

Benemérita Universidad Autónoma de Puebla

Facultad de Ciencias de la Computación

Licenciatura en Ciencias de la Computación



BUAP

Perspectivas Sísmicas en México Usando Machine Learning

Presenta:

Juan Adolfo Bustillos Alatorre

Director de Tesis:

María del Carmen Santiago Díaz

Asesor de Tesis

Gustavo Trinidad Rubín Linares

Puebla, México

08 / 04 / 2022

*Tesis presentada para obtener el grado de
Licenciatura en Ciencias de la Computación*

Agradecimientos

Mis agradecimientos están dedicados a mi familia por creer en mí, a mi institución que me tomó como estudiante y se convirtió en mi alma mater, a mis profesores que depositaron sus conocimientos y valores en mí, a mis asesores por sus virtudes, su paciencia, su motivación y constancia y agradezco a todas las personas que me dieron apoyo para realizar esta investigación.

Agradecimientos especiales a mis asesores María del Carmen Santiago Díaz y Gustavo Trinidad Rubín que estuvieron en todo el camino guiando mis esfuerzos, y cuya asistencia en este largo proceso fue fundamental para la culminación de esta tesis, agradezco a Bustillos Alatorre Victor Manuel y Gonzales del Real Marijose que apoyaron en la traducción del artículo de esta investigación y a mi madre por su siempre presente apoyo incondicional.

Desarrollo

La predicción de terremotos es un área de investigación desafiante y se han realizado modestos esfuerzos utilizando el aprendizaje automático para esta aplicación. En este trabajo, las características sísmicas se calculan utilizando conceptos sismológicos, como la ley de Gutenberg-Richter, tasa de cambio sísmico, frecuencia de anticipación, liberación de energía sísmica y tiempo de recurrencia total. Con esto, se construyó un modelo de clasificación de Redes Neuronales Artificiales (ANN) utilizando el método de ventanas deslizantes (a diferencia de los conjuntos fijos de ventanas vistos en otros estudios) para calcular sus instancias, y el retraso estadístico para reformular el problema como un problema de predicción de series de tiempo multivariable. , que luego se volvieron a muestrear como una medida para contrarrestar los datos desequilibrados con el fin de obtener predicciones de terremotos; el modelo predice más eventos sísmicos por encima de un umbral sin sacrificar significativamente las métricas

Contenido

Agradecimientos	1
Contenido	2
Introducción	3
1. Literatura Relacionada	12
1.1. Catálogo de estudios relacionados	14
1.2. Análisis de los estudios relacionados	22
2. Perspectiva General de México	26
3. Preparación de Catálogo de Terremotos	31
3.1. Extensión Geométrica del Catálogo	33
3.2. Extensión Temporal del Catálogo	38
3.3. Magnitud de corte	39
4. Cálculo de Indicadores Sísmicos	41
4.1. Parámetros	44
4.2. Indicadores	44
4.3. Etiquetas	48
5. Modelos de Predicción	50
5.1. Modelos predictivos	50
5.2. Redes Neuronales	50
5.3. Descripción del modelo de predicción	51
5.4. Descripción de la red Neuronal	54
5.5. Parámetros	57
5.6. Variaciones de Modelos de Predicción	57
6. Resultados	64
6.1. Métricas	64
6.2. Resultados	66
Conclusión	71
Referencias	72

Introducción

Los terremotos son movimientos y sacudidas repentinas dentro del suelo que producen energía en las rocas en el interior de la Tierra. La capa más superficial de la Tierra, denominada litósfera es una capa rígida compuesta por material que puede fracturarse al ejercer una fuerza sobre él y está formado por piezas llamadas Placas Tectónicas. La corteza terrestre está dividida en 16 grandes placas tectónicas y 42 menores o microplacas que se deslizan sobre el magma, se desplazan debido a movimientos convectivos en la capa intermedia de la Tierra o manto, esto es, material caliente del interior de la Tierra sube a la superficie liberando calor interno, mientras que el material frío baja al interior. Es justo en los límites entre placas, donde hacen contacto y se generan fuerzas de fricción produciendo un gran estrés en los materiales. Cuando dicho estrés sobrepasa la resistencia de la roca, o cuando se vence la fuerza de fricción, esta roca cede y se desliza a lo largo de fracturas llamadas fallas, la tensión se alivia y se libera repentinamente la energía acumulada en una ruptura violenta, generando así un temblor que irradia dicha energía en forma de ondas que se propagan en todas direcciones a través del medio sólido de la Tierra, denominadas ondas sísmicas. Al propagarse estas ondas sísmicas provocan el movimiento del terreno por donde pasan, a esto se le da el nombre de terremoto. [67][41]

Para registrar estos movimientos se utilizan equipos denominados sismógrafos y acelerógrafos, cuyo principio de operación, basado en la inercia de los cuerpos, consiste de una masa suspendida por un resorte que le permite permanecer en reposo por algunos instantes con respecto al movimiento del suelo. Los sismógrafos modernos utilizan este mismo principio de operación, solo que para su implementación utilizan componentes mecánicos y electrónicos para obtener una señal eléctrica proporcional al movimiento del suelo, la cual puede almacenarse en forma local o ser transmitida hasta un centro de adquisición. [67]

Para medir la fuerza de un terremoto se utiliza la escala de magnitud, la magnitud de un terremoto está relacionada con el área de la falla de ruptura, el desplazamiento de las placas, la fluidez del deslizamiento y la energía liberada en forma de ondas sísmicas que se propagan a través del interior de la Tierra durante un temblor, cuanto mayor sea el área de la falla, mayores serán los terremotos que puede producir. [41]

Para calcular esta energía y determinar la magnitud de un temblor se realizan cálculos matemáticos basados en los registros obtenidos por los sismógrafos de diferentes estaciones. En estos registros o sismogramas se miden algunas características de las ondas y la distancia a la que se encuentra la estación del epicentro. Estos valores son introducidos a una fórmula, obteniendo así la magnitud. [41]

Existen diferentes formas de medir la magnitud, esto quiere decir que existen diversas fórmulas matemáticas para calcularla. De hecho, actualmente ya no se usa la escala de Richter original, la cual es algo antigua y en su momento se hizo para ser utilizada con un tipo de sismómetro que ya no se usa y en regiones geográficas no generalizadas. [67]

Las magnitudes que usamos ahora son: La magnitud de coda (M_c), la magnitud de energía (M_e), la magnitud de momento sísmico (M_w), entre otras. Cada una de estas formas de calcular la magnitud tiene sus ventajas y sus limitaciones. Algunas son más rápidas de calcular pero menos precisas, otras son más precisas pero su cálculo necesita más tiempo; unas son más confiables para sismos pequeños, otras para sismos grandes. Esta es una de las razones por la cual, para algunos sismos, la magnitud preliminar — la primera que se reporta — a veces se cambia un tiempo después, cuando ya se pudieron realizar cálculos más precisos. [41]

También se puede utilizar la escala de *intensidad*, esta se asigna en función a los daños o efectos causados al hombre y sus construcciones, esta escala es menos objetiva y no está basada en fenómenos físicos. [67]

Los registros en los que son compiladas las magnitudes calculadas son llamados **catálogos sísmicos**, que funcionan como registros históricos de los eventos sísmicos informados de una región, sus magnitudes se registran junto con otras características del evento como magnitud, su ubicación geográfica, tiempo del evento ocurrido, profundidad, etc.

Los terremotos ocupan un gran espacio de interés de los científicos, debido a que son uno de los desastres naturales más peligrosos, como muchos desastres naturales, el terremoto causa muchos daños, pérdidas económicas, humanas y lesiones, así como otros efectos directos e indirectos, principalmente debido al hecho de que a menudo ocurren sin una advertencia explícita, sin dejar tiempo para reaccionar. [1][25]

En realidad, pequeños terremotos ocurren literalmente todos los días, pero no sentimos la mayoría de ellos, el tamaño de un terremoto se expresa usando una escala logarítmica llamada magnitud y la gente generalmente no puede sentir terremotos por debajo de una magnitud de 2 o 3. Cada vez que la magnitud aumenta en un grado, la amplitud de las ondas aumenta diez veces y el terremoto libera 32 veces más energía. Esto significa que un terremoto de grado 6 libera 32768 veces más energía que un terremoto grado 3.

Estos terremotos cotidianos pequeños tienden a originarse en fallas a menos de 70 kilómetros bajo tierra, por lo que se denominan terremotos superficiales, por lo general, son bastante inofensivos, aunque algunos terremotos superficiales pueden ser bastante severos. Como el terremoto de 19 de septiembre de 2017, un sismo a

51 km de profundidad de magnitud 7.1 se produjo a 55 kilómetros (34 millas) al sur de Puebla y a 100 kilómetros (62 millas) de la Ciudad de México que causó daños desde el sureste de la capital a través de los estados de Puebla y Morelos. Solo en la Ciudad de México, hubo 228 muertos y más de 40 edificios se derrumbaron. [40]

Cuando ocurre un sismo de magnitud considerable como éste, las rocas que se encuentran cerca de la zona de ruptura están sujetas a un reajuste. Durante este proceso se generan una serie de sismos en esta zona conocidos como réplicas, las cuales son de menor magnitud y pueden ocurrir minutos, días y hasta años después del evento principal. El número de estas réplicas puede variar desde unos cuantos sismos hasta cientos de eventos. [68]

Los terremotos en México que ocurren con mayor frecuencia se generan en alta mar cerca de la Costa del Pacífico donde las placas convergen. El Servicio Sismológico Nacional (SSN) define: México se encuentra ubicado en el contexto de cinco placas tectónicas: Caribe, Pacífico, Norteamérica, Rivera y Cocos; tiene un extenso litoral sobre el Cinturón de Fuego del Pacífico, se asienta sobre 5 placas tectónicas y cuenta con innumerables fallas. En promedio, un terremoto de magnitud mayor o igual a 7.0 ocurre en México cada dos años y medio. Esto convierte a México en uno de los países con mayor sismicidad, se encuentra entre los 10 países con más actividad sísmica. Durante el siglo XX, el 8% de todos los terremotos del mundo de magnitud mayor o igual a 7.0 han ocurrido en México. [10] Aunque, en el siglo XXI esta cifra se encuentra más cerca del 3%. En las regiones más sísmicas del país, es razonable experimentar a lo largo de la vida un mínimo de tres o cuatro terremotos de 6 grados o más. [87]

Los terremotos pueden dañar gravemente o destruir una región completa en segundos. Debido a sus efectos devastadores, los terremotos son una seria amenaza para la sociedad moderna, por ejemplo el terremoto de 1985 en la Ciudad de México con una magnitud de momento de 8.0, con réplicas de hasta 7.5. El número preciso de muertos, heridos y daños materiales nunca se conoció con precisión. En cuanto a las personas fallecidas, sólo existen estimaciones: 3192 fue la cifra oficial, mientras que 20,000 fue el dato resultante de los cálculos de algunas organizaciones. Provocó graves daños en la Ciudad de México 412 edificios se derrumbaron y otros 3124 resultaron gravemente dañados en la ciudad. [84]

Debido al potencial de destrucción de los terremotos como el sucedido en 1985 en la Ciudad de México hace que uno de los objetivos más ambiciosos de la sismología sea el de la predicción de terremotos a corto plazo, por lo que hay muchos tipos de investigación y esfuerzos de científicos que se dedican a este dominio. El riesgo que representan los terremotos hace que el problema de la predicción de terremotos sea extremadamente importante para la seguridad de la humanidad. Durante más de 100 años, los científicos han buscado métodos exitosos para la predicción de terremotos o precursores confiables, sin ninguna evidencia obvia de éxito. [34]

Los **precursores sísmicos** son los fenómenos anómalos que regularmente preceden los eventos sísmicos. [32] La comunidad científica ha estudiado muchos tipos diferentes de precursores de terremotos: movimientos del suelo, señales sísmicas, emisiones de fluidos o gases, señales eléctricas, señales térmicas, comportamiento animal, etc. [4] Existen recopilaciones y más información de precursores sísmicos conocidos hasta 2009. [2]

Los eventos sísmicos requeridos para calcular los indicadores sísmicos están bien cubiertos por la instrumentación sísmica de superficie, los eventos son compilados y accesibles en catálogos online, cada catálogo contiene las características de los eventos sísmicos (i.e. terremotos) de los que fueron informados.

Los **indicadores sísmicos** se utilizan para realizar análisis de actividad sísmica y describen características sísmicas del suelo, estos se obtienen al realizar ecuaciones estadísticas y matemáticas sobre los eventos de un catálogo sísmico de una región generados a partir de conceptos geofísicos y sismológicos.

La información relevante de los catálogos para el cálculo de los indicadores sísmicos son generalmente la fecha, la profundidad, la magnitud y la ubicación del evento. Usando esos datos se calculan relaciones no paramétricas (que no dependen de ningún parámetro ej. $\text{mean}(M)$) y relaciones paramétricas (que dependen de cualquier parámetro variable, generalmente otra relación previamente generada), en algunos casos tienen múltiples valores con base a diferentes variaciones de sus parámetros, a estas relaciones se les llama indicadores sísmicos, las relaciones que se utilizan en este modelo son los parámetros geofísicos y sismológicos relevantes para la predicción de terremotos disponibles en la literatura contemporánea, estos criterios se toman con el fin de retener la máxima información disponible sobre el estado geológico interno del suelo.

El potencial de destrucción y muerte de los terremotos hacen que el problema de la predicción de terremotos sea fundamental para la seguridad humana. Desde finales del siglo XIX, los investigadores en sismología y ramas de la ciencia afines han tratado de descubrir precursores. Se han estudiado muchos precursores posibles, incluidos los premonitorios (terremotos que ocurren antes de eventos sísmicos más grandes), anomalías electromagnéticas llamadas "luces de terremoto", cambios en los niveles de agua subterránea e incluso comportamiento animal inusual. [25] En algunos casos, la aparición de un precursor llevó a la evacuación oportuna de civiles. [70]

Es importante señalar que es difícil utilizar precursores para la predicción a corto plazo, ya que no solo son característicos de los terremotos (por ejemplo, las luces inusuales en la atmósfera pueden aparecer antes de las tormentas geomagnéticas o tener un origen tecnogénico). Además, diferentes precursores precedieron a los

terremotos, que fueron de diferente naturaleza, ocurrieron en diferentes zonas sísmicas e incluso temporadas. [25]

Durante décadas, la idea de que los terremotos pueden ser pronosticados se basó en el intervalo de recurrencia de los terremotos característicos, que se repiten casi periódicamente; esta idea fue abandonada, cuando los eventos predichos no ocurrieron. Desde entonces, se propusieron varios métodos de predicción, basados en la medición de señales relacionadas con los terremotos (geoeléctrica, radón, ionosférico, ...), o en parámetros derivados del análisis estadístico de sismicidad (b-value, dimensión fractal epicentral, etc.). Actualmente no existe una metodología general aceptada para la predicción de terremotos, es más preciso afirmar que en la actualidad no existe un método universalmente aceptado para predecir la ocurrencia de un terremoto de gran magnitud. [10] La plétora de métodos de predicción solo indica cuán complejos son los terremotos y cuán desafiante sigue siendo el problema de la predicción. [81] La predicción de terremotos es un área de investigación desafiante, además, la mera frase "predicción de terremotos" es tabú en muchos sectores y todavía no existe consenso en la comunidad científica sobre si es posible encontrar una solución a este problema.

Sin embargo, con el advenimiento y rápido desarrollo de los algoritmos inteligentes basados en la informática moderna y la aplicación exitosa de estos métodos a varios tipos de problemas indican que estas tecnologías podrían ayudar a extraer patrones ocultos y hacer predicciones precisas. [25] Se han logrado resultados significativos en diferentes campos de investigación, como la predicción meteorológica, la predicción de abandono y el diagnóstico de enfermedades. [36-38] Por lo tanto, al cerrar la brecha entre la informática y la sismología, se pueden lograr resultados sustanciales.

Investigadores de todo el mundo hacen todo lo posible para tratar de prevenir estos efectos severos e intentar predecir dónde y cuándo ocurre un terremoto en función de los indicadores sísmicos, señales eléctricas y otros precursores sísmicos utilizando técnicas como aprendizaje automático y algoritmos de optimización. [18]

Pero a pesar del continuo interés de la comunidad científica en este tema, sigue sin haber un consenso sobre si es posible encontrar la solución con suficiente precisión y se cree que los científicos solo pueden calcular la probabilidad de que ocurra un terremoto significativo en un área específica dentro de un cierto número de años. [67]

Agencias como USGS (United States Geological Survey) en U.S. y SSN (Servicio Sismológico Nacional) en México centra sus esfuerzos en la mitigación a largo plazo de los peligros sísmicos ayudando a mejorar la seguridad de las estructuras, en lugar de intentar lograr predicciones a corto plazo. Afortunadamente los ingenieros que trabajan en esto son mucho mejores en la protección de estructuras que

nosotros en la predicción de terremotos. Hoy en día, existen algunas formas principales en las que intentamos proteger los edificios de los terremotos, una es evitar que los edificios se muevan de lado a lado como sea posible y eso se puede hacer de varias maneras. Muchos rascacielos enormes utilizan bolas oscilantes masivas, también conocidas como amortiguadores de masa sintonizados; se colocan péndulos grandes colocados en lo alto dentro de los edificios y se balancean en respuesta a cualquier movimiento que haga el edificio que contrarreste lo que esté sucediendo afuera. Un ejemplo famoso de esta solución se encuentra en el edificio Taipei 101 en Taipei, Taiwán. Para edificios más cortos, los ingenieros a menudo eligen una ruta diferente, aíslan la base del edificio del suelo utilizando un sistema de amortiguadores.

Hace varias décadas, a mediados de la década de 1970, los sismólogos confiaban en que la predicción de terremotos a corto plazo se lograría en un corto período de tiempo. Esta confianza surgió en parte como resultado de la primera predicción exitosa de un gran terremoto, una predicción de terremoto en China basado en pequeños terremotos y una actividad animal inusual como precursores del terremoto de 1975 M7.4 Haicheng en China. Debido a esta predicción, se emitió una alerta dentro del período de 24 horas antes del choque principal, debido a esto muchas personas optaron por dormir fuera de sus hogares y, por lo tanto, se salvaron cuando ocurrió el terremoto principal, probablemente previniendo un mayor número de víctimas que las 1328 muertes que realmente ocurrieron por este evento. Sin embargo, este tipo de actividad sísmica rara vez es seguida por un gran terremoto y, desafortunadamente, la mayoría de los terremotos no tienen eventos precursores de ningún tipo. El siguiente gran terremoto en China 18 meses después no tuvo precursores y las víctimas de este terremoto se cuentan por cientos de miles, fue el terremoto M7.8 de Tangshan de 1976 que revivió la idea de la imposibilidad de predecir terremotos devastadores, esto fue un gran revés para el esfuerzo de predicción del terremoto [2].Lomnitz ofrece un resumen de estos eventos, así como de otros éxitos y fracasos en la predicción de terremotos. [3][67]

Se han estudiado muchos tipos diferentes de precursores de terremotos, en particular una encuesta de la literatura científica publicada para identificar y catalogar los precursores de terremotos observados donde entre los precursores de terremotos seleccionados para el análisis incluyeron campos eléctricos y magnéticos, emisiones de gas, cambios en el nivel del agua subterránea, emisiones de fluidos o gases, cambios de temperatura, deformaciones de la superficie, movimientos del suelo, comportamiento animal, sismicidad, etc. [2][4]

Otro fracaso controversial en la predicción de terremotos fue el de 6 de abril de 2009 cuando un terremoto golpeó cerca de la ciudad italiana de L'Aquila matando a casi 300 personas fue un desastre terrible en el que seis científicos italianos y un empleado del gobierno fueron juzgados y condenados más tarde por cargos de homicidio, por proporcionar información inexacta, incompleta y contradictoria sobre el peligro de un terremoto, fueron condenados a seis años de cárcel básicamente

por no predecir un terremoto, esto ha irritado a mucha gente en el mundo de la ciencia y con buena razón porque aunque sabemos mucho más que antes acerca de los terremotos todavía no hay una manera confiable de predecirlos. [83]

Criminalizar algo solo tendría sentido si realmente supiéramos cómo hacer esto y alguien lo hiciera mal” dice Seth Stein, profesor de Ciencias de la Tierra en la Universidad Northwestern en Illinois.

Hay diferentes términos relacionados con la predicción de terremotos que usa el USGS y el SSN que tienen diferentes significados como son, alerta temprana de terremotos, pronósticos de terremotos, probabilidades de terremotos y predicción de terremotos. Aquí hay descripciones detalladas de cada uno:

Los **sistemas de alerta temprana** de terremotos utilizan la ciencia de los terremotos y la tecnología de los sistemas de monitoreo para alertar a los dispositivos y a las personas cuando se espera que lleguen a su ubicación las ondas generadas por un terremoto, se emite después de que comienza un terremoto. En México existe un sistema de alerta sísmica, apropiadamente llamado SAS, que emite avisos en el Valle de México cuando una serie de estaciones sensoras localizadas a lo largo de la costa de Guerrero confirman la ocurrencia de un sismo de gran magnitud en esta zona. La utilidad del SAS radica en el principio que la velocidad de propagación de las ondas de radio (por medio de las cuales se transmite la alerta) es mayor a la velocidad de propagación de las ondas sísmicas. El aviso anticipado del inicio del sismo que ocurre en la costa de Guerrero a su arribo al Valle de México y Toluca es de aproximadamente 60 segundos, esta advertencia previa concede tiempo suficiente a las personas para llevar a cabo medidas de protección de vidas y propiedades que reduzcan la posibilidad de que se genere un desastre considerable de los temblores destructivos.

A pesar de que las palabras **pronóstico** y **predicción** a menudo se usan indistintamente, en la ciencia de los terremotos se acostumbra distinguirlas. En particular se expresa la idea de que una predicción de terremotos implica una mayor probabilidad que un pronóstico de terremotos; en otras palabras, una predicción es más definida que un pronóstico y requiere mayor precisión. [71]

Esta distinción es comparable con las probabilidades climáticas y los pronósticos meteorológicos.

Los **pronósticos de terremotos** son como las probabilidades pero para ventanas de tiempo más cortas, y generalmente aplicamos este término a las réplicas. Después de un gran terremoto, hay réplicas que suelen ser menos frecuentes y más pequeñas con el tiempo. La mayoría de las secuencias de réplicas siguen el mismo patrón, por lo que se puede determinar la probabilidad de una réplica en una ventana de tiempo después de un terremoto.

Las **probabilidades de terremotos** describen las posibilidades a largo plazo de que ocurra un terremoto de cierta magnitud durante una ventana de tiempo. La mayoría de las probabilidades de terremotos se determinan a partir de la tasa promedio de eventos históricos. Suponiendo que la tasa anual es constante, se puede hacer un enunciado de probabilidad sobre la probabilidad de que ocurra tal evento en los próximos años. Estas probabilidades pueden oscilar entre 1 en 30 y 1 en 300.

Una **predicción de terremoto** debe definir 3 elementos: la fecha y hora, la ubicación y la magnitud, mientras que las predicciones son más como declaraciones de cuándo, dónde y qué tan grande, lo que aún no es posible para los terremotos. [72]

Las predicciones de terremotos deben estar basadas en evidencia científica, algunas personas dicen que pueden predecir los terremotos pero desaciertan al usar ventanas muy grandes de alguno o varios elementos de la predicción o relacionan los terremotos con fenómenos que no tienen nada que ver como las nubes, los dolores y molestias corporales o las babosas. [67]

Es importante destacar que todos estos parámetros deben definirse de tal manera que se pueda afirmar objetivamente que algún terremoto futuro satisface o no la predicción. Es necesario tanto para usar como para evaluar predicciones. En particular, se requiere definir la "ubicación" claramente y determinar los límites espaciales exactos del área, ya que un terremoto no ocurre en un punto. [25]

Es un hecho innegable que la comunidad de sismólogos no ha tenido éxito en desarrollar métodos para predecir terremotos a pesar de más de dos siglos de esfuerzos. La predicción de terremotos siguió siendo un objetivo no logrado debido a varias razones. Una de las razones es la falta de tecnología para monitorear con precisión los cambios de tensión, las variaciones de presión y temperatura en las profundidades de la corteza a través de instrumentos científicos, lo que eventualmente resulta en la falta de disponibilidad de datos completos sobre las características sísmicas. Otra peculiaridad que complica el problema es que los terremotos muy grandes, que son aquellos que estamos interesados en predecir, ocurren con relativa poca frecuencia y, por lo tanto, están subrepresentados en los datos de entrenamiento. Eso significa que estos modelos funcionan mejor para los eventos menores, menos desastrosos y más comunes, mientras que los métodos de detección tradicionales pueden ser mejores para eventos más grandes y raros. Aunque existen esfuerzos para explorar formas de incorporar más modelos de ML para abordar los grandes eventos poco frecuentes y realizar más mejoras en el futuro. [28]

Otra razón probable es la brecha entre los sismólogos y los científicos de la computación para explorar los diversos lugares de la tecnología para cazar esta desafiante tarea; también tiene que ver con los factores que tienen relación con los terremotos ya que no se han descubierto todos los factores que pueden influir en la ocurrencia de un terremoto, e incluso los factores bien conocidos, como la tensión acumulada o la tasa de liberación de energía sísmica, no se pueden medir directamente (o es demasiado difícil hacerlo), y se muestra que las relaciones entre la ocurrencia de nuevos terremotos y estas características sísmicas son complicadas y altamente no lineales.

La distribución del resto de la tesis consiste de 6 capítulos:

En el capítulo 1 de literatura relacionada se mostrarán estudios similares y se describirá la metodología utilizada en ellos, se explicará la tarea de predicción y como diferentes estudios intentan abarcarlo, al final de esta sección se comparan las observaciones en la literatura relacionada y se sacarán conclusiones.

En el capítulo 2 de Perspectiva General de México se describe la situación sísmica histórica en México, como el país se ubica entre las regiones más sísmicas del mundo.

En la Sección 3 Preparación de Catálogo de Terremotos se explican diferentes conceptos de la sismología que justifican medidas que se deben de tomar antes de analizar los datos de un catálogo, y se describen los detalles sobre la preparación como la extensión geométrica del catálogo, la extensión temporal del catálogo y la magnitud de corte.

En la sección 4 de Cálculo de indicadores sísmicos, se describirán los indicadores y etiquetas usadas, así como los métodos y parámetros usados para calcularlos.

