

Monitoreo del nivel de prevención ante sismos por distritos del Perú*

*Profesora: Soledad Espezúa Llerena

Amy Antalu Checcullo Huachua
Ingeniería de la Información
Universidad del Pacífico
Lima, Perú
aa.checculloh@alum.up.edu.pe

Miguel Alexis Muñoz Chavez
Ingeniería de la Información
Universidad del Pacífico
Lima, Perú
ma.munozc@alum.up.edu.pe

Alexia Shariann Ríos Alarcón
Ingeniería de la Información
Universidad del Pacífico
Lima, Perú
as.riosa@alum.up.edu.pe

Alexandra Constanza Sanjinez Mendoza
Ingeniería de la Información
Universidad del Pacífico
Lima, Perú
ac.sanjinezm@alum.up.edu.pe

Abstract—Los sismos se caracterizan por ser impredecibles y destructivos; por ello, se busca implementar medidas preventivas para apaciguar los posibles efectos negativos que puedan ocasionar. Actualmente el Perú, se ubica en el puesto 13 del ranking de países más vulnerables a desastres naturales en América Latina, siendo caracterizado por la falta de capacidad y adaptabilidad del Estado para enfrentar desastres. Por ello, el objetivo de este proyecto es evaluar el nivel de prevención ante sismos de los distritos del Perú para brindar recomendaciones para la toma de decisiones y la mejora de estrategias de prevención. Para lograr esto, se emplearon datos del periodo 2012 - 2022 los cuales incluyen información sobre actividad sísmica histórica, tipos de viviendas, tipos de suelos e infraestructura de salud. Estos datos se centralizaron en solo un conjunto de datos para su posterior análisis a través de técnicas de Data Mining y Machine Learning. En cuanto a resultados, se obtuvieron tres grupos los cuales demuestran que el nivel de prevención ante movimientos sísmicos en el Perú es desigual. Al aplicar los modelos de Machine Learning se obtuvo que el mejor modelo fue ExtraTrees con una acuracia de 0,892.

Index Terms—Sismos, prevención, clustering, suelos, distritos.

I. INTRODUCCIÓN

Los sismos son eventos naturales que representan la liberación de energía acumulada en el interior de la Tierra, manifestándose en forma de vibraciones que pueden alcanzar diferentes magnitudes [1]. Estos fenómenos geológicos, a menudo impredecibles y destructivos, requieren una atención constante en términos de prevención y preparación para minimizar sus impactos.

El Perú, un país situado en una región geológicamente activa, se encuentra en el puesto 13 del ranking de naciones más vulnerables ante desastres naturales [13]. Esta vulnerabilidad no solo se limita a los sismos, sino que abarca una serie de amenazas naturales, como inundaciones, deslizamientos de

tierra y eventos climáticos extremos. Esta situación demanda una respuesta efectiva y un enfoque proactivo en la prevención y mitigación de estos eventos para salvaguardar vidas y bienes.

El objetivo principal de este trabajo es evaluar el nivel de prevención ante sismos de los distritos del Perú para brindar recomendaciones para la toma de decisiones y la mejora de estrategias de prevención. Para lograr este cometido, es imperativo no sólo reaccionar ante las consecuencias de estos eventos, sino también adoptar un enfoque integral que priorice la prevención y la preparación en todos los niveles de gobierno y en cada rincón del país. El conocimiento sobre el nivel de vulnerabilidad de los distritos del país y poder fortalecer la toma de decisiones son fundamentales para asegurar un futuro más seguro y sostenible para la población peruana.

La falta de capacidades por parte del Estado y su baja adaptabilidad a la realidad geológica y climática del Perú son los principales obstáculos para alcanzar la situación deseada. Asimismo, la poca inversión en infraestructuras resistentes a sismos, la limitada educación en prevención de desastres y la deficiente coordinación entre las diferentes instituciones gubernamentales son algunas de las insuficiencias que se deben abordar. Estos desafíos son cruciales para garantizar la seguridad y el bienestar de la población peruana.

A partir de este contexto, surge la siguiente pregunta de investigación: ¿Cuál es la situación de los distritos del Perú con respecto al nivel de prevención ante futuras actividades sísmicas en base a las variables de infraestructura de vivienda, tipo de suelo, actividad sísmica, disponibilidad de hospitales y proyectos de inversión?

A su vez, se plantea la siguiente hipótesis que se buscará confirmar al término de este trabajo: La situación de los distritos del Perú con respecto al nivel de prevención ante futuras actividades sísmicas, a través de la combinación de estas variables, se presenta de manera estratificada. Esto significa

que es posible agrupar los distritos en categorías o niveles de preparación, donde la interconexión de estos factores de manera única en cada región, contribuyen a la creación de distintos niveles de preparación que no siempre son evidentes a simple vista.

II. ESTADO DEL ARTE

Para la elaboración del presente trabajo, se tomaron como referencia dos informes técnicos. El primero fue realizado por la autoridad de CENEPRED con la colaboración de otras instituciones, como el Instituto Geofísico del Perú (IGP). En este informe, se evaluó el impacto de la ocurrencia de sismos de 8.8 Mw. Además de datos geológicos, se consideró un mapa de los establecimientos de salud [7]. En este estudio, se identificaron inicialmente los puntos geográficos más vulnerables a desastres naturales y, al combinarlos con el mapa de establecimientos de salud, se simulaban escenarios para determinar su nivel de riesgo. Se observó un alto riesgo en las áreas cercanas a la costa del Perú, con un nivel de riesgo muy alto del 44.6%. Las provincias con mayor concentración de población en esta categoría fueron Lima y Callao.

En un segundo informe, llevado a cabo por el Banco Mundial y el estado peruano, se buscó evaluar el impacto de los sismos en la infraestructura educativa y, a partir de estos datos, proponer estrategias para enfrentar esta situación. Se presentaron datos, como la distribución de instituciones educativas, mapas probabilísticos de ocurrencia de sismos, mapas de riesgos sísmicos y mapas de tipos de suelo, entre otros [8]. Como conclusión, se determinó que el 51% de las edificaciones tienen un alto riesgo de colapso, mientras que el 21% presenta un daño potencial. A partir de estos resultados, el estudio propuso programas correspondientes de sustitución, refuerzo y medidas de intervención contingente como estrategias para reducir el riesgo sísmico.

III. MARCO TEÓRICO

1) *Data mining*: La minería de datos es una técnica relativamente reciente para extraer conocimiento de vastas cantidades de datos. Esta práctica implica utilizar y procesar datos disponibles para tomar decisiones [4]. Implica la exploración de modelos en conjuntos de datos extensos utilizando técnicas que se sitúan en la intersección de aprendizaje automático, estadísticas y sistemas de bases de datos [5]. Facilita el análisis de patrones, como la categorización de datos a través de estudios de agrupamiento, la detección de registros anómalos, también conocida como detección de anomalías, y las reglas o dependencias asociadas

2) *Metodo KDD*: El proceso Knowledge Discovery Dictionary (KDD) abarca la recolección y el descubrimiento de datos e información, involucrando operaciones como el procesamiento, la selección y la preparación de datos, junto con la creación de información en conjuntos de datos y la interpretación de los enfoques más eficaces según los resultados observados [6]. Se caracteriza por una secuencia iterativa de integración de datos y la identificación de patrones en Minería de Datos (DM).

3) *Web Scraping*: Una técnica importante utilizada en el presente trabajo son las herramientas de web scraping que sirven para extraer automáticamente información específica de páginas web. El propósito es recolectar datos precisos y almacenarlos en un formato estructurado para su uso futuro. Estas herramientas permiten enfocarse en información concreta de un sitio web, ignorando los datos irrelevantes [9]. Este trabajo emplea la técnica de web scraping utilizando las bibliotecas 'requests', 'BeautifulSoup' y 're' para la extracción y procesamiento de datos. Requests facilita la solicitud de peticiones HTTP para adquirir el código HTML de los sitios web. A través de BeautifulSoup, se analiza y descompone este contenido HTML, permitiendo la selección y extracción de datos específicos. Por último, la biblioteca re se aplica para la búsqueda y manipulación de textos mediante expresiones regulares.

