



**BENEMÉRITA UNIVERSIDAD
AUTÓNOMA DE PUEBLA**



Reporte Fase 1: Descubrimiento del Proyecto

INTRODUCCIÓN A LA CIENCIA DE DATOS

Estimación Probabilística de Terremotos Basada en Datos Históricos

**Introducción a la Ciencia de Datos
Jaime Alejandro Romero Sierra**

**Samuel Naranjo Madrigal
202470537**

21.10.24

ÍNDICE

Objetivo del Proyecto_____	3
Descripción del Proyecto_____	4
Recursos Disponibles_____	5
Hipótesis Iniciales_____	7
Definición de Stakeholders Clave_____	8
Preguntas Clave_____	9
Fuentes de Datos Identificadas_____	10
Justificación del Proyecto_____	11
Cantidad y Tipos de Datos_____	12



Objetivo del Proyecto

Crear un Modelo que muestre un mapa con los lugares alrededor del mundo que son más propensos a sufrir un terremoto en los próximos años.

Esto con el fin de tomar medidas de prevención y mitigación de riesgos, actuar en casos de emergencia y mejorar la infraestructura y servicios de aquellos lugares que podrían sufrir un evento de este estilo. Además se busca identificar las zonas con mayor actividad sísmica de los últimos años para realizar análisis científicos que permitan descubrir patrones y comportamientos, con el fin de que estos estudios nos acerquen al siguiente avance de la ciencia sismológica: la capacidad de predecir un evento con gran precisión y anticipación.



Descripción del Proyecto

Los terremotos, protagonistas de numerosas tragedias a lo largo de la historia, continúan siendo un fenómeno natural cuyo estudio y comprensión son cruciales para mitigar sus devastadores efectos. Aunque la tecnología actual aún no permite predecir con exactitud cuándo y dónde ocurrirá un terremoto, existen otras herramientas valiosas para reducir los riesgos. Una de ellas es la *“estimación probabilística”*, un método que permite calcular la probabilidad de que un sismo ocurra en el futuro, basado en datos históricos y patrones sísmicos.

Este proyecto desarrolla precisamente esa metodología, utilizando datos recopilados durante los últimos 33 años. Con esta información, es posible estimar la probabilidad de que un terremoto ocurra en una región dentro de un rango temporal determinado. El objetivo es proporcionar un análisis de las tendencias sísmicas que permita a los gobiernos, científicos y comunidades tomar medidas preventivas.

Además, este método ayuda a identificar zonas de gran actividad sísmica, que con un estudio constante en los próximos años, podría contribuir significativamente a avances en la predicción de terremotos, lo que reduciría aún más el impacto de estos desastres.

Recursos Disponibles

Tecnología y herramientas

Para el trabajo se estará utilizando *Visual Studio Code (VSC)*, *Jupyter notebooks* y el lenguaje de programación “*Python*” junto a ciertas librerías como: “*Pandas*”, que permite visualizar y manipular la base de datos a utilizar; “*Matplotlib*” y “*Seaborn*”, las cuales serán de gran utilidad para la representación de datos en forma gráfica, y así, de manera preliminar, mostrar un conocimiento más específico para lograr comprender los datos a mayor profundidad.

Los datos utilizados son proporcionados por la comunidad de *Kaggle*, y extraídos de la “*United States Geological Survey (USGS)*”. Cuenta con datos de los terremotos registrados alrededor del mundo desde 1990 hasta 2023 (Últimos 33 años).

Es posible que el proyecto utilice *Power BI* en la fase de Dashboards. Además, para el modelo final y el mapa interactivo, esta considerado *Quantum GIS*, sin embargo *Python* es el enfoque principal.

Datos

Los datos proporcionados por la base de datos son los siguientes:

- **Time:** Fecha y hora del suceso representada en tiempo Unix.
- **Place:** Lugar (País, Estado o Región) donde ocurrió el evento.
- **Status:** Indica si el evento ha sido o no revisado.
- **Significance:** Muestra en una escala numérica la relevancia del suceso (Entre mayor el número, mayor relevancia).
- **Tsunami:** Indica de manera booleana si el evento generó o no, un tsunami.

- **Data Type:** Indica si es un terremoto, una explosión minera o algo más.
- **Magnitudo:** Expresa la intensidad del terremoto en escala de Richter.
- **State:** Muestra el estado del país donde sucedió el evento.
- **Longitude:** Coordenadas geográficas X.
- **Latitude:** Coordenadas geográficas Y.
- **Depth:** Indica la profundidad del suelo donde se generó el terremoto.
- **Date:** Cuándo y a qué hora se registró o sucedió el terremoto.



