varias etapas. En primer lugar, se incluye la recolección, limpieza y procesamiento de información proveniente de instituciones oficiales como el Instituto Geofísico (IG-EPN) y la API de Open-Meteo. Seguidamente, se realiza el entrenamiento del modelo predictivo, orientado a detectar posibles correlaciones entre factores climáticos y eventos sísmicos registrados.

Una vez entrenado y validado el modelo, el sistema ofrece una visualización interactiva a través de un mapa web que muestra los niveles de riesgo por cantón, permitiendo al usuario explorar los resultados de forma dinámica. Además, el proyecto contempla la generación automática de alertas cuando el nivel de predicción supera un umbral determinado y el almacenamiento de reportes históricos para su análisis posterior.

En síntesis, el alcance del proyecto busca proporcionar una herramienta accesible que apoye la gestión preventiva del riesgo sísmico, combinando información científica y visualización tecnológica para fortalecer la toma de decisiones en el contexto ecuatoriano.

2.5 Limitaciones de la solución

A pesar de su alcance, el sistema presenta ciertas limitaciones técnicas y operativas propias de la naturaleza de los datos y del entorno de desarrollo.

En primer lugar, la información utilizada no se transmite en tiempo real, ya que las fuentes de datos se actualizan de forma periódica. Esto significa que las predicciones dependen de registros históricos y no sustituyen a los sistemas oficiales de alerta y monitoreo, por lo tanto, su uso es complementario.

Otra limitación importante es que la precisión del modelo depende directamente de la calidad y cantidad de datos disponibles. Si las bases de datos presentan vacíos o inconsistencias, el rendimiento del modelo predictivo puede verse afectado.

Además, en caso de fallos en el acceso a Internet o interrupciones en la comunicación con los servicios externos, la plataforma puede experimentar demoras en la carga o actualización de información. Para mitigar este tipo de situaciones, el sistema intentará reconectarse automáticamente y, si no es posible, mostrará los últimos datos disponibles acompañados de una notificación informativa al usuario.

Para finalizar, cabe señalar que el proyecto ha sido diseñado específicamente para el territorio ecuatoriano. Su uso en otros países requeriría ajustes en las fuentes de datos, calibraciones del modelo y adaptaciones del sistema de visualización

2.6 Riesgos de la solución

Entre los principales riesgos de la solución se encuentran la escasez o calidad limitada de los datos históricos disponibles. El sobreajuste del modelo de aprendizaje que ocurre cuando el sistema se adapta excesivamente a los datos de entrenamiento y pierde la capacidad de generalización ante nuevas observaciones. Este escenario podría desencadenar predicciones poco confiables en el territorio ecuatoriano.

También existen riesgos asociados a la dependencia de servicios tecnológicos de terceros, como las API de visualización de mapas o las API meteorológicas para actualizar

variables en tiempo real. Una interrupción en alguno de estos servicios podría afectar temporalmente la visualización o la actualización de los resultados.

Para finalizar, se consideran posibles errores en la interpretación de las predicciones, especialmente si los usuarios finales no poseen conocimientos técnicos. Para mitigar estos riesgos, la solución incorpora mecanismos de respaldo de datos, monitoreo de conexión, actualizaciones automáticas del modelo y mensajes informativos que ayuden a interpretar correctamente los resultados mostrados en la interfaz.

2.7 Beneficios de la solución

Los beneficios de la solución incluyen la posibilidad de anticipar condiciones de riesgo sísmico a partir de patrones detectados por el modelo de machine learning, ofreciendo información que respalda la planificación preventiva y las acciones tempranas.

Asimismo, la propuesta integra fuentes de datos sísmicos y climáticos que tradicionalmente se analizan de manera independiente, proporcionando una visión más completa y multidimensional del riesgo. Esta integración facilita el estudio de correlaciones entre variables ambientales y la actividad sísmica, fortaleciendo la investigación a nivel nacional.

2.8 Usuarios

Entre los usuarios de la aplicación podemos encontrar:

- Usuario general: Ciudadanía en general
- Cliente: Municipio de Salinas

2.8.1 Historias de Usuario

Usuario General

Tabla 4. Historias de Usuario General

Criterio de Aceptación (Título)	Contexto	Evento	Resultado / Comportamiento esperado
Visualización del mapa de riesgo sísmico predictivo	En caso de que el usuario navegue hacia la opción de 'Mapa de riesgo'	Cuando se muestran los resultados actualizados del modelo de predicción sísmica	El sistema muestra un mapa interactivo de Ecuador con niveles de riesgo coloreados según nivel de probabilidad sísmica y magnitud.
Visualización del mapa de riesgo sísmico filtrado por ubicación	En caso de que el usuario filtre los resultados la ubicación de su interés	Cuando el usuario seleccione una ubicación específica en la opción de filtrado	El sistema resalta en el mapa la ubicación seleccionada por el usuario
Interacción con regiones	En caso de que el usuario seleccione una provincia	Cuando el usuario haga clic en la región a analizar	El sistema muestra la probabilidad de sismo, posible magnitud y las variables climáticas asociadas para la ubicación seleccionada.
Actualización automática del mapa de riesgo sísmico	En caso de que los datos del modelo predictivo se actualicen	Cuando se genere una nueva salida del modelo	El sistema muestra un mapa actualizado con los colores y niveles de riesgo
Fallo del servidor donde se aloja el modelo	En caso de que el modelo no responda	Cuando el usuario intenta visualizar el mapa	El sistema muestra el mapa con la última actualización guardada antes de el fallo en el servidor
Fallo en la visualización del mapa	En caso de que no se encuentre disponible la visualización del	Cuando el usuario intente visualizar el mapa y el mapa no se muestra	El sistema muestra un mensaje de error "Servicio del mapa no disponible temporalmente"

	mapa		
Visualización de predicciones filtradas por fecha y ubicación	En caso de que el usuario navegue hacia la opción 'Resultados'	Cuando el usuario seleccione un rango de fechas y/o ubicación	El sistema mostrará todas las predicciones realizadas durante las fechas ingresadas para la ubicación seleccionada
Visualización de predicciones sin filtrar	En caso de que el usuario navegue hacia la opción 'Resultados'	Cuando el usuario acceda a la opción 'resultados'	El sistema presentará todas las alertas generadas mediante paginación, mostrando 10 filas por página
Fallo del servidor donde se aloja la modelo	En caso de que el modelo no responda	Cuando el usuario intenta visualizar las últimas predicciones	El sistema muestra los últimos resultados guardados antes de el fallo en el servidor

Historias de Usuario del Cliente

Tabla 5. Historias de usuario del cliente

Criterio de Contexto	Evento	Resultado / Comportamiento esperado
Aceptación		
Se recibe una alerta en menos de un minuto tras detectar un índice de probabilidad mayor al 75%	En caso de que el sistema calcule el índice de probabilidad sísmica mayor al 75%	cuando el sistema registre un índice de probabilidad sísmica mayor al 75%
visualización de alertas recibidas	El administrador acceda al módulo de alertas	Cuando haga clic en la opción de 'Alertas'
		El sistema muestra el registro histórico de alertas emitidas mediante paginación, mostrando 10 filas por página

Descarga de registro histórico de alertas emitidas	En caso que el administrador acceda al al módulo de alertas	Cuando haga clic en el botón 'Descargar registro de alertas'	El sistema inicia la descarga del registro histórico de alertas emitidas
Visualización del mapa de riesgo sísmico	En caso que el usuario navegue hacia la opción de 'Mapa de riesgo'	Cuando se muestran los resultados del modelo	El sistema muestra un mapa interactivo de Ecuador con niveles de riesgo coloreados según nivel de probabilidad sísmica.
Visualización del mapa de riesgo sísmico filtrado por ubicación	En caso de que el usuario filtre los resultados la ubicación de su interés	Cuando el usuario seleccione una ubicación específica en la opción de filtrado	El sistema resalta en el mapa la ubicación seleccionada por el usuario
Interacción con regiones	En caso que el usuario seleccione una provincia	Cuando el usuario haga clic en la región a analizar	El sistema muestra la probabilidad de sismo y las variables climáticas asociadas para la ubicación seleccionada.
Actualización automática del mapa de riesgo sísmico	en caso de que los datos del modelo predictivo se actualicen o cambien	Cuando se genere una nueva ejecución del modelo	El sistema muestra un mapa actualizado con los colores y niveles de riesgo
Fallo del servidor donde se aloja la modelo	En caso de que el modelo no responda	Cuando el usuario intenta visualizar el mapa y el mapa no está actualizado	El sistema muestra el mapa con la última actualización guardada antes de el fallo en el servidor
Fallo en la visualización del mapa	En caso de que no se encuentre disponible la visualización del mapa	Cuando el usuario intente visualizar el mapa y el mapa no se muestra	El sistema muestra un mensaje de error "Servicio del mapa no disponible temporalmente"

Visualización de predicciones filtradas por fecha y ubicación	En caso de que el usuario navegue hacia la opción 'Resultados'	Cuando el usuario selecciona un rango de fechas o ubicación	El sistema mostrará todas las predicciones realizadas durante las fechas ingresadas para la ubicación seleccionada
Visualización de predicciones sin filtrar	En caso que que el usuario navegue hacia la opción 'Resultados'	Cuando el usuario acceda a la opción 'alertas'	El sistema presentará todas las alertas generadas mediante paginación, mostrando 10 por página
Descarga de resultados	En caso de que el usuario haya seleccionado un filtro o seleccionado un conjunto de predicciones	Cuando presione el botón descargar registro	El sistema genera un archivo en formato csv con las predicciones seleccionadas
Fallo del servidor donde se aloja la modelo	En caso de que el modelo no responda	Cuando el usuario intenta visualizar las últimas predicciones y los resultados no están actualizados	El sistema muestra los últimos resultados guardados antes de el fallo en el servidor
autenticación mediante token de autenticación	En caso de que el administrador haya recibido el enlace con el token de autenticación	Cuando el administrador ingrese al enlace recibido	El administrador ingresa con éxito al panel de administración y sus funcionalidades
Token no inválido	En caso de que el administrador no haya copiado el token correctamente	Cuando el administrador intente ingresar con un token no válido	El sistema responderá con un error 401 unauthorized

2.8.2 Pruebas de usuario

Las pruebas de usuario tuvieron como finalidad evaluar la usabilidad y funcionamiento general de la plataforma web QuakePredictEC desde la perspectiva de los futuros usuarios. Estas pruebas permiten verificar que el sistema no solo funcione correctamente, sino que también sea intuitivo y pertinente para quienes van a interactuar con él en un contexto real de análisis y/o toma de decisiones.

