

# Técnicas de predicción de terremotos usando machine learning: Una revisión sistemática de la literatura

## Earthquake prediction techniques using machine learning: A systematic review of the literature

Fermín Orlando Pinedo Delgado<sup>1,a</sup>, Nora Bertha La Serna Palomino<sup>1,b</sup>

<sup>1</sup> Universidad Nacional Mayor de Marcos, Facultad de Ingeniería de Sistemas e Informática. Lima, Perú

<sup>a</sup> Autor de correspondencia: [fermin.pinedo@unmsm.edu.pe](mailto:fermin.pinedo@unmsm.edu.pe), ORCID: <https://orcid.org/0009-0002-7492-0814>

<sup>b</sup> E-mail: [nlasernap@unmsm.edu.pe](mailto:nlasernap@unmsm.edu.pe), ORCID: <https://orcid.org/0000-0002-4292-344X>

### Resumen

El desarrollo tecnológico, ha despertado el interés por parte de la comunidad científica en realizar investigaciones para predecir terremotos empleando Inteligencia Artificial. Una de las líneas de la Inteligencia Artificial es Machine Learning, que, a través de algoritmos de regresión y clasificación han tratado de predecir terremotos empleando Big Data. Para pronosticar terremotos con modelos predictivos, muchos investigadores, han utilizado eventos sísmicos, técnicas de Machine Learning, herramientas de programación y metodologías de desarrollo. Según la búsqueda exhaustiva de la literatura, se encontraron que, la mayor cantidad de estudios emplean las características sísmicas magnitud, latitud, longitud y profundidad. Mientras que, las técnicas de Machine Learning más utilizadas son las Redes Neuronales Artificiales, Bosques Aleatorios, Maquina de Vectores de Soporte, Regresión Lineal Múltiple y Árbol de Decisiones. También se ha visto que, una de las herramientas más importantes para desarrollar modelos predictivos es, el Lenguaje de Programación Python. Y, para proyectos con grandes volúmenes de datos, la mayoría de los investigadores trabajan con las guías metodológicas CRISP-DM, KDD y SEMMA.

**Palabras clave:** Predicción sísmica, técnicas de machine learning, inteligencia artificial.

### Abstract

Technological development has awakened interest on the part of the scientific community in researching to predict earthquakes. The article's objective is to know what variables, techniques, tools, and methodologies have been used in the different studies to predict earthquakes using machine learning techniques. To carry out the study, the Kitchenham methodology was used, which consists of three development phases: review planning, conducting, and reporting. In the planning phase, four research questions were posed; for this purpose, an exhaustive literature search was carried out. After carrying out the selection and exclusion criteria, the questions posed were developed, of which it was found that 15% of the variables to predict earthquakes were latitude, longitude, and depth. In comparison, 13% were the seismic magnitude. 17% of the most used techniques were Random Forest, followed by Artificial Neural Networks with 17%. 65% used Python to develop algorithms, followed by MATLAB and R at 14%. 50% implemented the CRISP-DM methodology for data mining projects, followed by KDD with 33%.

**Keywords:** Seismic prediction, machine learning techniques, artificial intelligence.

Recibido: 02/02/2024 - Aceptado: 16/06/2024 - Publicado: 30/06/2024

#### Citar como:

Pinedo Delgado, F. & La Serna Palomino, N. (2024). Técnicas de predicción de terremotos usando machine learning: Una revisión sistemática de la literatura. *Revista Peruana de Computación y Sistemas*, 6(1):79-90. <https://doi.org/10.15381/rpcs.v6i1.28442>

© Los autores. Este artículo es publicado por la Revista Peruana de Computación y Sistemas de la Facultad de Ingeniería de Sistemas e Informática de la Universidad Nacional Mayor de San Marcos. Este es un artículo de acceso abierto, distribuido bajo los términos de la licencia Creative Commons Atribución 4.0 Internacional (CC BY 4.0) [<https://creativecommons.org/licenses/by/4.0/deed.es>] que permite el uso, distribución y reproducción en cualquier medio, siempre que la obra original sea debidamente citada de su fuente original.

## 1. Introducción

Los sismos, siempre han sido una amenaza constante para el hombre, puesto que pueden ocurrir en cualquier momento, y de acuerdo con la energía liberada pueden ocasionar grandes desastres naturales. Hoy en día, es el desastre natural de mayor riesgo que enfrenta la humanidad, por las graves consecuencias que pueden ocasionar, sobre todo en las ciudades más pobladas [1][2]. Debido al choque de placas tectónicas, liberan gran cantidad de energía interna originando fallas geológicas, tsunamis, volcanes, deslizamientos, entre otras catástrofes [3].

La mayoría de los eventos telúricos, se producen por donde atraviesan las placas tectónicas, principalmente en las zonas de subducción. En estas zonas se han producido grandes terremotos, como el de Valdivia en Chile con Magnitud de momento (Mw) de 9.5 grados, considerado hasta la fecha como el de mayor magnitud a nivel mundial [4]. Otros eventos devastadores que se han producido desde el año 1900 hasta el 2023 han tenido lugar en Alaska en el año 1964 con 9.2 Mw, Sumara-Andamán en 2004 con 9.1 Mw, Tohoku-Japón en 2011 con 9.0 Mw y Kamchatka-Rusia en 1952 con 9.0 Mw [5].

En la actualidad, gracias al avance tecnológico, ha despertado el interés por parte de muchos investigadores en realizar estudios para predecir terremotos empleando los avances innovadores en Inteligencia Artificial (IA), sobre todo en el campo de Machine Learning (ML) y Deep Learning (DL) [6]. En muchos estudios han empleado diferentes algoritmos de ML para predecir terremotos, entre ellos se encuentran los siguientes: Artificial Neural Network (ANN), Random Forest (RF), Decision Tree (DT), Support Vector Machine (SVM), K-Nearest Neighbour (K-NN), Long Short-Term Memory (LSTM), Multiple Linear Regression (MLR), Gradient Boost (GB), eXtreme Gradient Boosting (XGBoost), Deep Neural Network (DNN), Convolutional Neural Network (CNN), etc., para la predicción de los algoritmos, se han empleado grandes volúmenes de datos históricos extraídos de diferentes catálogos sísmicos, y para desarrollar los modelos predictivos han utilizado lenguajes de programación y/o software estadísticos. Una vez que los modelos han sido entrenados, son evaluados con las métricas de precisión Mean Square Error (MSE), Root Mean Square Error (RMSE), Mean Absolute Error (MAE).

Un avance muy importante de la IA es, la Alerta Temprana de Terremotos, por sus siglas en inglés Earthquake Early Warning (EEW), son sensores sísmicos que detectan en tiempo real las primeras ondas de un terremoto, estas a su vez son recepcionadas en los centros de procesamiento de información más cercanos [7]. En algunas ciudades ya existen estos tipos de sistemas y son alertados a los residentes antes de producirse un temblor [8]. Las técnicas de DL se adoptan con estos sistemas inteligentes y pueden realizar predicciones antes de producirse un desastre, uno de los algoritmos más empleados es ANN, lo cual puede entrenar en tiempo real datos de ondas sísmicas llegadas a través de sensores

en milésimas de segundos [9]. Estos sistemas, previenen principalmente las pérdidas humanas, lesiones, ansiedad emocional y pérdidas económicas [10]. Las soluciones de alerta temprana, se ha convertido en un avance tecnológico emergente y eficaz para prevenir desastres naturales. Actualmente son 9 países los que ya cuentan con este tipo de sistemas, y otros 13 están en etapa de implementación y pruebas [11].

Los sismos, han causado grandes pérdidas económicas a la sociedad. Entre los años 2000 y 2019, han ocasionado alrededor de 721,000 muertes y 636,000 millones de dólares en pérdidas económicas a nivel mundial. Existen muchos estudios para predecir pérdidas económicas ocasionadas por desastres naturales [12]. En [13], realizaron modelos empleando técnicas de ML para predecir pérdidas financieras por causa de terremotos, para el entrenamiento emplearon los algoritmos RF y DT, los datos fueron extraídos en el periodo de tiempo desde 1990 hasta el 2020. Otro trabajo de [14], estimaron pérdidas a causa de terremotos basadas en múltiples fuentes para diversas ciudades de China. El estudio se realizó desde el año 2007 hasta el 2021, los modelos se entrenaron con los algoritmos MLR, SVM y XGBoost.