En la Sección 5 de Modelos de predicción se describen los tres diferentes modelos construidos y sus características, se describe la red neuronal y los parámetros usados para su entrenamiento; se describen las métricas seleccionadas para evaluar el desempeño de los modelos y se muestran los resultados de las métricas del entrenamiento de los diferentes modelos y por último se compara el desempeño de los diferentes modelos y se discuten los resultados.

Y para terminar una sección de Conclusión donde se mencionan las observaciones más importantes sobre el estudio y sus resultados y se realiza una previsión sobre el estado de esta área de estudio.

1. Literatura Relacionada

En la literatura existen diferentes metodologías de investigación, indicadores sísmicos que han sido usadas junto con técnicas de ML para el análisis sísmico. En esta sección se mostrarán algunas de ellas.

Muchos estudios de análisis sísmicos utilizan indicadores sísmicos que describen las características sísmicas actuales del suelo. En particular los estudios de predicción suponen que existen variaciones en los indicadores sísmicos durante el pre-sismo (tiempo que antecede al de la ocurrencia de un sismo), y que existen relaciones en estas variaciones que un modelo puede aproximar mediante el aprendizaje.

Los indicadores sísmicos se obtienen al realizar ecuaciones estadísticas y matemáticas sobre el catálogo sísmico (datos históricos sobre la actividad sísmica informada) de una región generados a partir de conceptos geofísicos y sismológicos. Los indicadores sísmicos se utilizan para evaluar la probabilidad de actividad sísmica futura, describen características sísmicas del suelo actual, forman parte de los datos de entrada de los modelos con los que son entrenados, y son especialmente útiles en países con una larga historia de actividad sísmica como California, China, Japón y Turquía. [32]

Como se ha dicho anteriormente las técnicas de Machine Learning (ML) han surgido como una herramienta potente con ventajas innegables para tratar problemas complejos, no lineales y con uso intensivo de datos como es el caso de la predicción de terremotos. Estos métodos a menudo se basan en datos no paramétricos.

Investigadores han comenzado a utilizar algoritmos ML para resolver el problema de predicción de geociencias y sus aplicaciones se han expandido rápidamente durante las últimas dos décadas. La confluencia de nuevos algoritmos de aprendizaje automático, unidades de procesamiento gráfico rápidas y económicas y unidades de procesamiento tensorial, y la disponibilidad de conjuntos de datos masivos, a menudo continuos, han impulsado esta revolución en el análisis basado en datos. Esta rápida expansión ha visto la aplicación de herramientas ML nuevas y existentes a un conjunto de problemas geocientíficos que abarcan la detección de ondas sísmicas y la identificación y ubicación de fases, identificación de formaciones geológicas, alerta temprana de terremotos, algunos estudios de ellos solo se centran en el estudio de precursores; monitoreo de volcanes, radar de apertura sintética interferométrica de eliminación de ruido (InSAR) , imágenes tomográficas, caracterización de yacimientos, estimar la ubicación de la fuente y las dimensiones del epicentro de un terremoto, con ANN y GA respectivamente, descubrir patrones de réplicas, que son pequeños terremotos que siguen a un gran terremoto (conocido como sismo principal) y ocurren en la misma área; categorización de tweets

relacionados con terremotos utilizando SVM, determinar el mejor conjunto de indicadores de sismicidad para predecir terremotos y analizar la frecuencia de terremotos. [11][42-58][77-79] Cabe destacar que, durante los últimos 5 años, se ha dedicado un esfuerzo considerable a utilizar estos enfoques para caracterizar la física de fallas y pronosticar fallas. [59-66]

Sin embargo, a pesar de la indudable relevancia del problema de predicción de terremotos, sólo unos pocos autores han intentado sistematizar conocimientos de diversas fuentes. [75]

Se han observado algunas características generales para estos fenómenos precursoros. Primero, las anomalías precursoras de mayor amplitud tienden a ocurrir antes de los terremotos de mayor magnitud (por ejemplo, la brecha sísmica, es una ventana de tiempo en la que la actividad sísmica es extremadamente baja. Tal ventana a menudo se convierte en el asiento de un gran terremoto más tarde). [32]

Además, el número de anomalías precursoras tiende a aumentar cuanto más se acerca en el tiempo a la ocurrencia del terremoto. Finalmente, las anomalías precursoras tienden a ocurrir cerca del eventual epicentro del terremoto. [18]

Los modelos de predicción por lo general comparten esta estructura, mostrada en Fig. 1, las fases son la adquisición de datos; modelado de datos: preprocesamiento, extracción y normalización; entrenamiento y pruebas de los modelos; comparación y ajustes de los modelos.

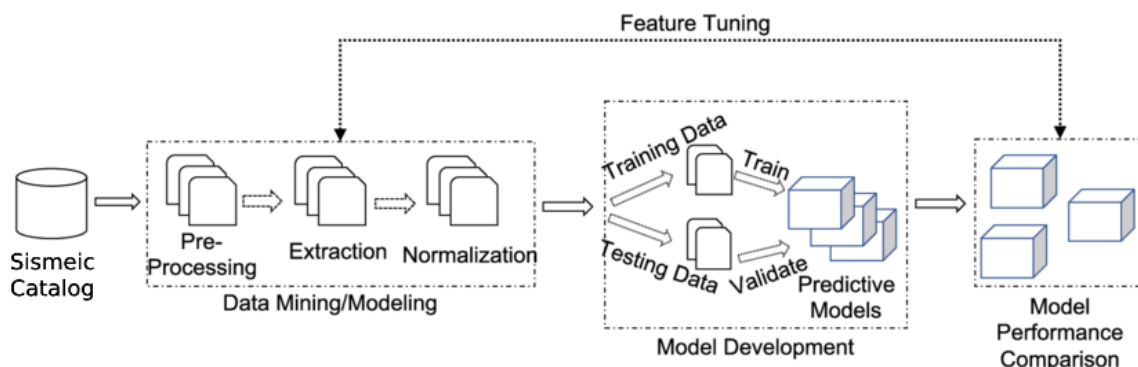


Fig. 1. Diagrama de la estructura general de los modelos de predicción.

En estos enfoques, la magnitud de un sismo se predice utilizando registros históricos obtenidos de catálogos sísmicos, a excepción de los estudios que utilizan indicadores como señales eléctricas sísmicas (SES). [35]

Estos hacen las modificaciones de limpieza y preprocesamiento necesarias al catálogo, después seleccionan los indicadores que serán utilizados y metodología para calcularlos, se calculan y se dividen en muestras de entrenamiento y de prueba, las muestras de entrenamiento se utilizan para entrenar el modelo predictivo, mientras que las muestras de prueba para calcular su desempeño en base a métricas seleccionadas, idealmente se desarrollan múltiples modelos y estos se comparan, se puede regresar a etapas anteriores para ajustar los parámetros y características de los modelos para mejorar su desempeño en la predicción.

Las técnicas de aprendizaje automático pueden predecir bien si se les proporciona un conjunto de datos adecuado. En algunos casos se puede observar una diferencia significativa al usar un catálogo de buena calidad y exhaustivo como el del sur de California en contraste con catálogos menos completos de otras regiones. [14]

Los estudios cubren el problema de diferentes maneras, con diferentes rangos de tiempo, rangos de magnitud, diferentes extensiones de las limitaciones geográficas, diferentes metas, y otros tipos de variedad. Por ello diferentes modelos y metodologías son usadas en los diferentes estudios.

A continuación se revisan varias publicaciones en las que se ha estudiado la aplicación de métodos de aprendizaje automático a la tarea de predicción de terremotos en varios intervalos temporales y espaciales. Debido a que, como se mencionó anteriormente, los procesos de ocurrencia de terremotos se consideran estocásticos y no lineales, las investigaciones más recientes en esta área están dedicadas a la aplicabilidad de las redes neuronales a este problema. También se revisan otras técnicas de aprendizaje automático, específicamente, varios algoritmos de regresión y clasificación.

1.1. Catálogo de estudios relacionados

En 2006 en que E. I. Alves utilizó algoritmos de inteligencia artificial, fue uno de los primeros en proponer redes neuronales (ANN) para la previsión de terremotos, se inspiró en la aplicación exitosa de enfoques similares a las tareas de previsión financiera, que, como él pensaba, son similares a la actividad sísmica en términos de naturaleza caótica de ambos sistemas. [30]

El pronóstico indicaba el tiempo y coordenadas geográficas de un terremoto dentro del espacio y ventanas temporales, así como el rango de magnitud. Sin embargo, no se calcularon medidas estadísticas, por lo que no se puede evaluar el desempeño de este enfoque de manera objetiva. A pesar de que las ventanas de tiempo eran demasiado amplias (el mes del evento sísmico se pronosticó dentro de ventanas de ± 5 meses), los resultados fueron "alentadores" y demuestran el potencial de usar redes neuronales para predecir terremotos. [25]

El siguiente año se publicaron dos estudios más en 2007 y 2009 respectivamente, por A. Panakkat y H. Adeli, en el primer estudio formularon el problema de predicción como una tarea de clasificación, este intentaba predecir la magnitud (en un rango de 0.5) del mayor evento sísmico en una ventana de tiempo y una región predefinida en el próximo mes. Propusieron ocho de los denominados **indicadores de sismicidad**: características calculadas matemáticamente, que se pueden utilizar para evaluar el potencial sísmico de una región. [20][16]

Y más tarde en el segundo estudio los mismos autores propusieron la arquitectura de una red neuronal probabilística (PNN) como una solución para el mismo problema donde utilizaron el mismo conjunto de indicadores de sismicidad como datos de entrada para el entrenamiento de la red. [34]

El modelo se probó con datos de la zona sísmica del sur de California y arrojó buenas precisiones de predicción para eventos de magnitud 4,5 a 6,0 (valores de puntuación R entre 0,62 y 0,78). Sin embargo, PNN no funcionó satisfactoriamente para terremotos de magnitudes superiores a 6.0, lo que arrojó puntuaciones R en el rango de 0.0 a 0.5, pero los autores propusieron el uso de otro modelo para abarcar este problema, se habla más tarde de éste. Estas investigaciones han sentado las bases de un enfoque científico para evaluar el riesgo sísmico potencial para diferentes regiones: el conjunto de ocho indicadores de sismicidad propuestos por ellos fue utilizado en varios estudios por investigadores de todo el mundo. La importancia de estos indicadores en el tema de predicción sísmica demanda que sean presentados, se pueden observar en Tabla 1.

Tabla 1. Parámetros de Panakkat y H. Adeli

Símbolo	Nombre	Expresión
ΔT	Tiempo transcurrido	$\Delta T = T_n - T_0$
M_{mean}	Promedio de magnitud	$M_{mean} = \frac{\sum_{i=1}^n (M_i)}{n}$
$dE^{1/2}$	La tasa de raíz cuadrada de la energía sísmica.	$dE \frac{1}{2} = \frac{\sum_{i=1}^n (10^{(11.8+1.5M_i)})^{1/2}}{\Delta T}$
β	valor b (pendiente de curva de Gutenberg-Richter)	$\beta = \frac{(n(\sum (M_i \log_{10} N_i) - \sum M_i \sum \log_{10} N_i))}{(\sum M_i)^2 - n \sum M_i^2}$
η	Desviación cuadrática media	$\eta = \frac{\sum_{i=1}^n (\log_{10} N_i - a - bM)^2}{n - 1}$ $b = \beta, a = \frac{\sum (\log_{10} N_i + b_{slq} M_i)}{n}$
ΔM	Deficit de magnitud	$\Delta M = M_{max} - M_{expected\ max}$ $M_{max} = \max(M)$ $M_{expected\ max} = \frac{a}{b}$
μ	Tiempo medio entre eventos	$\mu = \frac{\sum_{i=1}^n (\Delta T_i)}{n - 1}$
c	Coficiente de variación	$c = \sqrt{\frac{\sum_{i=1}^n (\Delta T_i - \mu)^2}{n - 1}}$

En 2011 se publicó Expert Systems and Applications, un estudio con el objetivo principal de evaluar la precisión de ANN para predecir la ocurrencia de terremotos utilizando datos de series de tiempo y señales eléctricas sísmicas (SES) en la región de Grecia. [35]

Más específicamente, el documento destaca dos áreas principales de investigación. El primer estudio de caso se refería a la predicción del evento sísmico más grande del día siguiente utilizando sólo datos de magnitud de terremotos de series de tiempo, y el segundo se refería al uso de las llamadas SES para predecir la magnitud del próximo evento sísmico así como el desfase temporal. Para el primer caso, se utilizó una red neuronal de retropropagación de retroalimentación. Un archivo de entrada contenía un valor de magnitud máximo para cada día. El modelo se entrenó utilizando un catálogo de terremotos para Grecia y el rendimiento se evaluó con una tasa de precisión, que se calculó con base en MAE. La tasa de

precisión promedio fue del 80,55% para todos los eventos, pero solo el 52,81% para lo que Moustra et al. considera "valores atípicos" (terremotos de magnitud superior a 5,2). El estudio del segundo caso que se refería a la predicción de terremotos utilizando SES constaba de dos partes principales. Cabe señalar que, en el momento del estudio, el equipo de VAN en Grecia solo registró y publicó 29 muestras de SES. A pesar de esto, los autores intentaron usar una ANN para estudiar la conexión entre SES y la ocurrencia de terremotos. Debido al hecho de que 29 muestras claramente no eran suficientes para entrenar redes neuronales, Moustra et al. había decidido construir los datos faltantes para el resto de eventos sísmicos del catálogo. En el primer caso, los SES se generaron aleatoriamente para todos los eventos; en el segundo, se utilizó la ANN para construir los datos faltantes utilizando series de tiempo de magnitud. La tasa de precisión de la predicción de magnitud fue un poco más del 60% en el primer conjunto de datos, y la ANN no encontró correlación entre el SES y el lapso de tiempo. El uso de datos construidos por la ANN mejoró significativamente el rendimiento: las tasas de precisión que resultaron de la predicción de la magnitud y el retraso fueron del 83,56% para la magnitud y del 92,96% para el retraso. Los resultados han llevado a los autores a la conclusión de que entrenar modelos con los datos adecuados es un factor clave que puede influir en gran medida en el rendimiento resultante. [25]

En 2012 se publicó un artículo por Alarifi et al. que usó ANN para la predicción de terremotos en la región del Mar Rojo, y los resultados mostraron la alta capacidad de pronóstico de ANN que cualquier otro modelo estadístico hasta la fecha. [80]

Esto se debe a que la red neuronal es capaz de capturar relaciones no lineales que los métodos estadísticos y otros métodos propuestos. Este estudio utilizó los siguientes indicadores: Número de secuencia del terremoto, se utiliza para seguir reservando el orden de los eventos sísmicos, y se utiliza en lugar del año, mes, día y hora; ubicación de cuadrícula, esta función identifica la ubicación en la que ocurrieron los eventos del terremoto, dividieron el área de estudio en mosaicos de 16×16 en lugar de latitudes y longitudes. De esta manera, cada ficha está representada por dos números con un rango de valor entre 0 y 15; magnitud del terremoto: La característica principal que debe considerarse ya que representa la fluctuación de la magnitud; y la profundidad de la fuente: la profundidad del evento, que se puede considerar como una característica. El objetivo de este estudio fue predecir la magnitud del próximo terremoto en lugar de la hora del terremoto, lo que significa que no es necesario mantener la fecha y la hora; el número de secuencia es suficiente para nuestro objetivo. Los autores explican que aunque la fecha y la hora pueden ayudar lógicamente a proporcionar una predicción más precisa, descubrieron después de muchos experimentos que la secuencia es más fácil de aprender para la configuración de la red neuronal que se propone.

En 2013 se publicó un artículo publicado en Applied Soft Computing en 2013, utilizó un modelo ANN para predecir la magnitud del terremoto durante los siguientes cinco días. [21]

La aplicación de pruebas y experimentos estadísticos mostró la mayor tasa de éxito de ese método que otros clasificadores de aprendizaje automático. [18]

El sistema está diseñado para proporcionar dos tipos de predicciones: a) la probabilidad de que ocurra un terremoto mayor que una magnitud umbral en cinco días y b) la probabilidad de que ocurra un evento sísmico dentro de un rango de magnitud predefinido. [25]

En este estudio se definieron nuevos parámetros sísmicos basados en la ley de Bath y la ley de Omori-Utsu , que describen las relaciones entre la conmovión principal y las réplicas, según su magnitud y frecuencia de ocurrencia, respectivamente. [73][74]

Se analizaron cuatro regiones sísmicas de Chile. Se aplicó una ANN diferente a cada área, aunque todas compartían la misma arquitectura. El prototipo predijo un terremoto cada vez que la probabilidad prevista fuera mayor que un valor umbral predefinido (los umbrales se ajustaron para reducir el número de falsas alarmas). La evaluación de los métodos propuestos se llevó a cabo utilizando medidas de rendimiento calculadas a partir de TP, TN, FP y FN. El análisis comparativo se realizó utilizando métodos estándar de clasificación como K vecinos más cercanos (KNN), máquinas de vectores de soporte (SVM) y clasificación mediante agrupación de K-medias. A pesar de la configuración individual de los parámetros, el desempeño de la ANN propuesta varió mucho según la región, desde 17.4% hasta 87%. [25]

En 2017 se publicó en Natural Hazards un artículo donde se estudia el problema de la predicción de terremotos como una tarea de clasificación binaria. [15]

Se hicieron predicciones para eventos de magnitud mayor o igual a 5,5 mensualmente. Se utilizaron ocho indicadores de sismicidad propuestos por Adeli & Panakkat como entrada para diferentes clasificadores de aprendizaje automático. [20]

Estos incluyeron la red neuronal recurrente (RNN), la red neuronal de reconocimiento de patrones (PRNN), el bosque aleatorio (RF) y el conjunto LPBoost de árboles de decisión. Además de la precisión de las predicciones, Asim et al. identificaron medidas de rendimiento como la sensibilidad y la especificidad, los valores predictivos verdaderos y falsos como los principales criterios para la comparación de los enfoques antes mencionados. Los clasificadores se utilizaron para predecir terremotos en la región de Hindukush. El conjunto LPBoost tendió a

tomar la delantera en precisión con un valor del 65%. Este clasificador también se desempeñó mejor en términos de sensibilidad a la ocurrencia de terremotos, arrojando 91% del valor S_n . Los autores también destacaron el resultado de PRNN, que produjo la menor cantidad de falsas alarmas como lo demuestra un alto nivel de valor predictivo positivo igual al 71%. Tras analizar los resultados, los autores afirmaron que todos los sistemas observados habían mostrado resultados satisfactorios de una forma u otra.

En 2018 el estudio realizado por G. Cortés et al, se publicó en Computers & Geosciences, se hizo un intento de predecir la magnitud del mayor evento sísmico en los próximos siete días. [69]

Está vez el problema de la predicción de terremotos se trató como una tarea de regresión: se aplicaron cuatro regresores: modelos lineales generalizados, máquinas impulsoras de gradiente, aprendizaje profundo y bosque aleatorio. Los indicadores de sismicidad propuestos por Panakkat & Adeli y Reyes et al. se utilizaron como datos de entrada. [20-21]

La característica principal del estudio es que el problema se observó en el contexto del análisis de big data: se utilizó un total de 1 GB de datos procesados mediante una información basada en la nube para entrenar y probar modelos de regresión. Para evaluar la eficacia de los enfoques propuestos, se utilizaron como medidas de rendimiento los errores medios absolutos (MAE) y relativos (RE). Además, debido a las particularidades de la tarea, también se tuvo en cuenta el tiempo dedicado a los modelos de entrenamiento. El regresor más eficaz fue el bosque aleatorio (RF), que arrojó un error absoluto medio de 0,74 en promedio. La RF también fue una de las más rápidas, ya que solo tardó 18 minutos en entrenar los modelos de regresión en todos los datos. En particular, las predicciones más precisas de RF se realizaron para terremotos moderados (magnitudes dentro de un rango en [4, 7); MAE \leq 0.26), mientras que los conjuntos de regresión se desempeñaron mejor en rangos de magnitud extrema ([0, 3) y [7, 8]). Con base en estos resultados, los autores concluyeron que el uso de conjuntos regresores más complejos mejoraría la precisión de las predicciones de terremotos de gran magnitud. [69][25]

En 2018 el sistema de predicción de terremotos (EPS) llamado EP-GP Boost se describió en el artículo que se publicó en Soil Dynamics and Earthquake Engineering en 2018. [14]

Este sistema es un clasificador basado en una combinación de programación genética (GP) y un algoritmo de impulso llamado AdaBoost . La aplicación de estos instrumentos al problema de la predicción de terremotos nunca se había estudiado antes de este artículo. Otra novedad del enfoque es una metodología de cálculo y uso simultáneo de indicadores de sismicidad, que se basa en la idea de obtener la

máxima información sobre las propiedades geológicas de las regiones observadas (en lugar de elegir los parámetros adecuados para cada zona de forma individual). Se calculó un total de 50 características, basadas en conceptos geológicos como la ley de Gutenberg-Richter, liberación de energía sísmica, frecuencia de premonición, etc. Algunos de estos parámetros se calcularon a través de diferentes enfoques (por ejemplo, el valor b mencionado anteriormente, que es una pendiente de una curva de Gutenberg Richter, se calculó utilizando dos métodos, a saber, el análisis de regresión de mínimos cuadrados y el método de máxima verosimilitud). Como resultado, se propuso un sistema para predecir eventos sísmicos de magnitud igual o superior a 5.0 para los próximos 15 días. El estudio de la aplicabilidad de EP-GP Boost se realizó utilizando datos de zonas sísmicas previamente utilizadas, a saber, Chile (32,5–36 S °, 70 –72,5 W °), Hindukush (35-39 N °, 69 –74,6 E °) y el sur de California (32 –36,5 N °, 114,75 –121 W °). Los experimentos han mostrado un desempeño sobresaliente en las tres regiones observadas, tanto en términos de un índice bajo de falsas alarmas (los valores de precisión fueron 74.3%, 80.2% y 84.2% para Hindukush, Chile Sur de California, respectivamente) y en términos de otras métricas consideradas para evaluación, como MCC y puntaje R. Los mejores resultados se obtuvieron para la región del sur de California (los autores afirmaron que la razón fue la calidad y exhaustividad del correspondiente catálogo de terremotos). Sin embargo, los resultados de todas las regiones muestran una mejora en comparación con los estudios anteriores. [16][18][23]

Otro artículo publicado en PLOS ONE en 2018 que fue escrito por Asim et al utilizó el enfoque para el uso de indicadores de sismicidad. [14][16]

Esta vez, se calcularon 60 parámetros sísmicos utilizando varios conceptos de sismología. Nuevamente, se calcularon algunas características específicas a través de diferentes enfoques para retener la información más completa sobre las zonas sísmicas observadas. Al igual que en su investigación anterior, los autores pretenden predecir los terremotos de magnitud igual o superior a 5,0 durante los próximos 15 días. El sistema propuesto es multipaso, a diferencia de otros predictores previos propuestos en la literatura que son principalmente simples. El sistema es una combinación de diferentes algoritmos de aprendizaje automático, y en cada paso, un algoritmo utiliza el conocimiento obtenido a través del aprendizaje de uno anterior. En primer lugar, la selección de características en dos pasos se utiliza para elegir los parámetros más relevantes para entrenar un modelo. Específicamente, se realizan verificaciones de relevancia y redundancia (se aplica el criterio de relevancia máxima de redundancia mínima, denotado como mRMR). El conjunto de parámetros resultante se pasa a un regresor de vector de soporte (SVR), y la tendencia predicha por SVR se usa luego como parte de los datos de entrada para una red neuronal híbrida (HNN). Una HNN propuesta aquí es una combinación de tres ANN diferentes y un algoritmo EPSO para la optimización del peso. El sistema resultante llamado SVRHNN se aplicó a regiones previamente estudiadas de Hindikush, Chile y el sur de California. El desempeño se evaluó con medidas como P_0 , P_1 , S_n , S_p , precisión, MCC y puntaje R. Los resultados también se compararon con los descritos en

investigaciones anteriores sobre estas zonas sísmicas. Los valores resultantes de las medidas de desempeño (por ejemplo, el puntaje R aumentó de 0.27 a 0.58 para Hindukush, de 0.344 a 0.603 para Chile, 0.623 de 0.5107 a 0.623 para el sur de California) mostraron que la metodología de múltiples pasos propuesta mejoró el desempeño de la predicción en comparación con técnicas individuales de aprendizaje automático. [25]

En 2021 M. Yousefzadeha, S. Ahmad Hosseini y MahdiFarnaghi muestran resultados en sus métricas entre .82 y .97 en diferentes métricas, resultados sin precedentes. [18]

Esta investigación propone dos modelos de predicción de terremotos utilizando indicadores sísmicos y técnicas híbridas de aprendizaje automático en la región del sur de California. Se calcularon matemática y estadísticamente siete indicadores sísmicos según los eventos sísmicos registrados previamente en el catálogo de terremotos de esa región. Estos indicadores son, a saber, los mismos propuestos en [20], a excepción del coeficiente de variación (c), en total son 7, una cantidad de indicadores significativamente menor que los demás estudios que lo preceden. Se proponen dos modelos híbridos de aprendizaje automático para predecir la magnitud del terremoto durante quince días. El primer modelo es FPA-ELM, que es un híbrido del algoritmo de polinización de flores (FPA) y la máquina de aprendizaje extremo (ELM). El segundo es FPA-LS-SVM, que es un híbrido de FPA y la máquina de vectores de soporte de mínimos cuadrados (LS-SVM). El rendimiento de estos dos modelos se compara y evalúa mediante cuatro criterios de evaluación: error cuadrático medio (RMSE), error absoluto medio (MAE), error porcentual absoluto medio simétrico (SMAPE) y error relativo medio porcentual (PMRE). Los resultados de la simulación mostraron que el modelo FPA-LS-SVM superó a los modelos FPA-ELM, LS-SVM y ELM en términos de precisión de predicción.

En México existen propuestas de proyectos de predicción sísmica como la del centro de investigación y de estudios avanzados del Instituto Politécnico Nacional (IPN), en la que planean usar algoritmos de random forest y deep learning. [81] No se encontraron otras publicaciones de predicción con el uso de ML específicas para la región mexicana.

1.2. Análisis de los estudios relacionados

A continuación en la Tabla 2 se muestra una comparación de los artículos revisados antes, y después se describen tendencias encontradas en los estudios.