Para ello, se elaboró un prototipo funcional de la interfaz web que integra las principales características del sistema, incluyendo la visualización del riesgo sísmico en un mapa interactivo, la consulta de los resultados generados por el modelo y el sistema de alertas.

Durante la prueba, los usuarios exploraron las dos funcionalidades principales de la plataforma sin presentar dificultades en la navegación. La estructura general fue percibida como intuitiva, y las opciones disponibles resultaron fáciles de identificar y utilizar. En la visualización del mapa los participantes no reportaron problemas para interpretar la información ni para acceder a los reportes del modelo. No obstante, los usuarios coincidieron en que la vista del mapa es considerablemente más atractiva e informativa que la sección de reportes, la cual percibieron como secundaria.

Una de las recomendaciones principales fue simplificar el esquema de semaforización empleado en la representación visual. Se sugirió reducir la cantidad de colores para facilitar la lectura rápida de los niveles de riesgo y mejorar la experiencia del usuario final al interpretar el mapa. En relación con las alertas, la funcionalidad fue bien recibida y destacada por el cliente, quien consideró que el flujo de uso es claro y coherente. Como sugerencia complementaria, propuso incorporar un mecanismo que permita compartir alertas con otros municipios o instituciones, fortaleciendo así la comunicación ante eventos potencialmente relevantes.

Finalmente, en cuanto al mapa, el cliente recomendó consultar los recursos disponibles en la Secretaría de Gestión de Riesgos, con el fin de integrar mapas más adecuados para el territorio ecuatoriano y que representen de forma más precisa las zonas de riesgo del país.

2.9 Diseño de la solución

Para el diseño de la solución se utilizó el Modelo “4+1” de Kruchten [9], el cual resalta las diferentes vistas de la arquitectura del sistema como se muestra en la figura

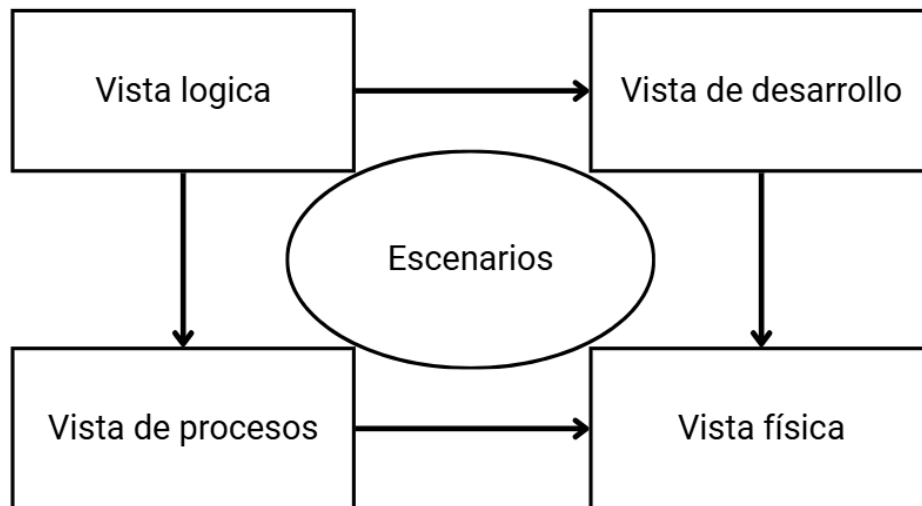


Figura 2. Modelo “4+1” de Kruchten [9]

2.9.1 Arquitectura general del sistema (vista lógica)

En la vista lógica presentamos el diagrama de Arquitectura General que representa el flujo completo de funcionamiento de la solución. Inicia con la obtención de datos sísmicos, desde la página del IG-EPN y datos climáticos, de la API de OpenMeteo, los cuales son procesados en el módulo ETL encargado de limpiar, estandarizar y preparar la información. Una vez integrados, estos datos conforman el dataset de entrenamiento que alimenta el modelo de machine learning que analiza correlaciones entre variaciones climáticas y actividad sísmica.

El modelo resultante se despliega a través de una API, que sirve como puente entre el sistema de predicción y la plataforma web, desde la cual los usuarios pueden consultar las predicciones representadas visualmente en un mapa de riesgo sísmico, visualizar alertas y acceder a reportes predictivos. Esta arquitectura modular permite separar la capa de procesamiento de datos del entorno de consulta, garantizando escalabilidad y mantenibilidad del sistema.

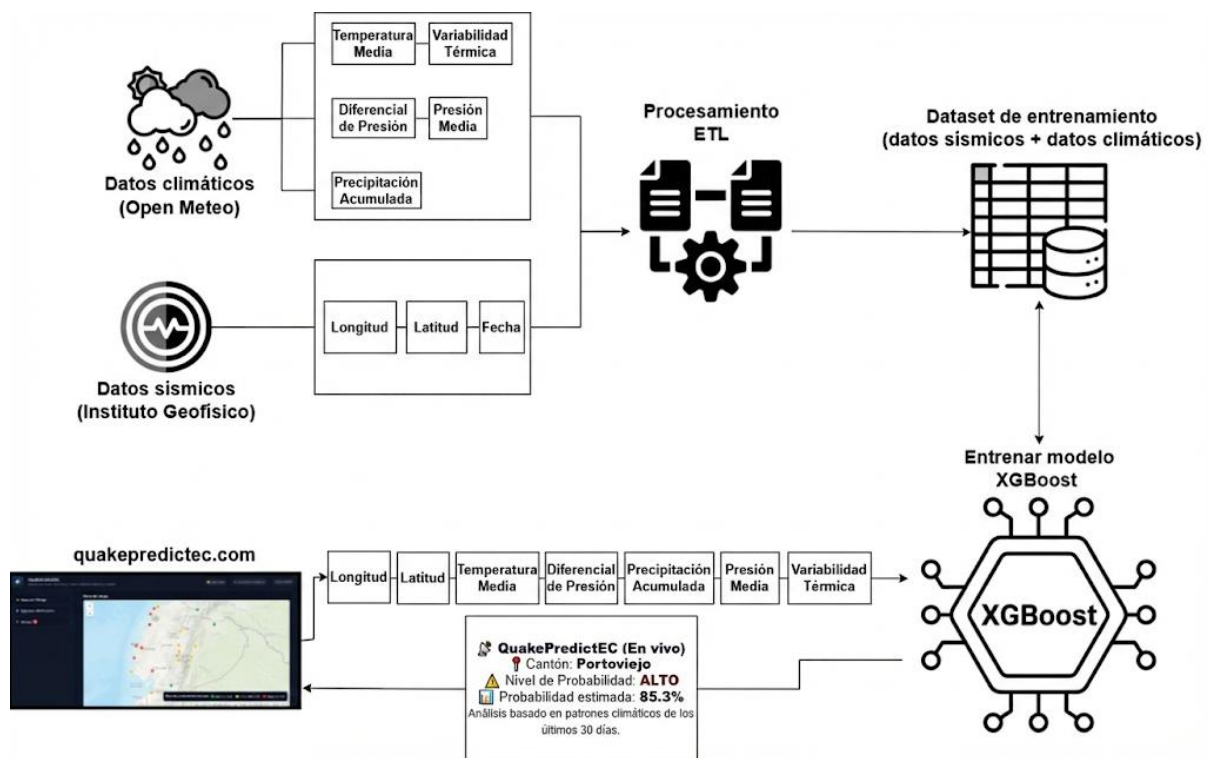


Figura 3. Arquitectura general del sistema

2.9.2 Estructura de procesamiento de datos (vista de desarrollo)

El diagrama de Procesamiento de datos ETL detalla las etapas de extracción, transformación y carga de los datos. En la fase Extract, el sistema obtiene información sísmica del IG-EPN y registros climáticos del Open-Meteo. Durante la fase Transform, los datos se someten a una limpieza que elimina registros faltantes o erróneos, se estandarizan las unidades, por ejemplo: temperatura, presión y precipitación, y se correlacionan los eventos sísmicos y climáticos en función de sus fechas para construir un conjunto de entrenamiento coherente. Por

último, en la fase Load, los datos transformados se almacenan en Google Drive el cual se puede integrar a Google Colab desde donde son utilizados para el entrenamiento del modelo predictivo.