Mientras que el mundo sigue siendo contaminado por la misma sociedad generando grandes gases de efecto invernadero dando origen al calentamiento global, muchos investigadores están buscando la manera de desarrollar métodos eficaces para predecir desastres naturales. Los sismólogos, de la mano con la tecnología, han desarrollado modelos predictivos empleando IA, también han extraído de catálogos sísmicos grandes volúmenes de datos, después de analizar y realizar la limpieza de la información, han seleccionado eventos sísmicos, técnicas de ML, herramientas y metodologías más importantes para desarrollar modelos predictivos. A pesar de los grandes esfuerzos realizados, la comunidad científica, todavía no se ha podido predecir con exactitud el lugar y fecha de la ocurrencia de un futuro terremoto.

El presente estudio tiene como objetivo, investigar las técnicas de ML empleadas en la predicción de terremotos. Para ello, se plantearon cuatro preguntas de investigación que permitieron analizar y discutir los trabajos relacionados con el objeto de estudio.

Este paper está organizado de la siguiente manera: sección 2 describe la metodología empleada en la investigación, sección 3 se analizan los resultados obtenidos, y en la sección 4 se realizan las conclusiones.

## 2. Metodología

El desarrollo de la revisión sistemática se realizó con la metodología propuesta por [15], que consiste en tres fases:

- Planificación de la revisión: se identifica la necesidad de la revisión y se elabora un protocolo de la búsqueda sistemática de la información.

- **Conducción de la revisión:** en esta fase, se realiza la identificación de la revisión, selección de los estudios primarios, evaluación de la calidad de los estudios, y aplicación de los criterios de inclusión y exclusión.
- **Reporte de la revisión:** se analizan los resultados estadísticos obtenidos.

### 2.1. Planificación de la Revisión

Se plantearon 4 preguntas de investigación, con la finalidad de determinar las técnicas de ML empleadas para la predicción de terremotos.

P1: ¿Qué variables sísmicas han sido empleadas para predecir terremotos?

P2: ¿Qué técnicas de ML han sido utilizadas para predecir terremotos?

P3: ¿Cuáles fueron las herramientas empleadas para entrenar algoritmos de ML?

P4: ¿Qué metodologías se ha utilizado para realizar proyectos de IA empleando minería de datos?

Para responder a las preguntas planteadas, se realizó la búsqueda sistemática de la información en las bases de datos Scie Direct, IEEE Explore, Directory of Open Access Journals (DOAJ), Springer, Scielo, Revista Peruana de Computación y Sistemas (RPCS), Instituto Geofísico del Perú (IGP).

Se consideraron los términos de búsqueda (“earthquake prediction”, “machine learning”, “artificial intelligence”, “data mining methodologies”) entre los años 2020 y 2024. Una vez que se encontraron las investigaciones relacionadas con el estudio, se establecieron criterios de selección y exclusión tal como se muestra en la tabla 1.

**Tabla 1**

#### Criterios de selección y exclusión

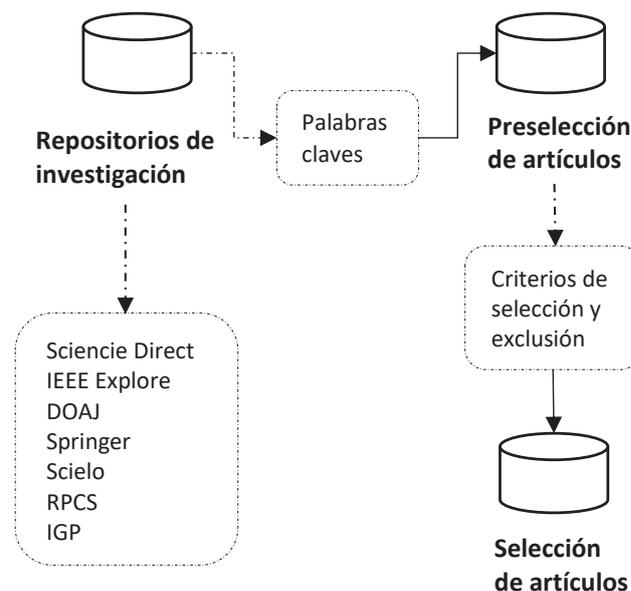
Inclusión	Exclusión
Los trabajos de investigación deben de estar dentro del rango de años desde el 2020 hasta el 2024	Los trabajos que no se encuentran en el rango de fecha establecida
Artículos que contengan variables, técnicas, herramientas y metodologías para predecir terremotos	Artículos que no tienen relación con las preguntas de investigación
Estar publicado en alguna revista de investigación	Artículos en proceso de revisión

### 2.2. Conducción de la Revisión

Primeramente, se realizó la búsqueda exhaustiva de la literatura en diferentes repositorios de revistas de investigación. Los trabajos fueron seleccionados de acuerdo con los criterios establecidos. El proceso de la selección de artículos se presenta en la figura 1.

**Figura 1**

#### Proceso de revisión de artículos científicos



### 2.3. Reporte de la Revisión

En la tabla 2, se muestran los 126 artículos relacionados con el objeto de estudio, de los cuales, aplicando los criterios de selección y exclusión, se preseleccionaron 88 trabajos y se seleccionaron 50. La mayor cantidad de estudios fueron seleccionados del repositorio Ciencia Direct, seguido de IEEE Explore, Scielo y RPCS.

**Tabla 2**

#### Artículos encontrados

Fuente	Artículos encontrados	Artículos preseleccionados	Artículos seleccionados
Ciencia Direct	90	72	44
IEEE Explore	20	8	3
DOAJ	5	2	
Springer	3	1	
Scielo	3	2	2
RPCS	3	2	1
IGP	2	1	

### 2.4. Análisis

En esta sección, se dará respuesta a las preguntas planteadas, para ello, se analizarán los diferentes artículos de investigación relacionados con el objeto de estudio.

#### 2.4.1. ¿Qué variables sísmicas han sido empleadas para predecir terremotos?

Las variables sísmicas, son extraídas de catálogos históricos en un determinado periodo de tiempo, una

vez que se realiza la extracción, limpieza y depuración de los datos, se seleccionan las principales características sísmicas para el estudio.

En [17], se extrajeron 21 características sísmicas del National Seismic Information Center (NEIC) entre los años 1965 hasta el 2016, de las cuales, se eligió las variables latitud, longitud, magnitud y profundidad. En la investigación de [18], se recopiló datos de los catálogos US Geological Survey (USGS) y del International Institute of Earthquake Engineering and Seismicity entre los años 1973 y 2019, después de realizar la limpieza de la información, se consideró como variables de entrada, latitud, longitud y profundidad, y como variable respuesta, magnitud sísmica. En los trabajos de [19] y [20] se compiló datos de la USGS para predecir terremotos en Japón y Bangladesh, para el estudio se empleó los eventos sísmicos: magnitud, latitud, longitud, profundidad, año, mes, día, hora, minuto y segundo.

Los sismos se producen debido al choque de las placas tectónicas, lo cual, al liberar grandes cantidades de energía generan fuertes terremotos y dan origen a la formación de fallas geológicas. Estos fenómenos han servido como estudio de muchos investigadores en el campo de la sismología. En el trabajo de [21], se recopiló diferentes eventos sísmicos entre las placas tectónicas de euroasiática y africana ubicadas en Chipre, cuyas características sísmicas fueron: longitud, latitud, año, mes, día, magnitud, profundidad, hora, minuto y segundo. En otro paper de [22], se efectuó estudios en la falla de Antolía del Norte en Turquía, la base de datos lo obtuvieron de la Presidencia de Gestión de Emergencias y Desastres del Ministerio del Interior de Turquía, las variables seleccionadas para entrenar modelos predictivos fueron: fecha, hora, latitud, longitud, profundidad y ubicación. En [23], se llevó a cabo investigaciones en el cinturón plisado de Zagros, es una zona altamente sísmica que abarca parte de los países de Irak e Irán, y se encuentra ubicado entre las placas tectónicas de arábica y euroasiática, para el estudio, se recurrió a los atributos sísmicos: ubicación, tiempo, latitud, longitud y magnitud.