4) *Geocodificación*: La geocodificación o geocoding en inglés es el proceso de convertir direcciones postales, nombres geográficos o puntos de interés en coordenadas geográficas o lo que es lo mismo, en un punto real de la superficie de la Tierra [10]. Esto se conoce como geocodificación directa la cual consiste en convertir direcciones a coordenadas (latitud y longitud); no obstante, también existe la geocodificación inversa la cual realiza el proceso inverso [11]. En el presente trabajo, se utiliza la geocodificación inversa con la librería 'Geopy', la cual facilita a desarrolladores de Python localizar las coordenadas de direcciones, ciudades, países y puntos de referencia en todo el mundo, y viceversa, mediante geocodificadores de terceros y otras fuentes de datos [11].

IV. METODOLOGÍA

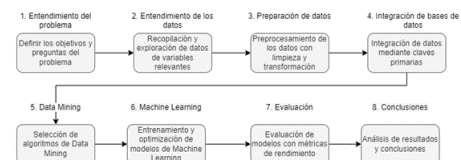


Fig. 1. Metodología del estudio

A continuación, se explicará cada uno de los pasos que se siguieron en la metodología como se muestra en el Figura 1 :

1) Entendimiento del problema:

- Definir claramente los objetivos y preguntas a responder con la investigación.
- Recopilar información de la literatura reciente sobre la problemática.

2) Entendimiento de los datos:

- Exploración de bases de datos provenientes de la Encuesta Nacional de Hogares (ENAHOG), Instituto Geofísico del Perú (IGP), Ministerio de Vivienda, Construcción y Saneamiento del Perú (VIVIENDA), Ministerio de Salud (MINSA) y Ministerio de Economía y Finanzas (MEF).
- Recopilación de datos de variables relevantes como el nivel de infraestructura de viviendas, la actividad

TABLE I
DESCRIPCIÓN DE LOS DATOS DEL IGP

Variable	Descripción
Fecha UTC	Fecha universal del sismo ocurrido.
Hora UTC	Hora universal del sismo ocurrido.
Latitud	Latitud del epicentro del sismo.
Longitud	Longitud del epicentro del sismo.
Profundidad	Profundidad (en kilómetros) del epicentro del sismo.
Magnitud	Magnitud del sismo en escala de Richter.

^aElaboración propia.

sísmica, el tipo de suelo, la disponibilidad de instalaciones hospitalarias, densidad demográfica y la presencia de proyectos de prevención de desastres naturales.

3) Preparación de los datos:

- Limpieza y preprocesamiento de los datos con técnicas de programación para garantizar la calidad para el análisis.

4) Integración de bases de datos:

- Integración de los datos en una única base utilizando claves formadas por el nombre del departamento, nombre de la provincia, nombre del distrito y año.

5) Modelado con Data Mining y Machine Learning:

- Aplicación de técnicas de clustering para identificar distritos con un nivel bajo de prevención y crear un mapa de la situación actual del país.
- Creación de un modelo de clasificación para los niveles de prevención con técnicas de Machine Learning.
- Diseñar un sistema de recomendación para asociar distritos.

6) Evaluación:

- Evaluación del rendimiento de los modelos para asegurar su precisión y capacidad de generalización.

7) Conclusiones:

- Análisis de los resultados y elaboración de conclusiones.
- Plantear recomendaciones estratégicas de identificación de los distritos con menor prevención.

V. RECOPIACIÓN Y PREPARACIÓN DE DATOS

Luego de decidir el tema, se procede a recopilar las fuentes de datos necesarias para poder realizar el análisis correspondiente. Se utilizaron cinco fuentes de datos diferentes.

A. Instituto geofísico del Perú (IGP)

Para el presente trabajo, se procedió a extraer los datos de los sismos del portal del IGP en donde se accede a la página de Descargar datos y se procede a descargar la data correspondiente al periodo de estudio.

El archivo descargado se encuentra en formato de excel en la cual se encuentran los siguientes atributos: fecha UTC, hora UTC, latitud, longitud, profundidad, magnitud. En la Tabla I, se muestran las descripciones de cada atributo.

Después de una breve descripción de los atributos de la base de datos del IGP, se procede a explicar la limpieza y preprocesamiento de estos datos que luego se unirán a una data centralizada.

Primero, se realiza una exploración general del dataset para identificar posibles incongruencias o datos nulos. En un primer alcance, no se presentan datos nulos y, el dataset tiene 12880 registros y 6 atributos. Sin embargo, la data debe adecuarse y transformarse de manera que brinde valor a la data centralizada. Por ello, los atributos latitud y longitud, que usualmente son eliminados debido a que no brindan mayor información o valor, deben ser transformados en información interpretable como direcciones.

Segundo, teniendo en cuenta lo anterior, se transforman los datos de latitud y longitud en 4 nuevos atributos los cuales son país, región, provincia, distrito. GeoPy facilita a desarrolladores de Python localizar las coordenadas de direcciones, ciudades, países y puntos de referencia en todo el mundo mediante geocodificadores de terceros y otras fuentes de datos [2].

Tercero, luego de la conversión de los atributos latitud y longitud las dimensiones del conjunto de datos cambia y se vuelve a explorar. Por un lado, los atributos pasan de ser 6 a 10 y se presentan datos nulos en 3 de las 4 columnas agregadas: Región, Provincia, Distrito. En el caso del atributo Región, se logra discernir que los datos faltantes corresponden a sismos que ocurrieron en el mar. Si bien es cierto, se puede etiquetar estos datos con el nombre "Mar" no brindaría mucho valor al análisis de las viviendas por distrito. Por ello, se eliminan aquellos registros. Luego, se realiza la revisión de datos nulos en los atributos de Provincia y Distrito. Para decidir qué hacer con estos datos, se evalúa si los registros corresponden a sismos con epicentro en Perú y no en otro país. Una vez realizada la revisión, se observa que los datos nulos corresponden a otros países; por ello, también se decide eliminarlos. Finalmente, se observan los datos nulos del atributo Distrito. En este caso, se realizaron dos acciones: eliminar los registros no nulos que corresponden a otros países o que no se lograron identificar e imputar los datos nulos de los distritos con el nombre de la provincia. Para el caso de la imputación, también se realiza la actualización de la latitud y la longitud de acuerdo a la combinación de región, provincia y distrito.

Cuarto, una vez realizado lo anterior, se obtiene un dataframe con 7697 registros y 10 atributos. De esta manera, se procede a crear la "key" que va a permitir que este conjunto de datos se una a la data centralizada. Esta clave se conforma por el nombre de la región, provincia, distrito y año.

Finalmente, se crea un nuevo dataset a partir de las estadísticas del conjunto de datos que ha sido tratado. Se crean las siguientes columnas: conteo de sismos, promedio de sismos, magnitud mínima de sismos, magnitud máxima de sismos. Con ello, se establece una base de información resumida que facilita la preparación para análisis avanzados y permite identificar tendencias, comparar regiones y evaluar riesgos de manera más eficiente.

TABLE II
DESCRIPCIÓN DE LOS DATOS DE LA ENAHO

Variable	Descripción
Year	Año en que se realizó la encuesta.
DOMINIO	Dominio geográfico de la zona de la vivienda.
Tipo Vivienda	Tipo de vivienda
Pared material	Material predominante de las paredes.
Piso material	Material predominante de los pisos.
Techo material	Material predominante de los techos.
Licencia	La vivienda cuenta con licencia de construcción.
Asistencia técnica	Vivienda construida con asistencia técnica de Ing.Civil o arquitecto.
Nivel propiedad vivienda	Nivel de propiedad de la vivienda.
Título propiedad	Presencia de título de propiedad.
Registro SUNARP	Vivienda registrada en la SUNARP.
Techo material	Material predominante de los techos.

^aElaboración propia.

B. ENAHO

La descarga de datos se obtiene del portal del Instituto Nacional de Estadística e Informática desde el año 2004 hasta el 2022. De manera complementaria, se obtienen datos de ubigeo para poder identificar la región, provincia y distrito de las viviendas encuestadas de la base de datos de la ENAHO.