Tabla 2. Indicadores del modelo

Ref.	Región	Datos	Modelos	Indicadores	Métricas	Rango	Extensión geométrica	Rango de magnitud
2006 [30]	Azores, Portugal	-	ANN	Financial oscillators.	-	5 meses	La latitud de un epicentro se predice dentro de 1°	-
2007 [20], 2009 [34]	South California, San-Francisco bay	Catálogo de SCEC	LMBP; RNN; RBFN	Indicadores sísmicos basados en la ley de G.R. y distribución de características.	S_n , FAR, FB, R score	1 mes	South California (32-36 N°, 114-120 W°) SanFrancisco bay (37.5-40 N°, 116-123.5 W°)	La magnitud se predice a la 0.5 más cercana.
2011 [35]	Grecia	SINOA, VAN team	ANN	Dos enfoques para modelar SES para todos los datos sísmicos.	MAE	1 día	La región predefinida de Grecia	-
2012 [80]	Sinai Peninsula, the Gulf of Aqaba, and the Gulf of Suez.	NCEDC	ANN, RNN	ANN	MAE, MSE	-	Latitud entre 28 y 32 y longitud entre 32 y 36	-
2013 [21]	Chile	Catálogo del Servicio Nacional de Sismología de la Universidad de Chile	ANN	Indicadores de [20] y nuevos basados en la ley de Bath y la ley de Omori-Utsu.	P_0, P_1, S_n, S_p	5 días	Los rangos de coordenadas varían de $0.5^\circ \times 0.5^\circ$ a $1^\circ \times 1^\circ$	Se predice la probabilidad de que la magnitud supere el valor umbral o esté en un rango predefinido
2017 [15]	Hindu Kush, Pakistan	Center for Earthquake Studies (Pakistan), USGS	LMBP-RNN; PRNN; RF; LPBoost	[20]	P_0, P_1, S_n, S_p , accuracy	1 mes	La región predefinida de Hindu Kush	Se predice la ocurrencia de eventos de magnitud ≥ 5.5
2018 [69]	California, USA	ANSS Composite Earthquake Catalog, NCEDS	GLM; GBM; DL	[20][21]	MAE, RE	7 días	Las coordenadas de un epicentro se predicen dentro de una celda de $0.5^\circ \times 0.5^\circ$	-
2018 [14]	Hindikush (Pakistan), Chile, South California (USA)	Catálogo USGS	GP, AdaBoost	[20][21], más variaciones de estos y nuevos, 50 indicadores sísmicos utilizados.	P_0, P_1, S_n, S_p , accuracy, MCC, R score	15 días	Southern California (32-36.5 N°, 114.75- 121 W°), Chile (32.5- 36 S°, 70-72.5 W°), Hindukush (35-39 N°, 69-74.6 E°)	Se predice la ocurrencia de eventos de magnitud ≥ 5.5
2018 [16]	Hindikush (Pakistan), Chile, South California (USA)	[14]	mRMR, SVR, HNN, EPSO	[20][21][14] y nuevos, 60 indicadores sísmicos utilizados.	[14]	[14]	[14]	[14]
2021 [18]	South California (USA)	Catálogo SCEC	ELM FPA-ELM, LS-SVM, FPA-LS-SVM	Los indicadores de [20] menos un indicador.	RMSE, MAE, SMAP E, PMRE	15 días	La región predefinida del sur de California.	-

Los desempeños de los estudios no son directamente comparables debido a la variedad de sus enfoques y objetivos. Por eso no se puede afirmar objetivamente que un modelo es mejor que el otro. [25]

Se pueden comparar directamente cuando utilizan las mismas métricas para medir su desempeño, el estudio es sobre la misma región, y tienen objetivos de tiempo y magnitud comparables. [14][16]

Una solución para este problema proponen crear un conjunto de datos que funcionen como “marco de referencia”, que los investigadores pueden utilizar con fines comparativos para diferentes algoritmos. [25]

El conjunto de datos puede contener datos de fuente abierta sobre zonas sísmicas utilizadas en estudios anteriores, como Chile, Hindukush y el sur de California, que han sido utilizados en distintas ocasiones. Para ser precisos, la definición de predicción de un terremoto dada por los sismólogos implica dar la definición exacta del tiempo y lugar de ocurrencia del terremoto, así como su magnitud. [25]

Sin embargo, la mayoría de los estudios observados se centran en un objetivo más amplio de predecir la magnitud para un área limitada y un rango temporal como se muestra en Tabla 2, tal vez sea más adecuado llamarlos pronóstico de terremotos para apegarnos mejor a las definiciones pero por lo general los investigadores en la literatura no son muy estrictos en hacer esta distinción. Los investigadores aún no pueden dar una predicción (en su manera estricta) por la extrema complejidad del proceso de ocurrencia de un terremoto, también se propone que el problema de predicción debe ser tratado intentando predecir simultáneamente la magnitud, el tiempo y el lugar de la ocurrencia de eventos sísmicos. [25]

Vale la pena señalar que en algunos trabajos se observa una tendencia a la disminución de la precisión con el aumento del umbral de magnitud. Esto es porque cuanto mayor sea el terremoto, más difícil será predecirlo. Dado que los grandes terremotos representan la mayor amenaza para la sociedad, es necesario realizar mayores esfuerzos en la tarea de predecir terremotos de gran magnitud (igual o superior a 5.5). [25]

Debido a esto autores han propuesto el uso de diferentes modelos y metodologías para magnitudes moderadas y magnitudes extremas o “valores atípicos” (adecuadamente llamados así debido a la baja frecuencia de estos eventos). [20][25][34-35][28]

Algunos autores proponen diferentes modelos de redes neuronales, estos proponen usar una RNN para predecir terremotos de gran magnitud y utilizar PNN para terremotos pequeños y moderados. [20][34]

Mientras que otros autores como proponen una metodología para este problema, para mejorar el rendimiento en terremotos atípicos, los autores entrenaron a la ANN en dos fases (primero en valores atípicos, luego en todo el conjunto de datos de entrenamiento), y la tasa de precisión mejoró de 52,81% a 58.02%. [35]

Los autores también remarcan la importancia de los datos de entrada de los modelos, hay dos estudios de caso que finalmente llevaron a la conclusión de que cuando se presentan los datos apropiados a NN, se puede predecir con mayor precisión. [35][18]

Mientras Maceda y col. aplicaron SVM al problema del terremoto y decidieron que SVM es perfecto para resolver problemas de clasificación utilizando un conjunto de datos de entrenamiento de pequeño tamaño. [77]

Existe un riesgo en tener realizar predicciones incorrectas o falsas alarmas, una alerta incorrecta provocará pérdidas innecesarias. Debido a esto los investigadores describen la baja generación de falsas alarmas como un criterio importante de evaluación del desempeño de un modelo de predicción de terremotos. Numerosos autores señalan que la predicción de terremotos es un tema delicado donde las falsas alarmas conllevan consecuencias particularmente negativas, como pérdidas económicas y pánico entre la población civil, que pueden ser críticas porque pueden generar desconfianza en el sistema. Por tanto, en algunos casos incluso podemos sacrificar la sensibilidad de un modelo a favor de reducir un número de falsas alarmas. [25]

Martínez Álvarez et al. utilizó varios precursores de sismicidad como datos de entradas para ANN en diferentes zonas sísmicas, y demostró que los indicadores sísmicos son las mejores características para predicción de terremotos, este estudio fue realizado en 2013. [79]

Prácticamente todos los estudios recientes utilizan indicadores sísmicos, y existe una tendencia de que el número de indicadores sísmicos utilizados crezcan con el tiempo, con excepción de un artículo, que logró alcanzar buenos resultados a pesar del número limitado de sus indicadores. [18]

Es necesario saber que por la naturaleza caótica de los eventos sísmicos y la complejidad de predecir eventos futuros, realmente no puede haber un modelo perfecto que resulte en el 100% de los eventos correctamente predichos, pero se hacen esfuerzos para mejorar la precisión tanto como sea posible, es importante considerar esto, para evitar casos controversiales como el del terremoto de 6 de abril de 2009 cerca de la ciudad italiana de L'Aquila del que se habló previamente. [27]

Actualmente no existe una metodología general para la predicción de terremotos. Además, todavía no existe consenso en la comunidad científica sobre si es posible encontrar una solución a este problema. Sin embargo, el rápido desarrollo de métodos de aprendizaje automático y la aplicación exitosa de estos métodos a varios tipos de problemas indican que estas tecnologías podrían ayudar a extraer patrones ocultos y hacer predicciones precisas. [25]

Los autores se centran especialmente en la magnitud, sin definir un intervalo de tiempo claro de ocurrencia. Las magnitudes predichas no son particularmente grandes, por lo tanto, no se hace ninguna distinción entre terremotos moderados y grandes en la predicción. El análisis espacial simplemente se evita en la mayoría de los casos y los autores usan, quizás, zonas demasiado amplias para reclamar una predicción de ubicación precisa. Finalmente, no se informa ninguna validación estadística ni comparación con métodos que se comportan intrínsecamente de manera diferente. [75]

No se espera que la predicción de eventos generalizada y efectiva sea posible a corto plazo, se necesita más investigación para desarrollar modelos ML que sean generalizables a distintos eventos sísmicos y entornos geológicos. Ciertamente, no podemos detener la ocurrencia de terremotos, los terremotos seguirán siendo desastres naturales destructivos. Pero el objetivo del pronóstico de terremotos es predecir terremotos repentinos y catastróficos que representan un peligro para la vida y la economía., pero un pronóstico exitoso, es decir que sea oportuno, confiable, que determine la ubicación geográfica, el tiempo y la magnitud de un terremoto, puede brindar la posibilidad de considerar medidas preventivas para mitigar los efectos devastadores de terremotos y aumentar el nivel de preparación pública.

La avenida de mejores y nuevos algoritmos y metodologías junto con la mejora constante de los de las redes sísmicas ha permitido una gran cantidad de datos y herramientas para manejarlos y, por lo tanto, ha renovado la esperanza de que la predicción de terremotos podría ser una tarea factible, si se combina con la mejora de los métodos de predicción. [81]

En este estudio se busca explorar la aplicación de algunas de estas técnicas de predicción al panorama mexicano utilizando algoritmos de clasificación con aprendizaje automático, para esto se calcularán distintos indicadores sísmicas del catálogo sísmico mexicano mediante el empleo de conceptos sismológicos, como la ley de Gutenberg-Richter, la tasa sísmica cambios, frecuencia de anticipación, liberación de energía sísmica, tiempo total de recurrencia y estos indicadores se introducirán a diferentes modelos de predicción de terremotos que se beneficiarán de la creciente calidad y cantidad de datos de este campo.

2. Perspectiva General de México

En el contexto social, la población mexicana ha enfrentado momentos desfavorables que le han mostrado que hay aprendizajes que son más lentos que otros, desde problemas de urbanismo, cambios ecológicos, políticos e inclusive adversidades económicas: recuperarse de desastres naturales, lamentablemente, no es una excepción. [68]

El primero de abril de 1904 se reunieron en Francia, dieciocho países, entre ellos México, con el fin de crear la Asociación Sismológica Internacional y mejorar la instrumentación sísmica a nivel mundial. Para cumplir con los compromisos adquiridos en esa reunión, el gobierno mexicano inició la instalación de una red sismológica en el territorio nacional. El 5 de septiembre de 1910 se inauguró la primera estación sismológica de la red en el Distrito Federal, ahora Ciudad de México. En ese momento, la red sismológica quedó a cargo del Instituto Geológico Nacional dependiente de la Secretaría de Minería y Fomento. Para 1923 se habían instalado otras 8 estaciones sismológicas, poco después en 1929 el Sistema Sismológico Nacional (SSN) pasó a ser parte de la Universidad Nacional Autónoma de México (UNAM). [41]

El SSN registra, almacena y distribuye datos del movimiento del terreno para informar sobre la sismicidad del país a las autoridades y a la población en general, promueve el intercambio de datos y coopera con otras instituciones de monitoreo e investigación a nivel nacional e internacional. Está integrada por 61 Observatorios Sismológicos que cuentan, cada uno, con un sismómetro, un acelerómetro y un digitalizador que permite almacenar de manera local la información generada y enviarla a través de distintos medios. De igual forma, cuenta con 43 sistemas GNSS para el monitoreo del desplazamiento. Los observatorios se localizan primordialmente en las costas del océano Pacífico, golfo de México y eje neovolcánico.

El catálogo sísmico del SSN en el presente contiene más de 200 mil eventos sísmicos que datan desde enero de 1900. Si bien el Servicio Sismológico Nacional comenzó a operar desde 1910, los registros sísmicos de las primeras décadas del siglo XX no son muy confiables debido a la pequeña cantidad de estaciones de registro y al mal funcionamiento de las mismas.

Fue hasta 1963 cuando se logró un mejor registro de la sismicidad mexicana, a partir de 1992 se inició la modernización de la Red Sismológica Nacional con la instalación de equipos de nueva tecnología que dio inicio a la *red de Observatorios Sismológicos de Banda Ancha*, la **red sismológica** que existe ahora en México que registra los eventos sísmicos en tiempo real desde diferentes puntos de la República Mexicana, que fue otro gran paso en mejorar el registro sismológico mexicano. [10][41]

Presumiblemente la completitud y calidad de los catálogos mejora cada década debido al incremento de esfuerzos y a la llegada de nuevas tecnologías, la completitud de un catálogo se define como la magnitud más baja en la cual todos los sismos son detectados exitosamente dentro de una región y periodo de tiempo, es esencial conocer la completitud de un catálogo para analizar la sismicidad de una región, un catálogo más completo significa un mayor desempeño en los métodos de análisis basados en datos como lo son los algoritmos de ML, un catálogo más completo es aquel que tiene una magnitud de completitud más baja, un catálogo con una magnitud de completitud más baja por lo general contendrá muchos más eventos sísmicos informados. [5][6]

La magnitud de completitud varía en el tiempo y el espacio y depende de muchos factores que afectan la capacidad de detección de una red sismológica, como son: la densidad y distribución de estaciones sísmicas, el tipo de instrumentación usada, la eficiencia del envío de datos desde las estaciones al centro de procesamiento, las prácticas y procedimientos de detección de sismos, entre otros factores [6-9].

Para satisfacer la completitud de una región se delimita la extensión geométrica de una región, se define un umbral temporal, y una magnitud de corte m^* , los eventos que sucedieron debajo de los umbrales y fuera de los límites de la región no son usados para el análisis, si el catálogo contiene todos los eventos sísmicos que sucedieron en una región, de magnitudes mayores o igual a m^* a partir de una fecha determinada entonces se puede decir que el catálogo de una región y un tiempo en particular está completo para magnitudes mayores a m^* .

Existen diferentes métodos para determinar la magnitud de corte, por ejemplo el *análisis de la curva de Gutenberg-Richter*, en este análisis se cuenta el número de eventos que suceden de cada magnitud y se selecciona el punto donde la curva se desvía del comportamiento exponencial, este se denomina como magnitud de corte m^* y solo se consideran los eventos mayores o iguales a esta magnitud para el análisis. [14]

Al aplicar este análisis en el catálogo de México para eventos a partir de 1974 la magnitud de corte resultante es $m^* \geq 3.9$. El catálogo sísmico de México está probablemente completo para $m^* \geq 7.7$ a partir de 1846, $m^* \geq 7.0$ a partir de 186 y $m^* \geq 4.3$ a partir de 1969. [10]

La calidad del catálogo incrementa drásticamente a partir de 1974, como se muestra en Fig.2 y la completitud se satisface con $m^* \geq 4.1$, como se muestra en Fig.3. Ésta es la fecha mínima que usaremos para analizar el catálogo sísmico de México. La ley de Gutenberg-Richter, describe la distribución de magnitud fractal y la ley de potencia de los terremotos en una región e intervalo de tiempo definidos. La ley de Gutenberg-Richter a menudo se conoce como una de las tres leyes básicas

de la sismología establecidas empíricamente; las otras dos son las leyes de Omori y Båth , la ley se describe con la siguiente expresión. [82]

$$\log_{10}N = a - bM$$

Aquí a representa la tasa de sismicidad total de la región, y b es la distribución relativa del tamaño de los terremotos y N es el número de terremotos. [12] Valores de b altos significan un predominio de pequeños terremotos; a la inversa, un valor b bajo significa que los terremotos grandes dominan sobre los terremotos más pequeños.

El parámetro b (comúnmente denominado "valor b ") suele estar cerca de 1,0 en regiones sísmicamente activas. Esto significa que para una frecuencia dada de eventos de magnitud 4.0 o más grandes, habrá 10 veces más terremotos de magnitud 3.0 o más grandes y 100 veces más terremotos de magnitud 2.0 o más grandes. Existe alguna variación de los valores b en el rango aproximado de 0,5 a 2, según el entorno de origen de la región. Un ejemplo notable de esto es durante los enjambres de terremotos cuando b puede llegar a ser tan alto como 2.5, lo que indica una proporción muy alta de terremotos pequeños a grandes.

Existe una disminución aparente del valor b para rangos de eventos de menor magnitud en todos los catálogos de terremotos, esto se puede observar en la Fig. 3. Este efecto se describe como "caída" del valor b , una descripción debido a que el gráfico de la versión logarítmica de la ley GR se vuelve más plano en el extremo de baja magnitud del gráfico. Esto puede deberse en gran parte a que cualquier conjunto de datos no esté completo debido a la incapacidad de detectar y caracterizar eventos pequeños. Es decir, muchos terremotos de baja magnitud no están catalogados porque menos estaciones los detectan y registran debido a la disminución de los niveles de señal instrumental a ruido. Por esto se considera la magnitud de corte, en la Fig. 3 la magnitud de corte es de 3.9, las magnitudes de los eventos sísmicos menores a la magnitud de corte no son tomados en cuenta para calcular a y b .

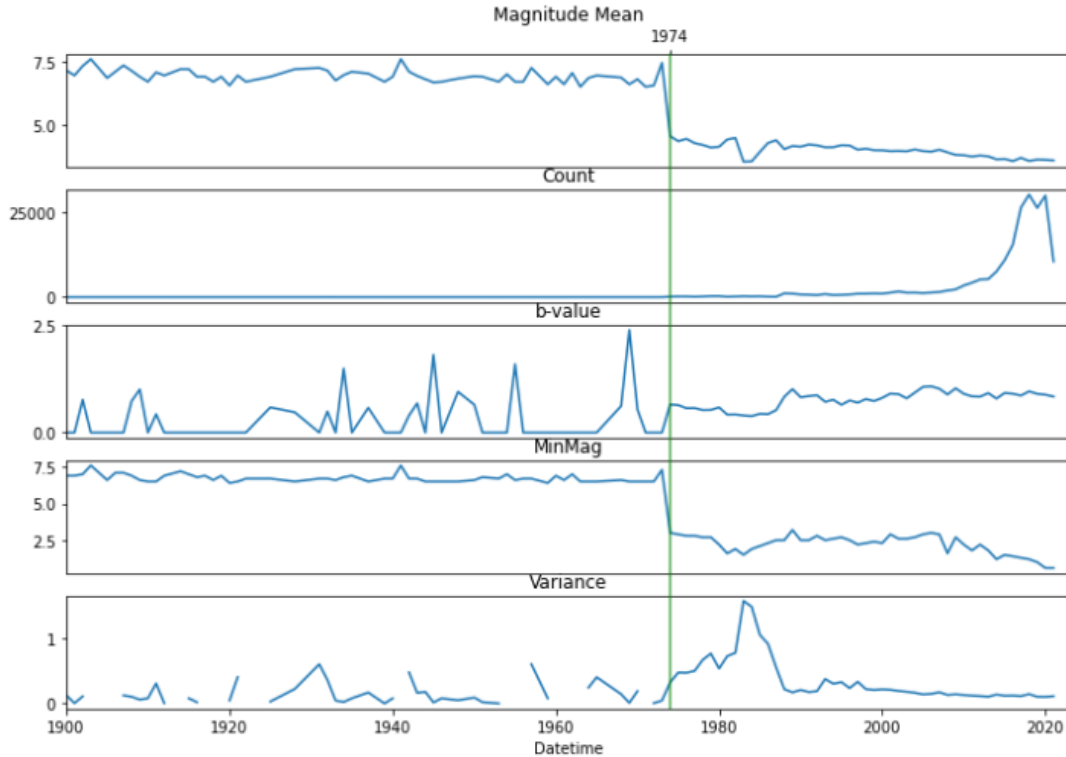


Fig. 2. Perspectiva de catálogo en México, antes de 1974 el catálogo es incompleto para magnitudes ≥ 3.8 , las brechas en la varianza, los ceros y valores atípicos en “b-value” y la varianza (variance) son causados por la falta de datos, la gráfica varianza está en función a la varianza de las magnitudes del año, y la gráfica “b-value” describe la frecuencia de la distribución del tamaño del terremoto con base en la ley de Gutenberg-Richter.

Las variaciones en los valores b tanto espacial como temporalmente se consideran generalmente como pistas para los precursores de grandes terremotos. [13]

Para determinar la magnitud de corte en este estudio se utilizó “b-slope analysis”, se cuentan el número de eventos para cada magnitud y se selecciona el punto donde la curva se desvía del comportamiento exponencial, este se denomina como magnitud de corte m^* y solo se consideran los eventos mayores o iguales a esta magnitud para el análisis. Al aplicar este análisis en el catálogo de México para eventos a partir de 1974 la magnitud de corte obtenida es $m^* \geq 3.9$, se puede decir que el catálogo mexicano está completo desde 1974 a partir de la magnitud $m^* \geq 3.9$.

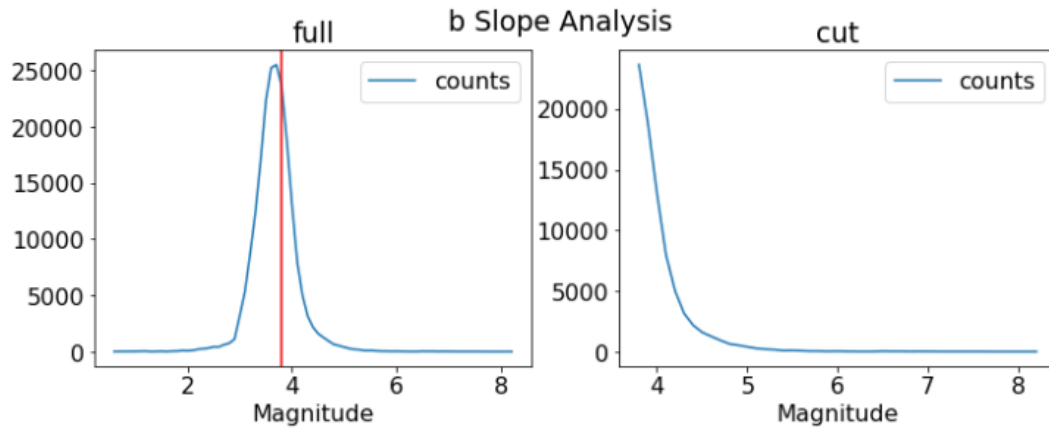


Fig. 3. Ley de Gutenberg-Richter aplicada en México, a partir de 1974 y para $m^* \geq 3.8$ con $a=8.326$ y $b=0.998$, el eje y contiene el número de eventos, y el eje x la magnitud de los eventos.

Al aplicar la ley de Gutenberg-Richter en el catálogo de México para sismos $m^* \geq 3.9$ a partir de 1974, resultó un valor para $a=8.326$ y $b=0.998$, se visualiza en Fig. 4.

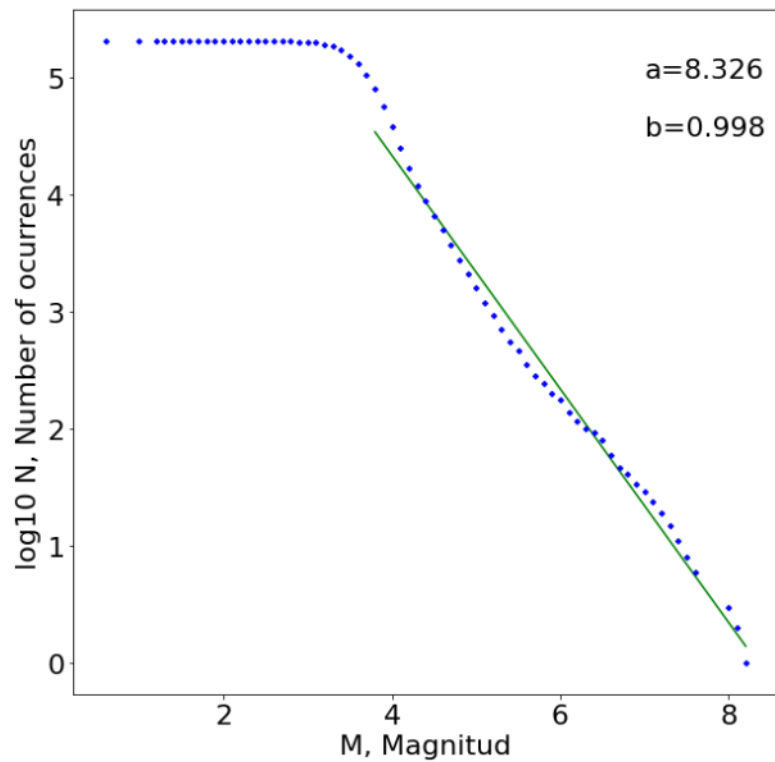


Fig. 4. Ley de Gutenberg-Richter aplicada en México, a partir de 1974 y para $m^* \geq 3.8$ con $a=8.326$ y $b=0.998$, se calcula usando el método de “linear least square regression” (lsq).

3. Preparación de Catálogo de Terremotos

En cuanto a la disponibilidad de datos, aquí se declara que todos los datos de este estudio son obtenidos de los catálogos de terremotos que fueron descargados del SSN (Servicio Nacional Sismológico), Los resultados se pueden reproducir utilizando el catálogo disponible públicamente. [41]

Existen diferentes métodos para hacer pronósticos, hay investigadores que estudian el comportamiento estadístico de ciertos parámetros físicos que podrían estar relacionados con el proceso de acumulación de estrés en la corteza terrestre. También se han aplicado el análisis fractal o análisis no extensivo de terremotos antes de la ocurrencia de grandes eventos como metodologías recientes para estudiar este complejo fenómeno. Y otros métodos que estudian las estructuras de los catálogos sísmicos tratando de encontrar patrones de sismicidad que se manifiestan antes de la ocurrencia de grandes terremotos. [10]

En este estudio se utilizarán solo los precursores de sismicidad, son indicadores que se obtienen al realizar ciertas ecuaciones sobre el catálogo sísmico de una región, los eventos sísmicos requeridos para calcular los indicadores están bien cubiertos por la instrumentación sísmica de superficie, son compilados y accesibles para todo el público en catálogos online, cada catálogo contiene las características de cada registro (evento sísmicos) de los que fueron informados, pueden contener más información pero aquí solo nos concierne la fecha, la magnitud y la ubicación geográfica de cada registro. Los indicadores dan información sobre el estado sísmico actual sobre el suelo y tienen comportamientos y relaciones complejas y no lineales, que pueden ser usados para entrenar algoritmos de aprendizaje automático para realizar predicciones. En este trabajo de tesis se seleccionaron los indicadores disponibles en la literatura contemporánea que contribuyan a mejorar el desempeño de las predicciones de terremotos.