La información que se obtiene luego de haberse producido un terremoto, son enviados por los diferentes sensores a las estaciones más cercanas para su posterior estudio y análisis de la información. En [24] se indagó 700 estaciones conectadas a sensores de avanzada tecnología en Japón, de las cuales, se obtuvo eventos sísmicos sobre la magnitud del terremoto, altitud de la estación, profundidad, latitud y longitud. Otro estudio de [25], se analizó 256 terremotos registrados en 39 estaciones en Italia, de donde se examinó las características sísmicas: magnitud, profundidad y epicentro.

En la tabla 3, se identificó a 14 eventos sísmicos de distintos trabajos de investigación. Las características sísmicas que más emplearon en los diferentes estudios fueron: latitud, longitud, profundidad y magnitud.

#### 2.4.2. ¿Qué técnicas de ML han sido utilizadas para predecir terremotos?

Las técnicas de ML, se ha convertido en una herramienta muy importante para la predicción de terremotos debido a su gran capacidad para predecir algoritmos con grandes volúmenes de datos sísmicos.

Existen muchos trabajos de investigadores que han empleado técnicas de ML para predecir terremotos. En [7], se predijo futuros terremotos en Turquía con magnitudes mayores a 4 Mw entre los años 1900 y 2019, para el entrenamiento de los modelos, se empleó los algoritmos RF, MRL y LSTM. En Malasia [27], para predecir terremotos se ejecutó los modelos predictivos ANN y RF. En otro trabajo de [27], se hizo investigaciones en Japón, con la finalidad de predecir terremotos con magnitudes mayores a 4 Mw durante los años 1997 hasta el 2019, para el estudio se empleó los algoritmos XGBoost, RF y DNN, mientras que en [28] se predijo en el mismo país 115,501 registros sísmicos con los algoritmos RF, GB y ANN. El grupo de [29], realizó estudios para prever el movimiento del suelo con las técnicas ANN, RF y SVM, para la predicción de los modelos, se consideró 4,528 registros de 374 movimientos telúricos registrados en 209 estaciones sísmicas de Texas, Oklahoma y Kansas.

Las predicciones sísmicas se pueden dar a corto y largo plazo. En [21], se vaticinó terremotos a corto plazo en la región de Chipre con modelos predictivos ANN, RF y SVM, se entrenó algoritmos para 5, 7, 10 y 15 días respectivamente. Mientras que en [30], se pronosticó terremotos a largo plazo con algoritmos LSTM, para el estudio, se empleó datos del Centro de la Red de Terremotos de China desde el 2012 hasta el 2022.

**Tabla 3**  
*Variables sísmicas empleadas para predecir terremotos*

Variables sísmicas	Fuentes
Latitud	[17][18][19][20][21][22][23][24]
Longitud	[17][18][19][20][21][22][23][24]
Profundidad	[17][18][19][20][21][22][24][25]
Magnitud	[17][19][20][21][23][24][25]
Hora	[19][20][21][22]
Minuto	[19][20][21]
Segundo	[19][20][21]
Año	[19][20][21]
Día	[19][20][21]
Mes	[19][20][21]
Ubicación	[22][23]
Fecha	[22]
Epicentro	[25]
Tiempo	[23]

Las placas tectónicas Sudamericana y Nazca convergen en la llamada fosa Perú-Chile, en estos países se han registrado grandes terremotos. En el estudio de [31], se previno terremotos en la cuenca de Santiago de Chile, para el estudio se empleó mapas de zonificación sísmica y los algoritmos MLR, RF, ANN y DT, los

datos se obtuvieron de terremotos producidos entre los años 1985 y 2010. En [32], se entrenó algoritmos para predecir terremotos en Chile con las técnicas RF, DT y K-NN. Mientras que en Perú [33], se empleó el algoritmo RF para predecir riesgos sísmicos en Pisco.

Producto de la subducción entre placas tectónicas, han dado origen a múltiples fallas terrestres [34], estas a su vez, han servido para que muchos investigadores realicen estudios sísmicos en esas zonas de fractura. En [35], se procedió a realizar estudios empleando aprendizaje profundo para predecir terremotos de laboratorio en la zona de falla, para realizar los modelos predictivos se utilizó los algoritmos LSTM y CNN. En el estudio realizado por [36], se predijo anomalías sísmicas en la falla de Antolia del Norte en Turquía, para realizar el estudio, se usó las técnicas de ML LSTM y SVM, siendo el modelo LSTM el que obtuvo la mejor precisión en los resultados.

Después que se produce un terremoto, según la magnitud sísmica, pueden ocasionar grandes destrucciones en las edificaciones. En [37], se auguró deterioros de viviendas en Nepal después de haber ocurrido un sismo de magnitud 7.8 Mw, para la predicción se utilizó los algoritmos RF, DT, XGBoost y ANN. En [38], se aplicó las técnicas ANN, RF, SVM y DT para predecir el comportamiento de las estructuras frente a terremotos.

En la tabla 4, se puede examinar a las técnicas de ML identificadas en diversos artículos para predecir

**Tabla 4**

*Técnicas de Machine Learning empleadas para predecir terremotos*

Técnicas de ML	Fuentes
RF	[7][27][28][29][21][31][32][33][37][38]
ANN	[28][29][21][31][37][38]
SVM	[29][21][36][38]
LSTM	[7][30][35][36]
DT	[31][32][37][38]
MRL	[7][31]
XGBoost	[27][37]
DNN	[27]
GB	[28]
K-NN	[32]
CNN	[35]

terremotos, de las cuales, se puede observar que, RF fue la más utilizada, seguido por ANN, SVM, LSTM y DT.

#### 2.4.3. ¿Cuáles fueron las herramientas empleadas para entrenar algoritmos de ML?

Para entrenar los algoritmos de ML, se emplean diferentes herramientas de desarrollo, entre las más comunes se encuentran, los diferentes lenguajes de programación y software estadísticos.

El lenguaje de programación Python, contiene librerías de IA para predecir modelos predictivos con algoritmos de ML. En [19] y [20], se utilizó Python y las librerías Sklearn, Keras, Tensorflow y Pandas en Jupyter Notebook como herramienta de desarrollo para la implementación del modelo predictivo LSTM, con la finalidad de predecir terremotos a largo plazo. En los trabajos de [7] y [39], se aplicó este programa para predecir terremotos empleando modelos de ML. En [31], se programó modelos con Python para predecir terremotos en Santiago de Chile. En el estudio de [25], se usó la biblioteca Keras Python para predecir sismos empleando ANN. Gracias a sus librerías de IA, en [17] y [38], se hizo uso de este mismo software para realizar predicciones de terremotos en ciudades inteligentes, así como también realizar sistemas de alerta temprana. Algunos investigadores [22], han empleado este lenguaje de programación para desarrollar algoritmos predictivos para investigar anomalías antes que se produzca un terremoto.

Otra herramienta de desarrollo empleada para predecir terremotos es MATLAB, este lenguaje de programación, facilita la elaboración de modelos predictivos gracias a sus funciones matemáticas con IA. En la región mediterránea de Chipre, un grupo de investigadores [21], previó sismos de baja magnitud a corto plazo, los modelos fueron entrenados con MATLAB y el código fuente fue puesto en línea para su posterior uso científico. En otro estudio de [40], se codificó modelos predictivos con este mismo lenguaje para predecir terremotos en tiempo real basada en IA.