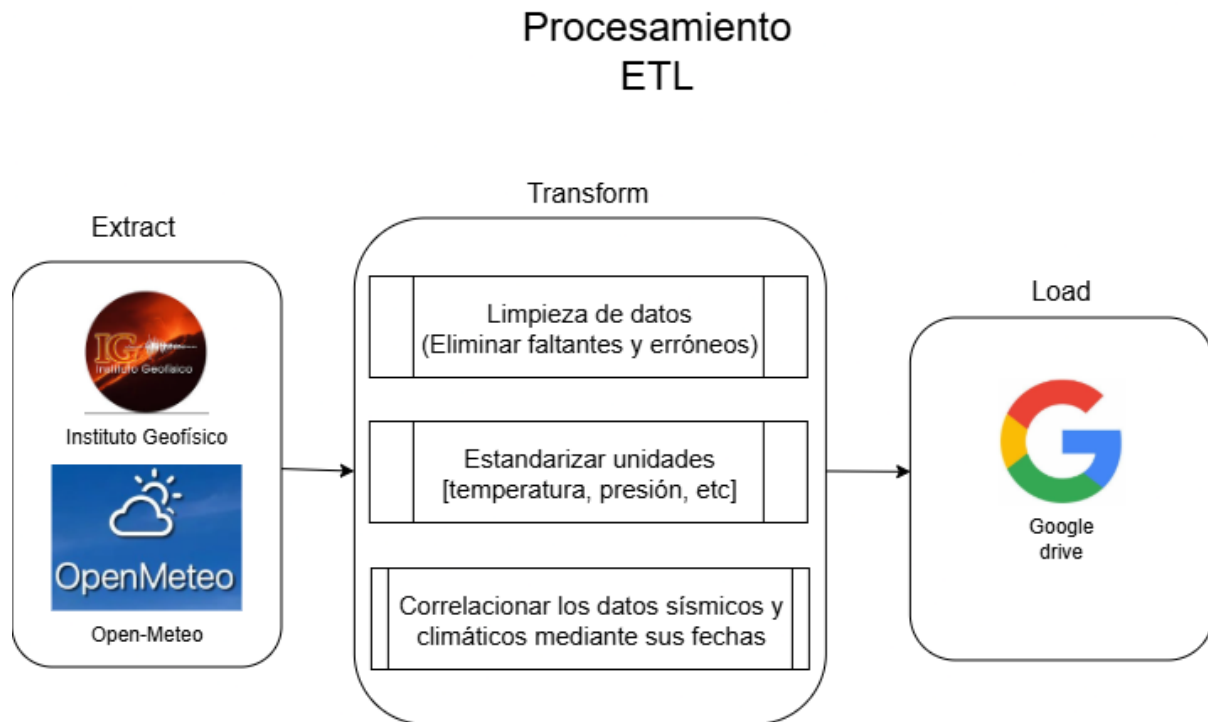


Figura 4. Diagrama de procesamientos de datos

2.9.3 Diagrama de despliegue (vista física)

El diagrama de despliegue ilustra cómo se distribuyen los componentes físicos y lógicos de la solución dentro de su infraestructura. El sistema se compone de tres nodos principales: el usuario, el servidor web y el servidor de base de datos.

El usuario accede a la aplicación desde un navegador web, donde se ejecuta la interfaz desarrollada en HTML, CSS y JavaScript. Esta interfaz se comunica, mediante el protocolo HTTPS, con el servidor web, encargado de alojar tanto el sitio principal de QuakePredictEC como la API del modelo de predicción. Dicha API gestiona las solicitudes provenientes de la

interfaz, procesando las peticiones y entregando los resultados de riesgo sísmico y generando las alertas correspondientes.

Por su parte, el servidor de base de datos MySQL almacena los reportes y las alertas diarias, asegurando la persistencia de la información procesada. La comunicación entre el servidor web y la base de datos se realiza mediante conexiones seguras, garantizando la integridad de los datos. Esta arquitectura distribuida permite un despliegue escalable, donde el modelo, la API y la interfaz pueden mantenerse o actualizarse de forma independiente sin afectar la disponibilidad general del sistema.

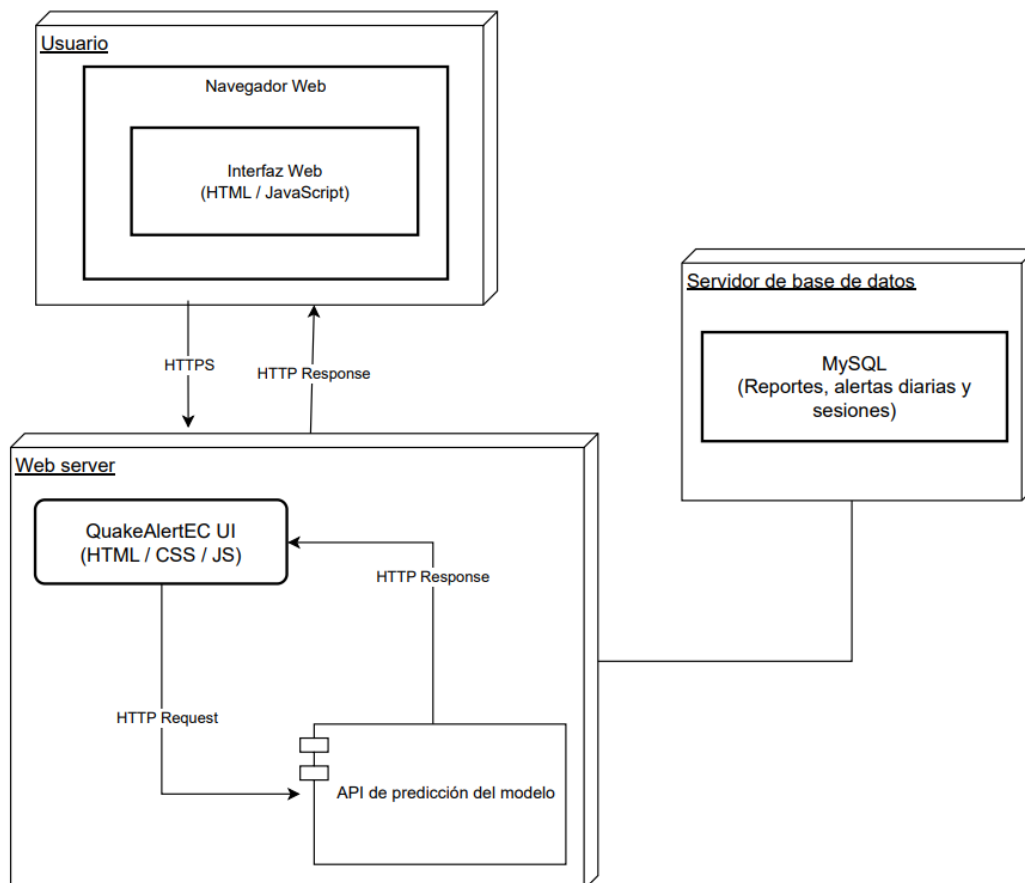


Figura 5. Diagrama de Despliegue

2.9.4 Diagrama de estados (vista de procesos)

El diagrama de estados muestra la secuencia de etapas por las que pasa la información dentro del sistema, desde su carga inicial hasta la consulta final del usuario.

El ciclo inicia con la carga de datos, donde se integran los registros sísmicos y climáticos provenientes de las fuentes oficiales. Posteriormente, los datos pasan al estado de procesamiento, donde son limpiados, estandarizados y preparados para su análisis. Una vez procesados, se ejecuta el análisis de correlaciones, etapa en la que el modelo identifica patrones entre las variables climáticas y la actividad sísmica.

Los resultados de este análisis se reflejan en la visualización del mapa interactivo, que constituye el principal medio de interacción con el usuario para la interpretación de los niveles de riesgo. A partir de esta representación, el sistema puede entrar en un estado de actualización automática de datos, cuando se incorporan nuevos registros, retornando así al ciclo de procesamiento. Al final, la información llega al estado de consulta del usuario, donde puede ser visualizada o descargada por los diferentes actores del sistema.

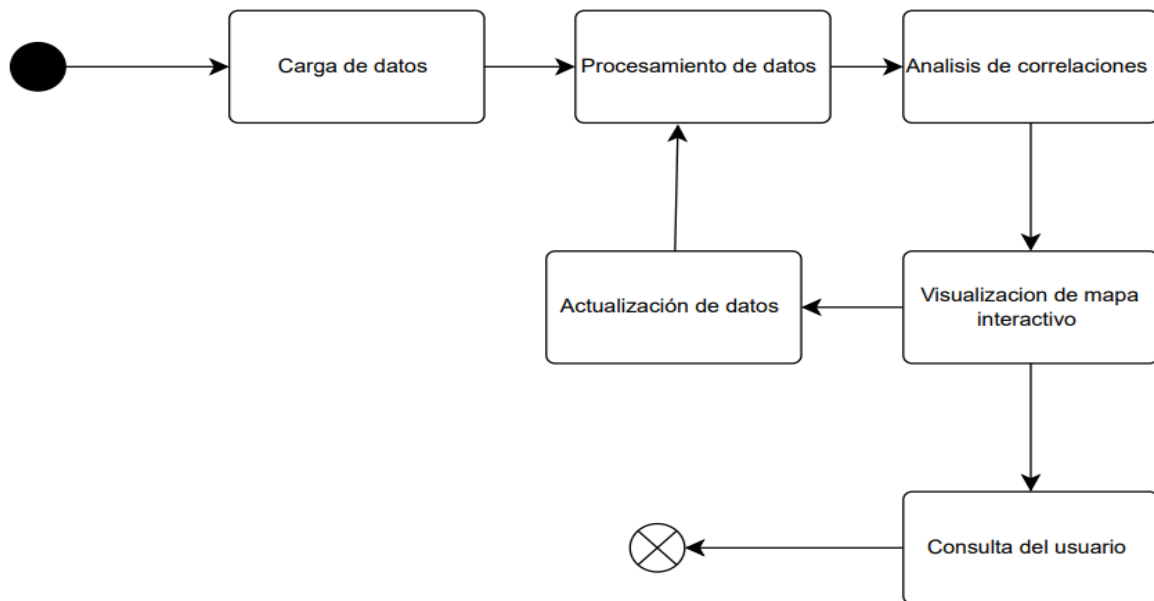


Figura 6. Diagrama de estados

2.9.5 Diagrama de casos de uso

El diagrama de casos de uso describe las interacciones entre los actores y las funcionalidades principales del sistema QuakePredictEC. Se identifican dos tipos de usuarios: el Usuario y el Administrador, cada uno con diferentes niveles de acceso y responsabilidades dentro de la aplicación.

El Usuario representa a la ciudadanía en general, cuya interacción con el sistema se limita a la consulta y visualización de información pública. Puede acceder al mapa interactivo de riesgo sísmico, filtrando por ubicación geográfica, y consultar los reportes de predicciones generados por el modelo, con la posibilidad de aplicar filtros por fecha o ubicación.

El Administrador representa a los actores institucionales, como los municipios, la Secretaría de Gestión de Riesgos u otras entidades públicas. Además de contar con las funciones de consulta disponibles para el usuario general, el administrador puede descargar reportes de predicciones en formato CSV y recibir alertas automáticas emitidas por el sistema

cuando se detectan condiciones de alto riesgo sísmico, definidas por un umbral de probabilidad establecido. Asimismo, dispone de un historial de alertas recibidas, donde puede aplicar filtros para su análisis y descargar los registros correspondientes.