Para reflejar con precisión las categorías preexistentes en la base de datos de la ENAHO dentro de la Tabla 2, se han identificado diversas clasificaciones en las variables. La variable dominio se divide en ocho categorías geográficas, 1(Costa Norte), 2(Costa Centro), 3(Costa Sur), 4(Sierra Norte), 5(Sierra Centro), 6(Sierra Sur), 7(Selva), 8(Lima Metropolitana). La variable Tipo_Vivienda se clasifica en siete categorías: 1(Casa independiente), 2(Departamento en edificio), 3(Vivienda en quinta), 4(Vivienda en casa de vecindad), 5(Chozas), 6(Vivienda improvisada), 7(Local no destinado para habitación humana). La variable Pared_material tiene las siguientes categorías: 1(ladrillo), 2(piedra, sillar), 3(adobe), 4(tapia), 5(Quincha: Caña con Barro), 6(Piedra con barro), 7(Madera), 8(Tripiley), 9(Otro). La variable Piso_material tiene las siguientes categorías: 1(madera pulida), 2(láminas asfálticas), 3(loseas), 4(madera), 5(Cemento), 6(Tierra), 7(Otro). La variable Techo_material se divide en las siguientes categorías: 1(concreto armado), 2(madera), 3(tejas), 4(calamina), 5(Triplay), 6(Paja), 7(Otro). La existencia de licencia y asistencia técnica en la construcción se evalúa en tres posibles estados: Sí (1), No (2) o No sabe (3). La variable Nivel_Propiedad_Vivienda tiene las siguientes categorías: 1(alquilada), 2(propia, pago completo), 3(propia, invasión), 4(propia, a crédito), 5(cedida por trabajo), 6(cedida), 7(Otra). La variable Título_propiedad tiene las siguientes opciones: 1(Si), 2(No), 3(En trámite). Por último, la variable Registro_SUNAP tiene las siguientes opciones: 1(Si) y 2(No).

En lo que respecta la limpieza y el preprocesamiento de los datos, se realiza la descarga de los datos de todos los años, para luego realizar su posterior concatenación.

Se divide el dataset original en dos partes por la inconsistencia en la disponibilidad de ciertas variables a lo largo de los años. En concreto, los conjuntos de datos de los años 2004 a 2011 no incluyen información sobre Licencia, Asistencia técnica y Registro SUNARP. Estas variables son cruciales para evaluar la seguridad de las viviendas, ya que proporcionan indicadores sobre si las construcciones se han realizado con la supervisión y el respaldo de profesionales, lo cual es un factor

TABLE III
DESCRIPCIÓN DE LOS DATOS DEL MINSA

Variable	Descripción
Nombre del establecimiento	Nombre del establecimiento
Clasificación	Clasificación del establecimiento
Tipo	Establecimiento con o sin internamiento
Departamento	Nombre del departamento
Provincia	Nombre de la provincia
Distrito	Nombre del distrito
Dirección	Dirección del establecimiento
Categoría	Categoría del establecimiento (I1,I2,I3 o I4)
Teléfono	Número de teléfono de contacto
Horario	Horario de funcionamiento
Año	Año de inicio de actividades

^aElaboración propia.

importante en la determinación de la seguridad estructural y la conformidad con las normativas vigentes. Por otra parte, los datos que abarcan los años 2012 a 2022 sí cuentan con estas variables.

El primer dataframe incluirá los datos de 2004 a 2011, donde estas variables específicas no están presentes, y el segundo abarcará de 2012 a 2022, incluyendo la información sobre Licencia, Asistencia técnica y Registro SUNARP. Esta división asegura que los análisis sean pertinentes y ajustados a la información disponible para cada intervalo de tiempo, facilitando así una interpretación correcta y relevante de la seguridad en las viviendas durante los diferentes periodos. En el presente trabajo, solo se utilizará el segundo dataframe que abarca los datos desde 2012 hasta 2022.

C. Ministerio de Salud (MINSA)

Se emplean los datos más recientes de los establecimientos de salud de primer nivel de atención en Perú, actualizados hasta el 25 de enero de 2023. Estos datos fueron proporcionados por el Ministerio de Salud del Perú y se emplean los registros hasta 2022. El archivo descargado se presenta en formato Excel y consta de las siguientes columnas: Nombre del establecimiento, clasificación, tipo, departamento, provincia, distrito, dirección, categoría, teléfono y horario. En total, se registraron 8,729 establecimientos de salud. Además, se incluyó una columna adicional, denominada "AÑO", la cual representa el año en que cada centro de salud comenzó sus operaciones. Esta columna se agregó ya que se utilizará en la concatenación con un conjunto de datos más amplio. La obtención de los años de inicio de actividades se realizó con técnicas de web scraping, que se explicaron anteriormente en la sección de marco teórico.

En la siguiente tabla, se muestra todas las variables junto a su descripción:

Es importante definir la variable categoría porque es la de mayor interés para el análisis de prevención. Existen 4 tipos de categorías: la categoría I-1 son los establecimientos de salud que cuentan con profesionales de la salud, pero no tienen médicos cirujanos, la categoría I-2 son los puestos de salud o posta de salud (con médico). Además de los consultorios médicos (con médicos con o sin especialidad), la categoría I-3 corresponde a los centros de salud, centros médicos, centros médicos especializados y policlínicos y la

categoría I-4 agrupa a los centros de salud y los centros médicos con camas de internamiento. La preparación de los datos incluyó la adición de una columna que reflejara el año de inicio de actividades de los establecimientos de salud. Para llevar a cabo este proceso, se empleó la técnica de web scraping, accediendo a la siguiente página web del Ministerio de Salud: <https://www.establishmentsofhealth.info/>. Una vez completado el dataset, se procedió a realizar dos ajustes importantes. Primero, se reemplazaron todos los años anteriores a 2004. Además, se corrigieron los 40 valores faltantes, que representaban un 0.5% del total de registros, asignándoles el año 2004, ya que este corresponde al año mínimo común con los demás conjuntos de datos utilizados. Posteriormente, se llevó a cabo la conversión de la variable "categoría" en variables dummy, lo que resultó en la creación de columnas con valores de ceros y unos para cada tipo de categoría (I1, I2, I3 e I4).

A continuación, se generó una tabla con los valores únicos de Departamento, Provincia y Distrito, junto con los años desde 2004 hasta 2022. Esta tabla tiene un total de 35 424 filas, resultado de la combinación de 1 866 valores únicos de Departamento, Provincia y Distrito durante un período de 19 años. Luego, se diseñó una función para calcular la suma de la cantidad de establecimientos en cada categoría a lo largo de los años. Se obtuvo la cantidad acumulada de establecimientos por categoría en cada año. Finalmente, se creó una variable "key" que representa la combinación de Departamento, Provincia, Distrito y Año. Esta variable se utilizó para facilitar la integración con otros conjuntos de datos. Las variables que contribuyeron a crear esta clave se eliminaron, y lo que permanece en la tabla son las sumatorias totales por categoría y key.

D. Tipos de suelo

Para la obtención de las etiquetas del peligro sísmico del tipo de suelo presentes en los distritos del Perú se utilizó la información pública del Mapa de la distribución del peligro sísmico en el Perú, publicado por el Ministerio de Vivienda, Construcción y Saneamiento del Perú en 2007. Utilizamos el programa QGIS, que es un Sistema de Información Geográfica (SIG). En este programa, abrimos los archivos en formato shapefile (de los suelos) y los interceptamos con el archivo de límite distrital, lo que nos permite obtener el resultado deseado nivel de distrito con los tipos de clasificación de suelos, el resultado es el que se ve en la tabla IV.

TABLE IV
DESCRIPCIÓN DE LA BASE DE DATOS DE SUELOS

Variable	Descripción
Ubigeo	Código Único de Identificación Geográfica
Zona	Zona sísmica de cada ubigeo que indica su vulnerabilidad.

En un principio, el archivo tenía unas dimensiones de 2048 filas y 12 columnas, que representan los atributos detallados en la tabla IV. No obstante, con el propósito de integrarlo con las demás bases de datos, se agregó una columna adicional que

funciona como clave primaria. Esta clave primaria se compone del nombre del departamento, nombre de la provincia, nombre del distrito y el año. En este caso, se asumió que la clasificación del suelo permanece constante durante el período de 2004 a 2022. Sin embargo, es importante señalar que esta suposición no siempre es precisa, ya que eventos naturales como huaycos, aluviones e inundaciones pueden modificar la topografía del terreno. Por lo que, la fuente de datos utilizada para extraer esta información tiene limitaciones temporales debido a que no se conoce cómo ha cambiado el suelo desde el año 2007. En ese sentido, para investigaciones futuras, sería beneficioso identificar y tener en cuenta estos cambios para mejorar el análisis.

A continuación, se tiene una descripción breve de los términos utilizados para la descripción de los suelos, los términos se obtuvieron de la leyenda del mapa.

- 1) Zona 1: Zona con menor vulnerabilidad ante actividad sísmica.
- 2) Zona 2: Zona con una vulnerabilidad moderada ante actividad sísmica.
- 3) Zona 3: Zona con una vulnerabilidad considerable ante actividad sísmica.
- 4) Zona 4: Zona con mayor vulnerabilidad ante actividad sísmica.