El software estadístico SPSS y el lenguaje de programación R, son muy empleados por muchos investigadores para predecir terremotos, sobre todo, por sus librerías de IA, gráficos y funciones estadísticas. En la tesis de [41], se encontró patrones frecuentes de datos con su aplicación en la predicción de terremotos, los registros sísmicos fueron entrenados con técnicas de ML, cuyo desarrollo, se realizó con R Studio, se obtuvo resultados óptimos para terremotos con magnitud mayor a 5 Mw. En [8], se analizó el comportamiento humano ante los sistemas de alerta temprana, se tomaron en cuenta 4,977 residentes de una determinada área en China, para el entrenamiento de los algoritmos se desarrolló con R Studio y con el software estadístico SPSS, se obtuvo resultados significativos, con índice de acierto del 95%.

En la tabla 5, se puede apreciar que, la herramienta más empleada para desarrollar algoritmos de ML es

**Tabla 5**

*Herramientas para entrenar algoritmos de ML*

Herramientas	Fuentes
Python	[19][20][7][39][31][25][17][38][22]
MATLAB	[21][40]
R	[41][8]
SPSS	[8]

el lenguaje de programación Python, seguido por MATLAB, R y SPSS.

**2.4.4. ¿Qué metodologías se ha utilizado para realizar proyectos de IA empleando minería de datos?**

Existen varias guías metodológicas para desarrollar proyectos de IA empleando grandes volúmenes de datos. Entre las más utilizadas se encuentran las siguientes: Cross Industry Standard Process for Data Mining (CRISP-DM), Knowledge Discovery in Databases (KDD) y Sample, Explore, Modify, Model, Assess (SEMMA).

**Metodología CRISP-DM**

Es una metodología que consiste en 6 fases de desarrollo, cuyos procesos se puede apreciar en la figura 2. La primera fase, consiste en la comprensión del negocio, en esta etapa se debe de entender los objetivos y requerimientos de la empresa, así como, los problemas que se desean resolver. La segunda fase, consiste en la comprensión de los datos, para lo cual, se realiza la limpieza, selección y depuración de atributos. En la tercera fase, se realiza la preparación de los datos, tanto para entrenamiento y para pruebas. En la cuarta fase, se desarrolla el modelado, se selecciona las técnicas y herramientas apropiadas para entrenar los modelos predictivos. En la quinta fase, se realiza la evaluación de

los modelos, para ello, se emplean métricas de precisión para medir los resultados. En la sexta fase se realiza la implementación del proyecto en el negocio.

Muchos trabajos de investigación que utilizan grandes volúmenes de datos emplean esta metodología para predecir modelos empleando técnicas de ML. En la tesis de maestría de [41], empleó la metodología CRISP-DM para predecir terremotos, los datos fueron extraídos del catálogo sísmico Advanced National Seismic System (ANSS) en el periodo de 10 años, desde el 2000 hasta el 2009. En [42], se utilizó esta metodología, para desarrollar un sistema basado en contenido para jueces utilizando procesamiento de lenguaje natural y aprendizaje profundo. También es muy aplicada en el sector industrial, en la investigación de [43], se implementó esta guía metodológica para la obtención de un modelo de segmentación geográfica sobre una base de datos pública en México. En otro trabajo de [44], se aplicó para analizar minería de datos en la industria de servicios financieros, mientras que en [45], se evaluó los riesgos de la industria Fintech basado en el marco de la guía CRISP-DM.

**Metodología KDD**

Es una metodología centrada en el usuario, es iterativa, puede contener bucles entre dos pasos cualquiera. En la figura 3, se puede observar los pasos que componen el proceso KDD.

**Figura 2**  
Procesos de la metodología CRISP-DM [43]

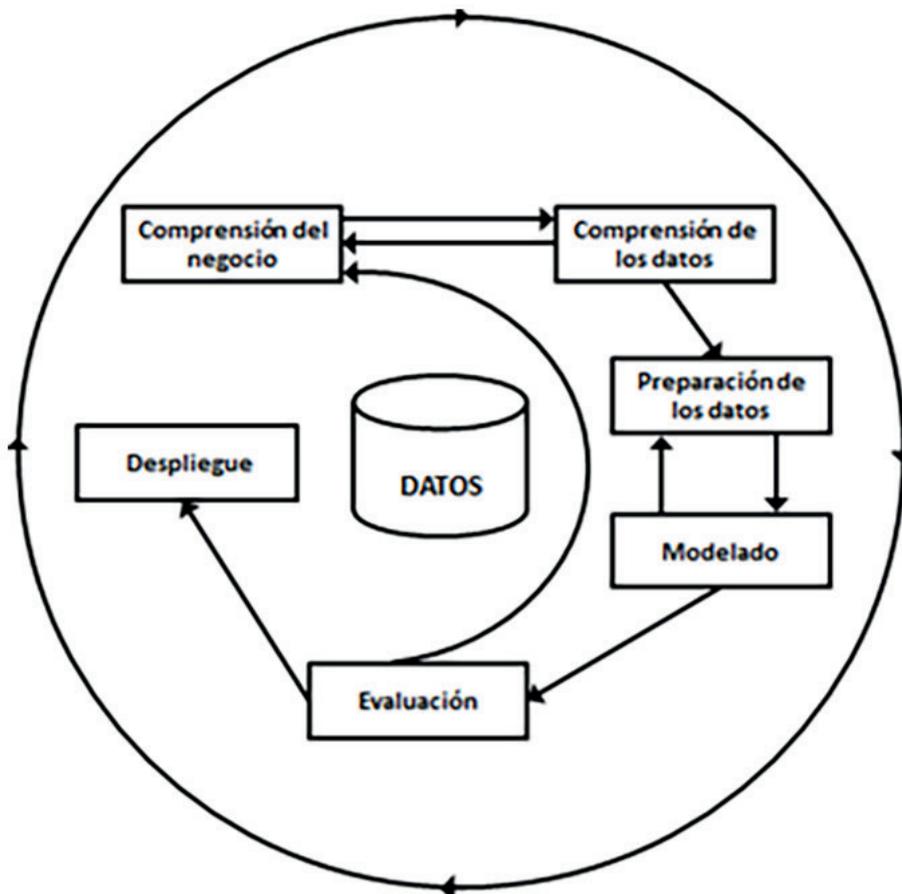
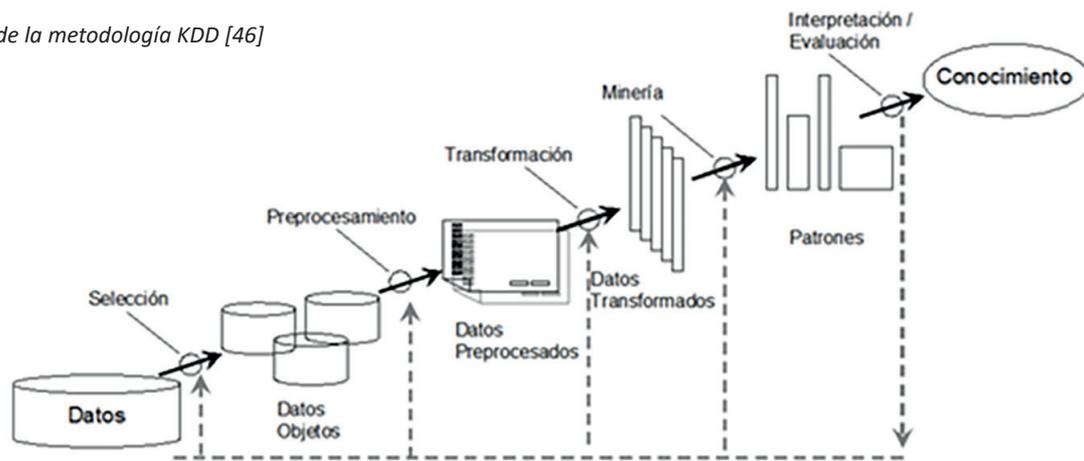


Figura 3

Procesos de la metodología KDD [46]



Primero, se realiza la selección del conjunto de datos, se seleccionan las variables de entradas y salidas del modelo. En el segundo paso, se realiza el análisis de los datos, se realiza la limpieza, y depuración de valores atípicos. En el tercer paso, se realiza la transformación de los datos, quedando listos para ser entrenados por algoritmos de ML. En el cuarto paso, se aplican las técnicas de ML, se desarrollan los modelos predictivos de regresión o clasificación. En el quinto paso, se realiza la extracción de conocimiento mediante la técnica empleada, se obtienen los resultados y se analizan si los modelos son válidos o necesitan aprender más y ser nuevamente entrenados. En el sexto paso, se interpreta y evalúa los resultados de los modelos, se debe de evaluar la precisión con métricas de medición y se busca el modelo que mejor se ajusta al problema planteado. Si ningún modelo alcanza los resultados esperados, se deben de alterar los pasos para tener nuevos resultados.