Varios investigadores han considerado utilizar algoritmos de ML (Machine Learning) para aumentar la precisión de las predicciones, estos modelos utilizan precursores sísmicos, que son los eventos que regularmente preceden un terremoto. La comunidad científica ha estudiado muchos tipos diferentes de precursores de terremotos: campos eléctricos y magnéticos, emisiones de gas, cambios en el nivel del agua subterránea, emisiones de fluidos o gases, cambios de temperatura, deformaciones de la superficie, movimientos del suelo, comportamiento animal, sismicidad, etc. [2][4]

No se pueden detener los terremotos pero se pueden adoptar medidas de protección y precauciones para minimizar los efectos nocivos al predecir la magnitud del terremoto utilizando técnicas de aprendizaje automático.

Estos algoritmos ML utilizan los indicadores calculados a partir de los catálogos sísmicos tratando de encontrar y aprender los patrones de sismicidad que se manifiestan antes de la ocurrencia de grandes terremotos para realizar futuras predicciones. [10]

Estos se benefician recientemente gracias a la creciente disponibilidad de grandes cantidades de eventos sísmicos informados de mejor precisión y resolución en los catálogos sísmicos, que permitieron la exploración del uso de enfoques de aprendizaje automático (ML) para encontrar tendencias y patrones para generar predicciones que pueden reducir los daños inducidos por los terremotos.

Para manejar, limpiar (por ejemplo, deshacernos de instancias con datos faltantes), preprocesar (como normalizar los datos) y procesar información se utilizaron librerías que hacen uso de técnicas y algoritmos estándares y bien conocidas. Se seleccionaron los indicadores disponibles en la literatura contemporánea que contribuyan a mejorar el desempeño de las predicciones de terremotos. Se construirán diferentes modelos, y se seguirá el proceso descrito en la metodología para entrenar los modelos y se utilizarán algoritmos postulados en la literatura, estos modelos serán optimizados, al final se comparan el desempeño de estos utilizando las mismas métricas que estudios similares de clasificación para tener un marco de comparación que sirva como marco de referencia para futuros estudios.

En este estudio nos enfocaremos en el último, el análisis de los catálogos sísmicos para encontrar patrones de sismicidad a través de indicadores sísmicos, esto se realizará con el apoyo de algoritmos ML, hay miles de algoritmos, pero ninguno siempre es adecuado para todos los problemas porque hay muchos factores que lo afectan, como la cantidad de indicadores de entrada, la cantidad de registros del conjunto de datos y el tipo de problema (clasificación o regresión), por lo que se aplicaran diferentes algoritmos de aprendizaje automático con diferentes parámetros, después se compararán sus resultados entre sí para determinar el algoritmo más adecuado para este problema.

El desempeño de cada algoritmo será calculado con métricas estándares del campo para poder comparar los resultados de los modelos con los de otros investigadores, idealmente en el mismo conjunto de datos y el mismo alcance de la predicción. [31]

Los datos sin procesar tenían algunos problemas. Había columnas con información que no necesitamos, y otras columnas redundantes, descartamos estas columnas y eliminamos eventos con información incompleta, y nos quedamos con 204806 eventos sísmicos y con las columnas Fecha, Magnitud, Latitud y Longitud y Profundidad, aunque la columna de profundidad no se usó en este análisis pero es posible integrarlo en futuros análisis.

Se dividió el conjunto de datos en conjuntos de prueba, validación y entrenamiento. El conjunto de validación se utiliza durante el ajuste del modelo para evaluar la pérdida y las métricas, sin embargo, el modelo no se ajusta a estos datos.

El conjunto de prueba no se usa durante la fase de entrenamiento y solo se utiliza al final para evaluar qué tan bien se generaliza el modelo a nuevos datos. Esto es especialmente importante con conjuntos de datos desequilibrados donde el sobreajuste es una preocupación importante por la falta de datos de entrenamiento.

3.1. Extensión Geométrica del Catálogo

Los terremotos son una de las mayores catástrofes y su imprevisibilidad causa aún más destrucción en términos de vidas humanas y pérdidas financieras.

Debido al alto potencial de destrucción y muerte, así como a los efectos directos e indirectos de los terremotos, estos ocupan un gran espacio de interés de los científicos, por lo que los investigadores han estado trabajando vigorosamente en este dominio con la idea de proponer diferentes enfoques para la predicción de terremotos. [1]

Se hace el mejor esfuerzo para predecir dónde y cuándo ocurre el terremoto en función de los indicadores sísmicos y otras señales eléctricas sísmicas utilizando técnicas de aprendizaje automático y algoritmos de optimización; esto para alertar a las personas a tiempo porque una alerta incorrecta provoca pérdidas innecesarias.

Si bien no se pueden detener los terremotos, al menos una predicción efectiva nos puede alertar para que se adopten medidas de protección y precauciones para minimizar los efectos nocivos.

La extensión geométrica del catálogo completo es de latitud 33.5 a 10.3 y longitud -120.5 a -85.5, es decir todos los eventos de este catálogo se encuentran dentro de estos límites, éste abarca todo el territorio mexicano y, los eventos se muestran en la Fig. 5, que es demasiado extenso en comparación con otras regiones de análisis en la literatura relacionada como. [14][15][16]



Fig. 5. Visualización geométrica de terremotos sobre México, el tamaño de los eventos están en función a su magnitud.

Para seleccionar una región más pequeña se separó el catálogo en dos grupos para reducir el tamaño de la región de análisis usando el algoritmo de KNN (un algoritmo de clasificación que considera los datos vecinos más cercanos) en función a su geometría en Fig. 6, aun en este caso las regiones obtenidas son demasiado grandes en comparación con la de otros estudios, por esto se usó KNN con 3 grupos para seleccionar regiones más pequeñas Fig. 7. Usar KNN y no seleccionar una región de manera completamente arbitraria aumenta la correlación de los datos (al tratarse de datos que ocurren en una misma región); Se podrían usar los datos de cada grupo pero se decidió delimitar una región geográfica que abarque aproximadamente los grupos para formalizar el espacio de la región de análisis, esto resulta en un área de análisis más consistente (con datos menos separados), facilita réplicas del análisis y comparación con otros análisis.

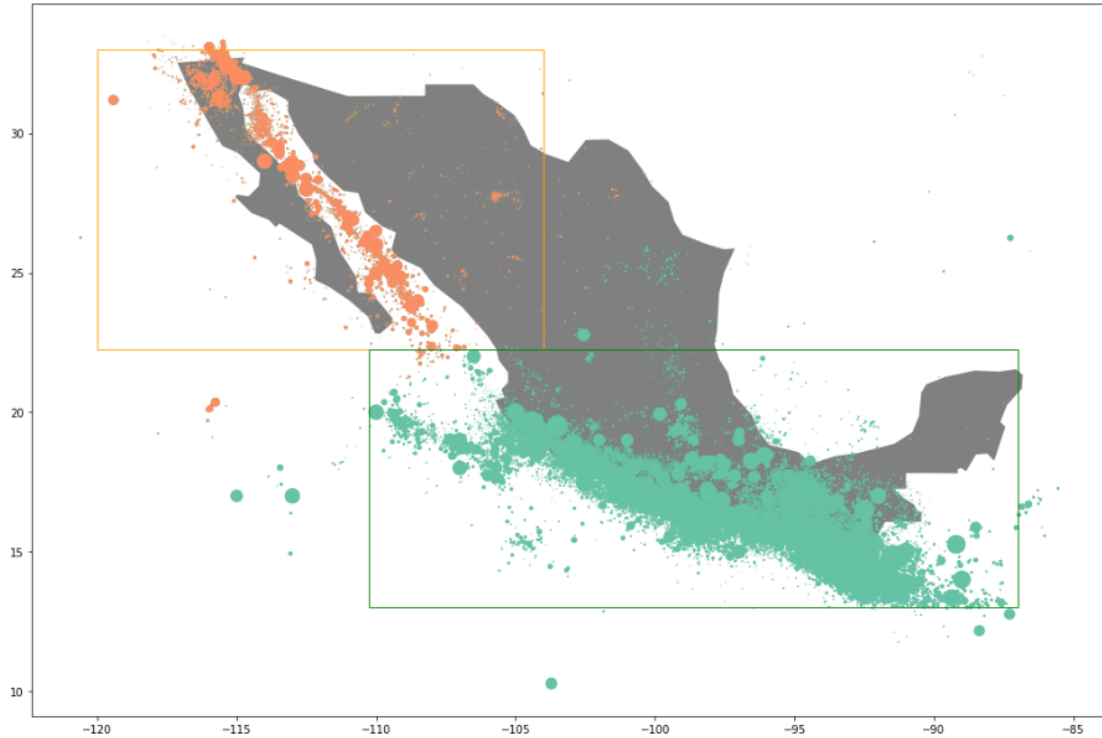


Fig. 6. División de catálogo por KNN en función a su geometría, y delimitación de región sur.

La división entre región noroeste y región sur resultantes son coherentes con las colindancias de las placas tectónicas que interactúan en México, ilustrado en Fig. 9 y también con las regiones sísmicas Fig. 8.

En la Fig. 6 la región noroeste se representa con color naranja, esta abarca el pacífico Norte, Golfo de California y Noroeste de México, desde longitud 120 hasta -104 y desde latitud 33.25 hasta 22.25, contiene 12184 eventos donde los valores $a=7.313$ y $b=0.988$ (en el capítulo 2 se explica el significado de estos valores), y las placas colindantes son Placa de Norteamérica y Placa del pacífico. La región sur se representa en la Fig. 6 con color verde, abarca el pacífico sur tropical y el occidente, centro y sureste de México, desde longitud -110.25 hasta -87.5 y desde latitud 22.25 hasta 13, contiene 192622 eventos donde $a = 8.221$ y $b = 0.987$ y las placas colindantes son Placa de Norteamérica, Placa de Rivera, Placa de Cocos y Placa del Caribe. Se realizarán más análisis sobre estas regiones pero, aunque los datos de estas regiones no son sobre las cuales los modelo hacen las predicciones.

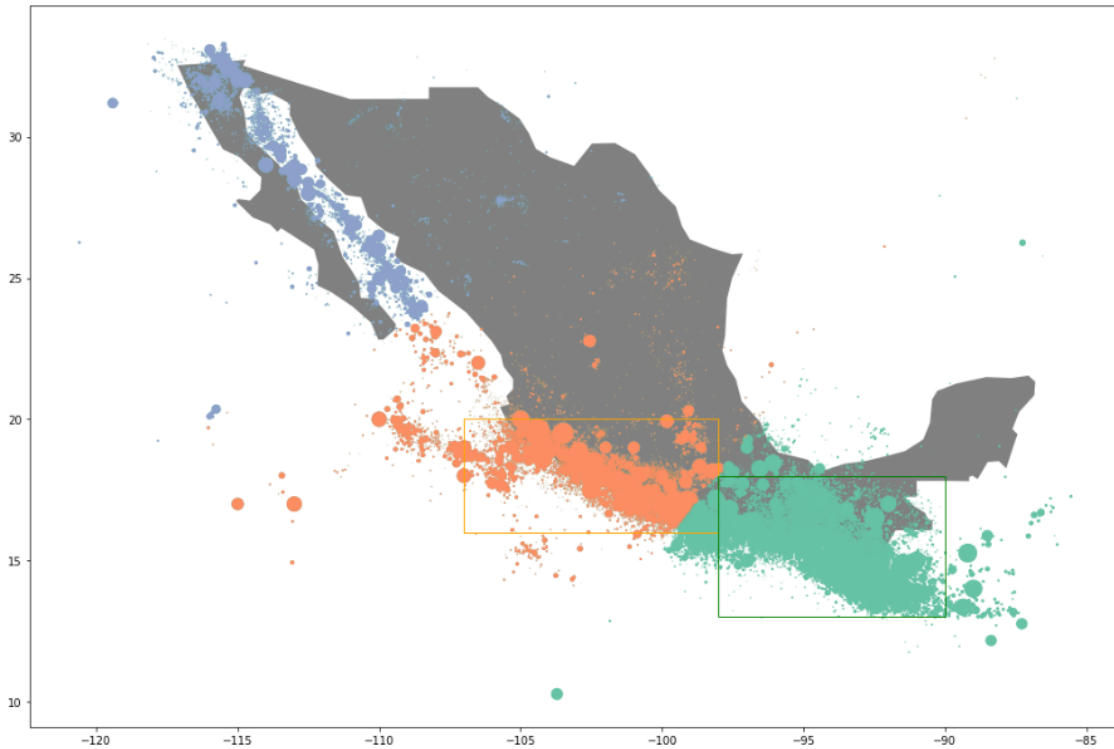


Fig. 7. División de catálogo por KNN en 3 grupos en función a su geometría, y delimitación de regiones suroeste (naranja) y sureste (verde).

En la Fig. 7 las regiones mostradas en el suroeste (naranja) y sureste (verde) son las regiones que se utilizan en el modelo de predicción, la región suroeste que abarca desde la latitud 16 hasta 20 y desde la longitud -98 hasta -107, y la región sureste que abarca desde la latitud 13 hasta 18 y longitud -90 hasta -98. Ambas zonas abarcan las regiones de alta sismicidad del sur de México, las regiones de sismicidad en México se pueden observar en Fig. 8.



Fig. 8. Regiones sísmicas de México (obtenido del Servicio Sismológico Nacional); D indica una zona sísmica con intensidad y frecuencia muy alta donde han ocurrido la mayor parte de los sismos, con las magnitudes más altas y generado más daños; C indica una zona sísmica con intensidad y frecuencia alta; B indica una zona sísmica con intensidad y frecuencia moderada; A indica una baja zona sísmica con poca intensidad y frecuencias.

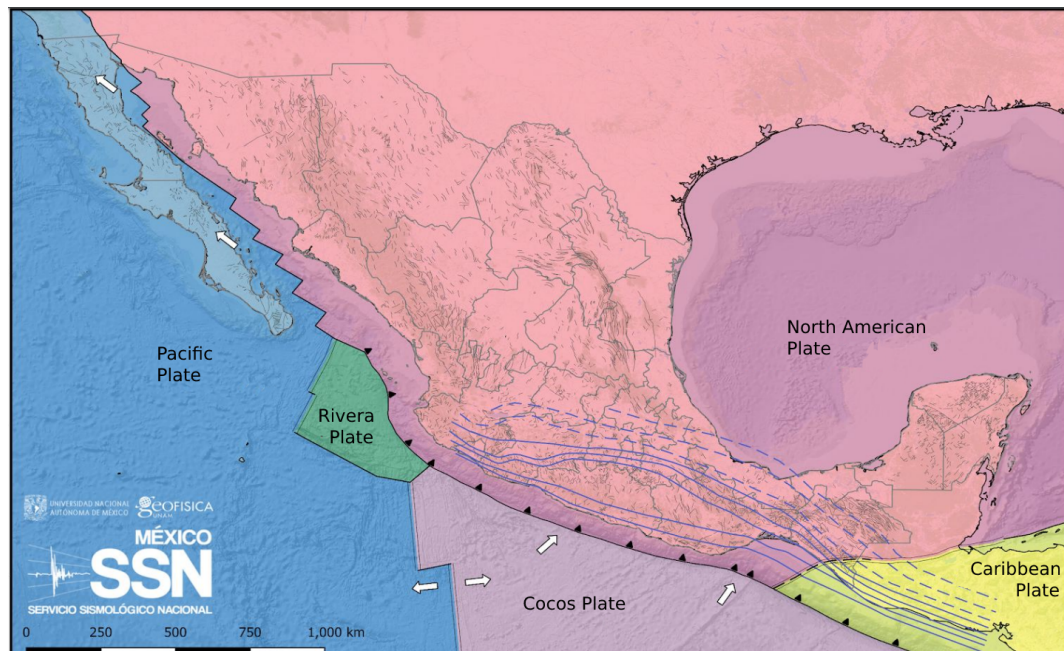


Fig. 9. Placas tectónicas que interactúan en territorio mexicano (Adaptada de SSN, Reporte especial: Sismo del 23 de junio de 2020, Oaxaca M7.4).

3.2. Extensión Temporal del Catálogo

El modelo está diseñado en base a la frecuencia de eventos con la que se informan los eventos, por eso es necesario que la frecuencia de cada año sea similar. Existe una mayor cantidad y calidad de los parámetros medidos sobre los eventos sísmicos en los años más recientes, y se alcanza una frecuencia constante a partir de 2010 para la región noroeste Fig. 10 y desde 2017 para la región sur Fig. 11. Se considerará 2017 también para las regiones de análisis suroeste y sureste.

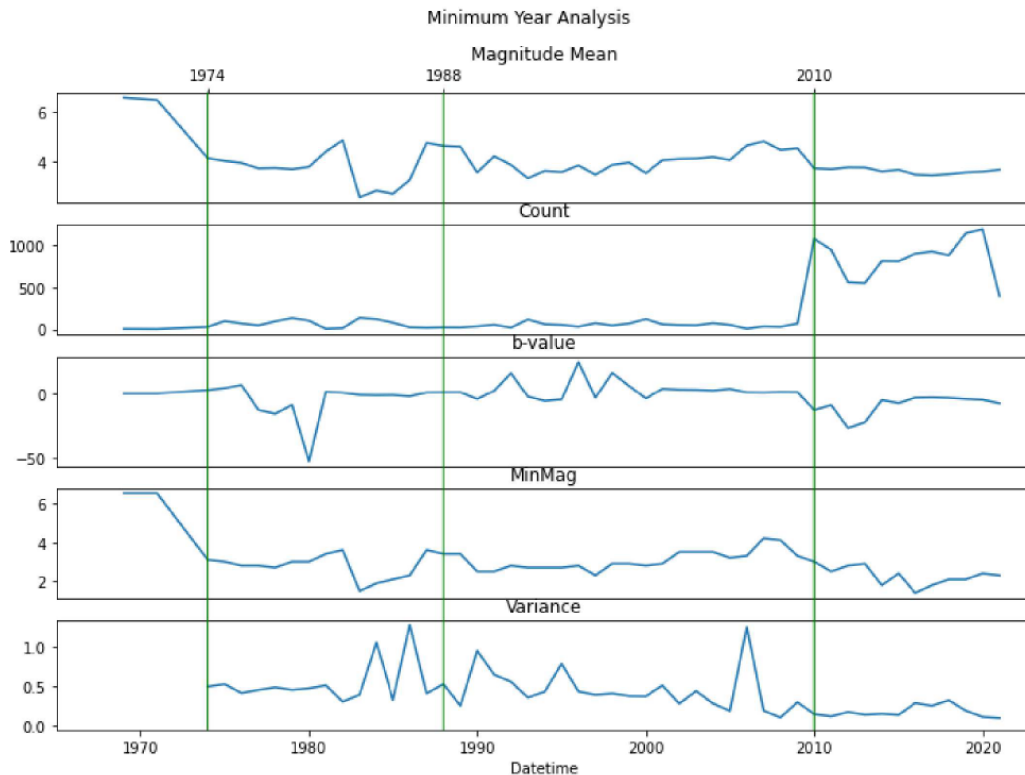


Fig. 10. Análisis de año mínimo para la región noroeste, aquí se visualiza la magnitud promedio, la ocurrencia de eventos, el valor b, la magnitud mínima y la varianza de cada año con el fin de obtener el año mínimo para el análisis.

Para la región noroeste en la Fig. 10 se utilizó 2010 para el año límite, ya que aquí se concentra la mayor cantidad de eventos con una mejor calidad, esto se demuestra en que la magnitud mínima, magnitud promedio y la varianza son menores, y además la ocurrencia de eventos es mayor, lo que resulta en la mayor concentración y calidad de los eventos.

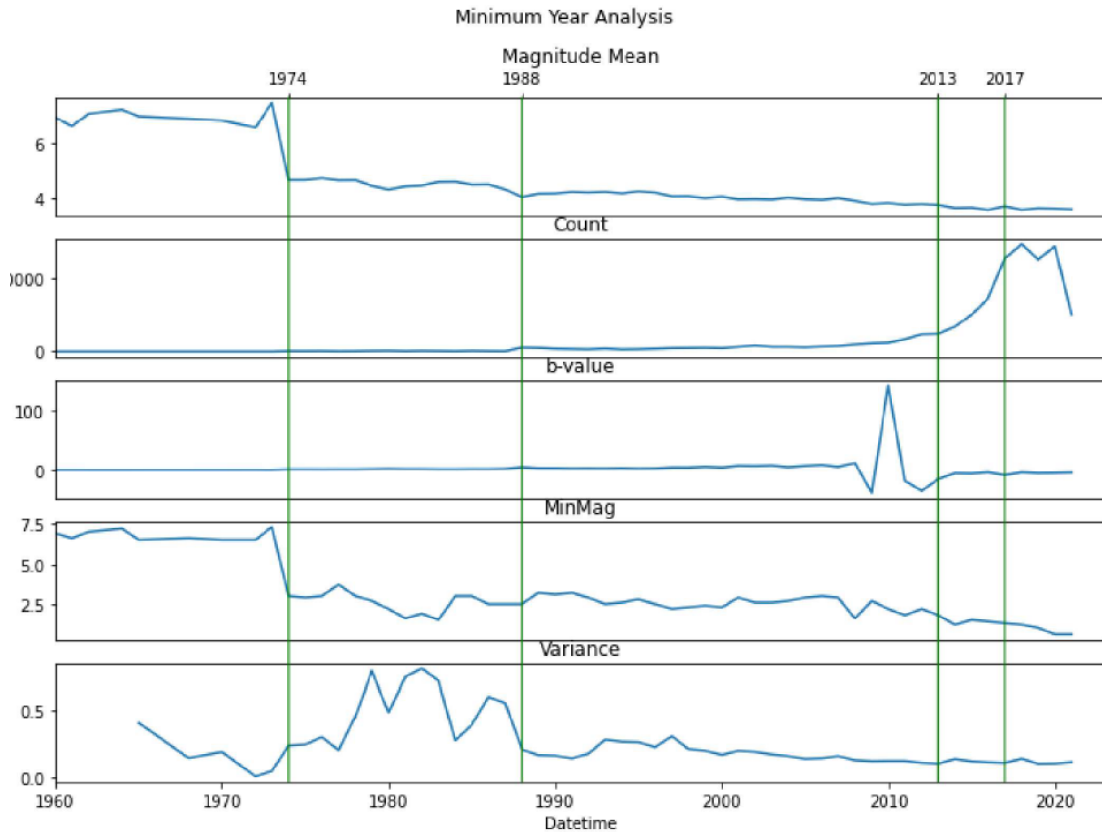


Fig. 11. Análisis de año mínimo para la región sur, aquí se visualiza la magnitud promedio, la ocurrencia de eventos, el valor b, la magnitud mínima y la varianza de cada año con el fin de obtener el año mínimo para el análisis.

Para la región sur en la Fig. 11 se utilizó 2017 para el año límite, a partir de 2010 incrementa la frecuencia de eventos informados, pero hasta 2017 la frecuencia de los eventos deja de crecer y se estabiliza, además aquí se concentra la mayor cantidad de los eventos con una mayor calidad.

3.3. Magnitud de corte

La magnitud de corte corresponde a la magnitud mínima en el catálogo por encima del cual el catálogo de terremotos se considera completo, es decir que a partir de esa magnitud no faltan eventos sísmicos, y se tiene que calcular para cada región y rango de tiempo. El valor de la magnitud de corte depende de la integridad del catálogo, que en sí depende de la instrumentación. Una mejor instrumentación (que obtenga la magnitud de más eventos sísmicos, incluyendo sismos de baja magnitud y con más precisión) en una región conduce a mejor integridad del catálogo con una magnitud de corte baja.

Para evaluar la magnitud de corte de este catálogo se utilizó el análisis de la curva de Gutenberg-Richter. En este análisis el punto donde la curva se desvía del comportamiento exponencial se selecciona como magnitud de corte para considerar sólo las magnitudes con eventos sísmicos informados completos. Todos los eventos por debajo de la magnitud de corte son descartados y no se utilizan en el análisis. [17]

Las magnitudes de terremotos y frecuencias de ocurrencias de una región se trazan como se muestra en la Fig. 12 y Fig. 13. En la sub-figura izquierda de cada figura la línea roja marca la magnitud de corte seleccionada, en ambas regiones noroeste y sur, con magnitud de corte $m^* \geq 3.8$. En la sub-figura derecha de cada figura se muestran las curvas resultantes después del corte, éstas siguen un comportamiento exponencial siguiendo la ley de Gutenberg-Richter, lo que asegura que cada catálogo esté completo con esta magnitud de corte.

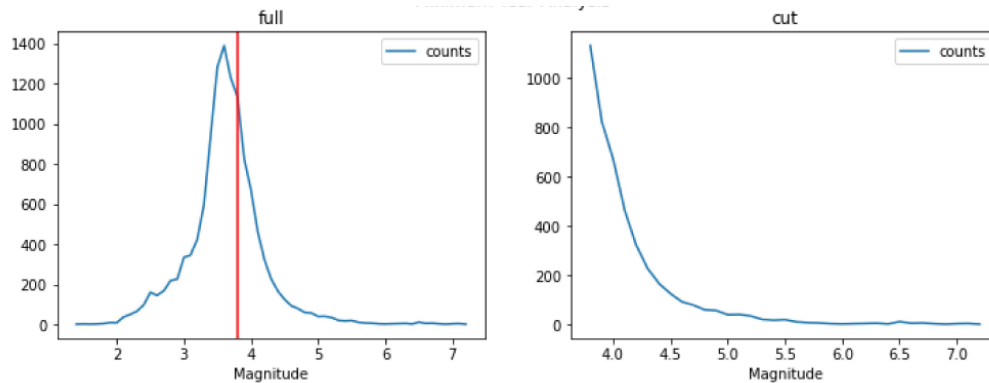


Fig. 12. Análisis de la curva de Gutenberg-Richter para la región noroeste .