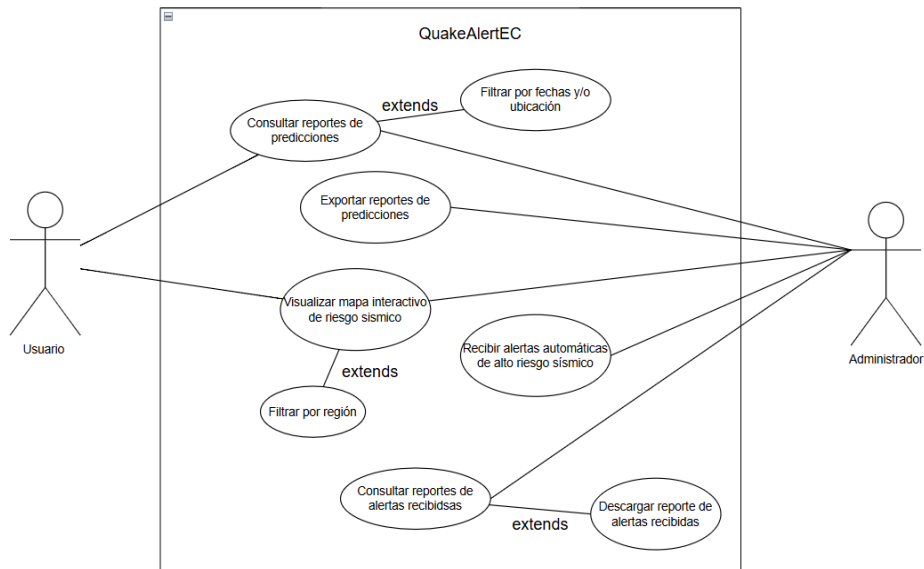


Figura 7. Diagrama de casos de uso

Capítulo 3

3. Resultados y análisis

En el presente capítulo se exponen los resultados obtenidos tras la ejecución del plan de evaluación del proyecto. El análisis se inició con el estudio de las variables climáticas seleccionadas para determinar su correlación con los sismos y justificar su uso en el desarrollo de los modelos predictivos de sismos. A partir de ello, se comparó el desempeño de los algoritmos de inteligencia artificial implementados, analizando sus curvas de aprendizaje y métricas de rendimiento para validar su eficacia. Para concluir, se presentan los resultados de las pruebas de usabilidad realizadas con usuarios finales, evaluando la funcionalidad y la experiencia de uso de la aplicación web QuakePredictEC.

3.1 Resultados de la implementación del modelo

La implementación del modelo predictivo se llevó a cabo siguiendo un flujo estructurado que incluye un previo análisis estadístico, exploración de datos y la evaluación de varias técnicas de aprendizaje automático. En una primera etapa, se realizó un análisis comparativo de las variables climáticas a utilizar con el fin de identificar las posibles diferencias en su comportamiento entre días con ocurrencia de sismos y días sin actividad sísmica. Posteriormente, se aplicó una prueba de hipótesis para evaluar si dichas diferencias eran estadísticamente significativas y, por tanto, no producto del azar. A partir de estos resultados, se analizó la relación entre las variables climáticas y la magnitud de los sismos mediante medidas de correlación, con el objetivo de determinar su potencial predictivo. De este modo, con esta información como base se inició el entrenamiento y evaluación de los modelos de aprendizaje automático: XGBoost, Random Forest y Multi Layer Perceptron (MLP) , permitiendo comparar sus capacidades para identificar los patrones relevantes y evaluar su desempeño predictivo.

3.1.1 Análisis de la distribución de las variables climáticas respecto a la ocurrencia de sismos

En el análisis exploratorio de datos, los diagramas de caja y bigotes (boxplots) son herramientas estándar que permiten visualizar de forma compacta cómo se distribuyen los valores de una variable numérica y comparar esas distribuciones entre categorías distintas. Un boxplot resume estadísticas descriptivas clave: muestra el valor mínimo y máximo (excluyendo outliers), el valor central o mediana (el punto medio de la distribución) y los cuartiles que dividen los datos en segmentos ordenados por percentiles, particularmente el primer cuartil (25 %) y el tercer cuartil (75 %) que encierran el 50 % de los datos centrales [14]. Las líneas que salen de la “caja” (los bigotes) muestran la extensión de los datos fuera de ese rango intercuartílico hasta los valores más extremos, y los puntos adicionales representan valores atípicos [15]. En este sentido, los boxplots permiten no solo comparar las posiciones centrales y la dispersión de las variables entre diferentes grupos, sino también identificar rápidamente diferencias en la forma de la distribución, como sesgos o presencia de valores extremos, lo que es particularmente útil cuando se quiere contrastar conjuntos de datos grandes y comparar categorías, como días con ocurrencia sísmica frente a días sin sismo [16].

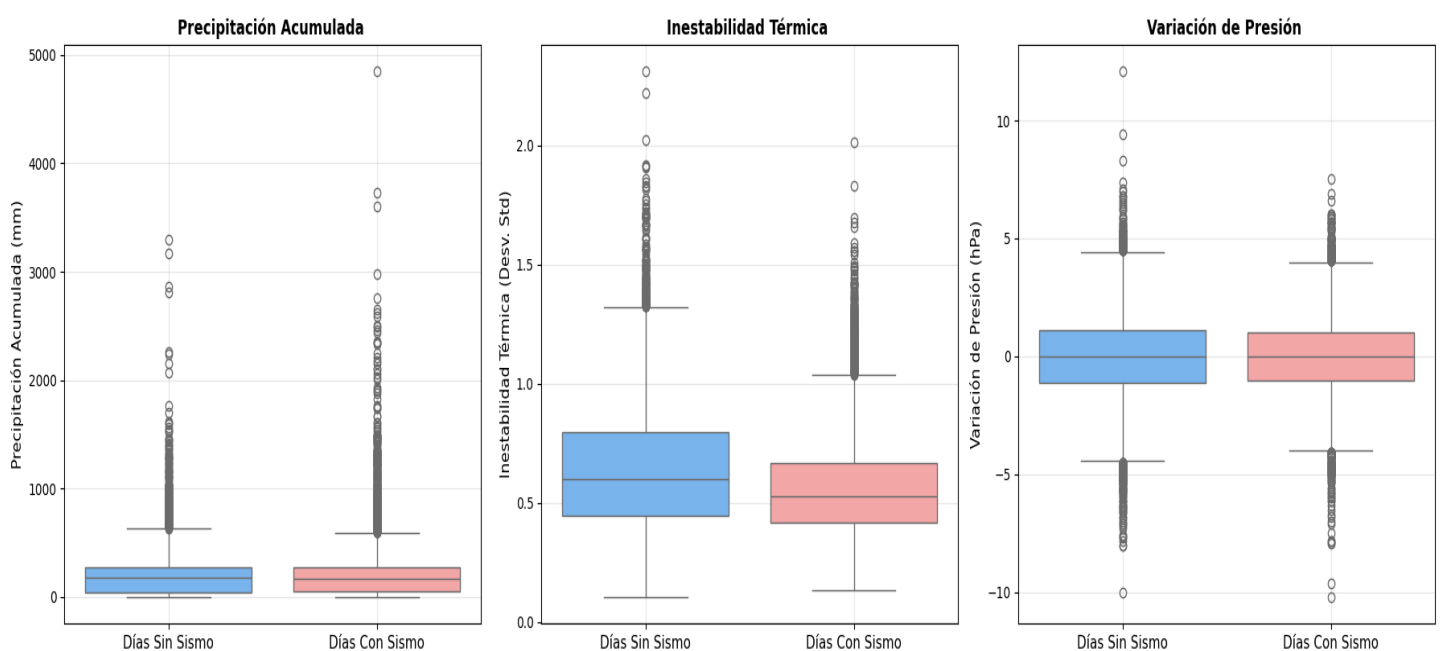


Figura 8. Distribución de las variables climáticas respecto a la ocurrencia de sismos

y una fuerte asimetría hacia valores altos, lo que indica que la mayor parte de los días presenta precipitaciones relativamente bajas, mientras que existen algunos eventos aislados con lluvias muy intensas (valores atípicos). Al observar la posición de la mediana y de la caja (rango intercuartílico), se aprecia que en los días con sismo estos valores tienden a ubicarse ligeramente por encima de los correspondientes a los días sin sismo, lo que sugiere que, en promedio, la precipitación acumulada previa a los sismos es algo mayor. Sin embargo, la superposición considerable entre ambas cajas indica que esta diferencia no es absoluta ni determinante por sí sola: existen muchos días sin sismo con niveles de precipitación similares a los de días con sismo. El gráfico sugiere una tendencia leve a mayores lluvias antes de los sismos, pero también deja claro que la precipitación (por sí sola) no permite distinguir de manera clara entre días con y sin ocurrencia sísmica.

En el boxplot de inestabilidad térmica se comparan los valores de la desviación estándar de la temperatura registrados en días sin sismo y días con sismo. En este caso, se observa una diferencia más clara en la posición central de ambas distribuciones, la mediana y el rango intercuartílico de los días sin sismo se sitúan ligeramente por encima de los correspondientes a los días con sismo. Esto indica que, en promedio, la variabilidad térmica diaria tiende a ser mayor en días sin sismos que en los días previos a la ocurrencia de sismos.

A pesar de esta diferencia, las cajas y los bigotes de ambos grupos siguen mostrando una superposición importante, lo que significa que existen muchos días con valores de inestabilidad térmica similares independientemente de que ocurra o no un sismo. Los valores atípicos presentes en ambos casos reflejan episodios puntuales de alta variabilidad térmica, pero no parecen concentrarse exclusivamente en uno de los grupos. Este gráfico sugiere que la inestabilidad térmica muestra una diferencia moderada entre días con y sin sismo, siendo ligeramente menor antes de los sismos.

En el boxplot de variación de presión atmosférica se observa que las distribuciones correspondientes a días sin sismo y días con sismo son muy similares. Las medianas de ambos grupos se encuentran cercanas a cero, lo que indica que, en promedio, no existe un cambio marcado en la presión atmosférica que diferencie claramente a un día con sismo de uno sin sismo. Además, los rangos intercuartílicos (las cajas) presentan tamaños comparables y una fuerte superposición, lo que refleja una variabilidad muy parecida en ambos escenarios.