La metodología KDD, es empleada en varios campos de la ciencia, por ejemplo, en [46] se implementó biomarcadores en pacientes con datos de ensayos clínicos, para el desarrollo del sistema predictivo, se elaboró una base de datos inductiva con el algoritmo DT. En [47] se implementó la metodología KDD con registros de una gran base de datos con indicadores ambientales, económicos y sociales, cuyos resultados arrojaron 86%

de correlación entre todos los sectores. Mientras que en [48], se empleó la guía metodológica KDD con la finalidad de realizar un estudio de características 3D basado en la densidad del kernel para nubes de puntos, para su análisis de datos, se realizó medidas topológicas en histogramas y se codificó imágenes en 2D y 3D, los resultados fueron favorables tanto para datos públicos como para las nubes de puntos de fragmentos logrando un rendimiento robusto.

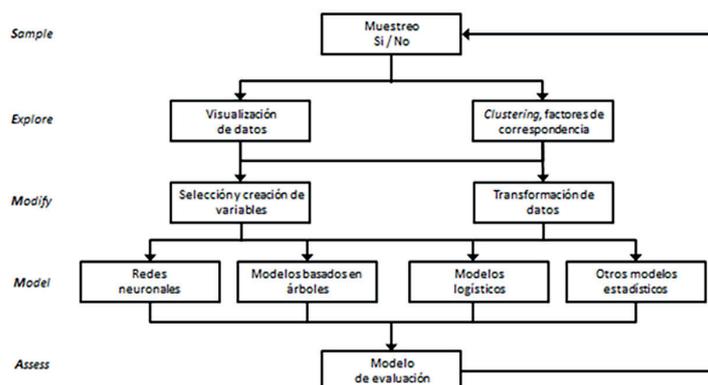
#### Metodología SEMMA

Esta metodología consta de cinco fases: Sample (Muestreo), Explore (Exploración), Modify (Modificación), Model (Modelado) y Assess (Valoración), estos procesos se muestran en la figura 4.

En la fase 1, se realiza la extracción de datos, se selecciona la muestra con la que se va a trabajar, la cual debe ser representativa de la población. Fase 2, se realiza el análisis de los datos, se hace la selección de las variables y limpieza de la información. Fase 3, se modifican los datos y se hace la preparación para ser entrenados. Fase 4, se modela el conjunto de datos, se debe de seleccionar las mejores herramientas y técnicas de ML para tener los resultados esperados. Fase 5, consiste en evaluar los resultados del modelo mediante métricas de precisión y seleccionar el modelo que obtuvo los mejores resultados para su posterior implementación.

Figura 4

Procesos de la metodología SEMMA [50]



Esta metodología es empleada para diferentes campos que utilicen minería de datos, por ejemplo, en [49] se utilizó la metodología SEMMA para analizar el sentimiento de la población hispanohablante en tiempo de COVID, para el estudio se usó grandes volúmenes de datos. Otra investigación de [50], se analizó grandes volúmenes de datos en el campo de las telecomunicaciones, gobernanza y arquitectura, para el estudio, se implementó las metodologías SEMMA, CRISP-DM y KDD.

En la tabla 6, se puede observar las diferentes metodologías que fueron implementadas en diferentes proyectos con Big Data. Siendo la metodología CRISP-DM la que más se empleó, seguido por KDD y SEMMA.

### 3. Resultados

En la figura 5, se muestran en total 126 trabajos encontrados, de los cuales, 90 fueron del repositorio Ciencia Direct, 20 de IEEE Explore, 5 de DOAJ, 3 de Springer, Scielo y RPCS, y 2 de IGP. De los trabajos seleccionados, se localizaron en total 88, de donde, 72 fue de Ciencia Direct, 8 de IEEE Explore, 2 de DOAJ, Scielo y RPC, y 1 de Springer e IGP. Finalmente, se eligió en total 50 artículos, 44 de Ciencia Direct, 3 de IEEE Explore, 2 de Scielo y 1 de RPCS.

En la figura 6, se aprecia los 50 trabajos seleccionados, de donde, el 88% fue del repositorio Ciencia Direct, 6% de IEEE Explore, 4% de Scielo y 2% de RPCS.

En la figura 7, se puede observar a las 14 variables encontradas en diversos estudios para predecir terremotos empleando técnicas de ML. De las cuales, las variables más utilizadas fueron: latitud, longitud y profundidad con el 15%, seguido de la magnitud con el 13% respectivamente.

En la figura 8, se pueden identificar las 11 técnicas de ML localizadas en diferentes trabajos de investigación. Las más utilizadas para prever terremotos fueron: RF con el 28%, ANN con el 17%, SVM, LSTM y DT con el 11%.

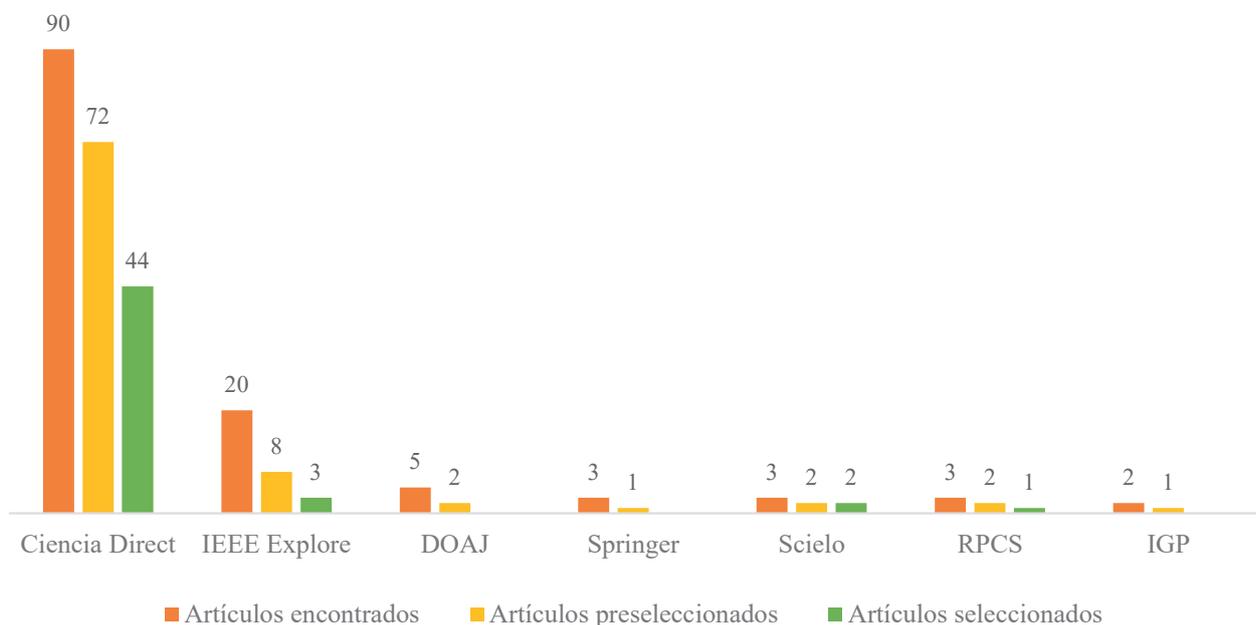
En la figura 9, se perciben las 4 herramientas más empleadas para desarrollar algoritmos de ML. El 65% de los trabajos ubicados usan Python, seguido por MATLAB y R con el 14% respectivamente.

En la figura 10, se presentan las metodologías que fueron implementadas en proyectos con minería de datos. El 50% de los artículos empleó CRISP-DM, el 33% KDD y el 17% SEMMA.