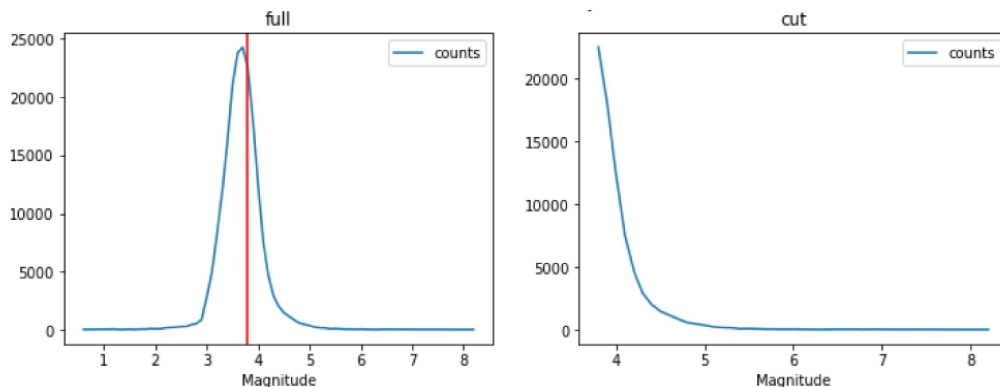


Fig. 13 Análisis de la curva de Gutenberg-Richter para la región sur.

4. Cálculo de Indicadores Sísmicos

Los indicadores son la parte más importante de un problema de clasificación. [16]

Se calculan dos tipos de indicadores, no paramétricos (que no depende de ningún parámetro por ejemplo $\text{mean}(M)$) e indicadores paramétricos (que dependen de cualquier parámetro variable) que en varios casos tienen múltiples versiones con base a diferentes variaciones de un parámetro, las indicadores que se utilizan en este modelo son los parámetros geofísicos y sismológicos relevantes para la predicción de terremotos disponibles en la literatura contemporánea, estos criterios se toman con el fin de retener la máxima información disponible sobre el estado geológico interno del suelo.

El pronóstico de series de tiempo puede enmarcarse como un problema de aprendizaje supervisado, el objetivo de un algoritmo de aprendizaje supervisado es tomar algunos datos con una relación conocida (magnitud y tiempo de los eventos sísmicos) y crear un modelo de esas relaciones. Los datos del catálogo son la información sobre el problema, a partir de estos datos se calculan los indicadores que conforman las instancias con las cuales el algoritmo será entrenado y el modelo contendrá las relaciones aprendidas resultantes para poder hacer futuras predicciones.

Para el cálculo de estos indicadores haremos uso de ventanas sobre series de tiempo, la predicción sobre series de tiempo puede enmarcarse como un problema de aprendizaje supervisado. En particular se utilizará una ventana móvil o deslizante que es una técnica que se utiliza para enmarcar un conjunto de datos de series de tiempo, la cual se muestra en Fig. 14 y es son generadas por el Algoritmo 1. Para calcular los indicadores de este modelo, se utiliza una ventana móvil multivariante de tamaño $n \times 2$ (bidimensionalidad) que captura la fecha y la magnitud de cada evento del catálogo, esta se desliza a través de los eventos del catálogo y se extraen subconjuntos del catálogo de tamaño $n \times 2$ en todas las posiciones de la ventana. Los subconjuntos resultantes serán de $E - n + 1$ subconjuntos de tamaño $n \times 2$, donde E es el número total de eventos.

Los indicadores se calculan para cada subconjunto generado, lo que resulta en $E - n + 1$ subconjuntos de tamaño $1 \times I$ donde I es el número de indicadores (que son las características con las que se entrena el modelo) y etiquetas (que son las variables que se buscan predecir) calculadas. El modelo utilizará los indicadores para entrenar los modelos e intentar predecir las etiquetas. Para calcular los indicadores y etiquetas de nuevos eventos sísmicos no es necesario esperar a que sucedan otros n^{**} nuevos eventos, ya que se pueden utilizar los j nuevos eventos y los $n^{**} - j$ últimos eventos. En este estudio, el problema de predicción de terremotos se modela como un problema de clasificación binaria, donde se busca predecir si un evento sísmico de magnitud mayor o igual que m ocurrirá en los próximos d días.

Cuando se usa clasificación, el resultado es una clase, entre un número limitado de clases. Con clases nos referimos a categorías arbitrarias según el tipo de problema, en este estudio las clases son clase 1 y clase 0, estas son mutuamente exclusivas. Si el modelo predice que ocurrirá un evento sísmico mayor a m en los próximos d días la clase será 1, en caso contrario, si el modelo predice que no ocurrirá un evento sísmico de magnitud mayor a m en los próximos d días entonces la clase será 0.

Algoritmo 1: Sliding window for creating sets

Data: E: This array contains the magnitude of the event and the time of occurrence.
n: Size of the sliding window

Result: Returns an array of sets, which everyone of these is going to be converted into an instance of the training and test dataset.

```
declare set_array [ 0.. (length(E) - n + 1) ]
```

```
for i ← 0 to (length(E) - n + 1) do
```

```
    declare temp_set [0..n]
```

```
    for j ← 0 to n do
```

```
        temp_set [j] ← E[i+j]
```

```
    set_array [i] ← temp_set
```

```
return set_array
```

Otros estudios en vez de utilizar una ventana deslizante simplemente dividen los datos y generan los conjuntos cada uno de n elementos, que resulta en E/n conjuntos en vez de los $E - n + 1$ conjuntos que se obtienen al usar la ventana deslizante, además de generar más información de esta manera los modelos son más sensibles a los cambios de los indicadores que se calculan, lo que es de interés para los intentos de predicción del modelo.

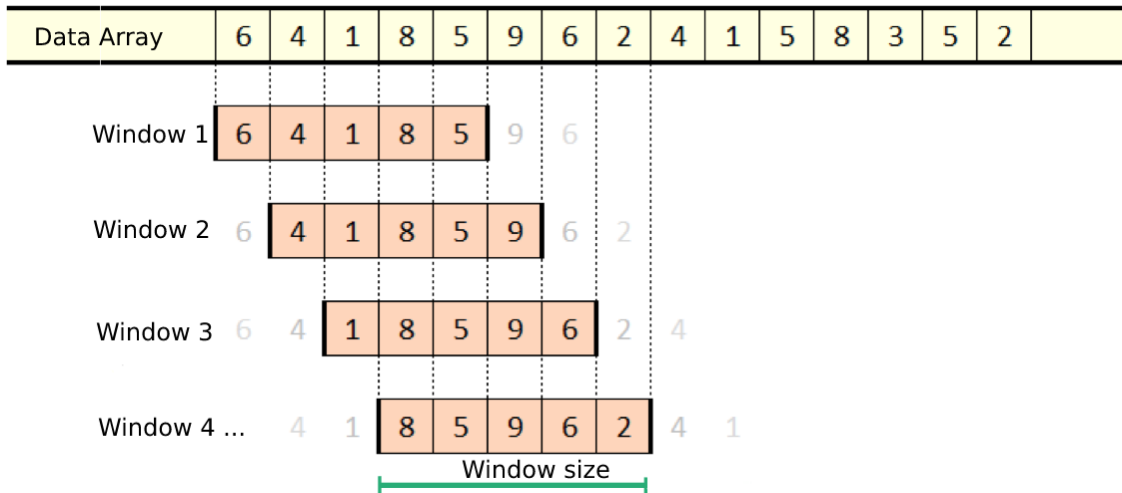


Fig. 14 Diagrama una ventana deslizante de tamaño 5 sobre un Data array unidimensional de tamaño 15. La ventana se desliza sobre los datos del data array capturando conjuntos de tamaño 5, el número total de conjuntos que obtendrá es $E - n + 1$, donde E es el número de elementos del array y n es el tamaño de la ventana, entonces en este caso se generarán 11 conjuntos con la ventana deslizante.

La capacidad de generalización para cualquier ANN disminuye con el tiempo, hecho que se refleja en la existencia del factor. En consecuencia, los pesos sinápticos deben calcularse periódicamente para que la ANN siempre se actualice, ya que las ANN solo se pueden aplicar para predecir el mismo número de elementos que forman el conjunto de entrenamiento, considerando este número como su límite superior. En otras palabras, el conjunto de prueba puede contener como máximo el mismo número de elementos existentes en el conjunto de entrenamiento. [21]

Se puede pensar que el alto rendimiento de las ANN diseñadas se debe a la influencia de la agrupación de terremotos. De hecho, Kagan declaró: "Cualquier técnica de predicción de terremotos prospectiva debe demostrar que su éxito no se debe a la influencia de la agrupación de terremotos (secuencias de premontaje-sismo principal-réplica)". [85]

Por otro lado, los terremotos se presentan inherentemente en grupos. Cualquier modelo que ignore este hecho podría ser señalado como incompleto. A este respecto, Zechar y Jordan afirmaron que: "Los terremotos se agrupan en el espacio y el tiempo, por lo tanto, cualquier pronóstico que capture este comportamiento de agrupación debería superar a un modelo de referencia uniforme". [86][21]

4.1. Parámetros

Los parámetros son variables que determinan el cálculo de los indicadores del modelo, en este trabajo se utilizaron los parámetros presentados en la Tabla 3.

Tabla 3. Parámetros del modelo

Símbolo	Nombre	Descripción
n	Tamaño de la ventana	Tamaño de la ventana, describe el número eventos de la ventana, el valor seleccionado es arbitrario y se puede ajustar para obtener mejores resultados, normalmente en la literatura relacionada es entre 50 y 100, el tamaño seleccionado en este estudio es 100.
m	Umbral de magnitud	Umbral de magnitud, ésta es la magnitud que va a funcionar como umbral para la clasificación, y varios parámetros y etiquetas dependen de esta variable, el valor usado normalmente es entre 5 y 6, la magnitud seleccionada en este estudio es 5.5 , la precisión disminuye en magnitudes altas. [26]
d_a	Días adelante	Días adelante, este valor puede cambiar dependiendo de cuántos días adelante se quiere predecir el próximo evento sísmico, el valor utilizado normalmente es entre 7 y 15.
d_b	Días atrás	Días atrás, el valor usado normalmente es 7, es usado para calcular los indicadores sísmicos.

4.2. Indicadores

En este estudio los indicadores sísmicos serán utilizados como las características introducidas al modelo como variables de entrada y son utilizados para entrenar el modelo, además todos los indicadores son utilizados simultáneamente.

Los indicadores se generan realizando cálculos sobre los conjuntos obtenidos por la ventana deslizante. Cada conjunto genera 23 indicadores y dos etiquetas (labels), estos 23 indicadores (son las características o atributos de la observación) que conforman una instancia (llamado también ejemplo o muestra). El objetivo de cada instancia es brindar la mayor cantidad de información del suelo en ese momento, a partir de los datos de un catálogo de una región.

Utilizaremos los indicadores propuestos en [20] y [21], más una selección de los indicadores propuestos en [14] y [16], los indicadores que no pasaron esta selección fueron el de “tiempo de recurrencia de terremotos”, que no agregaría sólo 1 indicador sí no 84 indicadores más, dos por cada grado de magnitud considerado (uno por cada variación de a y b , si fuese el caso que se consideran las magnitudes desde 3.5 a hasta 7.7 de un catálogo para una predicción habría un indicador por cada magnitud, es decir 84 nuevos indicadores).

Los valores de a y b se calculan numéricamente mediante dos métodos diferentes, con regresión lineal de mínimos cuadrados (lsq) y con Máximo-Mínimo (mlk). Las variaciones en los valores a y b tanto espacial como temporalmente se consideran generalmente como pistas para los precursores de grandes terremotos. En un estudio de predicción de terremotos para el sur de California, se propone un método basado en el análisis de regresión lineal de mínimos cuadrados [12][13][20]. Pero se prefiere el método de máxima verosimilitud para la predicción de terremotos en Chile [21]. Es decir, hay regiones en donde un método es preferible sobre el otro. Hay 7 indicadores que usan el valor de a y b como parámetro, estos indicadores tendrán dos versiones, la versión de los parámetros calculada con los a y b lsq y la versión calculada con a y b mlk, generando en total 14 indicadores.

Los indicadores utilizados en este modelo se pueden observar y son explicados a detalle en la Tabla 4. Entre corchetes se encuentra el nombre que se utiliza en el código (e.g. [meanM] para M_{mean}).

Tabla 4. Indicadores del modelo seleccionados

Nombre y expresión	Descripción
Primera fecha [firstT] T_1	Primera fecha de la ventana, esta es la fecha del primer evento sísmico de la ventana. donde T_i es la fecha del i -ésimo evento de la ventana, T tiene n elementos, uno por cada elemento de la ventana.
Última fecha [lastT] T_n	Última fecha de la ventana, ésta es la fecha del último evento sísmico de la ventana.
Tiempo transcurrido [elapsedT] $\Delta T = T_n - T_1$	El tiempo transcurrido de la ventana, se puede entender también como el tiempo durante los cuales han ocurrido los n eventos sísmicos de la ventana. Se calcula obteniendo la diferencia entre la primera fecha y la última fecha de la ventana, entre menor sea el tiempo transcurrido los eventos suceden con más frecuencia y viceversa, entre mayor sea el tiempo los eventos suceden con menos frecuencia.
Promedio de diferencia de tiempo transcurrido entre	Promedio de diferencia de tiempo transcurrido entre eventos de la ventana. Idealmente, el tiempo medio entre eventos característicos debería ser igual. [19]

<p>eventos [meanT]</p> $\mu = \frac{\sum_{i=1}^n (\Delta T_i)}{n - 1}$	
<p>Desviación estándar de las diferencia de tiempo transcurrido promedio entre eventos [meanTstd]</p> $c = \sqrt{\frac{\sum_{i=1}^n (\Delta T_i - \mu)^2}{n - 1}}$	<p>De manera similar, la desviación estándar de las diferencias de tiempo transcurrido entre eventos de la ventana también se toma como parámetro sísmico.</p>
<p>Valor a y b [a, b]</p> $b_{slq} = \frac{(n(\sum M_i \log_{10} N_i) - \sum M_i \sum \log_{10} N_i)}{((\sum M_i)^2 - n \sum M_i^2)}$ $b_{mlk} = \frac{\log_{10} e}{M_{mean} - M_{min}}$ $a_{slq} = \frac{\sum (\log_{10} N_i + b_{slq} M_i)}{n}$ $a_{mlk} = \log_{10} n + b_{mlk} M_{min}$ <p>donde M son las magnitudes de la ventana. N_i es el número total de eventos sísmicos correspondientes a Magnitud M_i.</p> <p>b es la pendiente de la curva y a es la intersección con el eje y.</p>	<p>Estos valores se basan directamente en la conocida ley geofísica conocida como Gutenberg-Richter. De acuerdo con esta ley, el número de terremotos aumenta exponencialmente al disminuir magnitud. [16]</p> $\log_{10} N_i = a - b M_i$ <p>Los valores de a y b Se calculan numéricamente mediante dos métodos diferentes, un método representa la regresión lineal de mínimos cuadrados (lsq), y el otro representa Maximum Maximum (mlk). Las variaciones en los valores a y b tanto espacial como temporalmente se consideran generalmente como pistas para los precursores de grandes terremotos. [12][13]</p> <p>En un estudio de predicción de terremotos para el sur california, se propone un método basado en el análisis de regresión lineal de mínimos cuadrados. [20]</p> <p>Pero se prefiere el método de máxima verosimilitud para la predicción de terremotos en Chile. [21]</p>
<p>Magnitud promedio [meanMag]</p> $M_{mean} = \frac{\sum_{i=1}^n (M_i)}{n}$	<p>Promedio de las magnitudes de los eventos de la ventana. Por lo general, la magnitud de los eventos sísmicos aumenta antes de cualquier evento sísmico grande.</p>
<p>Magnitud máxima [maxMag]</p> $M_{max} = \max(M)$	<p>Es el máximo valor de las magnitudes de los eventos de la ventana.</p>
<p>Máxima magnitud esperada [maxEMag]</p> $M_{expected\ max} = \frac{a}{b}$	<p>Es el máximo valor esperado de las magnitudes de los eventos de la ventana, este es calculado matemáticamente, es calculado para ambas versiones de a y b (mlk y sql).</p>
<p>Déficit de magnitud [magDef]</p> $\Delta M = M_{max} - M_{expected\ max}$	<p>El déficit de magnitud es la diferencia entre el valor máximo esperado máximo y el valor máximo actual, es calculado para ambas versiones de $M_{expected\ max}$ (mlk y sql).</p>
<p>Desviación estándar del valor b [bStd]</p> $\sigma b = 2.3b \sqrt{\frac{\sum_{i=1}^n (M_i - M_{mean})^2}{n(n - 1)}}$	<p>Desviación estándar del valor b, es calculado para ambas versiones de b (mlk y sql).</p>
<p>Mean square deviation [grcStd]</p> $\eta = \frac{\sum_{i=1}^n (\log_{10} N_i - a - b M_i)^2}{n - 1}$	<p>Desviación cuadrática media.</p>
<p>Tasa de raíz cuadrada de energía [rateSqrtEnergy]</p> $dE \frac{1}{2} = \frac{\sum_{i=1}^n (10^{(11.8+1.5M_i)})^{1/2}}{\Delta T}$	<p>La energía sísmica (dE) se sigue liberando del suelo en forma de pequeños terremotos. Si la liberación de energía se detiene, el fenómeno se conoce como quiescencia, que puede ser un precursor de un evento sísmico grande. El estado de</p>

	<p>reposito también puede conducir a la reducción de tasa sísmica para la región, disminuyendo así el valor b. [16]</p>
<p>Cambio de tasa sísmica B [bSeismicRateChange]</p> $\beta = \frac{N_a - \Lambda}{\sqrt{\Lambda}}$ <p>donde</p> $\Lambda = N_b \times \frac{\Delta T_a}{\Delta T_b}$ <p>N_b and T_b son de la ventana actual N_a and T_a son de la ventana anterior</p>	<p>El cambio de la tasa sísmica es el aumento o disminución general del comportamiento sísmico de la región durante dos intervalos de tiempo diferentes. Se proponen dos formas de calcular los cambios de velocidad sísmica.</p> <p>Este indicador es el cambio de la tasa sísmica β, mide el cambio de la tasa sísmica según lo propuesto por [22], donde las variables subíndice a y subíndice b corresponden a la tasa sísmica para dos intervalos diferentes. Las variables subíndice a se calculan con la ventana anterior a la actual, y las variables subíndice b se calculan con la ventana actual. T_a y T_b representan la duración de diferentes intervalos. N_a y N_b muestran el número de eventos sísmicos de cada intervalo a y b.</p> <p>Tenga en cuenta que se podría proponer una medida simétrica como</p> $\beta = \frac{\Lambda - N_b}{\sqrt{\Lambda}}$ <p>donde</p> $\Lambda = N_a \times \frac{\Delta T_b}{\Delta T_a}$ <p>que es igual a β cuando $\Delta t_a = \Delta t_b$</p>
<p>• Cambio de tasa sísmica Z [zSeismicRateChange]</p> $z = \frac{N_a \Delta T_b - N_b \Delta T_a}{\sqrt{N_a \Delta T_b^2 + N_b \Delta T_a^2}}$ <p>donde</p> <p>N_b and T_b son de la ventana actual N_a and T_a son de la ventana anterior.</p>	<p>Este es el otro valor que mide el cambio de la tasa sísmica, este es el cambio de tasa sísmica Z según lo propuesto por [24], aquí de nuevo T_a y T_b representan la duración de diferentes intervalos. N_a y N_b muestran el número de eventos sísmicos de cada intervalo.</p>
<p>Magnitud máxima en los últimos d días [lastDMaxMag]</p> $x6 = \max\{M\}$ <p>donde</p> $T \in [T_n - d_b, T_n)$ <p>cuando</p> $d_b = \text{days to look behind}$	<p>La magnitud máxima registrada en los d_b días previos a el último evento de la ventana también se considera un parámetro sísmico importante y se representa como $x6$. [21][23]</p> <p>La representación de este parámetro se mantiene igual que la de la literatura, para mantener una mejor continuidad.</p>
<p>Probability of event greater than m [pTMag]</p> $x7 = 10^{-bm}$	<p>Probabilidad de un terremoto igual o mayor que una magnitud m, es calculado para ambas versiones de b (mlk y sql). La representación de este parámetro también se mantiene igual que la de la literatura, para mantener una mejor continuidad.</p>

La correlación de los indicadores de este catálogo se puede observar en Fig. 15, los indicadores con mlk fueron calculados con a_{mlk} y b_{mlk} y los indicadores con lsq fueron calculados con a_{lsq} y b_{lsq} .

Describir cómo se calculan en pseudocódigo cada una de estos indicadores sería un proceso largo y tedioso y sin mucho valor, se puede apreciar cómo fue calculado en el repositorio proporcionado.

4.3. Etiquetas

En este estudio se les llama etiquetas a las variables que estamos tratando de predecir. Las etiquetas en este estudio inician con y_1 , que es continua (y no es usada para la predicción), pero se utiliza para calcular y_2 , si y_1 excede la magnitud umbral m su valor será 1, si y_1 no excede m su valor será 0, haciendo de este problema una tarea de clasificación binaria.

En este caso, el resultado es una categoría (terremoto de gran magnitud o terremoto no de gran magnitud) y llamamos a este tipo de problema un problema de clasificación (es clasificación binaria porque solo son dos categorías exclusivas). Si la salida fuera un valor numérico, lo llamaríamos un problema de regresión.

Es considerado que aún es muy ambicioso usar este tipo de análisis para predecir una magnitud (y tratarlo como un problema de regresión i.e. y_1). [16] La magnitud de umbral m seleccionada es de 5.5 que es la adoptada para una magnitud de “gran terremoto”. [25] Entonces este modelo intenta responder a la pregunta, ¿Ocurrirá un terremoto de gran magnitud en los próximos 14 días?, en la Tabla 5 se muestran más a detalle las etiquetas del modelo.

Tabla 5. Indicadores del modelo

Nombre y expresión	Descripción
<p>Magnitud máxima en los próximos D días [nextDMaxMag] $y_1 = \max\{M\}$ cuando $T \in [T_n, T_n + d_a)$ donde $d_a = \text{days to look ahead}$</p>	<p>Magnitud máxima en los próximos D días, se cuentan D días a partir del último evento de la ventana T_n y se toma la magnitud máxima que ocurrió en ese lapso de tiempo.</p>
<p>Ocurrencia de un evento > m $y_2 = \begin{cases} 1 & \text{if } y_1 \geq m \\ 0 & \text{if } y_1 < m \end{cases}$</p>	<p>Se calcula a partir del resultado de y_1, clasifica si y_1 se encuentra sobre el umbral o debajo de él, si este se encuentra por encima, el valor de y_2 es 1 si este se encuentra por debajo del umbral el valor de y_2 es 0.</p>

La correlación de las etiquetas de este catálogo también se pueden observar en Fig. 15.

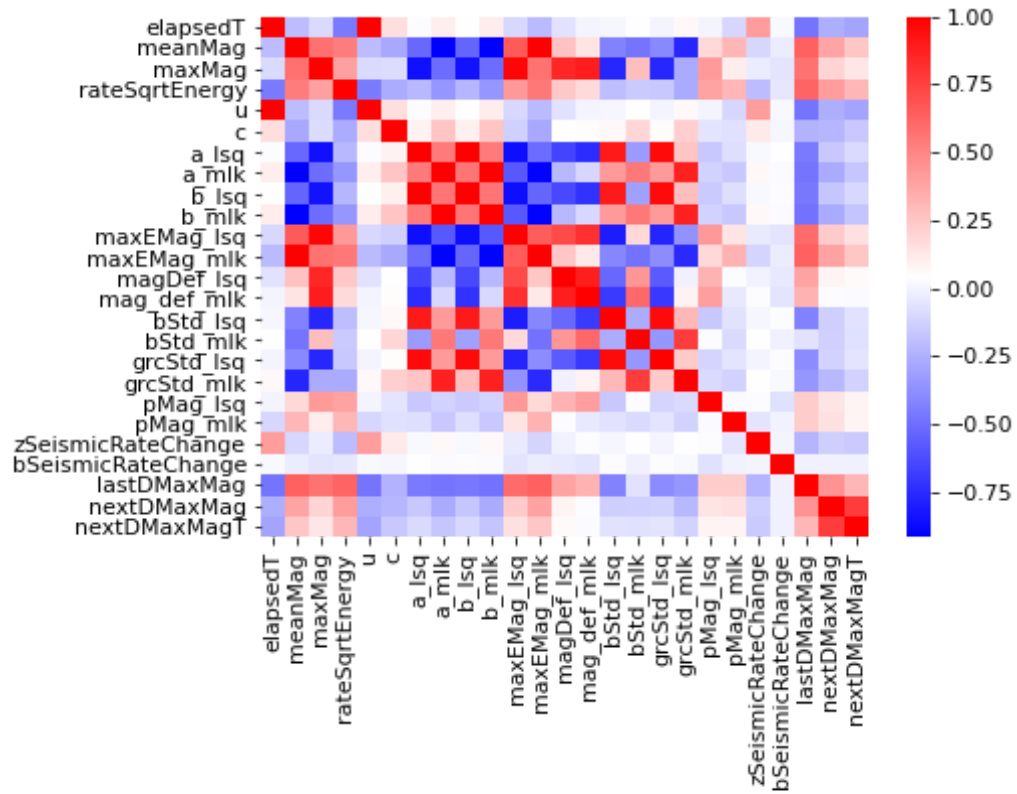


Fig. 15 Matriz de correlación de los indicadores y etiquetas, calculada con los indicadores generados de la región sur.

5. Modelos de Predicción

5.1. Modelos predictivos

Un modelo predictivo en el campo de Machine Learning utiliza algoritmos para predecir resultados. La mayoría de las veces, el evento que uno quiere predecir está en el futuro, pero el modelado predictivo se puede aplicar a cualquier tipo de evento desconocido, independientemente de cuándo ocurrió. En este estudio se utilizan modelos para la predicción de terremotos.

Estos modelos se utilizan para que el dominio de la predicción sea un conjunto de clasificadores y determinar la probabilidad de que un conjunto de datos pertenezca a otro conjunto. En este caso, se puede utilizar un modelo para determinar si un terremoto es de gran magnitud (mayor a 6) o de baja magnitud.

