UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO Posgrado en Ciencias de la Tierra
Instituto de Geofísica

Detección y clasificación automática de señales sísmicas del volcán Popocatépetl

TESIS

QUE PARA OPTAR POR EL GRADO DE:

Maestro en Ciencias de la Tierra

PRESENTA:

Karina Bernal Manzanilla

TUTOR

Dr. Marco Calò

Instituto de Geofísica, UNAM

Ciudad Universitaria, CDMX, julio 2022

Resumen

Los volcanes activos son ambientes en los que se registra una gran variedad de eventos sísmicos. Esto se debe a la diversidad de procesos que ocurren en su interior y que involucran la interacción de materiales en fases sólidas y fluidas. La descripción de esos procesos, mediante el análisis de registros sísmicos, requiere de una etapa de preprocesamiento en la que se seleccionan eventos que son de interés para el fenómeno que se estudia. Sin embargo, la búsqueda de señales específicas en registros continuos y ruidosos es una tarea compleja, que consume tiempos considerables. En ese sentido, el objetivo de este proyecto es elaborar una serie de catálogos de las señales más frecuentes que se registran en el volcán Popocatépetl; estos servirán para complementar las técnicas tradicionales que se emplean en la selección de eventos de interés. Para elaborar los catálogos se usaron métodos de aprendizaje automático o machine learning, que son capaces de ordenar los registros, automáticamente, en distintas clases. En particular, se aplicaron dos algoritmos: bosques aleatorios y máquinas de vectores de soporte. El esquema de organización que se empleó está basado en la clasificación de eventos sísmicos que realiza el Centro Nacional de Prevención de Desastres (CENAPRED), como parte del monitoreo del volcán. Las clases que se consideraron son: señales de periodo largo (LP), tremores (TR), explosiones (EX), sismos volcano-tectónicos (VT), sismos regionales (RE) y ruido (NO). El modelo clasificador que se encontró se aplicó en el análisis de señales continuas que van de noviembre del 2019 a julio del 2020. Para evaluar su desempeño y, en consecuencia, la calidad de los catálogos obtenidos, se comparó la clasificación de 2141 eventos con las etiquetas asignadas por personal especializado del CENAPRED. Se encontró que el modelo es capaz de identificar el 78 % de los eventos reportados y, además, clasifica varios eventos no reportados en los catálogos oficiales. Con respecto a su desempeño en las distintas clases, se encontró que tiende a sobrevalorar el número de eventos de las clases EX y VT; sin embargo, logra detectar la mayoría de las señales de esas clases. Además, es particularmente eficiente en la detección de las clases LP y TR, que son las dominantes en la sismicidad del volcán.

Abstract

Seismic records of active volcanoes present a wide variety of waveforms. These reflect the diversity of processes that occur within the volcanoes and involve the interaction of materials in solid and fluid phases. The description of these processes, through the analysis of seismic records, requires the selection of events relevant to the studied phenomena. However, detecting specific signals in continuous and noisy records is a complex and time-consuming task. In this sense, the aim of this project is to elaborate a series of catalogs of the most frequent signals recorded at the Popocatépetl volcano; these will complement traditional techniques used in the selection of events of interest that will be useful for future studies. We used machine learning methods to elaborate the catalogs, which can automatically sort the records into different classes. In particular, we explored two algorithms: random forests and support vector machines. The implemented workflow is based on the classification of seismic events performed by the National Center for Disaster Prevention (CENAPRED) as a part of the monitoring of the volcano. The classes considered are: long-period signals (LP), tremors (TR), explosions (EX), volcano-tectonic earthquakes (VT), regional earthquakes (RE) and noise (NO). We applied the best classifier model in the analysis of continuous signals from November 2019 to July 2020. In order to evaluate its performance and the quality of the catalogs, we compared the classification of 2141 events with the labels assigned by CENAPRED personnel. The model can identify 78 % of the reported events and finds several signals not reported in the official catalogs. Regarding its performance per class, our model overestimates the number of events in classes EX and VT; however, it detects most of the signals of these classes. In addition, it is highly efficient in detecting signals of the LP and TR classes, which are dominant in the volcano's seismicity.

Agradecimientos

Agradezco enormemente a mi tutor, Marco Calò, por sus enseñanzas, su apoyo incondicional y su guía durante el desarrollo de mis estudios. Igualmente, a mis sinodales, Sébastien Valade, Denis Legrand, Alejandra Arciniega Ceballos y Mathieu Perton, por enriquecer el trabajo con sus sugerencias y comentarios. Asimismo, aprovecho para agradecer a todos mis profesores del posgrado, por el valioso conocimiento que me transmitieron. Agradezco el apoyo recibido por el CENAPRED, en particular el de Gema Caballero Jiménez, quien facilitó el intercambio de los catálogos internos de clasificación de eventos y compartió su experiencia en el tema. También guardo un agradecimiento especial a Leonarda Esquivel Mendiola, quien participó en la organización de los registros sísmicos que se usaron en el proyecto y, además, me aconsejó en varias ocasiones. A todas las personas que ayudaron en campo, con la recolección de los datos y el mantenimiento de las estaciones, gracias por su tiempo y generosidad. Mucho agradezco a mis compañeros del posgrado por el apoyo, las pláticas, sesiones de estudio y su solidaridad; aún a distancia, mi experiencia no hubiera sido la misma sin ustedes. Agradezco al posgrado y a su personal administrativo por su constante apoyo, amabilidad y eficiencia, fueron indispensables durante mis estudios. El financiamiento económico de mi maestría se lo debo al pueblo de México, a través del Sistema de Becas del CONACYT y al apoyo parcial brindado por el proyecto PT5.2 de GEMex, Convocatoria CONACYT-SENER S0019, 2015-04 proyecto No: 267084. Finalmente, gracias a mi mamá, a mi hermana, a mi familia y amigos; les debo todo.