E. Proyectos de inversión ante fenómenos naturales

Se integran datos sobre la cantidad de proyectos de inversión ante fenómenos naturales que un distrito ha realizado a lo largo del año. En esta medida, obtuvimos el dataset del Gasto del Fondo para Intervenciones ante la Ocurrencia de Desastres Naturales (FONDES) elaborado cada año por el Ministerio de Economía y Finanzas. La data extraída se encuentra a partir del 2017 hasta el 2022. Debido a que en el registro consolidado de cada año se muestran todos los proyectos de inversión propuestos por los distritos, se realiza un agrupamiento a nivel distrital para realizar el conteo de proyectos realizados por cada uno de ellos. Como resultado, el dataset cuenta con las siguientes características descritas en la tabla V.

TABLE V
DESCRIPCIÓN DE LA BASE DE PROYECTOS DE INVERSIÓN

Variable	Descripción
Departamento/Provincia/Distrito	Identificación geográfica del distrito
Año	Año de realización del proyecto
Proyectos _{inversion}	Cantidad de proyectos por distrito

VI. EXPLORACIÓN DE DATOS

Después de integrar los datos según se muestra en la Figura 2 se continuo con la exploración y visualización de los datos. Previamente, se realiza un análisis por cada conjunto de datos antes de la data centralizada.

A. Dataset Sismos

En el caso de los registros de sismos, se evaluará principalmente el comportamiento que han tenido en cuanto a ocurrencia y magnitudes en el periodo de estudio.

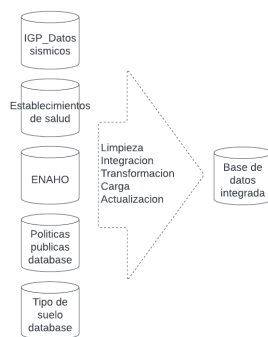


Fig. 2. Integración de los dataset

Primero, se evalúa la situación inicial y final de la cantidad y magnitud de los sismos registrados.

En la Figura 9, se muestra cómo los sismos se encuentran distribuidos en el año 2004 a lo largo del país siendo el rango detectado de la magnitud entre 4 y 5,8. Teniendo esto en cuenta, se observa que hay varios sismos que se registraron en la parte norte del país; sin embargo, teniendo en cuenta la escala de magnitud la mayoría de los registros son sismos de magnitud 5 o menor. Asimismo, si bien es cierto que en la parte sur del país se registraron pocos sismos, la magnitud suele de ser de 5 o mayor. Por lo tanto, no necesariamente si una parte del país presenta mayor ocurrencia de sismos significa que las magnitudes son altas o de manera contraria, si un lugar presenta pocos sismos puede darse que las magnitudes de esos pocos registros sea medianamente alta.

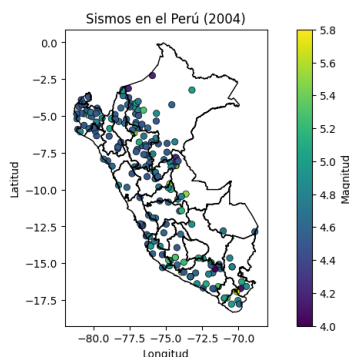


Fig. 3. Sismos en el Perú (2004)

Por otro lado, en la Figura 4, se puede visualizar cómo los sismos se encuentran distribuidos en el año 2022 siendo el rango de magnitud entre 3,4 y 7,3. A diferencia del gráfico anterior, en este caso, los sismos presentan una mayor concentración en la parte sur del país con magnitudes registradas que suelen ser mayores a 4. En el caso de la parte norte del país, se logra observar que la magnitud de los sismos varían mayormente entre 3,4 y 6,3. Por lo que, si bien es cierto, el sismo de mayor magnitud en este año se registró en la parte sur del país, también se debe considerar analizar los registros de la parte norte. De esa manera, se puede deducir

que si bien no se pueden presentar muchos sismos en zona norte, no necesariamente indica que las magnitudes pueden ser bajas; sino que, puede darse el caso que ocurran sismos de importante magnitud sobre todo aquellos que fueron mayores a 6 los cuales suelen causar daños severos [12].

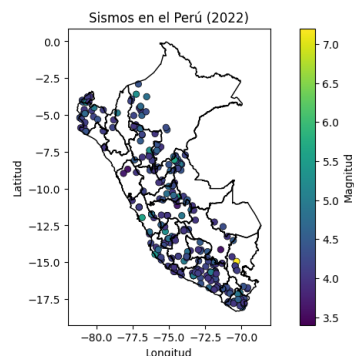


Fig. 4. Sismos en el Perú (2022)

En la Figura 5, se puede observar la evolución de la cantidad de sismos por año del periodo de estudio. A su vez, se puede evidenciar que en este rango de años, han habido más de 100 sismos, pero menos de 700. Si bien es cierto no se puede observar una tendencia clara, se puede visualizar que a partir del año 2010 se han registrado las más altas ocurrencias de sismos de todo el periodo mencionado; mientras que los años anteriores se evidenciaron las ocurrencias más bajas.

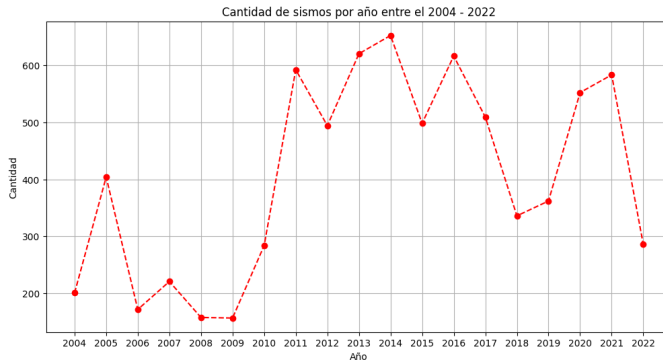


Fig. 5. Cantidad de sismos por año entre el 2004 - 2022

Teniendo en cuenta la evolución de la ocurrencia de sismos, se decide evaluar cómo evolucionan los registros sísmicos de los cuatro años con mayor y menor ocurrencia para identificar si presentan un comportamiento similar.

En la Figura 6, se observa la cantidad de sismos por mes de los cuatro años con mayor ocurrencia donde destaca el año que presentó el mayor número de registros. Si bien es cierto, no se muestra una tendencia clara para los cuatro años se logra discernir que hay meses donde presentan una ocurrencia bien cercana, como ejemplo se pueden mencionar el mes siete y once donde los valores son bien cercanos.

Por otro lado, en la Figura 7, se visualiza la cantidad de sismos por mes de los cuatro años con menor ocurrencia donde

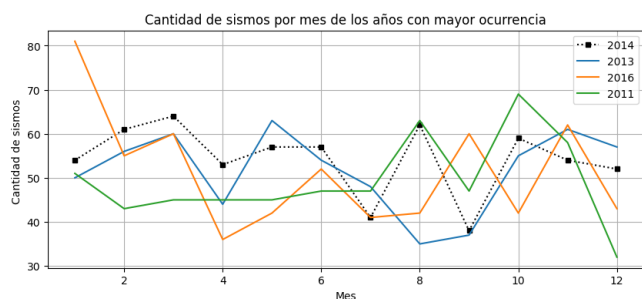


Fig. 6. Cantidad de sismos por mes de los años con mayor frecuencia

destaca el año que presentó el menor número de registros. De manera similar, tampoco se observa un comportamiento similar entre estos cuatro años; no obstante, a diferencia del gráfico anterior presentan un comportamiento más aleatorio.

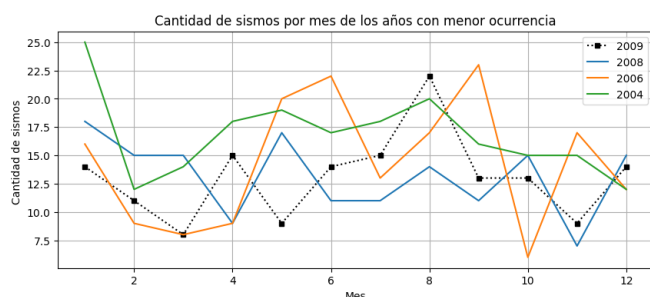


Fig. 7. Cantidad de sismos por mes de los años con menor frecuencia

Finalmente, teniendo como base lo anterior, en la Figura 8, solo se muestran los sismos que tienen magnitud mayor a seis los cuales son conocidos por causar más daños.