Existen dos tipos de aprendizaje automático, el aprendizaje supervisado supone que partimos de un conjunto de datos etiquetado previamente, es decir, conocemos el valor del atributo objetivo (el dato que se quiere predecir) para el conjunto de datos que disponemos; y el aprendizaje no supervisado parte de datos no etiquetados previamente.

5.2. Redes Neuronales

Las redes neuronales, también conocidas como redes neuronales artificiales (ANN) o redes neuronales simuladas (SNN), son un subconjunto del aprendizaje automático. Su nombre y estructura están inspirados en el cerebro humano, imitando la forma en que las neuronas biológicas se transmiten entre sí.

Las redes neuronales artificiales se componen de capas de nodos, que contiene una capa de entrada, una o más capas ocultas y una capa de salida. Cada nodo, o neurona artificial, se conecta a otros y tiene un peso y un umbral asociados. Si la salida de cualquier nodo individual está por encima del valor de umbral especificado, ese nodo se activa y envía datos a la siguiente capa de la red. De lo contrario, no se transmiten datos a la siguiente capa de la red. Se puede visualizar un diagrama representativo de una red neuronal, con sus capas de entrada, capas escondidas y capas de salida en la Fig. 16.

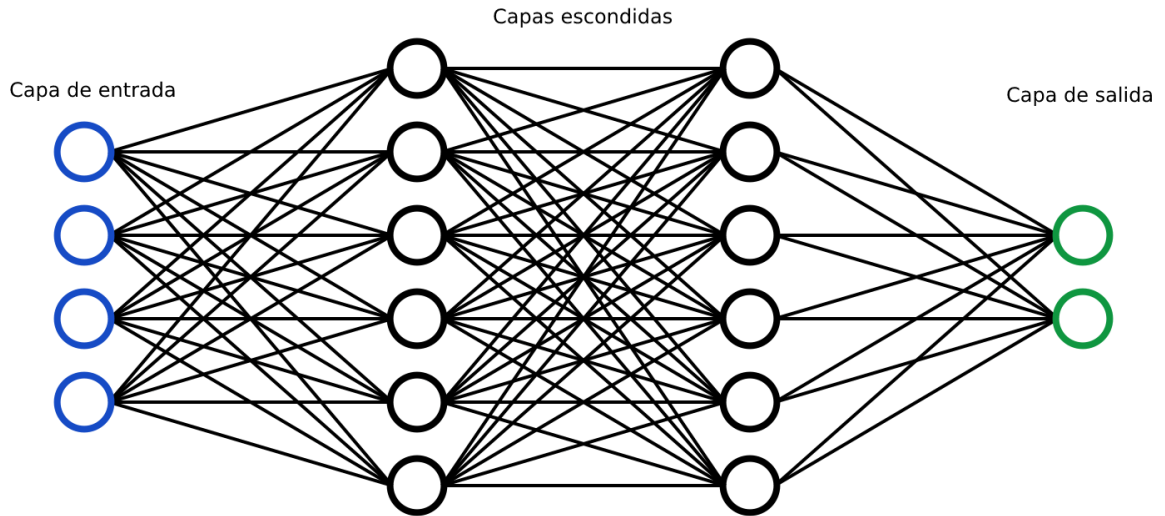


Fig. 16. Diagrama de redes neuronales.

5.3. Descripción del modelo de predicción

La tarea presente es un problema de aprendizaje supervisado donde se tienen variables de entrada (la muestra X, cada instancia con 23 indicadores) y una variable de salida (la muestra Y, las etiquetas respectivas de cada instancia) y se usa un algoritmo ANN para aprender la función de mapeo desde la entrada hasta la salida.

El objetivo de un modelo de predicción es aproximar el mapeo suficientemente bien para que cuando tenga nuevos datos de entrada (datos de prueba), pueda predecir las variables de salida para esos datos.

Conocemos las respuestas correctas; el algoritmo hace predicciones de forma iterativa sobre los datos de entrenamiento y se corrige mediante actualizaciones. El aprendizaje se detiene cuando el algoritmo alcanza un nivel aceptable de rendimiento.

Para realizar la predicción se usó una red neuronal y se introdujeron los 23 indicadores calculados. La estructura del modelo se puede observar en Fig. 17. Primero se selecciona un catálogo, se prepara, se seleccionan los parámetros del modelo, a partir de estos parámetros se calculan las ventanas con las cuales se generan los indicadores sísmicos, los indicadores se preparan y se dividen para entrenamiento y pruebas, para finalizar se usan estos indicadores para entrenar la red neuronal.

La clasificación de datos desequilibrados es una tarea intrínsecamente difícil, ya que hay muy pocas muestras de las que aprender, pero al usar ventanas deslizantes el modelo tiene más información y es más sensible a los cambios y diferencias de los indicadores sísmicos, aunque aumenta la probabilidad de sobre ajuste, pero existen medidas que se pueden tomar como la capa de dropout.

Al usar una capa dropout durante el entrenamiento, básicamente un porcentaje aleatorio de las neuronas de una capa en particular se desactiva. Esto mejora la generalización porque obliga a las capas de la red a aprender las mismas relaciones con diferentes neuronas. El valor del dropout seleccionado aquí fue de 0.5, los valores recomendados del dropout son de 0.5 a 0.8, dónde 1.0 significa que no hay dropout (funciona normalmente) y 0.0 significa que no hay salidas de la capa, al igual que otros métodos de regularización, dropout es más eficaz en aquellos problemas en los que hay una cantidad limitada de datos de entrenamiento. Al utilizar ventanas deslizantes hace que los datos se parezcan más, incluyendo los datos de entrenamiento y los datos de prueba, logrando que sea menos visible el sobreajuste.

Es común que las redes más grandes (más capas o más nodos) se sobreajusten más fácilmente a los datos de entrenamiento.

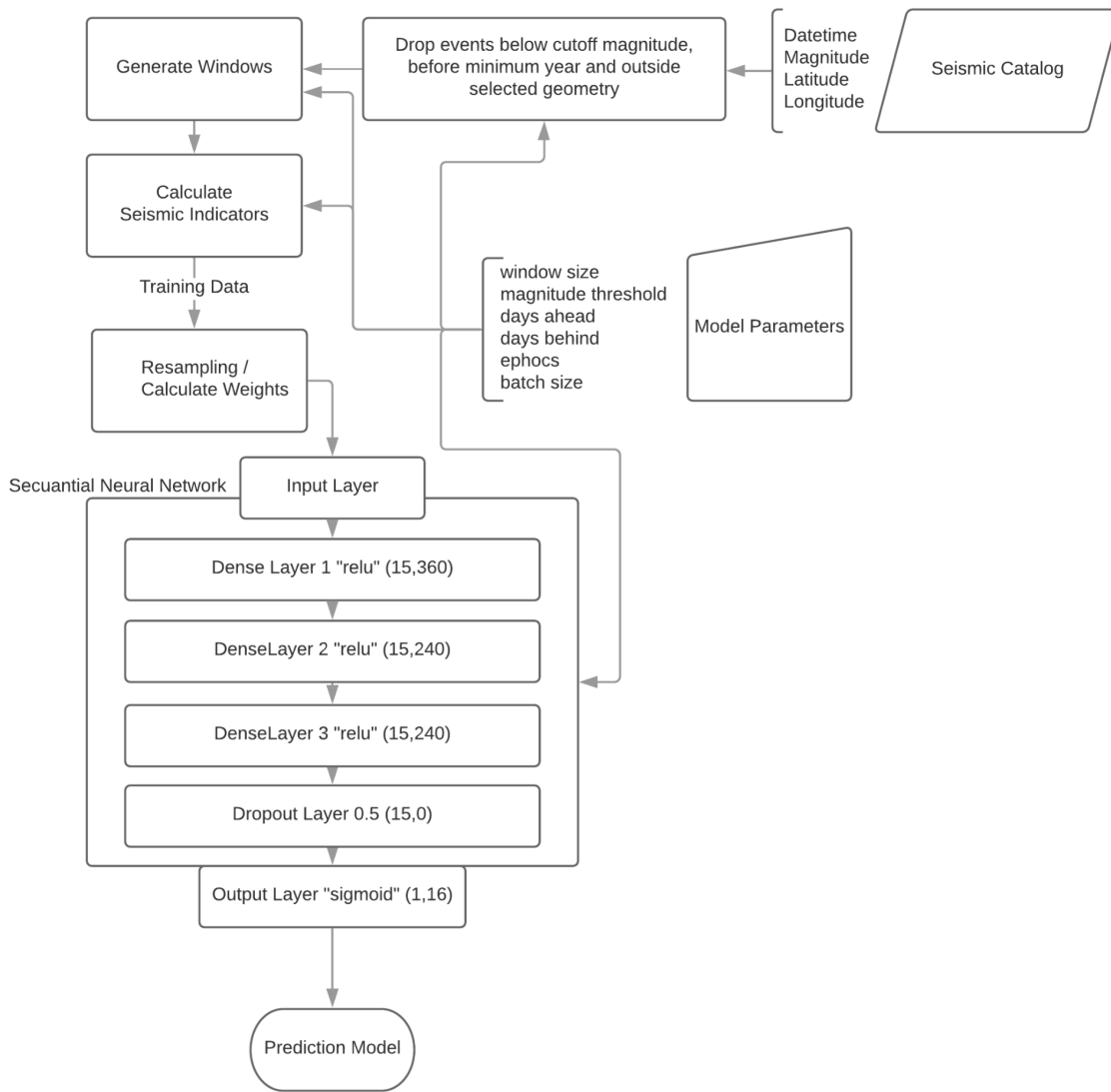


Fig. 17. Estructura del modelo de predicción, hay diferentes variaciones del modelo, donde la etapa Resampling no se utiliza o es sustituida por una etapa Weighting, estas variaciones son explicadas posteriormente.

5.4. Descripción de la red Neuronal

En este modelo se utilizó una red neuronal secuencial. El modelo consiste de tres capas densas con 15 neuronas cada una, conectadas profundamente. Una capa densa significa que cada neurona en la capa recibe información de todas las neuronas de su capa anterior, además estas capas ocultas tienen activación “relu”; una capa de dropout que ofrece un método de regularización económico y eficaz desde el punto de vista computacional para reducir el sobreajuste y mejorar el error de generalización en redes neuronales profundas; y una capa de salida con activación sigmoideal.

La estructura de la red neuronal se muestra en la Tabla 6. Como ya se dijo anteriormente se decidió utilizar una capa dropout para evitar sobre ajustes como es habitual usar en problemas con datos desbalanceados, al usar pesos muy grandes o al hacer resampling.

Que los datos sean desbalanceados en una tarea de clasificación significa que existe un número de elementos en una clase que superan en gran medida a los ejemplos en otra. En el caso de los terremotos, según la ley de Gutenberg Richter, los terremotos de baja magnitud, (aquellos de clase 0), son exponencialmente más comunes que los de gran magnitud, (aquellos de clase 1).

Al aplicar pesos (ponderaciones) a las clases desbalanceadas, naturalmente una clase tendrá un peso mayor en proporción a ese desbalance y favorecerá riesgo de sobreajuste, lo mismo pasa cuando se hace un resampling (remuestreo) de las mismas muestras de la clase minoritaria que antes no aparecía con la misma frecuencia, ahora lo hará corriendo el mismo tipo de riesgo de sobreajuste.

Para la región seleccionada del sur de México después de 2017 existen 39371 eventos registrados con una magnitud mayor a 3.5 (la magnitud de corte seleccionada para el conjunto de datos), de los cuales 39325 son terremotos con magnitudes menores a 5.5 y solo 46 son terremotos con magnitudes mayores a 5.5 considerados de gran magnitud.

Después de utilizar las ventanas para la región del sur de México se generan 26341 conjuntos capturados sin un evento sísmico de gran magnitud (mayor a 5.5) en sus próximos 15 días y 12714 conjuntos capturados con un terremoto de gran magnitud en sus próximos 15 días, al usar una ventana de predicción de 15 días (si esta fuera mayor, las métricas tendrían mejores resultados) y la ventana deslizante se generan 400 veces más muestras positivas que eventos sísmicos individuales de terremotos de gran magnitud, esto disminuye significativamente el problema de los datos desbalanceados.

Para la región seleccionada del suroeste de México, después de 2017 existen 7082 eventos registrados con una magnitud mayor a 3.5 (la magnitud de corte), de los cuales 7073 son terremotos con magnitudes menores a 5.5 y 9 son terremotos con magnitudes mayores a 5.5 considerados de gran magnitud. Después de utilizar la ventana deslizante para esta región se generan 6915 conjuntos en total, 6470 conjuntos capturados sin un evento sísmico de gran magnitud (mayor a 5.5) en sus próximos 15 días y 445 conjuntos capturados con un terremoto de gran magnitud en sus próximos 15 días, lo que equivale a 50 veces más muestras positivas que eventos sísmicos individuales de terremotos de gran magnitud.

Para la región seleccionada del sureste de México, después de 2017 existen 30864 eventos registrados con una magnitud mayor a 3.5, de los cuales 30829 son terremotos con magnitudes menores a 5.5 y 35 con magnitudes mayores, considerados de gran magnitud. Después de utilizar la ventana deslizante sobre esta región se generan 30620 conjuntos capturados en total, 22174 sin un evento sísmico de gran magnitud (mayor a 5.5) en sus próximos 15 días y 8446 de gran magnitud en sus próximos 15 días, equivalente a 335 veces más muestras positivas que eventos sísmicos individuales de terremotos de gran magnitud.

Hay una diferencia significativa en la actividad sísmica de las dos sub regiones del sur seleccionadas, ya que al aplicar el modelo de predicción en estas muestra mejor cómo esta metodología se desempeña en diferentes regiones de diferente intensidad sísmica.

En este trabajo se les llama muestras a las instancias que serán introducidas al modelo para entrenarlo. El número de las muestras en este estudio depende del número de eventos sísmicos considerados, el tamaño de la ventana y el tiempo de la ventana de predicción.

Tabla 6. Resumen del modelo de la red neuronal

Layer (type)	Shape (activation)	Param #
Layer 1 (dense)	23 (relu)	552
Layer 2 (dense)	23 (relu)	552
Dropout Layer		
Layer 6 (dense)	1 (sigmoid)	24

Total Parameters: 2,784
Trainable Parameters: 2,784
Non-trainable Parameters: 0

Para entrenar un modelo con datos desbalanceados, se ajusta con un tamaño de lote (que es el conjunto de ejemplos usados en una iteración, es decir una actualización de gradiente durante el entrenamiento del modelo) suficientemente grande para garantizar que cada lote tenga una probabilidad suficiente de contener algunas muestras positivas. Se les llama muestras positivas a las muestras de clase 1 (en las que ocurre un terremoto) y muestras negativas a las muestras de clase 0 (en las que no ocurre un terremoto). Hay otras soluciones como class weighting y oversampling, que se pueden aplicar para el problema de trabajar con datos desbalanceados y serán exploradas más adelante.

5.5. Parámetros

Los parámetros usados para todos los modelos de predicción son los siguientes: los tamaños de las ventanas son de 100 elementos, la magnitud umbral es de 5.5; (tratará de clasificar las ventanas con 1 si la ventana tendrá un sismo mayor o igual a 5.5 en sus próximos d_a días, ó con 0 si no sucede un sismo mayor a 5.5), el valor de d_a seleccionado es 14. Finalmente se utilizaron 600 Epochs para el entrenamiento.

El tamaño de lote seleccionado fue de 4096, el cual es lo suficientemente grande para garantizar que cada lote tenga una probabilidad suficiente de contener algunas muestras positivas. Si el tamaño del lote es demasiado pequeño, probablemente no tendrían muestras de clase 1 de las que aprender.

5.6. Variaciones de Modelos de Predicción

El modelo *Baseline* es el modelo generado más simple, en este las instancias son introducidas sin modificar, y no hay parámetros especiales ni procesamiento adicional en ningún momento para el modelo ni para las instancias, por lo tanto, el modelo se entrena con los datos de las instancias “tal como son”.

Las curvas de las métricas por *epoch* del modelo *baseline* se pueden visualizar en Fig. 18 para la región suroeste y Fig. 19 para la región sureste.

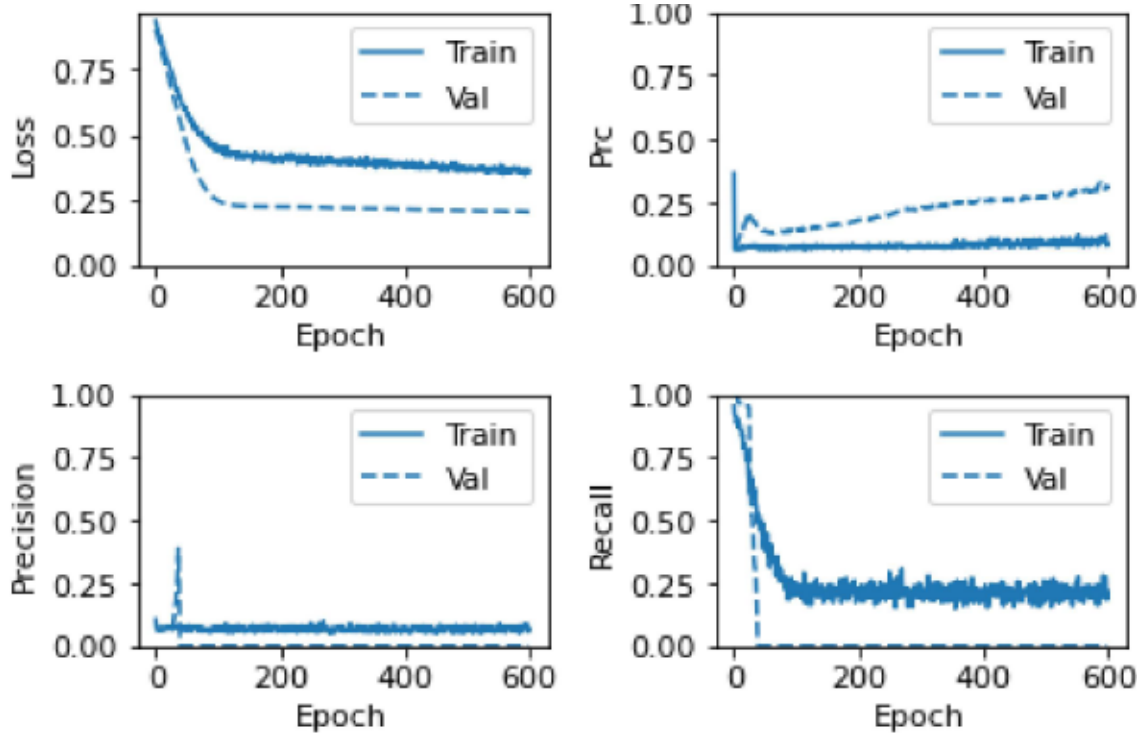


Fig. 18. Las curvas de las métricas por epoch durante el entrenamiento del modelo Baseline al aplicarse en la región suroeste.

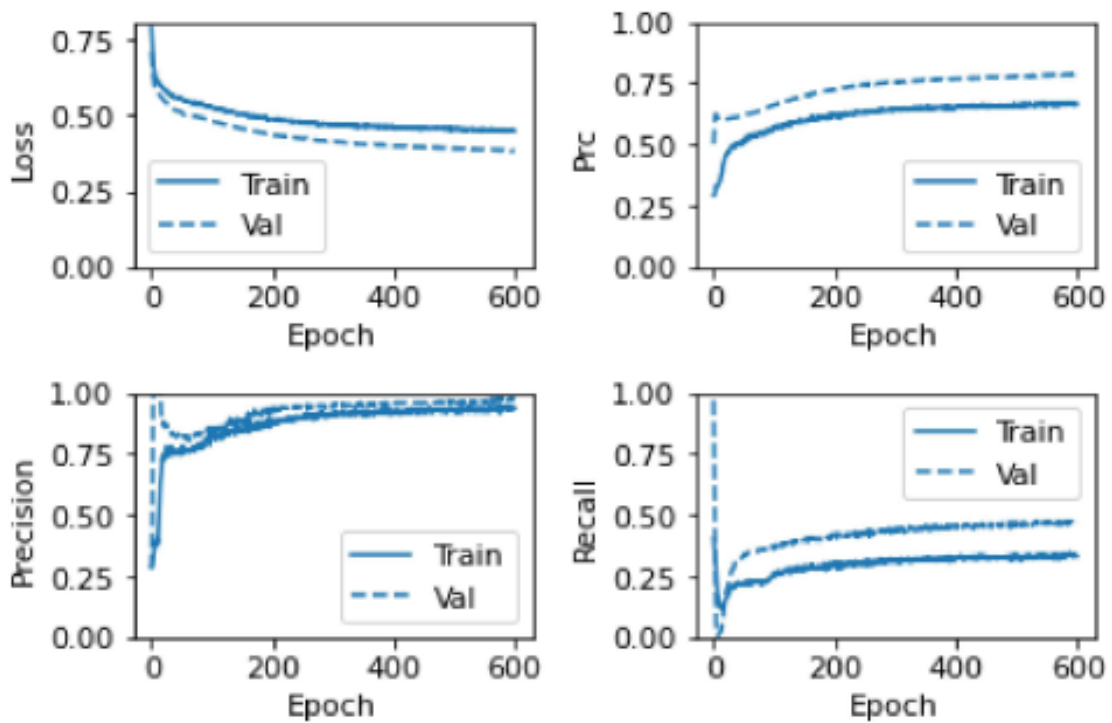


Fig. 19. Las curvas de las métricas por epoch durante el entrenamiento del modelo Baseline al aplicarse en la región sureste.

Será recurrente en todas las figuras de las gráficas que sean más “suaves” para la región del sureste que la región suroeste durante el entrenamiento debido al tamaño de la muestra de entrenamiento y la cantidad de muestras positivas.

Queremos predecir los casos en los que ocurre un terremoto de gran magnitud (muestras positivas) pero ya que no se tienen tantas muestras positivas para trabajar, entonces conviene que el clasificador pondere en gran medida los pocos ejemplos disponibles para lograr este objetivo. Se pueden pasar pesos para cada clase a través de un parámetro, esto ocasionará que el modelo les dé un mayor peso estadístico a los ejemplos de una clase subrepresentada, como se muestra en el siguiente modelo.

En el modelo Weighted los indicadores son introducidos sin modificar y se agrega el parámetro peso para que las clases 1 y 0 tengan diferente peso al clasificar. El objetivo es identificar sismos sobre el umbral, pero según la ley de Gutenberg–Richter no encontraremos muchas de esas muestras de clase 1 en comparación con las de clase 0, por esta razón debemos darle más valor estadístico a la clase 1. Con estas ponderaciones de clase, puede que la exactitud (accuracy) y la precisión (precision) sean menores porque habrá más fp (falsos positivos), pero, el recall y el AUC son más altos porque el modelo también encontró más tp (verdaderos positivos) y menos fn (falsos negativos). A pesar de tener una precisión más baja, este modelo tiene por lo general una mayor sensibilidad (recall) e identifica más sismos sobre el umbral.

Para calcular los pesos se utilizaron las siguientes expresiones:

$$\text{Peso para clase 0} = (1/neg) * (total/2.0)$$

$$\text{Peso para clase 1} = (1/pos) * (total/2.0)$$

Para la región del suroeste los pesos resultantes son 0.53 para la clase 0 y 7.77 para la clase 1.

Para la región del sureste los pesos resultantes son 0.69 para la clase 0 y 1.81 para la clase 1.

Las curvas de las métricas por epoch del modelo weighted se pueden visualizar en Fig. 20 para la región suroeste y Fig. 21 para la región sureste.

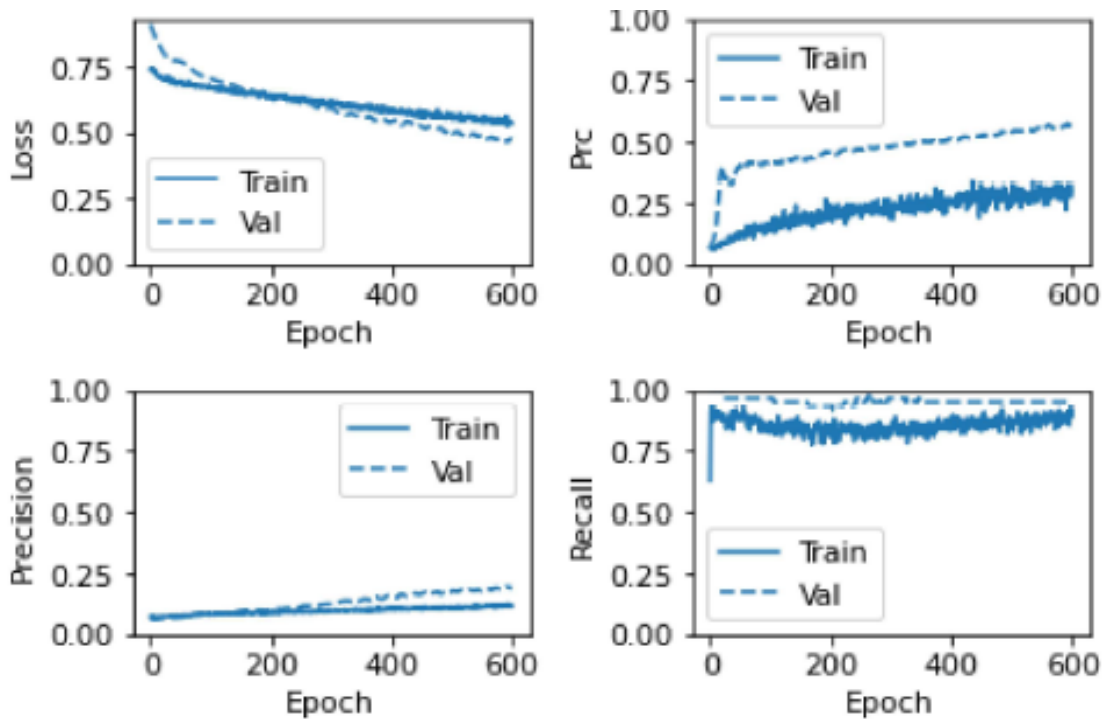


Fig. 20. Las curvas de las métricas por epoch durante el entrenamiento del modelo Weighted al aplicarse en la región suroeste.