Aunque se identifican valores atípicos tanto positivos como negativos en los dos grupos correspondientes a episodios puntuales de cambios bruscos de presión estos no se concentran de forma predominante en los días con sismo. Este gráfico sugiere que la variación de la presión atmosférica no muestra un patrón distintivo previo a la ocurrencia de sismos.

3.1.2 Prueba de hipótesis para validación estadística de patrones entre las variables climáticas y la ocurrencia de sismos

La prueba de hipótesis Mann-Whitney U es un método estadístico diseñado para comparar dos grupos independientes cuando no se puede asumir que los datos sigan una distribución normal, lo que hace que pruebas como el t de Student no sean apropiadas. En lugar de comparar medias, esta prueba combina ambos conjuntos de datos, los ordena de menor a mayor y asigna rangos a cada valor, luego suma los rangos dentro de cada grupo y calcula una estadística U en función de estas sumas para evaluar si las distribuciones de los dos grupos parecen diferentes. Bajo la hipótesis nula, se asume que los dos grupos provienen de la misma población o tienen la misma tendencia central, y valores de U o p bajos sugieren evidencia en contra de esa hipótesis. Es muy utilizada porque funciona con datos ordinales o continuos sin normalidad, es robusta a valores atípicos y útil para muestras pequeñas, lo que la convierte en una alternativa flexible frente a pruebas paramétricas cuando sus supuestos no se cumplen. [17]

Tabla 6. Resultados de las pruebas de hipótesis U de Mann-Whitney

Indicador	Mediana Con Sismos	Mediana Sin Sismos	Tendencia Observada	P-Value	Relación con Sismos
Precipitación Acumulada (mm)	165.00	177.10	Sin sismos > Con sismos	0.042	Existe Relación
Variabilidad Térmica (σ)	0.53	0.60	Sin sismos > Con sismos	0.00	Existe Relación
Presión Atm. Promedio (hPa)	979.93	986.79	Sin sismos > Con sismos	0.00	Existe Relación
Variación de Presión (hPa)	0.00	0.00	Sin sismos > Con sismos	0.03	Existe Relación
Temperatura Promedio (°C)	22.86	23.61	Sin sismos > Con sismos	0.00	Existe Relación

Los resultados de la prueba U de Mann-Whitney muestran que las distribuciones de todas las variables climáticas analizadas son significativamente diferentes entre días con ocurrencia sísmica y días sin sismo. En el caso de la precipitación acumulada, aunque la mediana en días con sismo es ligeramente menor que en días sin sismo, el valor p (0,04169) indica que esta diferencia no es resultado del azar, lo que sugiere una relación estadística con la ocurrencia de sismos. La variabilidad térmica, medida como desviación estándar, presenta una mediana notablemente menor en días con sismo y un p prácticamente cero, lo que confirma que las distribuciones de temperatura difieren entre ambos tipos de días. De forma similar, la

presión atmosférica promedio muestra una diferencia significativa entre grupos, aunque las medianas son cercanas, reflejando un patrón consistente. La variación de presión también exhibe un valor p bajo (0,03021), lo que indica diferencias en las distribuciones pese a que las medianas coinciden, y finalmente la temperatura promedio tiene una diferencia notable en medianas acompañada de un p estadísticamente significativo. En conjunto, si bien las diferencias son estadísticamente significativas, el tamaño del efecto observado es moderado, por lo que estas variables deben ser interpretadas como factores auxiliares y no determinantes individuales de la ocurrencia sísmica, lo que refuerza la idea de que estas variables pueden contener información útil para análisis predictivo, aunque ninguna por sí sola determine de manera exclusiva la ocurrencia sísmica.

3.1.3 Correlación entre las variables climáticas y la magnitud de los sismos.

La prueba de correlación de Pearson es un método estadístico que se utiliza para cuantificar la fuerza y la dirección de la relación lineal entre dos variables cuantitativas continuas. Conceptualmente, esta prueba calcula un coeficiente, comúnmente denotado como r , que toma valores entre -1 y +1. Los valores cercanos a +1 indican una correlación positiva fuerte (ambas variables tienden a aumentar juntas), los valores cercanos a -1 indican una correlación negativa fuerte (una variable aumenta mientras la otra disminuye), y valores alrededor de 0 sugieren ausencia de relación lineal significativa entre las variables [18].

Tabla 7. Coeficientes de correlación de las variables climáticas respecto a la magnitud

Variables climáticas	Coefficiente de Correlación (r) con la Magnitud Sísmica
Presión Atmosférica Media	0.261014
Variabilidad Térmica (Desv. Estándar)	0.066132
Diferencial de Presión Atmosférica	-0.007861

Precipitación Acumulada	-0.099623
Temperatura Media	0.267197

La matriz de correlación entre las variables climáticas y la magnitud de los sismos muestra coeficientes muy cercanos a cero en la mayoría de los casos, lo que indica que no existe una relación lineal fuerte entre estas condiciones y la magnitud del sismo. La presión atmosférica promedio presenta la correlación positiva más alta, pero aún débil. La inestabilidad térmica, la lluvia acumulada y el cambio de presión tienen correlaciones prácticamente nulas o débiles. Esto sugiere que, aunque el clima pueda mostrar diferencias entre días con y sin sismo, no influye directamente en la magnitud de los eventos sísmicos de forma lineal.

3.2 Resultados de la evaluación de los modelos

En esta sección se presentan los resultados obtenidos tras la evaluación de las tres arquitecturas de aprendizaje automático consideradas: Random Forest, XGBoost y Perceptrón Multicapa (MLP). El análisis comparativo se centró en medir la eficacia de cada algoritmo, priorizando métricas como la sensibilidad (capacidad de detección) y la precisión (gestión de falsas alarmas). Tras examinar el desempeño de cada modelo, se seleccionó XGBoost como el modelo definitivo para ser integrado al sistema de predicción de sismos, debido a que ofreció el equilibrio más robusto entre capacidad predictiva y estabilidad, superando la tendencia a las falsas alarmas de Random Forest y la complejidad de convergencia de la Red Neuronal.

3.2.1 Modelo 1: XGBoost

XGBoost, abreviatura de eXtreme Gradient Boosting, es un algoritmo y una biblioteca de aprendizaje automático supervisado diseñado para tareas de clasificación y

regresión que combina múltiples modelos base de árboles de decisión en un solo modelo fuerte mediante la técnica de gradient boosting. [19]

La arquitectura de XGBoost se basa en un ensamblado “aditivo” de árboles de decisión: el modelo comienza con una predictora inicial simple y, en cada iteración, se construye un nuevo árbol que intenta corregir los errores residuales dejados por los árboles anteriores. El entrenamiento está guiado por una función de pérdida diferenciable, y se optimiza utilizando información del gradiente para reducir esa pérdida progresivamente. [19]

- **Aprendizaje del modelo**



Figura 9. Gráfica de curva de aprendizaje del modelo XGBoost

La curva de pérdida demuestra un entrenamiento robusto. El modelo converge rápidamente a una solución óptima y mantiene la estabilidad en el conjunto de validación sin mostrar signos de sobreajuste a medida que aumenta la complejidad, gracias al ajuste de la tasa

de aprendizaje y la profundidad máxima. La brecha final entre entrenamiento y validación es la esperada, indicando que el modelo ha extraído la máxima señal generalizable posible de los datos disponibles.

- **Evaluación de precisión del modelo**

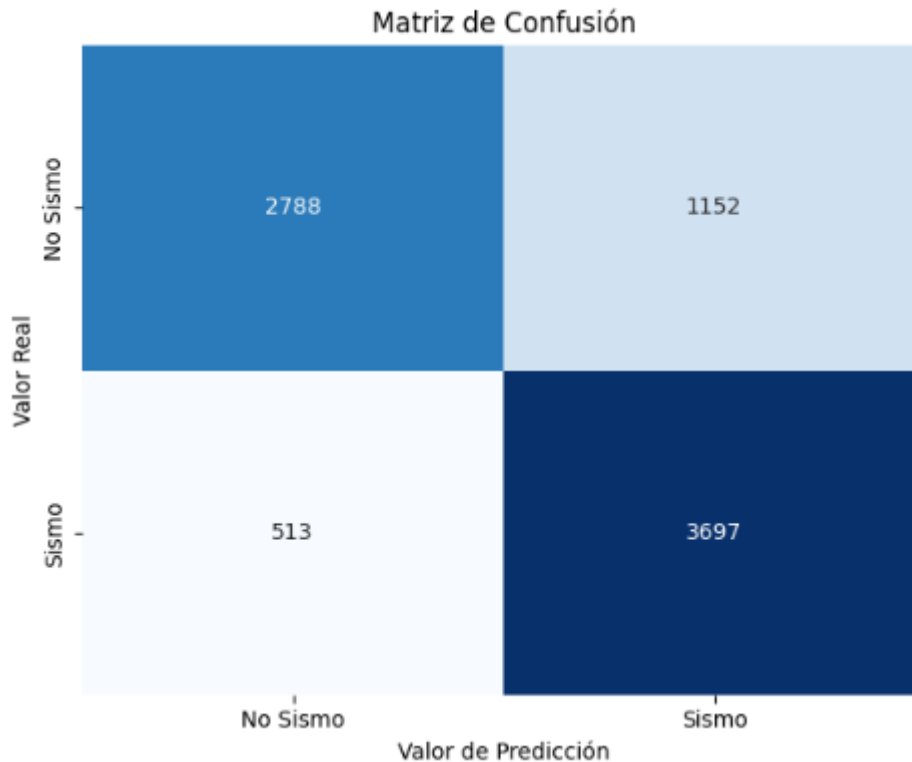


Figura 10. Matriz de confusión del modelo XGBoost

El análisis cuantitativo del modelo XGBoost revela un comportamiento más estable y discriminante en comparación con otros algoritmos. Según la matriz de confusión resultante, el desempeño se desglosa de la siguiente manera:

- Exactitud Global (Accuracy): El modelo logró una exactitud del 79.6% (6485 aciertos de 8150 casos). Este valor indica una alta capacidad general para clasificar correctamente tanto días sin sismo como días de actividad sísmica.
- Sensibilidad (Recall): El sistema alcanzó un 87.8% de sensibilidad, logrando detectar correctamente 3697 eventos sísmicos y omitiendo 513. Aunque su sensibilidad es

ligeramente inferior a la del Random Forest, sigue siendo lo suficientemente alta para un sistema de alerta, capturando la gran mayoría de los eventos sísmicos.