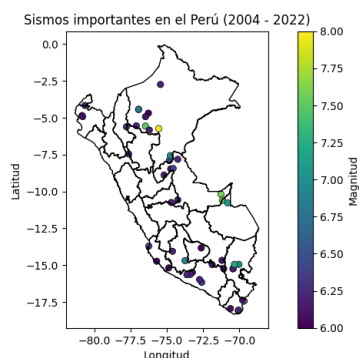


Fig. 8. Sismos importantes en el Perú (2004 - 2022)

Al observar este gráfico, se puede discernir que los sismos con mayor magnitud suelen ocurrir en las zonas donde también se evidenciaron mayor concentración de sismos anteriormente, los cuales son la zona sur y norte del país.

B. Dataset Establecimientos de salud

En la Figura 9, se puede observar que en el Perú la mayor cantidad de establecimientos de salud pertenecen a

la categoría I-1, lo cual indica un bajo nivel de calidad de los establecimientos de salud del primer nivel de atención. Asimismo, este gráfico es importante porque expresa que se puede tener una cantidad alta de establecimientos de salud; sin embargo, eso no significa que todos sean de una alta calidad.

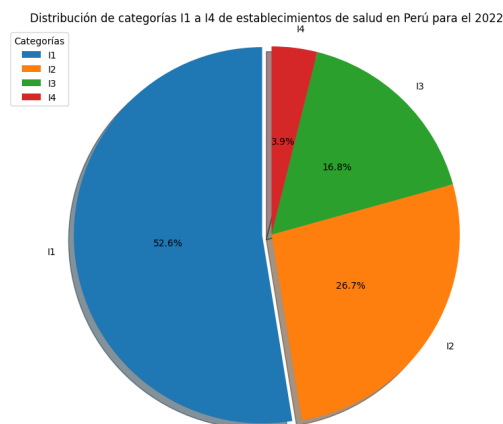


Fig. 9. Gráfico de pie chart de las categorías de los establecimientos

La siguiente Figura 10 muestra los 10 departamentos con la menor proporción de establecimientos de salud por población. Este gráfico determina los departamentos cuyos distritos podrían tener un menor nivel de prevención por la baja cantidad de establecimientos disponibles para su cantidad de población. Se muestra que Lima es el departamento con una menor proporción con 0.005%. Esto se debe a que Lima es el departamento con la mayor población del país, pero no tiene una cantidad proporcional de establecimientos de salud.

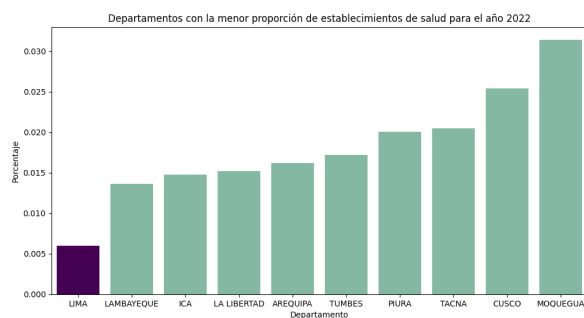


Fig. 10. Gráfico de barras de los 10 departamentos con la menor proporción de establecimientos de salud para el año 2022

La siguiente Figura 11 muestra los 10 departamentos con la mayor proporción de establecimientos de salud por población, lo que puede indicar un mayor nivel de acceso a la salud en dichos departamentos. Este gráfico indica que el departamento de Amazonas tiene la mayor proporción de establecimientos de salud con 0.11%. Es decir, entre todos los departamentos, es el que tiene una mayor cantidad de establecimientos de salud para su población. Esto se debe a que Amazonas es un departamento con una población relativamente pequeña, pero

tiene una cantidad proporcionalmente alta de establecimientos de salud.

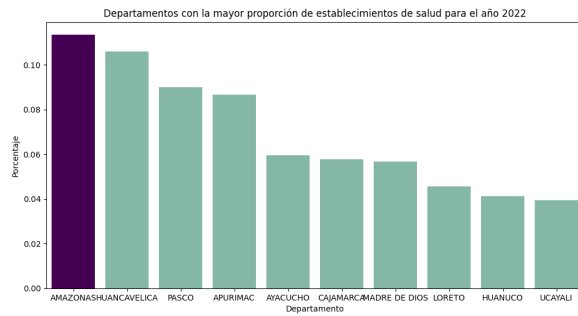


Fig. 11. Gráfico de barras de los 10 departamentos con la mayor cantidad de establecimientos de salud para el año 2022

Finalmente, la Figura 12 muestra la evolución de la cantidad de establecimientos de salud en el departamento de Cajamarca, el cual tiene la mayor cantidad de establecimientos en el país. Este gráfico analiza cómo la cantidad de establecimientos ha ido aumentando a lo largo de los años, lo que sugiere una mejora en el acceso a la atención médica en el departamento.

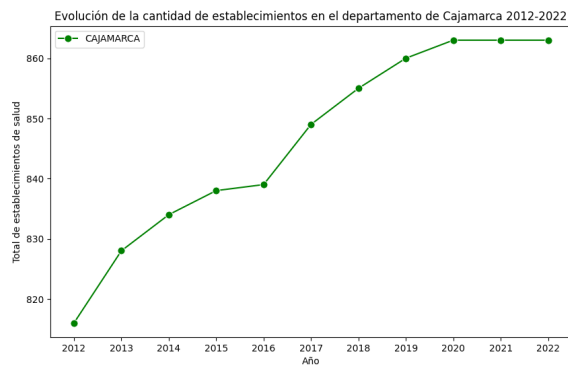


Fig. 12. Gráfico de líneas de la evolución de establecimientos de salud en Cajamarca

C. Dataset Suelos

La Figura 13 muestra el mapa de la distribución de las zonas de peligro sísmico del Perú, distinguiendo con un color diferente las zonas del 1 al 4. Se puede apreciar que la separación es por zonas geográficas como costa, sierra o selva.

La Figura 14 muestra que la zona 3 es la más extendida en el Perú, con un 42.3% de la superficie. Esto indica que la mayoría del país tiene una vulnerabilidad considerable ante la actividad sísmica. La zona 2, con un 36.2%, corresponde a zonas con vulnerabilidad moderada, mientras que la zona 1, con solo un 1.47%, es la que presenta menor vulnerabilidad. En otras palabras, solo una pequeña parte del Perú tiene una baja vulnerabilidad ante los sismos.

D. Dataset ENAHO

Las regiones extraídas son las que cuentan con una mayor densidad poblacional y por ende son de mayor interés, pues su

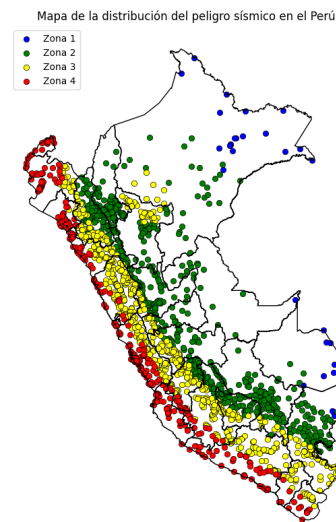


Fig. 13. Mapa de la distribución del peligro sísmico en el Perú

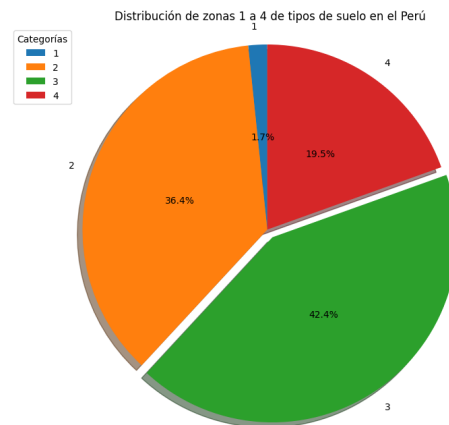


Fig. 14. Distribución de zonas 1 a 4 de los tipos de suelo en el Perú

proporción de viviendas en situación de Asistencia o registro en SUNARP ocupan una mayor muestra de viviendas. En la Figura 15, se muestra la proporción de viviendas construidas con Asistencia Técnica entre 2012 y 2022. Se muestra que Ica y Lambayeque son los únicos departamentos que superan el veinticinco por ciento de viviendas con asistencia técnica de arquitecto o ingeniero civil. Lo cual indica que a pesar de que tengan la mayor proporción es preocupante que este sea el máximo que se pueda alcanzar.