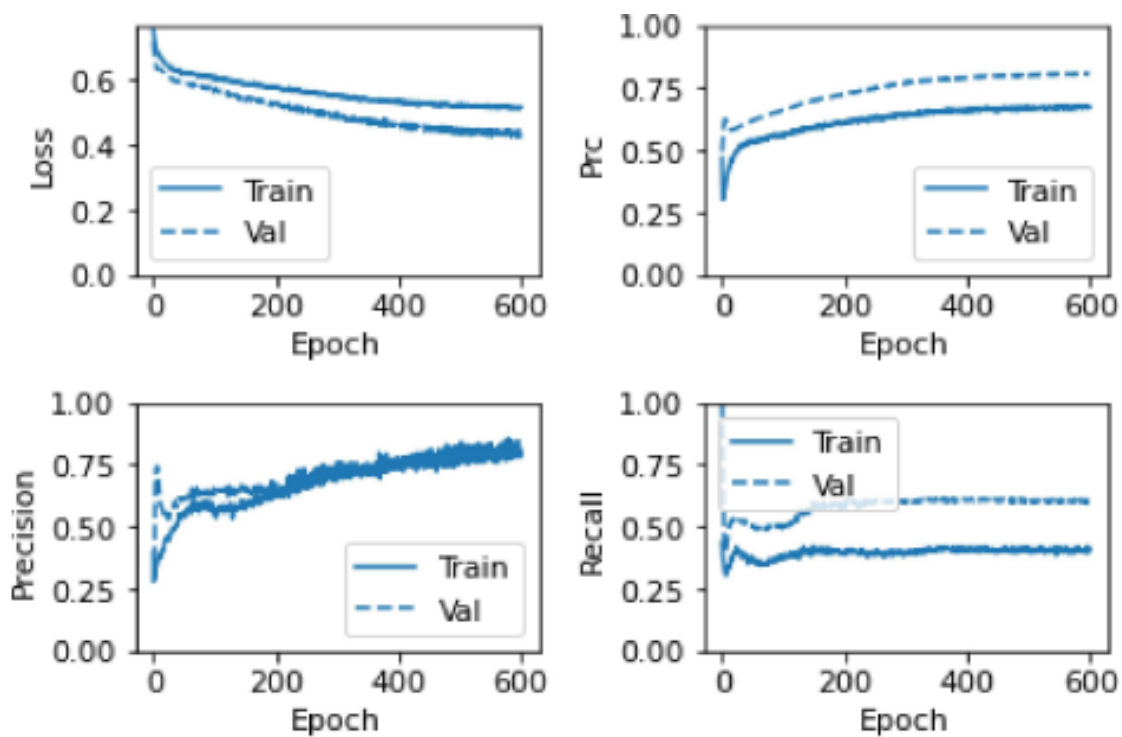


Fig. 21. Las curvas de las métricas por epoch durante el entrenamiento del modelo Weighted al aplicarse en la región sureste.

En el modelo Resampled los indicadores son modificados, y también se modificaron los parámetros. El enfoque está relacionado con el modelo anterior, pero en lugar de calcular pesos para cada clase, se realiza un oversampling que es una técnica de remuestreo utilizada para ajustar la distribución de clases de un conjunto de datos sobre una clase minoritaria y, así equilibrar el conjunto de datos, el resultado será aproximadamente 50% de datos clase 1 y 50% de datos clase 0, En lugar de 6% y 94% en el caso de la región del suroeste y 27% y 63% en el caso de la región del sureste.

Finalmente se modificó la cantidad de pasos por epoch, es decir la cantidad de lotes necesarios para ver cada ejemplo negativo una vez, obteniéndose que en lugar de que los datos de clase 1 se muestran en un lote con un gran peso, estos se muestran en muchos lotes diferentes con un pequeño peso. Esta señal de gradiente más suave facilita el entrenamiento del modelo, debido al incremento de instancias y al tamaño del lote, ocasionando que este modelo requiera más tiempo de entrenamiento y memoria.

Las curvas de las métricas por *epoch* para el modelo *resampled* durante el entrenamiento del modelo se pueden visualizar en Fig. 22 para la región suroeste y Fig. 23 para la región sureste.

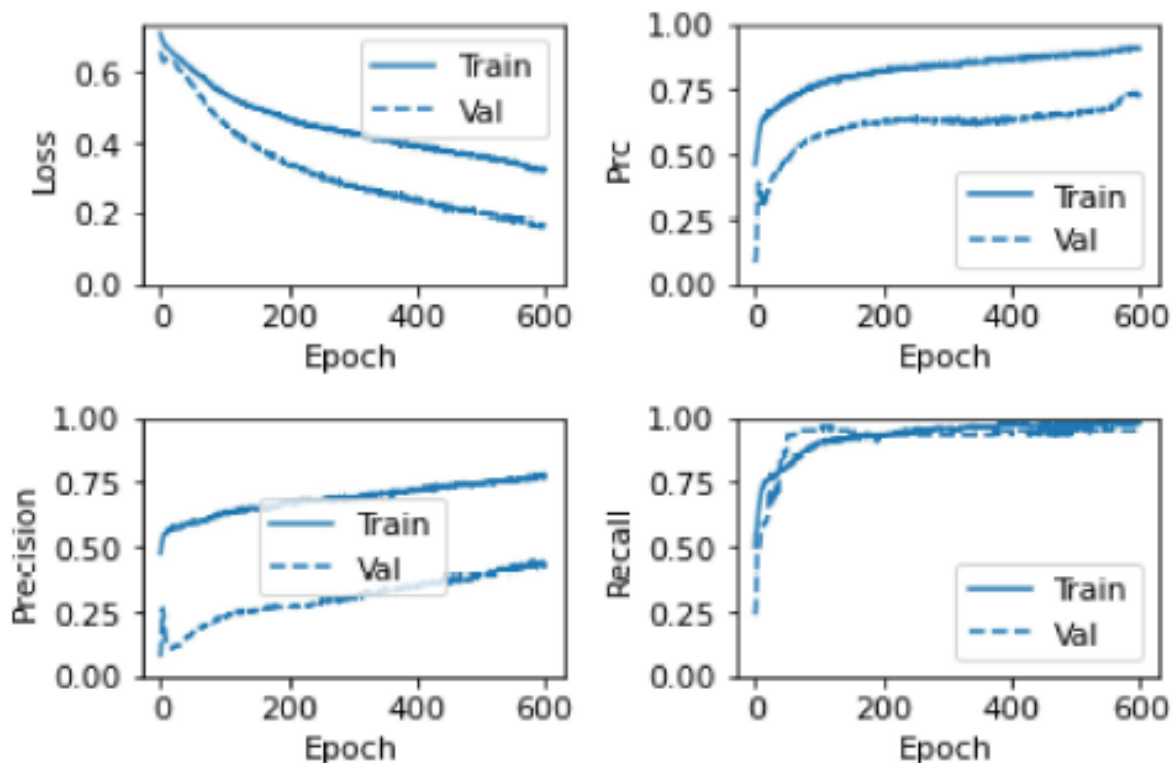


Fig. 22. Las curvas de las métricas por epoch durante el entrenamiento del modelo Resampled al aplicarse en la región suroeste.

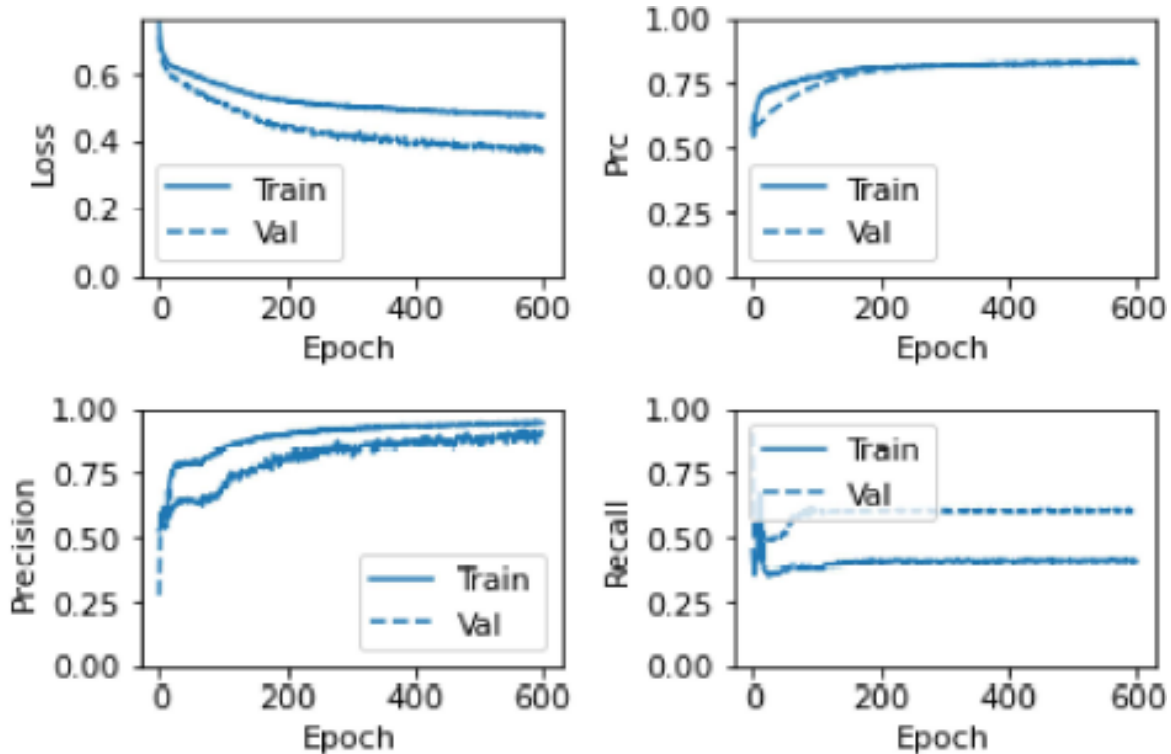


Fig. 23. Las curvas de las métricas por epoch durante el entrenamiento del modelo Resampled al aplicarse en la región sureste.

Las curvas PRC y AUC del modelo y sus variaciones se pueden observar en las figuras Fig. 24 y Fig. 25.

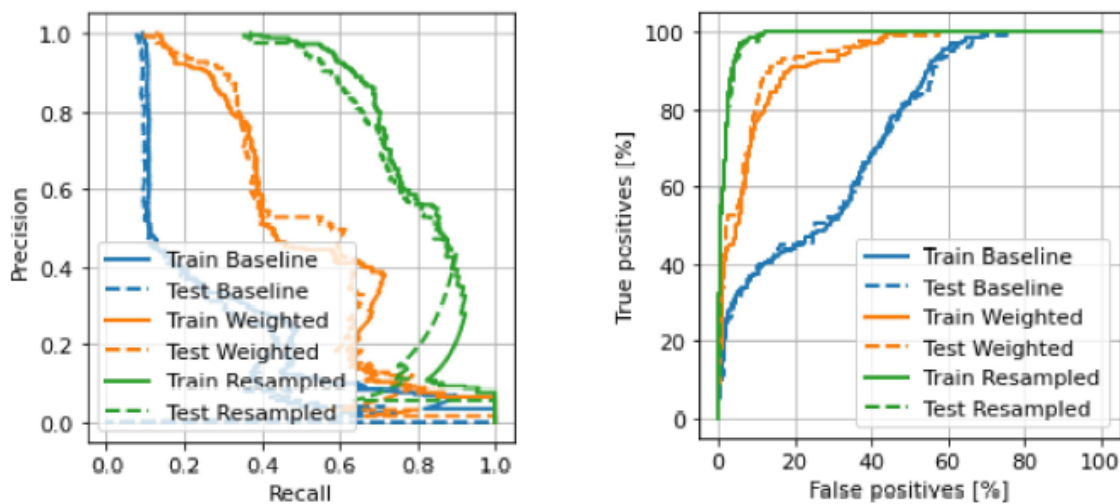


Fig. 24. Las curvas de las métricas de la región suroeste, ROC a la derecha y AUPRC a la izquierda.

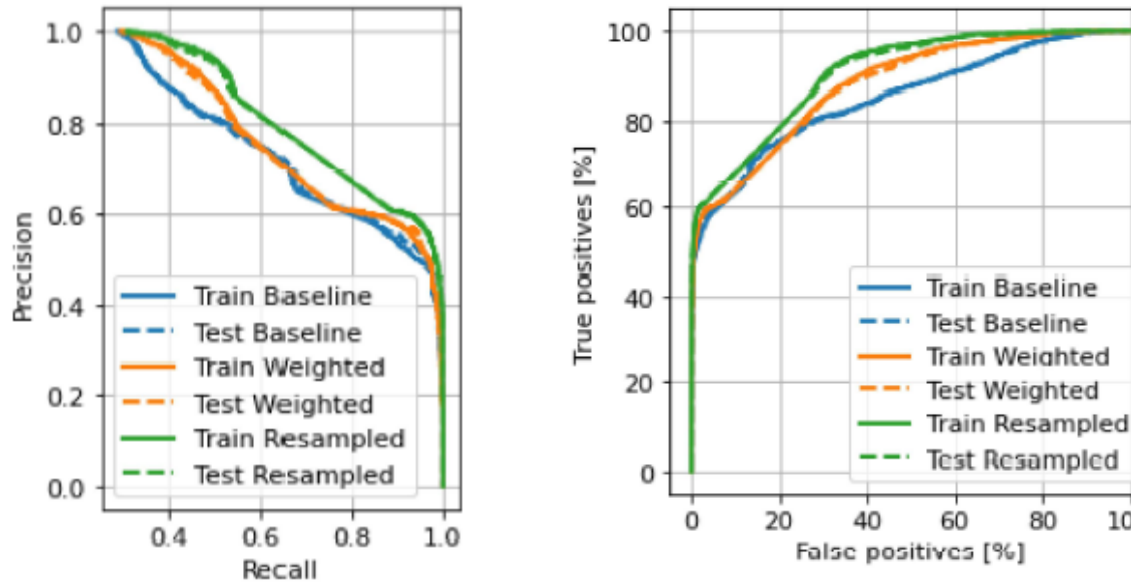


Fig. 25. Las curvas de las métricas de la región sureste, ROC a la derecha y AUPRC a la izquierda.

En cuanto a la sensibilidad, los modelos Weighted y los modelos Resampled en mayor grado lograron un mejor desempeño que el modelo Baseline.

La media fue de 65% para todos los modelos y 82% para el modelo preferido, tal desempeño puede considerarse exitoso considerando el dominio de la aplicación y el resto de los resultados. [21]

6. Resultados

El desempeño de los modelos se midió utilizando las mismas métricas de los otros modelos de predicción usando clasificación binaria.

6.1. Métricas

El desempeño del modelo se midió usando la muestra de prueba obtenida del 30% de los datos, y el entrenamiento se realizó sobre la muestra de entrenamiento obtenida del 70% de los datos (también se usó 20% de esta muestra de entrenamiento para validación), con las métricas, tp, fp, tn, fn, Accuracy, Precision/P1, Recall/Sn, FPR/Sp, NPV/P0, MCC, F1 Score y R score, son las métricas que se usan en la literatura como en [14] y [16], más otras métricas que se pueden usar para evaluar el desempeño de modelos clasificadores. Todas las métricas utilizadas y su forma de evaluación se encuentran en la Tabla 7.

A menudo es posible calibrar los parámetros de un modelo y mejorar los resultados de una métrica a expensas de otra. En este caso nos interesa reducir las falsas alarmas como ya se explicó anteriormente. El modelo tiene 3 variaciones que se superan entre sí en diferentes métricas. El propósito de analizar los resultados a través de estas métricas mencionadas es que cada desempeño métrico destaca cierto aspecto de los resultados. Por tanto, el propósito es resaltar todos los méritos y deméritos de los resultados obtenidos a través de las diferentes variaciones de modelos de predicción propuestos.

Tabla 7. Métricas de la red neuronal

Nombre y expresión	Descripción
True Positives $tp = \text{True Positives}$	Un terremoto mayor al umbral ocurrió y también fue predicho por el modelo.
False Positives $fp = \text{False Positives}$	No ocurrió un terremoto mayor al umbral pero fue predicho falsamente por el modelo.
True Negatives $tn = \text{True Negatives}$	No ocurrió un terremoto mayor al umbral y tampoco fue predicho por el modelo
False Negatives $fn = \text{False Negatives}$	Un terremoto mayor al umbral ocurrió pero no fue predicho por el modelo.
Exactitud $\text{Accuracy} = \frac{tp + tn}{tp + tn + fp + fn}$	Mide la frecuencia con la que el algoritmo clasifica un terremoto correctamente.
Negative predictive value / P0 $\text{NPV} = \frac{tn}{tn + fn}$	La probabilidad de que un terremoto predicho por debajo del umbral esté realmente por debajo de él.
Precision / P1 $\text{Precision} = \frac{tp}{tp + fp}$	La precisión y la sensibilidad se utilizan juntos, la precisión se define es el porcentaje de eventos sísmicos sobre el umbral predichos que se clasificaron correctamente.
Sensibilidad (recall) / Sn $\text{Recall} = \frac{tp}{tp + fn}$	La sensibilidad (recall) es el porcentaje de eventos sísmicos sobre el umbral que realmente sucedieron y que fueron clasificados correctamente.
False positive rate / Especificidad / Sp $\text{FPR} = \frac{fn}{tn + fp}$	La tasa de falsos positivos se calcula como la relación entre el número de veces que un terremoto se categoriza incorrectamente como por encima del umbral y el número total de terremotos por debajo del umbral (independientemente de la clasificación).
F1 Score $F = \frac{2 \text{ precision} \times \text{recall}}{\text{precision} + \text{recall}}$	Combina precisión y sensibilidad utilizando la media armónica.
MCC $\text{MCC} = \frac{(tp \times tn) - (fp \times fn)}{\sqrt{(tp + fp) \times (tp + fn) \times (tn + fp) \times (tn + fn)}}$	El coeficiente de correlación de Matthews generalmente se considera una de las mejores maneras de describir la matriz de confusión de verdaderos y falsos positivos y negativos con un solo número.
R score $\text{R score} = \frac{(tp \times tn) - (fp \times fn)}{(tp + tn) \times (fp + tn)}$	R es otra medida equilibrada para evaluar la matriz de confusión de verdaderos y falsos positivos y negativos con un solo número.
AUC	AUC se refiere al área bajo la curva de una curva característica de funcionamiento del receptor (ROC-AUC). Esta métrica es igual a la probabilidad de que un clasificador clasifique una muestra positiva aleatoria por encima de una muestra negativa aleatoria.
PRC	Se refiere al área bajo la curva de la curva de recuperación de precisión. Esta métrica calcula pares de recuperación de precisión para diferentes umbrales de probabilidad.

6.2. Resultados

Los 3 modelos son aplicados con los mismos parámetros para ambas regiones del suroeste y sureste.

Los resultados de las métricas y la matriz de confusión de la región del suroeste se puede observar respectivamente Fig. 26 y Tabla 8, y de la región del sureste en Fig. 27 y Tabla 9 de la misma manera. Los resultados de estos modelos no son directamente comparables con la de los otros estudios porque aquí se están usando ventanas que capturan más muestras y más similares en vez de divisiones de los datos completamente diferentes.

Decimos similares porque al usar ventanas de tamaño 100 la primera instancia va a ser 99% similar a la segunda instancia porque va a compartir 99 de los eventos sísmicos para calcular los indicadores, aún en la ventana número cincuenta esta seguirá compartiendo 50% de la información que la primera instancia, ya que compartirán 50 eventos sísmicos para calcular los indicadores.

En este caso, todas las matrices muestran que tiene relativamente muchos TN (verdadero negativo), lo que significa que hubo relativamente muchos eventos sísmicos con magnitud mayor a 5.5 considerados correctamente como no un gran terremoto.

Es probable que se desee tener menos falsos negativos a pesar del costo de aumentar el número de falsos positivos. Esta compensación puede ser preferible porque los falsos negativos significan que un terremoto no fue previsto y sucedió, mientras que los falsos positivos pueden hacer que se tomen medidas preventivas aunque no suceda el terremoto previsto.

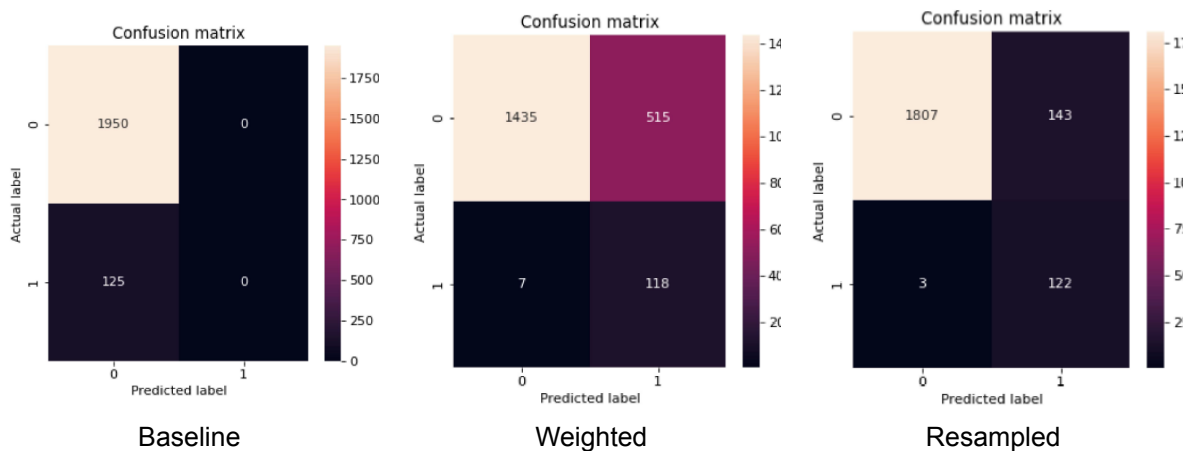


Fig. 26. Matriz de confusión del modelo Baseline, Weighted y Resampled al aplicarse en la región suroeste.

Tabla 8. Métricas del modelo de la red neuronal suroeste

Region Suroeste			
Métricas	Baseline	Weighted	Resampled
tp	0	118	122
fp	0	515	143
tn	1950	1435	1807
fn	125	7	3
Accuracy	0.93	0.74	0.92
P0 / NPV	0.93	0.99	0.99
P1 / Precision	0.00	0.18	0.46
Sn / Recall	0.00	0.94	0.97
Sp / FPR	1.00	0.73	0.92
R score	0.00	0.67	0.90
MCC	0.00	0.35	0.64

En el caso de la región del suroeste el modelo *Baseline* no realizó ninguna predicción positiva de terremotos, y aun así obtuvo un puntaje “alto” en algunas de sus métricas, a saber estas métricas fueron *Accuracy*, P_0 y FPR, estas métricas por lo general en la aplicación de predicciones de terremotos siempre serán relativamente altas, esto lo explica la ley de Gutenberg-Richter, siempre habrán muchos más terremotos pequeños que grandes y hace que los datos estén desbalanceados, significa que habrá más muestras de una clase que de otra (en este caso son más muestras negativas que positivas), y el modelo les prestará más “atención” a esas muestras de la clase con mayor exposición y más presente.

Los modelos *Weighted* y *Resampled* tratan de contrarrestar el problema de los datos desbalanceados usando las técnicas diseñadas para esto, se usaron pesos para clases y remuestreo respectivamente.

La mayoría de las métricas sólo evalúan el desempeño de un aspecto de las predicciones, debido a esto es ideal que se utilicen diferentes métricas para evaluar diferentes aspectos de las predicciones realizadas. Existen métricas que intentan evaluar el desempeño de un modelo considerando todos los ángulos como MCC.

El objetivo de construir diferentes modelos es que compitan entre sí comparando los resultados con sus métricas.

En esta región suroeste todos los modelos tuvieron una puntuación baja en la precisión y en MCC, la mejor puntuación de precisión fue del modelo *Resampled* con 46%, este resultado en esta área de aplicación de los métodos de inteligencia artificial se puede considerar un desempeño “exitoso”. [21]

El modelo *Weighted* y en mayor grado el modelo *Resampled* obtuvieron mejores resultados en las métricas, el modelo *Resampled* obtuvo un número similar de TP (predicciones efectivas) al modelo *Weighted*, pero logró incrementar el número de TN y reducir las predicciones FP y FN (predicciones erróneas). A excepción de Recall y NPV todas las demás métricas aumentaron considerablemente del modelo *Weighted* al modelo *Resampled*.

Debido a la pequeña cantidad de datos en esta región suroeste, en especial la pequeña cantidad de las muestras positivas, hay 7082 eventos sísmicos, 7073 son terremotos con magnitudes menores a 5.5 y 9 terremotos con magnitudes mayores a 5.5 (terremoto de gran magnitud), aun después de utilizar la ventana deslizante y calcular las instancias resultan 6915 instancias en total, 6470 instancias negativas y 445 instancias positivas. Esto resulta en una razón de 1 instancia positiva por cada 14 instancias negativas (1:14), esto no parece tan malo, pero se debe considerar que debido a las ventanas deslizantes unas múltiples instancias pueden estar relacionadas con el mismo evento sísmico de gran magnitud, si solo se considerarán los eventos sísmicos individuales habría un evento sísmico de magnitud mayor a 5.5 por cada 786 eventos sísmicos de magnitud menor a 5.5 (1:786). En el caso de la región sureste este problema es mucho menor porque las proporciones son menores.

A pesar de que el modelo no “conoce” los datos de prueba, con los que son obtenidas las métricas a evaluar, debido al uso de las ventanas deslizantes que hace que todas las muestras de datos se parezcan más, a las demás medidas tomadas en el modelo *Resampled* y *Weighted* (que le dan más valor estadístico a las muestras positivas para reducir el problema de los datos desbalanceados), y a la pequeña cantidad de muestras, especialmente muestras positivas; estos factores pueden actuar de forma contraproducente y producir un problema de falta de generalización, esto significa que al modelo le será difícil hacer predicciones para eventos con “patrones” diferentes, que pueden aparecer cuando los eventos están muy apartados temporalmente de las muestras con las que fue entrenado, esto se puede solucionar haciendo un entrenamiento periódico, para que la fecha actual nunca esté demasiado lejos de la fecha del último evento con el que fue entrenado; que los eventos sean de otra región no debe considerarse un problema, porque a pesar de que al modelo se le dificulte hacer predicciones sobre eventos de otra región sobre la que no fue entrenado, esta característica no es parte del alcance propuesto de estos modelos, los modelos de predicción de terremotos son entrenados para regiones específicas, y el mismo modelo no se usa para predecir eventos sísmicos de diferentes regiones.

Se considera que en este caso el modelo no tiene un número suficiente de muestras y en especial muestras positivas para poder ser juzgado efectivamente y los resultados de las métricas están comprometidos.