**Tabla 6**  
*Metodologías para realizar proyectos empleando minería de datos*

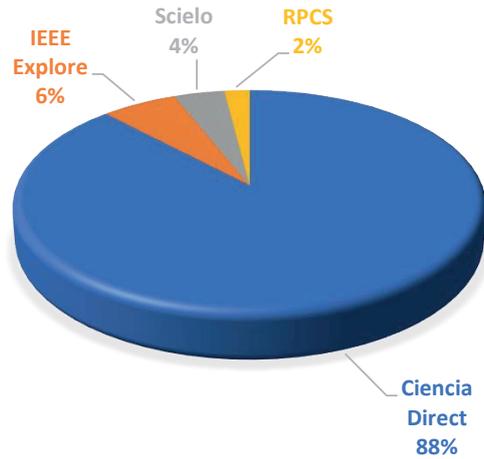
Metodologías	Fuentes
CRISP-DM	[41][42][43][44][45][50]
KDD	[46][47][48][50]
SEMMA	[49][50]
SPSS	[8]

**Figura 5**  
*Total de artículos encontrados, preseleccionados y seleccionados*



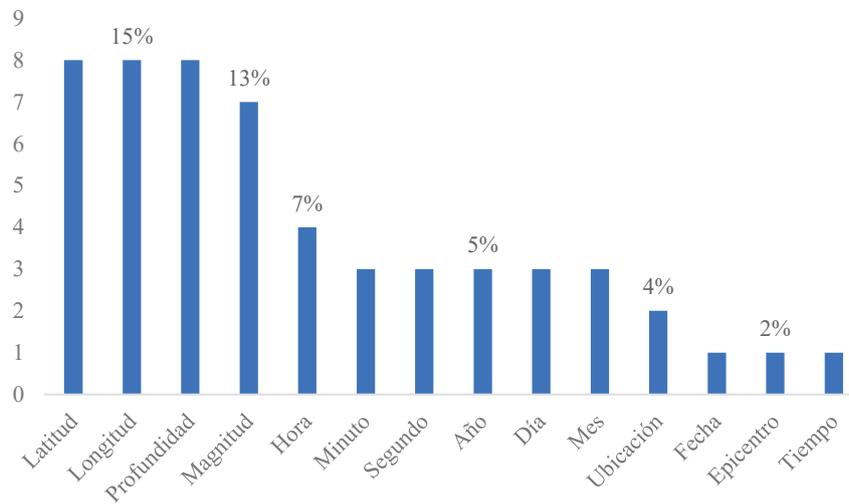
**Figura 6**

*Distribución de los artículos seleccionados*



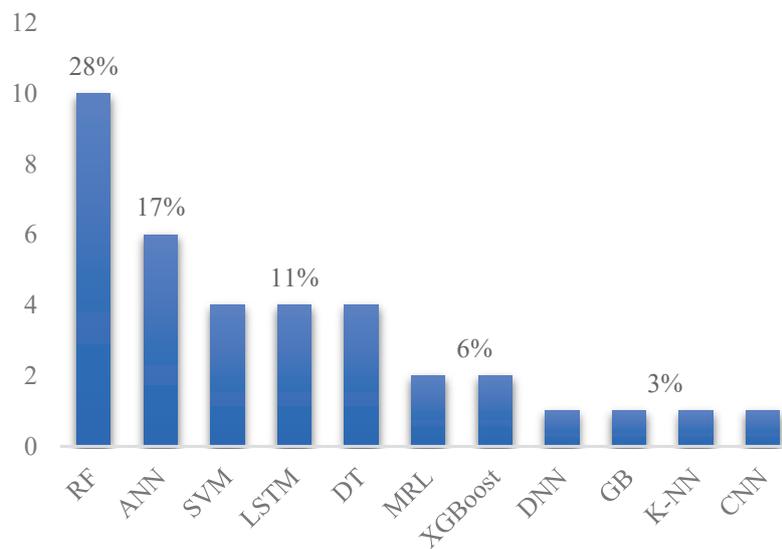
**Figura 7**

*Variables empleadas para predecir terremotos*

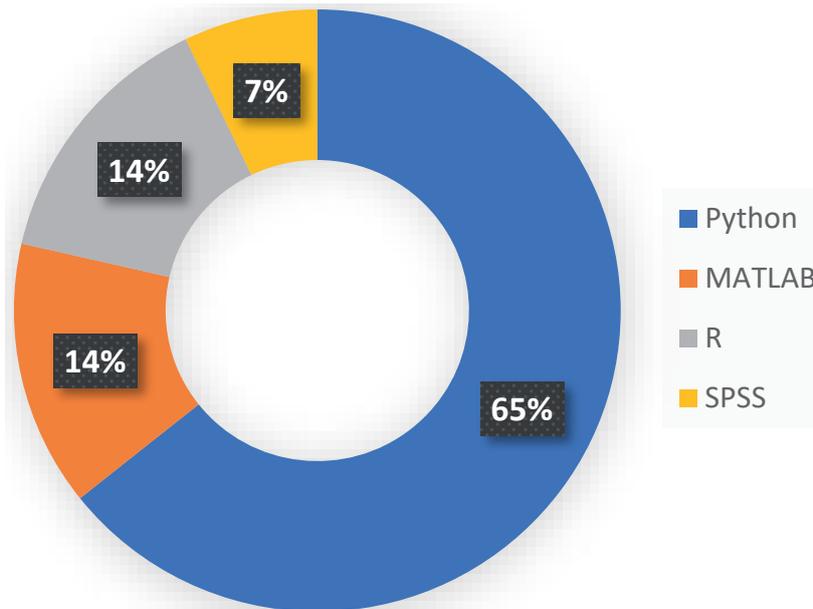


**Figura 8**

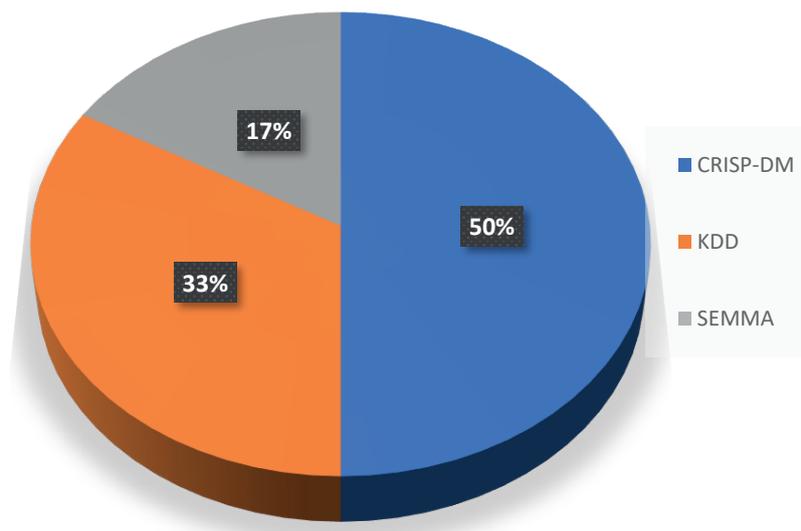
*Técnicas de ML usadas para desarrollar modelos predictivos*



**Figura 9**  
*Herramientas para desarrollar algoritmos de ML*



**Figura 10**  
*Metodologías implementadas en proyectos con minería de datos*



#### 4. Conclusiones

Este trabajo presenta una revisión sistemática de la literatura sobre la predicción de terremotos empleando técnicas de ML. Para ello, se recurrió a las bases de datos indexadas, encontrándose 126 artículos que tenían relación con el objetivo de estudio, después de realizar los criterios de inclusión y exclusión se seleccionaron en total 51 papers para el presente trabajo.

En los artículos seleccionados, se encontraron las variables, técnicas, herramientas y metodologías empleadas para predecir terremotos. Las variables sísmicas más utilizadas en los diferentes estudios fueron: latitud, longitud, profundidad y magnitud. Las técnicas de ML más empleadas para desarrollar modelos predictivos en los diferentes trabajos, se encontraron los algoritmos: RF,

ANN y SVM. El lenguaje de programación Python, es el que más usaron los diferentes autores para desarrollar modelos predictivos, mientras que, la guía metodológica CRISP-DM, es la que más implementaron al momento de desarrollar proyectos con minería de datos.