En la Figura 16, se muestra la proporción de viviendas registradas en la SUNARP. Se muestra que Lambayeque e Ica tienen la mayor proporción de casas en registros públicos, sin embargo, al igual que con el indicador de Asistencia técnica, los porcentajes no superan más de la mitad de las viviendas.

VII. RESULTADOS

A modo de resultados, se obtiene el análisis de todos los grupos de datos y la integración de todos en una sola data centralizada para su posterior análisis a través de clustering.

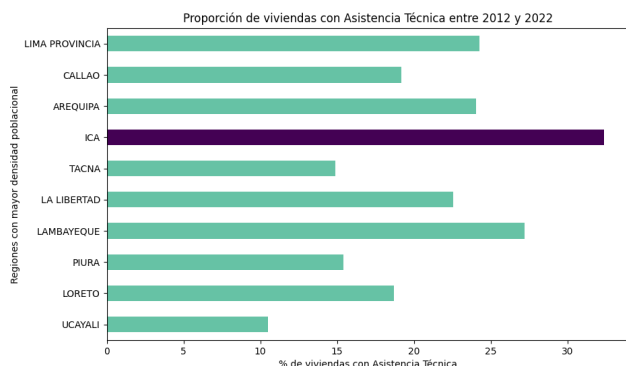


Fig. 15. Proporción de viviendas con Paredes de ladrillo o cemento entre 2012 y 2022

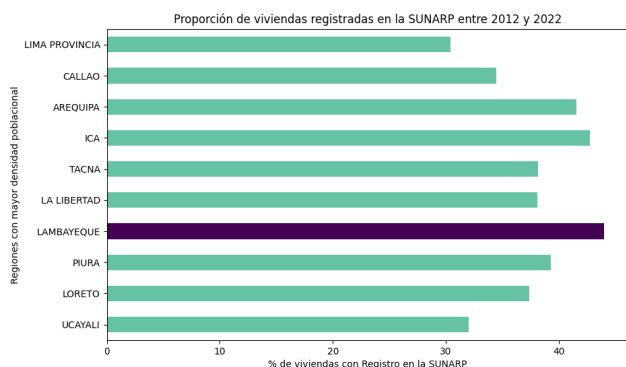


Fig. 16. Proporción de viviendas construidas con Asistencia Técnica entre 2012 y 2022

Se selecciona un año de muestra con el propósito de llevar a cabo pruebas de agrupación en relación al nivel de prevención ante sismos de los distritos analizados. En este proceso se consideran los siguientes criterios:

- El año de muestra debe ser igual o posterior a 2017, ya que a partir de este año se cuenta con información disponible en todas las bases de datos requeridas para el análisis.
- El año de muestra debe ser aquel en el cual se registró la mayor cantidad de sismos. Esto se debe a que, al utilizar un año con un alto número de sismos, es posible examinar de manera más efectiva la frecuencia de ocurrencia de eventos sísmicos en áreas que presentan recurrencia en este tipo de fenómenos.
- Se ha decidido no incluir la Base de Datos de Tipo de Suelo en este análisis. La razón detrás de esta decisión es que, a pesar de contar con información sobre los tipos de suelo en la región, no se dispone de una referencia exacta que vincule estos tipos de suelo con el nivel de vulnerabilidad de las viviendas. Esta información es esencial para lograr una interpretación más detallada y precisa en el contexto de la prevención ante sismos.

A continuación, se observa la imagen de la Silloutte Score en la Figura 17 para obtener la cantidad de clusters adecuados, se decide por $k=3$.

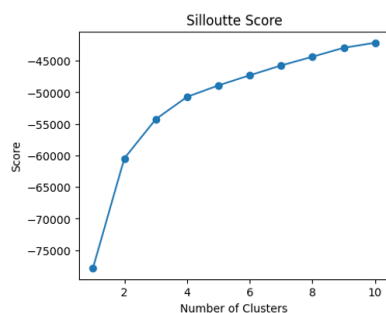


Fig. 17. Silloutte score

Luego, en el entrenamiento del modelo de Clustering, se agrupa los resultados obtenidos en relación del clúster asignado para analizar la media de los valores de las variables para cada uno de los tre clústers asignados. En la Figura 18, se evidencia que existe diferencias en los centroides, sin embargo en algunas variables son más amplias que en otras. Es así que en la siguiente sección analizaremos cuáles son aquellas características clave que definen el sentido de los distritos registrados.

	cluster_std	1	2	3
DOMINIO	5.026892	5.244073	4.549364	
ESTRATO	5.957171	6.186941	5.517740	
Registro_SUNARP_SI	10.788513	5.177225	9.033166	
Registro_SUNARP_NO	2.064409	1.073649	1.100270	
Asistencia_Tecnica_SI	4.434595	2.116401	5.402430	
Asistencia_Tecnica_NO	25.097942	10.983171	17.686271	
Asistencia_Tecnica_NOSABE	3.088313	1.660707	3.015812	
Licencia_SI	5.965471	3.243296	6.673544	
Licencia_NO	23.365870	14.373688	16.345931	
Licencia_NOSABE	3.289509	1.743296	3.085037	
Tipo_casa_independiente	32.390106	18.777108	24.633629	
Tipo_casa_Departamento	0.498340	0.549942	1.924219	
Tipo_casa_Quinta	0.151062	0.187136	0.328963	
Tipo_casa_Vicindad	1.206175	0.920132	0.722137	
Tipo_casa_Chozas	0.809097	0.258647	0.589086	
Tipo_casa_improvisada	0.020916	0.010688	0.033552	
Tipo_casa_No_habitable	0.001992	0.001166	0.002700	
Tipo_casa_Otro	0.000332	0.001166	0.000578	
Pared_Ladrillo	16.778884	7.191216	14.527189	
Pared_Piedra_sillar	0.090305	0.243101	0.082144	

Fig. 18. Tabla de agrupación por cada clúster

A. Descripción de los Grupos

De acuerdo al análisis para la obtención adecuada de los grupos, se obtuvieron 3 grupos para el agrupamiento. En la Figura 19, se puede visualizar cómo se agruparon los grupos a nivel geográfico.

A continuación, se describen las características más resaltantes de los grupos formados:

Como se observa en la Figura 20, los grupos tienen un comportamiento muy similar a través de los años seleccionados, pero en cuanto al volumen de viviendas promedio, el Cluster 1 es el que presenta mayores valores, seguido del Cluster 2 y, por último, el Cluster 3. Además, se puede observar que en el año 2020, en ambos grupos hubo una subida que difiere del comportamiento histórico y luego regresa al comportamiento usual.

En la Figura 21, se observa en porcentaje la distribución de los grupos en las distintas regiones (Costa, Sierra, Selva y Lima Metropolitana). Donde el Cluster 3 tiene predominancia

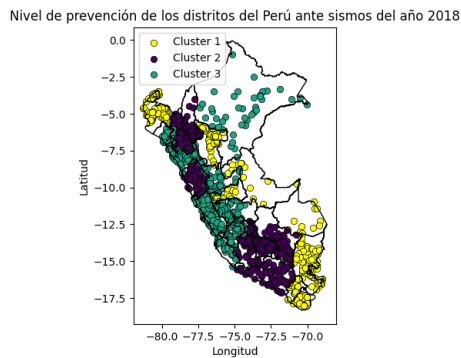


Fig. 19. Localización de los departamentos con su respectivo cluster

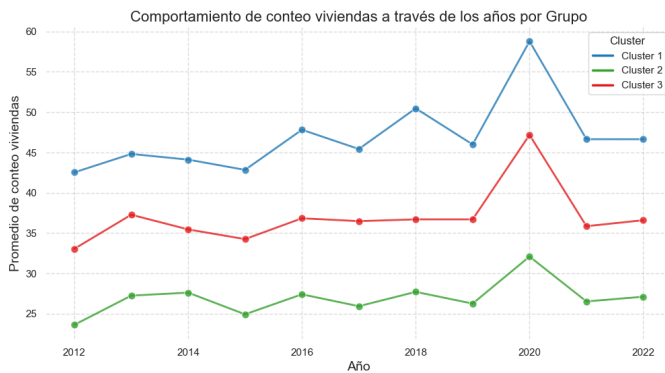


Fig. 20. Comportamiento de la cantidad promedio de viviendas por año

en la región de Lima Metropolitana y en la Costa. En el caso de la región Selva, el Cluster 1 tiene mayor presencia, y para la región Sierra, el Cluster 2 es el predominante.