- Precisión de sismos: XGBoost generó 1152 falsas alarmas (Falsos Positivos), una cifra significativamente menor que los otros modelos. Con una precisión del 76.2%, este modelo demuestra ser más inteligente al detectar mejor la variabilidad climática no asociada a sismos, reduciendo el número de alertas innecesarias sin sacrificar drásticamente la seguridad.

3.2.2 Modelo 2: Random Forest

Random Forest, es un algoritmo de aprendizaje supervisado que combina múltiples árboles de decisión para obtener predicciones más robustas y precisas que un único árbol individual. Cada árbol en el “bosque” actúa como un estimador débil, y el conjunto (el bosque completo) converge hacia un predictor fuerte mediante la agregación de sus resultados. [20]

En el proceso de entrenamiento, se generan múltiples subconjuntos de los datos originales mediante muestreo con reemplazo (bootstrap). Cada árbol se entrena de forma independiente sobre uno de estos subconjuntos, utilizando además un subconjunto aleatorio de características en cada división (split) interna. Esta doble aleatorización reduce la correlación entre árboles y limita el sobreajuste, mejorando la generalización del modelo. [20]

Una vez entrenados, los árboles del bosque hacen predicciones de manera individual. Para problemas de clasificación, la predicción final se obtiene por votación mayoritaria entre los árboles, para regresión se promedian sus salidas. Este enfoque de promedio o voto colectivo

estabiliza el resultado final frente a variaciones individuales de cada árbol, logrando mejor rendimiento y menor varianza que modelos de árbol de decisión simples. [21]

- **Aprendizaje del modelo**

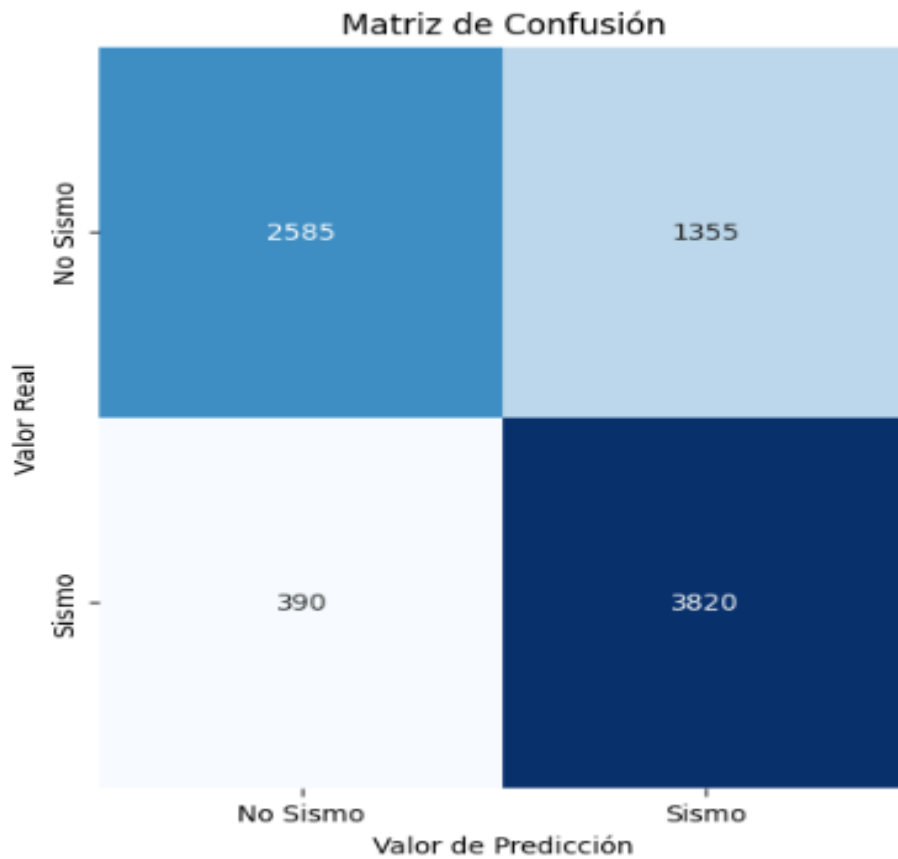


Figura 11. Matriz de confusión del modelo Random Forest

El comportamiento del modelo Random Forest refleja un aprendizaje orientado a la minimización de riesgos. Al analizar la distribución de sus predicciones, observamos que el algoritmo aprendió eficazmente a identificar las características climáticas complejas asociadas a los sismos, logrando capturar la gran mayoría de eventos positivos (3820 aciertos).

Sin embargo, su aprendizaje muestra un sesgo hacia la alta sensibilidad, el modelo ha aprendido que es preferible emitir una alerta en situaciones de duda (lo que genera 1355 falsos positivos) antes que ignorar un posible evento real. Esto indica que el modelo ha interiorizado

bien los patrones de peligro, aunque todavía tiene dificultades para discriminar la variabilidad climática en días donde fluctúan las variables climáticas, pero no hay sismo.

- **Evaluación de precisión del modelo**

Basándonos en la matriz de confusión resultante, el desempeño cuantitativo se resume en:

- Exactitud Global: El modelo acierta aproximadamente el 79% de las veces en total (6405 aciertos sobre 8150 casos).
- Sensibilidad (Recall): El modelo alcanza un 90.7% de sensibilidad (3820 / 4210). Esto significa que de cada 10 sismos reales, el sistema es capaz de detectar 9.
- Precisión de sismos: La confiabilidad de las alertas es del 73.8% (3820 / 5175). Esto implica que cuando el sistema predice un sismo, hay una probabilidad de 3 sobre 4 de que realmente ocurra.

3.2.3 Modelo 3: Multi Layer Perceptron (MLP)

El Multilayer Perceptron (MLP) es una arquitectura clásica de red neuronal artificial de tipo feedforward utilizada en aprendizaje supervisado para tareas de clasificación, regresión y reconocimiento de patrones. A diferencia del perceptrón simple, que tiene una sola capa de procesamiento, un MLP comprende múltiples capas de neuronas interconectadas, lo que le permite aprender relaciones no lineales complejas en los datos. [22]

La estructura de un MLP típicamente incluye: la capa de entrada, que recibe las características o variables del conjunto de datos, una o más capas ocultas, donde se realiza la transformación interna de la información mediante funciones de activación no lineales (como ReLU o sigmoide) y la capa de salida, que produce la predicción final del modelo. Cada neurona en una capa está conectada con pesos a todas las neuronas de la capa siguiente, formando una red completamente conectada (fully connected). [23]

Durante el entrenamiento, los pesos y sesgos de estas conexiones se ajustan a través de un algoritmo de retropropagación (backpropagation), que minimiza una función de pérdida comparando las predicciones con los valores reales. Esta combinación de conexiones densas y activaciones no lineales permite al MLP aproximar funciones arbitrariamente complejas cuando se diseña con suficientes neuronas y capas ocultas. [22]

- **Aprendizaje del modelo**

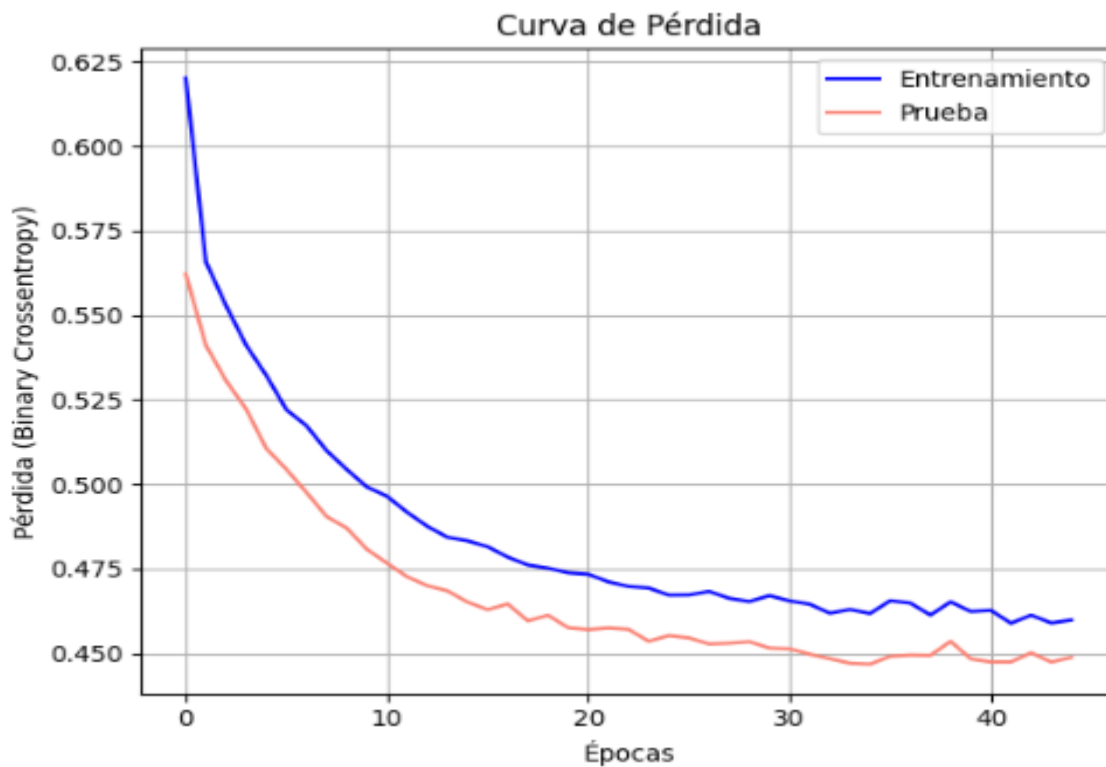


Figura 12. Gráfica de curva de aprendizaje del modelo Multi Layer Perceptron

El comportamiento de la Red Neuronal (MLP) durante el entrenamiento evidenció una capacidad progresiva para modelar las relaciones no lineales entre el clima y la sismicidad. A diferencia de los modelos de árboles que toman decisiones rápidas, la red mostró una curva de aprendizaje más gradual, logrando reducir el error consistentemente a medida que avanzaban las épocas.