En un trabajo futuro, se recomienda realizar modelos de ML para predecir terremotos con las variables, técnicas, herramientas y metodologías encontradas.

#### Referencias

- [1] He, M., Ren, S., Tao, Z., "Cross-fault Newton force measurement for Earthquake prediction", Rock Mechanics Bulletin, 2022, Vol. 1, No. 100006, pp. 1-19. DOI: 10.1016/j.rockmb.2022.100006.
- [2] Hamdy, O., Gaber, H., Abdalzaher, M., Elhadidy, M., "Identifying Exposure of Urban Area to Certain Seismic Hazard Using Machine Learning and GIS: A Case Study of Greater Cairo", Sustainability, 2022, Vol. 14, No. 10722, pp. 1-32. DOI: 10.3390/su141710722.

- [3] Pwavodi, J., Intégration des données de géophysique en forage et des propriétés pétrophysiques sur carottes pour modéliser les propriétés hydrogéologiques de la zone de subduction de Nankai”, Tesis Doctoral, Universidad Grenoble Alpes, 2023. <https://theses.fr/2023GRALU001>
- [4] Tréhu, A., Bangs, N., Contreras-Reyes, E., Davenport, K., Geersen, J., “Imaging the source region of recent megathrust earthquakes along the Chile subduction zone: A summary of results from recent experiments”, *Journal of South American Earth Sciences*, 2023, Vol. 17, No. 104313, pp. 1-2. DOI: 10.1016/j.jsames.2023.104313.
- [5] Fernandez, R., “Statista - Ranking de los terremotos más fuertes según la escala de Richter ocurridos en el mundo de 1900 a 2024”. URL: <https://goo.su/0s8Ks5>
- [6] Pwavadi, J., Umar, A., Coston, P., Al-Turjman, F., Mohand, A., “The role of artificial intelligence and IoT in prediction of earthquakes: Review”, *Artificial Intelligence in Geosciences*, 2024. Vol. 5, No. 100075, pp. 2-15. DOI: 10.1016/j.aiig.2024.100075.
- [7] Abebe, E., Kebede, H., Kevin, M., Demissic, Z.: “Earthquakes magnitude prediction using deep learning for the Horn of Africa”, *Soil Dynamics and Earthquake Engineering*, 2023, Vol. 170, No. 107913, pp. 2-11. DOI: 10.1016/j.soildyn.2023.107913.
- [8] Chanda, S., Raghucharan, M., Karthik, K., Chaudhari, V., Nadh, S., “Duration prediction of Chilean strong motion data using machine learning”, *Journal of South American Earth Sciences*, 2021, Vol. 109, No. 103253, pp. 1-7. DOI: 10.1016/j.jsames.2021.103253.
- [9] Fayaz, J., Galasso, C., “Interpretability and spatial efficacy of a deep-learning-based on-site early warning framework using explainable artificial intelligence and geographically weighted random forests”, *Geoscience Frontiers*, 2024, Vol. 15, No. 101839, pp. 1-14. DOI: 10.1016/j.gsf.2024.101839.
- [10] Nakayacha, K., Yokoi, R., Goltz, J., “Human behavioral response to earthquake early warnings (EEW): Are alerts received on mobile phones inhibiting protective actions?”, *International Journal of Disaster Risk Reduction*, 2024, Vol. 105, No. 104401, pp. 1-9. DOI: 10.1016/j.ijdr.2024.104401.
- [11] Chun, F., Ren, T., Liu, Z., Zhong, Z., “Toward earthquake early warning: A convolutional neural network for rapid earthquake magnitude estimation”, *Artificial Intelligence in Geosciences*, 2023, Vol. 4, pp. 40-45. DOI: 10.1016/j.aiig.2023.03.001.
- [12] Bostrom, A., McBride, S., Becker, J., Goltz, J., Groot, R., Peek, L., Terbush, B., Dixon, M., “Great expectations for earthquake early warnings on the United States West Coast”, *International Journal of Disaster Risk Reduction*, 2022, Vol. 82, No. 103296, pp. 2-20. DOI: 10.1016/j.ijdr.2022.103296.
- [13] Wei, Y., Li, X., Wang, Z., “Deep learning for magnitude prediction in earthquake early warning”, *Gondwana Research*, 2023, Vol. 123, pp. 165-171. DOI: 10.1016/j.gr.2022.06.009.
- [14] Chen, W., Zhang, L., “An automated machine learning approach for earthquake casualty rate and economic loss prediction”, *Reliability Engineering & System Safety*, 2022, Vol. 255, No. 108645, pp. 1-14. DOI: 10.1016/j.res.2022.108645.
- [15] Gu, Z., Li, Y., Zhang, M., Liu, Y., “Modelling economic losses from earthquakes using regression forests: Application to parametric insurance”, *Economic Modelling*, 2023, Vol. 125, No. 106350, pp. 1-8. DOI: 10.1016/j.econmod.2023.106350.
- [16] Peng, J., Li, X., Chen, S., Liu, C., “Multi-source driven estimation of earthquake economic losses: A comprehensive and interpretable ensemble machine learning model”, *International Journal of Disaster Risk Reduction*, 2024, Vol. 106, No. 104377, pp. 1-21. DOI: 10.1016/j.ijdr.2024.104377.
- [17] Dey, B., Dikshit, P., Sehgal, S., Trehan, V., Kumar, V., “Intelligent solutions for earthquake data analysis and prediction for future smart cities”, *Computers & Industrial Engineering*, 2022, Vol. 170, No. 108368, pp. 1-11. DOI: 10.1016/j.cie.2022.108368.
- [18] Yousefzadeh, M., Ahmad, S., Farnaghi, M., “Spatiotemporally explicit earthquake prediction using deep neural network”, *Soil Dynamics and Earthquake Engineering*, 2021, Vol. 144, No. 106663, pp. 1-9. DOI: 10.1016/j.soildyn.2021.106663.
- [19] Berhich, A., Belouadha, F., Issam, M., “An attention-based LSTM network for large earthquake prediction”, *Soil Dynamics and Earthquake Engineering*, 2023, Vol. 165, No. 107663, pp. 2-11. DOI: 10.1016/j.soildyn.2022.107663.
- [20] Al, H et al., “Attention-Based Bi-Directional Long-Short Term Memory Network for Earthquake Prediction”, *IEEE Access*, 2021, Vol. 9, pp. 56589 - 56603, DOI: 10.1109/ACCESS.2021.3071400.
- [21] Asim, K., Moustafa, S., Azim, I., Elawadi, E., Iqbal, T., Martínez, F., “Seismicity analysis and machine learning models for short-term low magnitude seismic activity predictions in Cyprus”, *Soil Dynamics and Earthquake Engineering*, 2020, Vol. 130, No. 105932, pp. 1-10. DOI: 10.1016/j.soildyn.2019.105932.
- [22] Mahoma, A., Kùlahcı, F., Birel, S., “Investigating radon and TEC anomalies relative to earthquakes via AI models”, *Journal of Atmospheric and Solar-Terrestrial Physics*, 2023, Vol. 245, No. 106037, pp. 2-9. DOI: 10.1016/j.jastp.2023.106037.
- [23] Ommi, S., Hashemi, M., “Machine learning technique in the north zagros earthquake prediction”, *Applied Computing and Geosciences*, 2024, Vol. 22, No. 100163, pp. 1-8. DOI: 10.1016/j.acags.2024.100163.
- [24] Wang, X., Wang, Z., Wang, J., Miao, J., Dang, H., Li, Z., “Machine learning based ground motion site amplification prediction”, *Frontiers in Earth Science*, 2023, Vol. 11, pp. 1-12. DOI: 10.3389/feart.2023.1053085.
- [25] Jozinovic, D., Lomax, A., Stajduhar, I., Michelini, A., “Transfer learning: improving neural network-based prediction of earthquake ground shaking for an area with insufficient training data”, *Geophysical Journal International*, 2022, Vol. 229, No. 1, pp. 704-708. DOI: 10.1093/gji/ggab488
- [26] Essam, Y., Kumar, P., Najah, A., Ary, M., El-Shafie, A., “Exploring the reliability of different artificial intelligence techniques in predicting earthquake for Malaysia”, *Soil Dynamics and Earthquake Engineering*, 2021, Vol. 147, No. 106826, pp. 2-11. DOI: 10.1016/j.soildyn.2021.106826.
- [27] Dang, H., Wang, Z., Zhao, D., Wang, X., Li, Z., Wei, D., Wang, J., “Ground motion prediction model for shallow crustal earthquakes in Japan based on XGBoost with Bayesian optimization”, *Soil Dynamics and Earthquake Engineering*, 2024, Vol. 177, No. 108391, pp. 2-13. DOI: 10.1016/j.soildyn.2023.108391.
- [28] Kim, S., Hwang, Y., Se, H., Kim, B., “Ground motion amplification models for Japan using machine learning techniques”, *Soil Dynamics and Earthquake Engineering*, 2020, Vol. 132, No. 106095, pp. 1-16. DOI: 10.1016/j.soildyn.2020.106095.
- [29] Khosravikia, F., Clayton, P., “Machine learning in ground motion prediction”, *Computers & Geosciences*, 2021, Vol. 148, No. 104700, pp. 1-9. DOI: 10.1016/j.cageo.2021.104700.
- [30] Zhang, B., Hu, Z., Wu, P., Huang, H., Xiang, J., “EPT: A data-driven transformer model for earthquake prediction”, *Engineering Applications of Artificial Intelligence*, 2023, Vol. 123, No. 106176, pp. 1-11. DOI: 10.1016/j.engappai.2023.106176.
- [31] Díaz, J., Sáez, E., Monsalve, M., Candia, G., Aron, F., Gonzáles, G., “Machine learning techniques for estimating seismic site amplification in the Santiago basin, Chile”, *Engineering Geology*, 2022, Vol. 306, No. 106764, pp. 1-11. DOI: 10.1016/j.enggeo.2022.106764.
- [32] Barra, S., Morenob, M., Orega, F., Benavente, R., Araya, R., Bedford, J., Calisto, I., “A supervised machine learning approach for estimating plate interface locking: Application to Central Chile”, *Physics of the Earth and Planetary Interiors*, 2024, No. 107207. DOI: 10.1016/j.pepi.2024.107207.
- [33] Izquierdo, L., Zevallos, J., Yépez, Y., “An integrated approach to seismic risk assessment using random forest and hierarchical analysis: Pisco, Peru”, *Heliyon*, 2022, Vol. 8, No. 10, pp. 1-8. DOI: 10.1016/j.heliyon.2022.e10926.
- [34] Wang, W., Xu, S., Zhuo, Y., Liu, P., Shengli, M., “Unraveling the roles of fault asperities over earthquake cycles”, *Earth and Planetary Science Letters*, 2024, Vol. 636, No. 118711, pp. 1-9. DOI: 10.1016/j.epsl.2024.118711.