Fig. 21. Distribución de zonas por región y grupo

Según la Figura 22, en el Cluster 2 se observa una predominancia de zonas de nivel 1; en contraste, en el Cluster 3 se nota una mayor presencia de zonas de nivel 4. En cambio, para el Cluster 1, no se logra discernir una sola zona predominante. En cuanto al conteo de las viviendas, los grupos

tienen una gran densidad de las mismas entre 0 y 300. Sin embargo, en comparación con el Cluster 1 y el Cluster 3, para valores mayores de 300, el Cluster 2 tiene menor densidad de observaciones en ese rango.

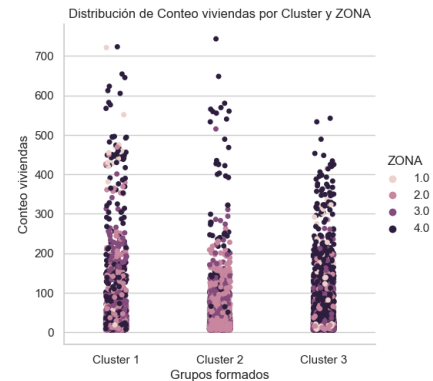


Fig. 22. Distribución de viviendas por Cluster y ZONA

1) *Cluster 1*: Identificado como aquel con una mayor orientación hacia la prevención, se distingue por poseer viviendas erigidas con la asistencia técnica de ingenieros civiles o arquitectos. Este conjunto de residencias no solo cuenta con licencia de construcción, sino que también se caracteriza por ser compuesto por casas independientes, predominantemente construidas con ladrillo, y respaldadas por un título de propiedad.

Ubicado en zonas con una vulnerabilidad moderada-alta ante la actividad sísmica, el cluster 1 exhibe una propensión a experimentar sismos, con una magnitud promedio superior a la observada en los demás conglomerados. Sin embargo, se destaca por haber sido objeto de una mayor cantidad de proyectos de inversión destinados a prevenir desastres naturales.

En líneas generales, el cluster 1 alberga viviendas más seguras y resilientes frente a los terremotos en comparación con las de otros conglomerados. Este atributo se atribuye a la utilización de materiales y técnicas de construcción apropiadas, respaldadas por los permisos y títulos necesarios. Además, la significativa inversión en proyectos de prevención de desastres naturales fortalece la capacidad del cluster para enfrentar su tendencia a experimentar sismos de mayor magnitud.

2) *Cluster 2*: Identificable como el de zonas con un bajo nivel de prevención, se caracteriza por albergar viviendas que han sido erigidas sin la asistencia técnica de ingenieros civiles o arquitectos. Estas residencias carecen de licencia de construcción, son menos propensas a ser independientes, no predominan en su construcción con ladrillo y no ostentan título de propiedad.

Situado en zonas con una vulnerabilidad moderada ante la actividad sísmica, el cluster 2 ha experimentado una cantidad intermedia de sismos, con un promedio de magnitud inferior al observado en el cluster 1. No obstante, se destaca por haber sido objeto de una menor cantidad de proyectos de inversión destinados a prevenir desastres naturales.

En términos generales, las viviendas del cluster 2 exhiben menor seguridad y resistencia frente a los terremotos. Esta característica se atribuye a que no han sido construidas con materiales y técnicas apropiadas, carecen de los permisos y títulos necesarios. Además, la menor inversión en proyectos para la prevención de desastres naturales, junto con una vulnerabilidad moderada a los sismos, identifica a este conglomerado como una zona con bajo nivel de prevención.

3) *Cluster 3*: Denominado el de zonas con nivel moderado de prevención, se caracteriza por tener viviendas que han sido construidas con asistencia técnica de un ingeniero civil o arquitecto. Además, cuentan con licencia de construcción, son casas independientes, están construidas principalmente de ladrillo y tienen título de propiedad. Este cluster se encuentra ubicado en zonas con una vulnerabilidad inferior ante actividad sísmica que los demás clusters. Además, ha sido afectado por una menor cantidad de sismos, con un promedio de magnitud inferior al del cluster 1. Por otro lado, ha sido objeto de una cantidad intermedia de proyectos de inversión en prevención de desastres naturales. En general, las viviendas del cluster 3 son más seguras y resistentes a los terremotos que las viviendas del cluster 2, pero menos que las viviendas del cluster 1. Esto se debe a que han sido construidas con materiales y técnicas adecuadas, y cuentan con los permisos y títulos necesarios. Además, son zonas afectadas por sismos de menor magnitud y tienen una cantidad intermedia de proyectos de prevención ante desastres.

B. Sistema de Recomendación

En la sección de Sistema de Recomendación, hemos optado por aplicar un filtro colaborativo tipo 'User-Based', utilizando los perfiles de los distritos de diferentes lugares del Perú como registros. El objetivo es encontrar distritos similares en características. En este caso, no contamos con registros que tengan valores faltantes, ya que todos han pasado por un proceso de preprocesamiento.

Siguiendo la línea de nuestro objetivo, buscamos un perfil de distrito con mayor propensión a no estar preparado ante un movimiento sísmico. Para esto, elegimos el filtro colaborativo basado en usuarios (distritos) similares. Identificamos un distrito base al considerar las características de los clusters. El Cluster 2 tiene el perfil de viviendas vulnerables (menos seguras y más propensas a sismos). Limitamos nuestro análisis al último año disponible de los datos (año 2022), resultando en 478 instancias.

Nos interesa encontrar el distrito con la mayor cantidad de sismos registrados y baja asistencia técnica. Después de realizar la búsqueda, seleccionamos el registro con el índice '8996', el cual tiene 8 viviendas con asistencia técnica y 3 sin ella. Además, 4 viviendas tienen techo de calamina y 7 paredes de ladrillo. Este registro se encuentra en Arequipa, Provincia de Caravelí, y es el que presenta el mayor número de sismos contados en el año 2022, sin proyectos de inversión.

Luego, seleccionamos las columnas con valores numéricos representativos de las características más importantes del distrito. Aplicamos medidas de similitud, como el coseno y la

correlación de Pearson, entre nuestro distrito de interés y el subconjunto de distritos. Encontramos que los 3 distritos más similares según estas métricas son los registros con los índices '9019', '8986' y '8984', coincidiendo en los resultados obtenidos.

Los registros similares comparten la misma ubicación en el departamento de Arequipa. Sin embargo, difieren en la característica de conteo de sismos, con valores de 3, 1 y 0, respectivamente. En términos generales, comparten una alta proporción de viviendas que no han recibido asistencia técnica, lo cual es comparable al distrito base. Además, en cuanto a la presencia de proyectos de inversión, solo el segundo distrito más similar cuenta con 4 proyectos, mientras que los demás no tienen ninguno. Todos estos distritos muestran una mayor vulnerabilidad geográfica ante desastres sísmicos.

C. Aplicación de Modelos de Machine Learning

Para esta sección, se aplicarán modelos de clasificación para entrenar modelos de Machine Learning que permitan clasificar los registros en sus respectivos grupos anteriormente descritos.

Primero, se comenzó con la partición para los conjunto de entrenamiento (train) y prueba (test), y la distribución de los registros se ven reflejados en la Tabla 6. Según la Tabla VI, la cantidad de datos no está correctamente balanceada equitativamente para las tres clases. Solo los Cluster 2 y 3 están balanceados correctamente.

TABLE VI
DISTRIBUCIÓN DEL TAMAÑO DE DATOS EN LOS CONJUNTOS DE ENTRENAMIENTO Y PRUEBA PARA EL ESTUDIO.

Subconjunto	Tamaño	Cluster 1	Cluster 2	Cluster 3
Train set	10,675	2,404	4,127	4,144
Test set	2,669	608	1,019	1,042

Se probaron cuatro modelos de clasificación basados en árboles más SVM, dos modelos de Boosting (AdaBoost y Gradient Boosting) y dos modelos de Bagging (Random Forest y Extra Trees). En la Tabla VII, se observan resultados sólidos para los modelos de Bagging y el modelo Gradient Boosting. El clasificador SVM muestra un rendimiento deficiente, seguido por AdaBoost con resultados que no son ni tan buenos ni tan malos. El modelo Extra Trees es el que presenta el mejor rendimiento según las tres métricas presentadas. Por lo tanto, recomendamos utilizar el modelo Extra Trees, ya que obtuvo los mejores resultados con nuestra base de datos.