Sin embargo, el aprendizaje de la MLP demostró ser más sensible a la escala de los datos y requirió mecanismos de regulación (como Dropout) para evitar la memorización excesiva. Esto indica que, si bien la arquitectura de Deep Learning logra capturar los patrones complejos del fenómeno, exige una mayor cantidad de datos y ajustes para estabilizarse en comparación con la robustez inmediata de algoritmos como XGBoost.

- **Evaluación de precisión del modelo**

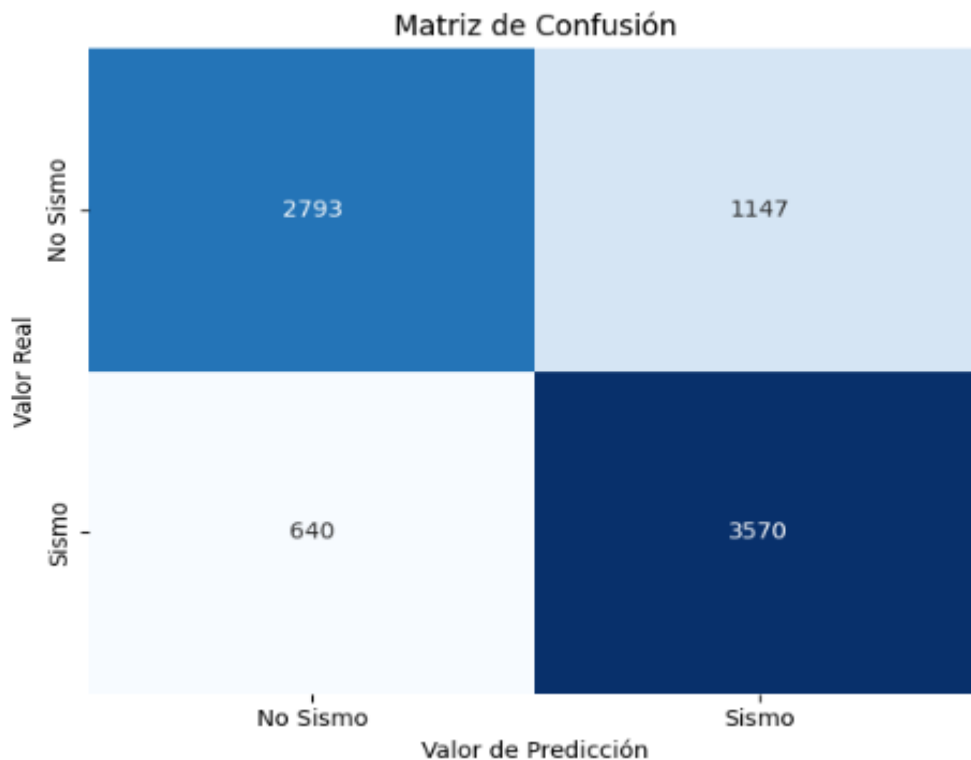


Figura 13. Matriz de confusión del modelo Multi Layer Perceptron

El desempeño del MLP demuestra que el Deep Learning es viable para este problema, aunque con resultados ligeramente inferiores a los modelos de árboles en la detección de patrones climáticos previos a los sismos:

- **Exactitud Global (Accuracy):** El modelo acertó en el 78.4% de los casos (6395 aciertos de 8150). Es un rendimiento muy competente, demostrando que la arquitectura neuronal converge correctamente.

- Sensibilidad (Recall): El modelo detectó 3664 sismos reales, lo que equivale a una sensibilidad del 87.0%. Aunque sigue siendo un número alto, es el más bajo de los tres modelos.
- Precisión de sismos: Generó 1209 falsos positivos. Generó menos ruido que el Random Forest, pero fue menos preciso que el XGBoost. Esto confirma que la red neuronal es robusta, pero para datos tabulares de este tipo, requiere mucho más ajuste para superar la eficiencia de los algoritmos basados en árboles.

3.2.4 Tabla comparativa de los resultados de los modelos

En la siguiente tabla se comparan los resultados obtenidos de cada modelo evaluado y sus respectivas observaciones:

Tabla 8. Tabla de los resultados de los modelos

Modelo	Exactitud Global	Precisión de Sismos	Recall	Observación
XGBoost	80%	76%	88%	Fue el modelo seleccionado por su mejor capacidad de generalización. Su arquitectura de Gradient Boosting con regularización penaliza la complejidad del modelo, permitiéndole ignorar la variabilidad climática no asociada a sismos mejor que Random Forest, logrando así una alta detección (88%) sin disparar las falsas alarmas.
Random Forest	79%	74%	91%	Alcanza un 91% de Recall debido a su capacidad para capturar relaciones no

				lineales complejas entre las variables climáticas. Esto le permite detectar sismos con señales débiles, aunque en consecuencia también genera el mayor número de falsos positivos.
MultiLay	78%	75%	87%	Muestra menor capacidad de generalización en este conjunto de datos tabulares. Esto indica que la red neuronal requiere un volumen de datos históricos mucho mayor para encontrar las relaciones no lineales con la precisión que logran los modelos basados en árboles (XGBoost y RF) al segmentar variables climáticas.
er				
Perceptro				
n				

Aunque el modelo Random Forest alcanzó el mayor valor de recall, evidenciando una alta capacidad para detectar sismos reales, este desempeño se obtuvo a costa de un incremento en las falsas alarmas, lo que implica un mayor consumo de tiempo y recursos para la verificación de eventos que al final no corresponden a actividad sísmica. En contraste, el modelo XGBoost presentó un desempeño más equilibrado, combinando una precisión global superior con un recall elevado, lo que permitió detectar la mayoría de los sismos sin generar un número excesivo de falsos positivos. Considerando que el sistema propuesto busca no solo maximizar la detección de eventos sísmicos, sino también optimizar la gestión operativa del sistema de alertas, XGBoost se identifica como el modelo más adecuado y balanceado para su implementación.

3.3 Evaluación de usabilidad de la página web

En esta sección se presentan los resultados obtenidos a partir de la evaluación del prototipo funcional de la página web QuakePredictEC, mediante la realización de pruebas de usuarios orientadas a analizar el nivel de comprensión, interpretación y uso adecuado de la información presentada por la plataforma.

La evaluación se llevó a cabo con una muestra de 20 participantes, conformada por usuarios con un rango de edad entre 18 y 45 años con el fin de obtener una percepción diversa y representativa sobre el uso de la aplicación. El perfil de los participantes incluyó estudiantes universitarios, profesionales jóvenes y trabajadores de distintos sectores, todos con experiencia básica e intermedia en el uso de las herramientas digitales. El objetivo principal de las pruebas fue identificar si los usuarios lograban interpretar correctamente los elementos visuales de la interfaz, especialmente el mapa interactivo y el módulo de alertas, sin confundir los niveles de riesgo mostrados ni asumir erróneamente la ocurrencia de eventos sísmicos reales.

Asimismo, se evaluó la capacidad de los participantes para navegar por la plataforma, comprender el carácter predictivo del modelo implementado y reconocer la existencia de la incertidumbre asociada a los resultados generados por el modelo.

3.3.1 Resultados de la evaluación del mapa interactivo

Para la evaluación del mapa interactivo, los participantes realizaron una exploración guiada de la interfaz y ejecutaron tareas específicas, como hacer clic en distintos puntos del mapa, interpretar los colores y describir el nivel de riesgo mostrado con el fin de analizar su comprensión del carácter predictivo del sistema.

Los resultados evidencian que aproximadamente el 80% de los participantes logró interpretar de manera adecuada la información presentada en el mapa interactivo. De los 20 usuarios evaluados, 17 identificaron correctamente que los puntos del mapa representan niveles

de riesgo estimado por el modelo y comprendieron que es necesario hacer clic sobre los puntos dentro del mapa para acceder a información más detallada.

Asimismo, 16 usuarios interpretaron correctamente los colores de los puntos y que representaba cada color según la leyenda que se encuentra en el mapa sin haber leído previamente la información desplegada. No obstante, 4 usuarios presentaron confusión inicial, principalmente al intentar descifrar el significado de los colores sin haber leído anteriormente la información más detallada. Finalmente, ante la pregunta sobre si un nivel de probabilidad alta de un sismo garantiza la ocurrencia del mismo, 18 usuarios respondieron correctamente que no se trata de un evento confirmado sino más bien de una predicción, mientras que 2 usuarios manifestaron dudas creyendo firmemente en que se daría lugar a la ocurrencia del sismo.

3.3.2 Resultados de la evaluación del módulo de alertas

Para la evaluación de la interfaz del módulo de alertas, los participantes interactuaron con una alerta simulada dentro de la página web QuakePredictEC, correspondiente a un evento con probabilidad mayor o igual al 70%. Se les solicitó interpretar el significado de la notificación, explicar si consideraban que el sismo estaba ocurriendo o si se trataba de una predicción del sistema y describir que acciones tomarían ante la alerta. Con el fin de analizar su comprensión del carácter predictivo de la interfaz, la identificación del umbral de activación y el reconocimiento de la incertidumbre asociada a los resultados mostrados.

Los resultados obtenidos a partir de la evaluación del módulo de alertas indican que cerca del 85% de los usuarios comprendieron correctamente que las alertas mostradas por la página web son parte de una predicción generada automáticamente por el modelo y no a un sismo real en curso. Durante las pruebas realizadas, 17 de los 20 participantes interpretaron

que la alerta representa una probabilidad alta y una ubicación aproximada de un sismo, reconociendo que la información mostrada no es exacta ni definitiva.

De igual forma, al analizar las respuestas verbales de los usuarios, se observó que 18 de los 20 participantes distinguieron correctamente entre un evento sísmico real y una predicción probabilística identificando explícitamente la existencia de un margen de incertidumbre asociado al resultado presentado. Asimismo, 17 estudiantes lograron relacionar la activación de la alerta con el umbral probabilístico mayor o igual al 70% evidenciando una comprensión adecuada del criterio de generación de alertas de la página web.