La solución sencilla a este problema sería aumentar el número de muestras (especialmente muestras positivas), a continuación se muestra la misma arquitectura de los modelos pero aplicados a una región con más eventos sísmicos, y se podrá observar como ahora el modelo se desempeña con conjunto de datos más favorable a continuación en la Fig. 27 y Tabla 9.

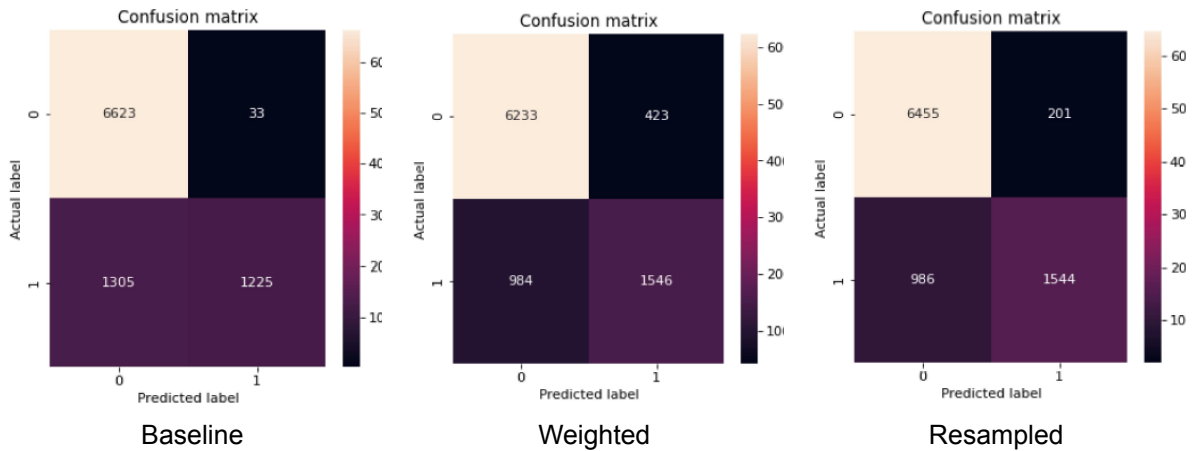


Fig. 27. Matriz de confusión del modelo Baseline, Weighted y Resampled al aplicarse en la región sureste.

Tabla 9. Métricas del modelo de la red neuronal sureste

Region Sureste			
Métricas	Baseline	Weighted	Resampled
tp	1225	1546	1544
fp	33	423	201
tn	6623	6233	6455
fn	1305	984	986
Accuracy	0.85	0.84	0.87
P0 / NPV	0.83	0.86	0.86
P1 / Precision	0.97	0.78	0.88
Sn / Recall	0.48	0.61	0.61
Sp / FPR	0.99	0.93	0.96
R score	0.47	0.54	0.58
MCC	0.62	0.59	0.66

El modelo *Baseline* aplicado sobre la región sureste se desempeña mejor que al aplicarlo sobre la de la región suroeste presumiblemente por tener mayor cantidad de datos. En algunas métricas en la región suroeste sigue siendo mejor que la región sureste pero se debe recordar que para juzgar efectivamente el desempeño de un modelo de predicción se deben considerar diferentes o todas las métricas en conjunto. En el suroeste el modelo no hizo ninguna predicción positiva mientras que en el sureste hizo aproximadamente el mismo número de predicciones positivas (positivo no quiere decir correcto, positivo significa que predijo que ocurriría un terremoto independientemente de si este ocurrió o no) que predicciones incorrectas.

Para el caso de los modelos *Weighted* y *Resampled* el desempeño general también aumentó, presumiblemente también por el aumento de la cantidad de datos.

Ambos modelos realizaron aproximadamente el mismo número de predicciones positivas correctas, y tienen la misma sensibilidad (*recall*) de 61%; pero el modelo *Resampled* logró hacerlo con un número menor de predicciones incorrectas lo que aumenta considerablemente su precisión, de 78% en el modelo *Weighted* a 88% en el modelo *Resampled*.

Ambos modelos *Weighted* y *Resampled* son medidas para contrarrestar los datos desbalanceados, ambos lo logran en cierta medida pero en ambas regiones el modelo *Resampled* se desempeñó mejor. Los modelos *Weighted* y *Resampled* logran aumentar el número de predicciones positivas correctas (TP) y reducir las predicciones incorrectas (FP y FN, esto al menos solo en la región del sureste).

Las métricas de diferentes estudios no son directamente comparables, a diferencia de otros estudios, aquí usamos una ventana deslizante de un tamaño fijo para calcular las instancias que se introducirán en el modelo, otros estudios dividen el conjunto de datos en fragmentos de un tamaño fijo para calcular los casos, esto se traduce en casi la mitad de los casos y menos sensibilidad a los cambios de indicador.

Conclusión

En este análisis interdisciplinario, la predicción de terremotos se ha realizado mediante la interacción de precursores de terremotos y técnicas de aprendizaje automático computacional.

Se presentaron tres modelos de predicción de terremotos a través de las relaciones entre indicadores sísmicos y técnicas computacionales de Machine Learning. De los tres modelos modelados, el modelo *Resampled* se desempeñó mejor en esta tarea ya que predice más eventos sísmicos por encima del umbral sin sacrificar significativamente las otras métricas, dando los resultados más deseables de los tres modelos, tiene un desempeño de 0.61 hasta 0.96 en diferentes métricas, es la variación del modelo más adecuado para este tipo de predicción.

En el modelo *Resampled* se enfrentó el problema como un problema de clasificación utilizando series de tiempo con una ventana deslizante y se trató también como un problema de datos desbalanceados, estas consideraciones lograron que los modelos *Weighted* y en mayor grado *Resampled* lograran un mejor desempeño que el modelo *Baseline*. Estas consideraciones hacen que el modelo *Resampled* sea comparable con el de otros modelos de predicción de otros estudios a pesar de ser una red neuronal simple.

En el futuro, se habrán hecho más progresos y habrá más datos disponibles, una propuesta para trabajo futuro sería añadir conceptos de series de tiempo para considerar múltiples ventanas a la vez; el aprendizaje automático ha demostrado ser muy útil en muchos campos y se están realizando muchos esfuerzos para adecuar el uso del aprendizaje automático para la predicción de terremotos.

Referencias

1. Asim KM, Martínez-Álvarez F, Basit A, Iqbal T, (2017) Earthquake magnitude prediction in Hindukush region using machine learning techniques. *Hazards*, vol. 85, no. 1, pp 471–486
2. Robert D. Cicerone, John E. Ebel, James Britton (2009) A systematic compilation of earthquake precursors, *Tectonophysics*, vol. 476, Issues 3–4, pp 371-396
3. Lomnitz C (1994) *Fundamentals of Earthquake Prediction*. John Wiley & Sons, New York. pp 326
4. Géli L, Piau JM, Dziak R, Maury M, Fitzenz D, Coutellier Q, Henry P (2014) Seismic precursors linked to super-critical fluids at oceanic transform faults. *Nature Geoscience*
5. Mignan, A., and Woessner, J. 2012. Estimating the magnitude of completeness for earthquake catalogs. *Community Online Resource for Statistical Seismicity Analysis*, 4: 1-45
6. Arroyo, Mario; Godínez, Kevin; Linkimer, Lepolt COMPLETITUD DEL CATÁLOGO DE LA RED SISMOLÓGICA NACIONAL DE COSTA RICA DURANTE 1975-2014 *Boletín de Geología*, vol. 39, núm. 3, septiembre-diciembre, 2017, pp. 87-98 Universidad Industrial de Santander Bucaramanga, Colombia
7. Chouliaras, G. 2009. Investigating the earthquake catalog of the National Observatory of Athens. *Natural Hazards and Earth System Sciences*, 9: 905-912.
8. Mignan A, Woessner J (2012) Estimating the magnitude of completeness for earthquake catalogs. *Community Online Resource for Statistical Seismicity Analysis*, no. 4, pp 1-45
9. Wiemer, S., and Wyss, M. 2000. Minimum magnitude of completeness in earthquake catalogs: Examples from Alaska, the Western United States, and Japan. *Bulletin of the Seismological Society of America*, 90(4): 859-869.
10. Rudolf-Navarro, A., Muñoz-Diosdado, A., & Angulo-Brown, F. (2010). Seismic quiescence patterns as possible precursors of great earthquakes in México.
11. Gutenberg, B., & Richter, C. (1944). Frequency of earthquake in California. *Bulletin of the Seismological Society of America*, 34(4), 185–188.
12. Wu, Y.-M., Chen, S. K., Huang, T.-C., Huang, H.-H., Chao, W.-A., & Koulakov, I. (2018). Relationship between earthquake b-values and crustal stresses in a young orogenic belt. *Geophysical Research Letters*, 45, 1832– 1837. <https://doi.org/10.1002/2017GL076694>
13. Smith, W. D. (1981). The b-value as an earthquake precursor. *Nature*, 289(5794), 136–139. <https://doi.org/10.1038/289136a0>
14. Khawaja M. Asim, Adnan Idris, Talat Iqbal, Francisco Martínez-Álvarez, Seismic indicators based earthquake predictor system using Genetic Programming and AdaBoost classification, *Soil Dynamics and Earthquake Engineering*, Volume 111, 2018, Pages 1-7, ISSN 0267-7261, <https://doi.org/10.1016/j.soildyn.2018.04.020>.

15. Earthquake magnitude prediction in Hindukush region using machine learning techniques, K. M. Asim and F. Martínez-Álvarez and A. Basit and T. Iqbal, *Natural Hazards*, 2016, v 85, p 471-486
16. Asim KM, Idris A, Iqbal T, Martínez-Álvarez F (2018) Earthquake prediction model using support vector regressor and hybrid neural networks. *PLoS ONE* 13(7): e0199004. <https://doi.org/10.1371/journal.pone.0199004>
17. Wiemer S, Wyss M. Minimum magnitude of completeness in earthquake catalogs: examples from Alaska, the western United States, and Japan. *Bull Seismol Soc Am* 2000;90:859–69.
18. Salam, M.A., Ibrahim, L.A., & Abdelminaam, D. (2021). Earthquake Prediction using Hybrid Machine Learning Techniques.
19. Marsan, D., and M. Wyss (2011), Seismicity rate changes, Community OnlineResource for Statistical Seismicity Analysis, doi:10.5078/corssa-25837590. Available at <http://www.corssa.org>.
20. Panakkat A, Adeli H. Neural network models For earthquake magnitude prediction using multiple seismicity indicators. *International Journal of Neural Systems*. 2007; 17(01):13±33. <https://doi.org/10.1142/S0129065707000890> PMID: 17393560
21. Reyes J, Morales-Esteban A, Martínez-Álvarez F. Neural networks to predict earthquakes in Chile. *Applied Soft Computing*. 2013; 13(2):1314±28.
22. Matthews MV, Reasenberg PA. Statistical methods for investigating quiescence and other temporal seismicity patterns. *Pure and Applied Geophysics*. 1988; 126(2±4):357±72.
23. Habermann R. Precursory seismic quiescence: past, present, and future. *Pure and Applied Geophysics*. 1988; 126(2±4):279±318.
24. Morales-Esteban A, Martínez-Álvarez F, Reyes J. Earthquake prediction in seismogenic areas of the Iberian Peninsula based on computational intelligence. *Tectonophysics*. 2013; 593:121±34. <http://dx.doi.org/10.1016/j.tecto.2013.02.036>.
25. Galkina, A., & Grafeeva, N. (2019). Machine learning methods for earthquake prediction: A survey.
26. Mohsen Yousefzadeh, Seyyed Ahmad Hosseini, Mahdi Farnaghi, Spatiotemporally explicit earthquake prediction using deep neural network, *Soil Dynamics and Earthquake Engineering*, Volume 144, 2021, 106663, ISSN 0267-7261, <https://doi.org/10.1016/j.soildyn.2021.106663>. (<https://www.sciencedirect.com/science/article/pii/S0267726121000853>)
27. T. L. Chin, C. Y. Huang, S. H. Shen, Y. C. Tsai, Y. H. Hu, and Y. M. Wu, “Learn to Detect: Improving the Accuracy of Earthquake Detection,” *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 8867–8878, 2019, doi: 10.1109/TGRS.2019.2923453.
28. Yeck, W. (2021, May 12). Improving Earthquake Monitoring with Deep Learning. *Usgs*. https://www.usgs.gov/natural-hazards/earthquake-hazards/science/improving-earthquake-monitoring-deep-learning?qt-science_center_objects=0#qt-science_center_objects

29. S. M. Mousavi, Y. Sheng, W. Zhu, and G. C. Beroza, "Stanford Earthquake Dataset (STEAD): A Global Data Set of Seismic Signals for AI," *IEEE Access*, vol. 7, pp. 179464–179476, 2019, doi: 10.1109/ACCESS.2019.2947848.
30. E. I. Alves, "Earthquake forecasting using neural networks: results and future work," *Nonlinear Dynamics*, vol. 44(1-4), pp. 341-349, 2006.
31. M. Maya and W. Yu, "Short-term prediction of the earthquake through neural networks and meta-learning," 2019 16th Int. Conf. Electr. Eng. Comput. Sci. Autom. Control. CCE 2019, pp. 1–6, 2019, doi: 10.1109/ICEEE.2019.8884562.
32. Tsuneji Rikitake, Kazuo Hamada, *Earthquake Prediction*, Editor(s): Robert A. Meyers, *Encyclopedia of Physical Science and Technology (Third Edition)*, Academic Press, 2003, Pages 743-760, ISBN 9780122274107, <https://doi.org/10.1016/B0-12-227410-5/00195-2>.
(<https://www.sciencedirect.com/science/article/pii/B0122274105001952>)
33. A. Vahaplar, B. T. Tezel, R. Nasiboglu, and E. Nasibov, "A monitoring system to prepare machine learning data sets for earthquake prediction based on seismic-acoustic signals," 9th Int. Conf. Appl. Inf. Commun. Technol. AICT 2015 - Proc., pp. 44–47, 2015, doi: 10.1109/ICAICT.2015.7338513.
34. H. Adeli and A. Panakkat, "A probabilistic neural network for earthquake magnitude prediction," *Neural Networks*, vol. 22, no. 7, pp. 1018–1024, 2009, doi: 10.1016/j.neunet.2009.05.003.
35. M. Moustra, M. Avraamides, and C. Christodoulou, "Artificial neural networks for earthquake prediction using time series magnitude data or Seismic Electric Signals," *Expert Syst. Appl.*, vol. 38, no. 12, pp. 15032–15039, 2011, doi: 10.1016/j.eswa.2011.05.043.
36. Nayak DR, Mahapatra A, Mishra P. A survey on rainfall prediction using artificial neural network. *International Journal of Computer Applications*. 2013; 72(16).
37. Idris A, Rizwan M, Khan A. Churn prediction in telecom using Random Forest and PSO based data balancing in combination with various feature selection strategies. *Computers & Electrical Engineering*. 2012; 38(6):1808–19. <http://dx.doi.org/10.1016/j.compeleceng.2012.09.001>.
38. Cosma G, Brown D, Archer M, Khan M, Pockley AG. A survey on computational intelligence approaches for predictive modeling in prostate cancer. *Expert systems with applications*. 2017; 70:1– 19
39. FRANCO, Sara I.; CANET, Carles; IGLESIAS, Arturo y VALDES-GONZALEZ, Carlos. Seismic activity in the Gulf of México. A preliminary analysis. *Bol. Soc. Geol. Mex [online]*. 2013, vol.65, n.3, pp.447-455. ISSN 1405-3322.
40. Melgar, D., Pérez-Campos, X., Ramirez-Guzman, L., Spica, Z., Espíndola, V. H., Hammond, W. C., & Cabral-Cano, E. (2018). Bend faulting at the edge of a flat slab: The 2017 Mw7.1 Puebla-Morelos, México Earthquake. *Geophysical Research Letters*, 45, 2633– 2641. <https://doi.org/10.1002/2017GL076895>

41. SSN, Últimos sismos: Universidad Nacional Autónoma de México, Instituto de Geofísica, Servicio Sismológico Nacional, México. Extraído el 9 de Septiembre de 2021 desde <http://www.ssn.unam.mx/sismicidad/ultimos>
42. K. J. Bergen, P. A. Johnson, M. V. de Hoop, G. C. Beroza, Machine learning for data-driven discovery in solid Earth geoscience. *Science* 363, eaau0323 (2019).
43. B. Moseley, L. Krischer J. S. Dramsch, “Chapter one - 70 years of machine learning in geoscience in review” in *Advances in Geophysics*, B. Moseley, L. Krischer, Eds. (Elsevier, 2020), vol. 61, pp. 1–55.
44. Q. Kong et al., Machine learning in seismology: Turning data into insights. *Seismol. Res. Lett.* 90, 3–14 (2019).
45. C. E. Yoon, O. O’Reilly, K. J. Bergen, G. C. Beroza, Earthquake detection through computationally efficient similarity search. *Sci. Adv.* 1, e1501057 (2015).
46. S. M. Mousavi, W. Zhu, Y. Sheng, G. C. Beroza, Cred: A deep residual network of convolutional and recurrent units for earthquake signal detection. *Sci. Rep.* 9, 10267 (2019).
47. T. Zhao, F. Li, K. J. Marfurt, Constraining self-organizing map facies analysis with stratigraphy: An approach to increase the credibility in automatic seismic facies classification. *Interpretation* 5, T163–T171 (2017).
48. M. J. Cracknell, A. M. Reading, Geological mapping using remote sensing data: A comparison of five machine learning algorithms, their response to variations in the spatial distribution of training data and the use of explicit spatial information. *Comput. Geosci.* 63, 22–33 (2014).
49. A. Lomax, A. Michelini, D. Jozinović, An investigation of rapid earthquake characterization using single-station waveforms and a convolutional neural network. *Seismol Res. Lett.* 90, 517–529 (2019).
50. S. Valade et al., Toward global volcano monitoring using multisensor sentinel missions and artificial intelligence: The MOUNTS monitoring system. *Rem. Sens.* 11, 1528 (2019).
51. C. X. Ren et al., Machine learning reveals the seismic signature of eruptive behavior at Piton de la Fournaise volcano. *Geophys. Res. Lett.* 47, e2019GL085523 (2020).
52. N. Anantrasirichai, J. Biggs, F. Albino, D. Bull, A deep learning approach to detecting volcano deformation from satellite imagery using synthetic datasets. *Remote Sens. Environ.* 230, 111179 (2019).
53. B. Rouet-Leduc, C. Hulbert, I. W. McBrearty, P. A. Johnson, Probing slow earthquakes with deep learning. *Geophys. Res. Lett.* 47, (2020).
54. N. Anantrasirichai, J. Biggs, F. Albino, D. Bull, The application of convolutional neural networks to detect slow, sustained deformation in InSAR time series. *Geophys. Res. Lett.* 46, 11850–11858 (2019).
55. S. Diersen, E. J. Lee, D. Spears, P. Chen, L. Wang, “Classification of seismic windows using artificial neural networks” in *Proceedings of the International Conference on Computational Science, ICCS 2011 (Procedia Computer Science* 4, Elsevier B. V., Amsterdam, Netherlands, 2011), vol. 4, pp. 1572–1581.

56. Y. Wu, Y. Lin, Z. Zhou, *Inversionet: Accurate and Efficient Seismic-Waveform Inversion with Convolutional Neural Networks* (Society of Exploration Geophysicists, 2018), pp. 2096–2100.
57. S. Karra, D. O'Malley, J. D. Hyman, H. S. Viswanathan, G. Srinivasan, Modeling flow and transport in fracture networks using graphs. *Phys. Rev. E* 97, 033304 (2018).
58. A. Alhakeem, "3D seismic attribute analysis and machine learning for reservoir characterization in Taranaki Basin, New Zealand," PhD thesis, Missouri Institute of Mining and Technology, Rolla, Missouri (2018).
59. B. Rouet-Leduc et al., Machine learning predicts laboratory earthquakes. *Geophys. Res. Lett.* 44, 9276–9282 (2017).
60. C. Hulbert et al., Similarity of fast and slow earthquakes illuminated by machine learning. *Nat. Geosci.* 12, 69–74 (2019).
61. B. Rouet-Leduc, C. Hulbert, P. A. Johnson, Continuous chatter of the Cascadia subduction zone revealed by machine learning. *Nat. Geosci.* 12, 908–1752 (2019).
62. F. Corbi et al., Predicting imminence of analog megathrust earthquakes with machine learning: Implications for monitoring subduction zones. *Geophys. Res. Lett.* 47, e2019GL086615 (2020).
63. B. Moseley, L. Krischer C. X. Ren, C. Hulbert, P. A. Johnson, B. Rouet-Leduc, "Chapter two - machine learning and fault rupture: A review" in *Advances in Geophysics*, B. Moseley, L. Krischer, Eds. (Elsevier, 2020), vol. 61, pp. 57–107.
64. C. Hulbert, B. Rouet-Leduc, R. Jolivet, P. Johnson, An exponential build-up in seismic energy suggests a months-long nucleation of slow slip in Cascadia. *Nat. Commun.* 11, 4139 (2020).
65. F. Corbi et al., Machine learning can predict the timing and size of analog earthquakes. *Geophys. Res. Lett.* 46, 1303–1311 (2019).
66. N. Lubbers et al., Earthquake catalog-based machine learning identification of laboratory fault states and the effects of magnitude of completeness. *Geophys. Res. Lett.* 45, 13.269–13.276 (2018).
67. U.S. Geological Survey, 2021, United States Geological Survey, accessed September 14, 2021 at URL <https://www.usgs.gov/>
68. Silva, L. M. (2019). *Crónica De Seis Siglos De Sismos En México: lecciones aprendidas y perspectivas*. Asociación Mexica
69. G. Cortés, A. Morales-Esteban, X. Shang, and F. Martínez-Álvarez, "Earthquake Prediction in California Using Regression Algorithms and Cloud-based Big Data Infrastructure," *Computers & Geosciences*, vol. 115, pp. 198-210, 2018.
70. K. Wang, Q. Chen, S. Sun, and A. Wang, "Predicting the 1975 Haicheng earthquake," *Bulletin of the Seismological Society of America*, vol. 96, pp. 757–795, 2006.
71. W. Marzocchi, and J. D. Zechar, "Earthquake forecasting and earthquake prediction: different approaches for obtaining the best model," *Seismological Research Letters*, vol. 82(3), pp. 442–448, 2011.

72. D. D. Jackson, "Hypothesis testing and earthquake prediction," *Proc. Natl. Aca. Sci. USA*, vol. 93, pp. 3772–3775, 1996.
73. M. Bath, "Lateral inhomogeneities in the upper mantle," *Tectonophysics*, vol. 2, pp. 483–514, 1965.
74. T. Utsu, "A statistical study of the occurrence of aftershocks," *Geophysical Magazine*, vol. 30, pp. 521–605, 1961.
75. Florido, Emilio & Aznarte, José & Morales-Esteban, Antonio & Martínez-Álvarez, Francisco. (2016). Croatian Operational Research Review Earthquake magnitude prediction based on artificial neural networks: A survey. *Croatian Operational Research Review*. 7. 159-169. 10.17535/corr.2016.0011.
76. M. Monterrubio-Velasco, J. C. Carrasco-Jimenez, O. Castillo-Reyes, F. Cucchiatti, and J. De La Puente, "A Machine Learning Approach for Parameter Screening in Earthquake Simulation," *Proc. - 2018 30th Int. Symp. Comput. Archit. High Perform. Comput. SBAC-PAD 2018*, pp. 348–355, 2019, doi: 10.1109/CAHPC.2018.8645865.
77. L. MacEda, J. Llovido, and A. Satuito, "Categorization of earthquake related tweets using machine learning approaches," *Proc. - 2018 Int. Symp. Comput. Consum. Control. IS3C 2018*, no. May 2011, pp. 229– 232, 2019, doi: 10.1109/IS3C.2018.00065.
78. G. Rajguru, Y. S. Bhadauria, and S. Mukhopadhyay, "Estimation of Earthquake Source Parameters Using Machine Learning Techniques," *2018 9th Int. Conf. Comput. Commun. Netw. Technol. ICCCNT 2018*, pp. 1–7, 2018, doi: 10.1109/ICCCNT.2018.8493922.
79. F. Martínez-Álvarez, J. Reyes, A. Morales-Esteban, and C. RubioEscudero, "Determining the best set of seismicity indicators to predict earthquakes. Two case studies: Chile and the Iberian Peninsula," *Knowledge-Based Syst.*, vol. 50, pp. 198–210, 2013, doi: 10.1016/j.knosys.2013.06.011.
80. A. S. N. Alarifi, N. S. N. Alarifi, and S. Al-Humidan, "Earthquakes magnitude prediction using artificial neural network in northern Red Sea area," *J. King Saud Univ. - Sci.*, vol. 24, no. 4, pp. 301–313, 2012, doi: 10.1016/j.jksus.2011.05.002.
81. Telesca, L., & Yu, W. (n.d.). Development of machine LEARNING-BASED innovative methods of seismic forecasting in Italy and México. Development of machine learning-based innovative methods of seismic forecasting in Italy and México | Consiglio Nazionale delle Ricerche. Retrieved September 27, 2021, from <https://www.cnr.it/en/bilateral-agreements/project/3243/development-of-machine-learning-based-innovative-methods-of-seismic-forecasting-in-italy-and-México>.
82. Jure Žalohar, Chapter 14 - Gutenberg-Richter's Law, Editor(s): Jure Žalohar, *Developments in Structural Geology and Tectonics*, Elsevier, Volume 2, 2018, Pages 173-178, ISSN 2542-9000, ISBN 9780128145807, <https://doi.org/10.1016/B978-0-12-814580-7.00014-9>.
83. Pappas, S. (2012, October 22). Italian scientists sentenced to 6 years for earthquake statements. *Scientific American*. Retrieved October 5, 2021, from <https://www.scientificamerican.com/article/italian-scientists-get/>.

84. Terremoto, México, 1985 Conmemoración del terremoto de México de 1985. Gobierno de México. (n.d.). Retrieved October 6, 2021, from <https://www.gob.mx/siap/es/articulos/terremoto-México-1985?idiom=es>.
85. Y.Y. Kagan, VAN earthquake predictions – an attempt at statistical evaluation, *Geophysical Research Letters* 23 (11) (1996) 1315–1318.
86. J.D. Zechar, T.H. Jordan, Testing alarm-based earthquake predictions, *Geophysical Journal International* 172 (2008) 715–724.
87. Juan Villoro, 8.8: El miedo en el espejo. Almadía, 2010