TABLE VII
RENDIMIENTO DE MODELOS DE MACHINE LEARNING

Model	Accuracy	Precision	ROC AUC
1. SVM	0.411	0.504	0.707
2. AdaBoost	0.677	0.695	0.840
3. GradientBoosting	0.815	0.816	0.943
4. RandomForest	0.880	0.892	0.976
5. ExtraTrees	0.892	0.899	0.978

Se propuso una arquitectura de MLP que consta de una capa de entrada con 128 neuronas, seguida por una capa dropout del 20% para evitar el sobreajuste del modelo. Luego, se incluye

una capa oculta con 32 neuronas y funciones de activación ReLU. Finalmente, se añade una capa de salida con neuronas y función de activación softmax. En total, la arquitectura de la red neuronal calcula 12,291 parámetros.

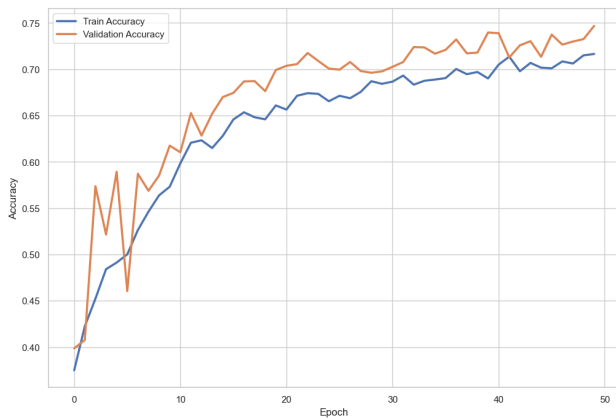


Fig. 23. Gráfico del desempeño de la MLP

Según la Figura 23, se observa que a medida que aumentan las épocas del entrenamiento de nuestra red básica MLP, mejora tanto en la etapa de entrenamiento como en la etapa de prueba. Se puede afirmar que el modelo presenta buenos resultados tanto para el conjunto de entrenamiento como para el de prueba. En concreto, logramos un 74.6% de precisión. Si se compara con los resultados de los modelos de ensamble, estos muestran valores superiores. Sin embargo, es importante señalar que nuestra MLP tiene solo una capa oculta con pocas neuronas. Sería ideal explorar la modificación de parámetros para hacer más sofisticada la red neuronal, pero, a pesar de su diseño básico y la inclusión de una capa de dropout, que previene el sobreajuste, se observan resultados decentes, tal y como se puede apreciar en la Figura 24, el cual se observa el desempeño por clase mediante una matriz de confusión.

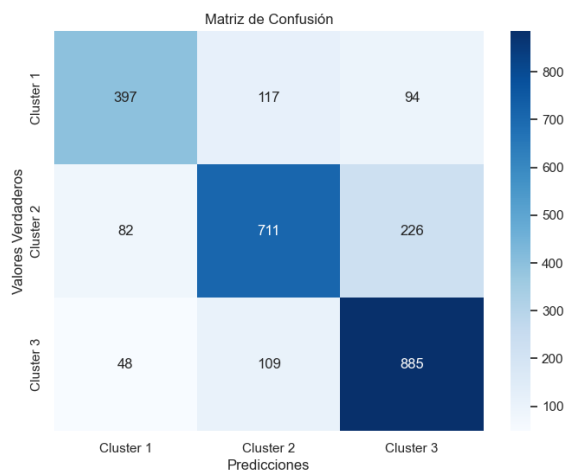


Fig. 24. Matriz de confusión multiclasa (MLP)

VIII. CONCLUSIONES

En conclusión, las variables de infraestructura de vivienda, tipo de suelo, actividad sísmica, disponibilidad de establecimientos de salud y políticas públicas influyen en la detección de patrones que reflejan el nivel de preparación ante sismos en los distritos del Perú.

Al analizar por partes, se logra discernir las zonas que suelen ser más afectadas por los sismos. Principalmente, se observa que la zona norte y sur, además en estas zonas son las que suelen tener sismos de mayor magnitud.

En los establecimientos de salud, se ha logrado identificar que la mayoría de establecimientos se agrupan en la categoría I-1 que hace referencia a establecimientos que cuentan con profesionales de la salud, pero no tienen médicos cirujanos. Es el nivel más bajo de todas las categorías, es decir, en Perú predominan los establecimiento de primer nivel con una baja calidad.

A nivel de los tipos de suelo, se concluye que conocer el nivel de vulnerabilidad de la zona ante sismos dependiendo del tipo es clave para poder tener un mejor panorama de la situación en la que se encuentra en el distrito.

En el caso de la encuesta de la ENAHO, se observó que la proporción de viviendas que han recibido asistencia técnica es muy baja, incluso en las regiones donde cuentan con un mayor número de población. Al igual que el porcentaje de viviendas que se encuentran registradas en la SUNARP. Estos dos factores son clave para que la municipalidad tenga seguimiento de la calidad de las viviendas en su territorio, sin embargo, si no se tiene registro de ello, se vuelve muy difícil plantear políticas de mejoramiento de infraestructura inmediata.

Como trabajos futuros, se propone elaborar rutas de evacuación, estimar la expansión de los sismos y conocer cómo afecta a zonas cercanas. Asimismo, se propone una aplicación del estudio a otros desastres naturales como huaicos o inundaciones.

REFERENCES

- [1] Instituto Geofísico del Perú. (s.f.). Descargar Datos. <https://ultimosismo.igp.gob.pe/descargar-datos-sismicos>
- [2] Morales, A. (s.f.). Cómo realizar geocodificación con GeoPy.
- [3] Relación de establecimientos de salud de primer nivel de atención en el Perú. (2023). [www.gob.pe. https://www.gob.pe/institucion/minsa/informes-publicaciones/391864-establecimientos-de-salud-de-primer-nivel-de-atencion-en-el-peru](https://www.gob.pe/institucion/minsa/informes-publicaciones/391864-establecimientos-de-salud-de-primer-nivel-de-atencion-en-el-peru)
- [4] K.G. Al-Hashedi, P. Magalingam, Financial fraud detection applying data mining techniques: a comprehensive review from 2009 to 2019, Comput. Sci. Rev. 40 2021.
- [5] J.R. Saura, D. Palacios-Marqu´es, D. Ribeiro-Soriano, Using data mining techniques to explore security issues in smart living environments in Twitter, Comput. Commun. 179 2021.
- [6] P. Butka, P. Bednár, y J. Ivančáková. Methodologies for Knowledge Discovery Processes in Context of AstroGeoInformatics. In Knowledge Discovery in Big Data from Astronomy and Earth Observation (pp. 1-20). Elsevier, 2020.
- [7] "Escenario de riesgo por sismo de gran magnitud seguido de tsunami frente a la costa central del Perú," Informe Técnico, Centro Nacional de Estimación, Prevención y Reducción del Riesgo de Desastres, San Isidro - Lima - Perú, 2020.
- [8] "Estrategia de Reducción del Riesgo Sísmico de Edificaciones Escolares Públicas del Perú," Informe Técnico, Banco Mundial, Washington, DC, 20433 EE. UU, 2017.

- [9] Martínez, R., Rodríguez, R. A., Vera, P., Parkinson, C. (2019). Análisis de técnicas de raspado de datos en la web aplicado al Portal del Estado Nacional Argentino. In XXV Congreso Argentino de Ciencias de la Computación (CACIC)(Universidad Nacional de Río Cuarto, Córdoba, 14 al 18 de octubre de 2019).
- [10] Soriano, P. (2021). Geocodificación: Definición y principales servicios web de geocoding. <https://geoinnova.org/blog-territorio/geocodificacion-servicios-geocoding/>
- [11] Morales, A. (s.f.). Cómo realizar geocodificación con GeoPy. <https://mappinggis.com/2018/11/geocodificacion-con-geopy/>
- [12] El Mundo. (2023). Cuántos grados tiene la escala de Richter y cómo se mide un terremoto. <https://www.elmundo.es/como/2023/02/06/63e0c261fdddf1228b45b2.html>
- [13] Riolo, P. (17 de marzo de 2023). Perú es de los países más vulnerables ante desastres naturales, ¿y el resto de la Latinoamérica? - Vigilante. <https://vigilante.pe/2023/03/17/peru-es-de-los-paises-mas-vulnerables-ante-desastres-naturales-y-el-resto-de-la-latinoamerica/>