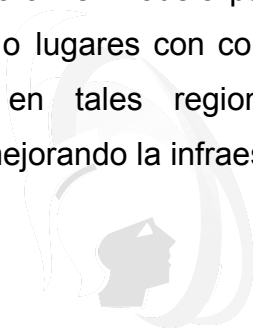
Hipótesis Iniciales

1. Las zonas cercanas a placas tectónicas son más propensas a sufrir mayor cantidad de terremotos, y de mayor magnitud.
2. La mayoría de los terremotos son casi imperceptibles por su bajo nivel de fuerza. Son pocos los significativos, los que causan mayores daños.
3. La profundidad y la ubicación del epicentro de un terremoto tiene un gran impacto en sus consecuencias: entre mayor profundidad y más alejado de la civilización, menor será su efecto o relevancia.



Definición de Stakeholders Clave

- **Población General:** Cualquier persona puede revisar el modelo para obtener información sobre las probabilidades de que ocurra un terremoto en cualquier lugar del planeta, todo depende del interés y la curiosidad por el tema.
- **Científicos sismológicos:** Podrán utilizar el modelo para identificar los lugares de mayor actividad sísmica y realizar una observación y análisis de patrones o comportamientos para el estudio y desarrollo de un modelo de predicción a futuro.
- **Gobiernos mundiales:** Usarán el modelo para conocer los lugares con mayor riesgo a un posible sismo o lugares con constante actividad. Esto para tomar medidas de prevención en tales regiones, educando a la población, proporcionando servicios, mejorando la infraestructura, etc.



Preguntas Clave

- ¿Qué partes del mundo sufren más sismos?
- ¿Dónde se han presentado los sismos de mayor magnitud?
- ¿Qué medidas de prevención son tomadas en zonas con alta actividad?
- ¿Qué nivel de la escala Richter se considera significativo?
- ¿Qué choques de placas son más comunes?
- ¿Cuál es el promedio de magnitudes por año y en los últimos 33 años?
- ¿Cuál es el promedio de la profundidad del epicentro?
- ¿Qué instrumentos se usan para detectar terremotos?
- ¿Cómo va el desarrollo de la tecnología de predicción?
- ¿Qué tan común es un tsunami en un terremoto?
- ¿Ha habido damnificados en un evento? ¿cuáles y cuántos?
- ¿Qué tanto afectan los terremotos a la vida humana?
- ¿En qué año se registraron más eventos?
- ¿En qué año se registraron menos eventos?

Fuentes de Datos Identificadas

- Datos de latitud, longitud y lugar.
- Información de placas tectónicas, localización y movimiento.
- Información de daños y pérdidas en terremotos significativos.
- Datos de escala richter.
- Datos de profundidad del epicentro.
- Información de civilización cerca de placas tectónicas.
- Información de medidas preventivas de distintos países en relación a sismos.
- Datos de tsunamis.



Justificación del Proyecto

Estudiando los datos históricos, tendencias y patrones de terremotos, se pueden identificar zonas sísmicas de alta actividad y desarrollar modelos predictivos para pronosticar futuros sismos. También es importante para la implementación de medidas preventivas como infraestructuras más resistentes, reduciendo daños materiales y protegiendo vidas humanas.

Por otro lado, el análisis de los registros históricos permite comprender aún más los procesos geológicos que conducen a la actividad sísmica. Analizando patrones tectónicos, se consigue identificar actividad sísmica estacional o a largo plazo, comprendiendo el comportamiento de las placas.

Finalmente, el público también puede acceder a la información del modelo, permitiéndoles observar la actividad sísmica en regiones de su interés, promoviendo la conciencia y la preparación ante cualquier evento sísmico.

El modelo de estimación probabilística será una gran herramienta de información para avances científicos, prevención de desastres y conciencia social sobre los terremotos alrededor de todo el mundo.

Cantidad y Tipos de Datos

La base de datos utilizada para el análisis cuenta con 3,563,492 de terremotos alrededor del mundo, registrados en un lapso de 33 años (1990 - 2023).

Cada sismo cuenta con 12 parámetros que proporcionan información del mismo, los datos representados en estos parámetros son de dos tipos: objetos y flotantes.

0	time	object
1	place	object
2	status	object
3	tsunami	float64
4	significance	float64
5	data_type	float64
6	magnitudo	float64
7	state	object
8	longitude	float64
9	latitude	float64
10	depth	object
11	date	object