Por otro lado, en cuanto a las reacciones reportadas, 16 de 20 usuarios manifestaron conductas preventivas y racionales, como informarse mejor, revisar el mapa interactivo o mantenerse atentos a nuevas actualizaciones, sin evidenciar respuestas de pánico inmediato.

No obstante, 4 usuarios asociaron inicialmente la alerta como un peligro inmediato, aunque esta percepción tendió a corregirse tras una explicación adicional o una interacción más profunda al visualizar el mapa interactivo.

En conclusión, los resultados muestran que el módulo de alertas comunica de manera efectiva el carácter predictivo e incierto del sistema, permitiendo que la mayoría de los usuarios interprete correctamente la información y adopte decisiones preventivas adecuadas, con un bajo riesgo de generar pánico o interpretaciones alarmistas.

Capítulo 4

4. Conclusiones

A continuación, se presentan las principales conclusiones que se identificaron a partir del desarrollo de la página web QuakePredictEC. Estas se fundamentan en el análisis técnico del sistema, la integración de datos sísmicos y climáticos, el diseño de la arquitectura tecnológica y la validación realizada con usuarios. Además, reflejan el cumplimiento de los objetivos planteados, así como el aporte del proyecto en el ámbito del análisis y visualización del riesgo sísmico en Ecuador.

- A partir de uno de los objetivos planteados, el cual se basó en la creación de una plataforma web que analice datos sísmicos y climáticos de Ecuador, se logró entrenar un modelo capaz de realizar predicciones de sismos a partir del análisis de las variables climáticas mencionadas obteniendo un porcentaje de recall del 86%.
- También, se aplicaron técnicas estadísticas y de minería de datos que evidenciaron diferencias estadísticamente significativas en las distribuciones de las variables climáticas durante los días de ocurrencia sísmica. Sin embargo, el análisis de correlación demostró que la magnitud sísmica no presenta una dependencia lineal con estas variables, mas bien, está principalmente determinada por las condiciones geológicas del área de estudio, específicamente por la presencia de las fallas geológicas existentes en la ubicación analizada.
- Se usaron modelos de Aprendizaje Automático para el entrenamiento de variables sísmicas y climáticas, en el cual se observó que la inclusión de las variables climáticas mejora el reconocimiento de patrones y la predicción de la ocurrencia de sismo. Además, se eligió el modelo XGBoost al obtener un porcentaje de recall del 86% con respecto de los demás modelos evaluados.

- Se desarrolló una interfaz gráfica interactiva que permite la visualización clara de los resultados del modelo en tiempo real, facilitando la comprensión de la información y apoyando la interpretación correcta de las predicciones por parte de los usuarios.
- Se integró el modelo XGBoost con la plataforma web, dando como resultado un sistema funcional que permite el procesamiento, predicción y visualización de datos, facilitando el análisis de la actividad sísmica desde un enfoque combinado sísmico-climático.

5. Recomendaciones

- Se recomienda ampliar el conjunto de datos sísmicos utilizados en el entrenamiento del modelo, incorporando registros históricos de mayor cobertura temporal y espacial, con el fin de fortalecer la capacidad de generalización y mejorar la estabilidad de las predicciones obtenidas
- Se recomienda continuar incorporando variables climáticas como la presión atmosférica, temperatura y precipitación en el proceso del entrenamiento de los modelos de Aprendizaje Automático, dado que se evidenció que estas aportan el 40% del aprendizaje del modelo, mejorando la identificación de patrones asociados a la ocurrencia de eventos sísmicos
- Es conveniente desarrollar modelos diferenciados para la predicción de la ocurrencia de sismo y para el análisis de la magnitud sísmica. Esto se debe a que ambos fenómenos corresponden a dinámicas diferentes; mientras la ocurrencia de un sismo puede analizarse a partir de múltiples variables, como las variables climáticas ya mencionadas, la magnitud está principalmente asociada a los factores geológicos estructurales, como la presencia de fallas activas y la acumulación de energía tectónica. Dado que los resultados demostraron que no existe una relación directa entre las variables climáticas y la magnitud del evento, resulta estratégico abordarlos mediante enfoques independientes.

Referencias

- [1] Palma Aguiar, I. A., & Guillén Bernal, J. E. (2018). Revisión del estado del arte de los sistemas de monitoreo para la prevención de desastres naturales en el Ecuador (Bachelor's thesis, Quito: Universidad de las Américas, 2018).
- [2] Quinde Martínez, Pablo, & Reinoso Angulo, Eduardo. (2016). Estudio de peligro sísmico de Ecuador y propuesta de espectros de diseño para la Ciudad de Cuenca. Ingeniería sísmica, (94),1-26. Recuperado de http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0185-092X2016000100001&lng=es&tlng=es.
- [3] Bragato, P. L. (2021). Statistical relationship between the decrease of major seismicity and drought in southern California since 1900. *Frontiers in Earth Science*, 9, 790412.
- [4] Hajikhodaverdikhan, P., Nazari, M., Mohsenizadeh, M., Shamsirband, S., & Chau, K. W. (2018). Earthquake prediction with meteorological data by particle filter-based support vector regression. *Engineering Applications of Computational Fluid Mechanics*, 12(1), 679-688.
- [5] Ommi, S., & Hashemi, M. (2024). Machine learning technique in the north zagros earthquake prediction. *Applied Computing and Geosciences*, 22, 100163.
- [6] Li, Z., Meier, M. A., Hauksson, E., Zhan, Z., & Andrews, J. (2018). Machine learning seismic wave discrimination: Application to earthquake early warning. *Geophysical Research Letters*, 45(10), 4773-4779.
- [7] Moghadamnejad, A., Moghaddasi, M. A., Hamidia, M., Mohammadi, R. K., & Zare, M. (2026). Ranking Earthquake Prediction Algorithms: A Comprehensive Review of Machine Learning and Deep Learning Methods. *Soil Dynamics and Earthquake Engineering*, 200, 109740.

- [8] Beauval, C., Marinière, J., Laurendeau, A., Singaicho, J. C., Viracucha, C., Vallée, M., ... & Alvarado, A. (2017). Comparison of observed ground-motion attenuation for the 16 April 2016 M w 7.8 Ecuador megathrust earthquake and its two largest aftershocks with existing ground-motion prediction equations. *Seismological Research Letters*, 88(2A), 287-299.
- [9] P. Kruchten, "The 4+ 1 view model of architecture," *IEEE software* vol. 12, no. 6, pp. 42-50, 1995.
- [10] Instituto Geofísico – Escuela Politécnica Nacional, "Catálogos Sísmicos," *IG-EPN*, Disponible en: <https://www.igepon.edu.ec/catalogos-sismicos>
- [11] T. Perol, M. Gharbi, and M. Denolle, "Convolutional neural network for earthquake detection and location," *Science Advances*, vol. 4, no. 2, e1700578, 2018.
- [12] L. Trani, G. A. Pagani, J. P. P. Zanetti, C. Chapeland, and L. Evers, "DeepQuake—An application of CNN for seismo-acoustic event classification in The Netherlands," *Computers & Geosciences*, vol. 159, 104980, 2022.
- [13] S. H. R., "A Comparative Analysis on Decision Trees, Random Forest and XGBoost," Medium, 10-Oct-2020. [Online]. Available: medium.com/codepth/a-comparative-analysis-on-decision-trees-random-forest-and-xgboost-f74b8fb716c7
- [14] "Diagrama de caja," Wikipedia, The Free Encyclopedia. [Online]. Available: https://en.wikipedia.org/wiki/Box_plot
- [15] M. Rodríguez, "Interpretación de los gráficos de caja en el análisis descriptivo e inferencial," Qvision, 30-Mar-2015. [Online]. Available: <https://www.qvision.es/blogs/manuel-rodriguez/2015/03/30/interpretacion-de-los-graficos-de-caja-en-el-analisis-descriptivo-e-inferencial>

- [16] “Exploratory Data Analysis: Box Plot,” JMP. [Online]. Available:
<https://www.jmp.com/es/statistics-knowledge-portal/exploratory-data-analysis/box-plot>
- [17] “Prueba U de Mann-Whitney: Qué es y cómo funciona,” QuestionPro. [Online].
Available: <https://www.questionpro.com/blog/es/prueba-u-de-mann-whitney>
- [18] “¿Qué es el coeficiente de correlación de Pearson?” QuestionPro. [Online]. Available:
<https://www.questionpro.com/blog/es/coeficiente-de-correlacion-de-pearson>
- [19] “XGBoost,” Wikipedia, The Free Encyclopedia. [Online]. Available:
<https://www.ibm.com/es-es/think/topics/xgboost>
- [20] “Random forest,” IBM Think. [Online]. Available: <https://www.ibm.com/mx-es/think/topics/random-forest>
- [21] “Random Forest Algorithm in Machine Learning,” GeeksforGeeks. [Online]. Available:
<https://www.geeksforgeeks.org/machine-learning/random-forest-algorithm-in-machine-learning>
- [22] “Multilayer Perceptron,” Wikipedia, The Free Encyclopedia. [Online]. Available:
https://en.wikipedia.org/wiki/Multilayer_perceptron
- [23] “What is a Multilayer Perceptron (MLP)?” Milvus AI Quick Reference. [Online].
Available: <https://milvus.io/ai-quick-reference/what-is-a-multilayer-perceptron-mlp>

ANEXOS

- **Enlace al backend de QuakePredictEC:**

<https://quakepredictec-backend.onrender.com/docs>

- **Enlace a la aplicación web de QuakePredictEC:**

<https://alexisloor.github.io/quakepredictec/>

- **Enlace al notebook de colab con los modelos implementados:**

<https://colab.research.google.com/drive/1GSJixsFUG-kM8mDPwwLjPyxLMiaIIHk5?usp=sharing>

- **Enlace al código de backend y frontend:**

<https://github.com/alexisloor/quakepredictec.git>

https://github.com/alexisloor/quakepredictec_backend.git