- [35] Laurenti, L., Tinti, E., Galasso, F., Franco, L., Marone, C.: "Deep learning for laboratory earthquake prediction and autoregressive forecasting of fault zone stress", *Earth and Planetary Science Letters*, 2022, Vol. 598, No. 117825, pp. 1-12. DOI: 10.1016/j.epsl.2022.117825.
- [36] Mahoma, A., K ulahcı, F., Birel, S., "Investigating radon and TEC anomalies relative to earthquakes via AI models", *Journal of Atmospheric and Solar-Terrestrial Physics*, 2023, Vol. 245, No. 106037, pp. 1-9. DOI: 10.1016/j.jastp.2023.106037.
- [37] Li, Z., Lei, H., Ma, E., Lai, I., Qiu, J.: "Ensemble technique to predict post-earthquake damage of buildings integrating tree-based models and tabular neural networks", *Computers & Structures*, 2023, Vol. 287, No. 107114, pp. 1-15. DOI: 10.1016/j.compstruc.2023.107114.
- [38] Cosgun, C.: "Machine learning for the prediction of evaluation of existing reinforced concrete structures performance against earthquakes", *Structures*, 2023, Vol. 50, pp. 1995-2002. DOI: 10.1016/j.istruc.2023.02.127.
- [39] Meng, F., Ren, T., Liu, Z., Zhong, Z., "Toward earthquake early warning: A convolutional neural network for rapid earthquake magnitude estimation", *Artificial Intelligence in Geosciences*, 2023, Vol. 4, pp. 40-45. DOI: 10.1016/j.aiig.2023.03.001.
- [40] Bhatia M., Ahamed, T., Manocha, A., "Artificial intelligence based real-time earthquake prediction", *Engineering Applications of Artificial Intelligence*, 2023, 120, (105856), pp. 1-13. DOI: 10.1016/j.engappai.2023.105856.
- [41] Mamani, Y., "Una metodolog a para encontrar patrones frecuentes de datos, con su aplicaci n en la predicci n de terremotos", 2021.
- [42] Julca, W., Paucar, H., "Sistema de Recomendaci n basado en Contenido para Jueces de Programaci n utilizando Procesamiento de Lenguaje Natural y Aprendizaje Profundo", *Revista Peruana de Computaci n y Sistemas*, Vol. 5, No. 1, pp. 26-30. DOI: 10.15381/rpcs.v5i1.25802.
- [43] Espinoza, J., "Implementation of the CRISP-DM methodology for geographical segmentation using a public database", *Ingenier a, Investigaci n y Tecnolog a*, 2020, Vol. 11, No. 1, pp. 2-13. DOI: 10.22201/fi.25940732e.2020.21n1.008.
- [44] Plotnikova, V., Dumas, M., Milani, F., "Applying the CRISP-DM data mining process in the financial services industry: Elicitation of adaptation requirements", *Data & Knowledge Engineering*, 2022, Vol. 139, No. 102013, pp. 1-15. DOI: 10.1016/j.datak.2022.102013.
- [45] Cheng, A., "Evaluating Fintech industry's risks: A preliminary analysis based on CRISP-DM framework", *Finance Research Letters*, 2023, Vol. 55, No. 103966, pp. 1-7. DOI: 10.1016/j.frl.2023.103966.
- [46] Bresso, E., Ferreira, J., Girerd, N., Kobayashi, M., Preud, G., Rossignol, P., Zannad, F., Dominique, M., Smail, M., "Inductive database to support iterative data mining: Application to biomarker analysis on patient data in the Fight-HF project", *Journal of Biomedical Informatics*, 2022, Vol. 135, No. 104212, pp. 1-9. DOI: 10.1016/j.jbi.2022.104212.
- [47] Llatas, C., Soust-Verdaguer, B., Castro, L., Cagigas, D., "Application of Knowledge Discovery in Databases (KDD) to environmental, economic, and social indicators used in BIM workflow to support sustainable design", *Journal of Building Engineering*, 2024, Vol. 91, No. 109546, pp. 1-14. DOI: 10.1016/j.job.2024.109546.
- [48] Zhang Y., Li, C., Guo, B., Guo, C., Zhang, S., "KDD: A kernel density based descriptor for 3D point clouds", *Pattern Recognition*, 2021, Vol. 111, No. 107691, pp. 1-12. DOI: 10.1016/j.patcog.2020.107691.
- [49] Condor, E., Loa, E., Huarcaya, J., "Miner a de datos en Twitter: an lisis del sentimiento del desempleo en la poblaci n hispanohablante en tiempos del COVID-19", 2021, pp. 195-209. URL: <https://acortar.link/ZCrzpG>.
- [50] Zouheir, M., Ait, A., "Big data analytics in telecommunications: Governance, architecture and use cases", *Journal of King Saud University - Computer and Information Sciences*, 2022, Vol. 34, pp. 2759-2769. DOI: 10.1016/j.jksuci.2020.11.024.