Contenido

Índice de figuras

Índice de tablas

Capítulo 1

Introducción

México es un país de volcanes. De acuerdo con Espinasa-Pereña y col., 2021 existen al menos 46 centros volcánicos en nuestro país que se consideran como activos o potencialmente activos. Esto es consecuencia de la complejidad del contexto tectónico del territorio de la República y, naturalmente, presenta un peligro para la población. Según Brown y col., 2015, México es el cuarto país en el mundo con el mayor número de personas expuestas a peligros volcánicos; después de Indonesia, Filipinas y Japón. Se estima que alrededor de 60 millones de mexicanos –cerca de la mitad de la población– viven a menos de 100 km de un volcán activo. Por consiguiente, el monitoreo volcánico, la evaluación del peligro y la creación de políticas y planes de acción para mitigar el riesgo asociado son cuestiones de interés para la seguridad nacional. Para que estas acciones se lleven a cabo de forma certera y eficiente es necesario entender el tipo de procesos que ocurren en el interior de los volcanes y mantener un registro de su evolución en el tiempo.

Sin embargo, el estudio de los procesos volcánicos es altamente complejo, pues dentro de un punto de vista físico, se presentan en un amplio rango de condiciones de temperatura y presión. Pueden ocurrir desde decenas de kilómetros por debajo de la superficie, como la formación y migración del magma, hasta decenas de kilómetros por arriba de la superficie terrestre, ya que los productos volcánicos de eventos importantes pueden llegar hasta la estratósfera. Por lo mismo, su descripción físico-química debe de incluir la interacción dinámica de elementos en fases sólidas, líquidas y gaseosas; así que hacer modelos realistas de estos procesos es un reto importante. Además, el hecho de que la observación directa de estos procesos sea complicada y frecuentemente imposible hace que el grado de dificultad aumente.

Por lo anterior, el estudio de los volcanes se lleva a cabo con la observación de manifestaciones en superficie de los procesos que ocurren a profundidad. Las técnicas de monitoreo volcánico que se usan con mayor frecuencia incluyen técnicas de análisis sísmico, geodésico, de exhalación de gases, de la temperatura en superficie y análisis geoquímicos en cuerpos de agua, entre otras. Mantener un registro de los cambios temporales de estas mediciones y la integración de la información brindada por todas ellas nos permiten inferir lo que ocurre dentro del volcán y hacer pronósticos sobre su comportamiento en el futuro.

1.1. Organización de texto

Esta tesis está dividida en 6 capítulos. En lo que resta del capítulo 1 se discute la relevancia del estudio de las señales volcano-sísmicas, se destaca la utilidad de las técnicas de machine learning en la sismología volcánica y se mencionan los objetivos del proyecto. El capítulo 2 está centrado en el volcán Popocatépetl, se plantea su contexto geológico, se hace un resumen de su historia eruptiva y se mencionan algunos trabajos que han caracterizado su sismicidad en el periodo de actividad actual. En el capítulo 3 se hace una breve introducción al aprendizaje automático, donde se presentan conceptos clave necesarios para entender la metodología que se usó en esta tesis, se discute el funcionamiento de los algoritmos de aprendizaje que se utilizaron y se hace un recuento de los trabajos publicados sobre la detección y clasificación automática de señales volcano-sísmicas. En el capítulo 4 se describe la metodología que se siguió para obtener un modelo capaz de detectar los eventos sísmicos más comunes que se presentan en el Popocatépetl y su aplicación para el análisis de los registros sísmicos entre noviembre 2019 y julio 2020. En el capítulo 5 se presentan los resultados obtenidos, donde se evalúan el desempeño y las limitaciones del modelo clasificador que se encontró; se hace una comparación con la detección manual de eventos realizada por el CENAPRED en el periodo de estudio y se hacen recomendaciones para trabajos futuros. Finalmente, el capítulo 6 presenta las conclusiones del proyecto.

1.2. Señales sísmicas en ambientes volcánicos

La sismología es el estudio de las ondas mecánicas que viajan por el interior de la Tierra y que se registran midiendo el movimiento del suelo en la superficie. El análisis de este fenómeno puede tener distintos objetivos. Por ejemplo, conocer la estructura interna de la Tierra, describir las fuentes que dan origen a los sismos o disminuir el riesgo asociado a los terremotos.

A diferencia de los ambientes predominantemente tectónicos, la presencia de fluidos en volcanes activos (magma, volátiles, fluidos hidrotermales y sus mezclas) introduce fenómenos que producen una gran variedad de formas de onda. El estudios estas señales puede usarse para conocer la estructura interna del volcán, para caracterizar los procesos físico-químicos que las generan y para monitorear el estado de actividad del volcán, entre otros. A modo general, es posible clasificar las señales volcano-sísmicas en dos grupos: en el primero, la energía sísmica se libera de la interacción entre materiales sólidos y en el segundo, la interacción es entre sólidos y fluidos (Chouet, 1996; Zobin, 2016).

Las señales del primer grupo se asocian a procesos en los que la roca se rompe o hay deslizamiento en fallas locales debido a un cambio del tensor local de esfuerzos. Esto ocurre cuando el magma abre camino mientras se mueve por la corteza terrestre y también se pueden presentar cuando sismos regionales modifican el tensor local de esfuerzos. A este tipo de eventos se les llama sismos volcano-tectónicos o VT (Chouet & Matoza, 2013). Son eventos de corta duración, por lo general, es posible identificar las fases P y S en los registros y su contenido espectral tiene energía por arriba de los 5 Hz. Este tipo de eventos son particularmente útiles en

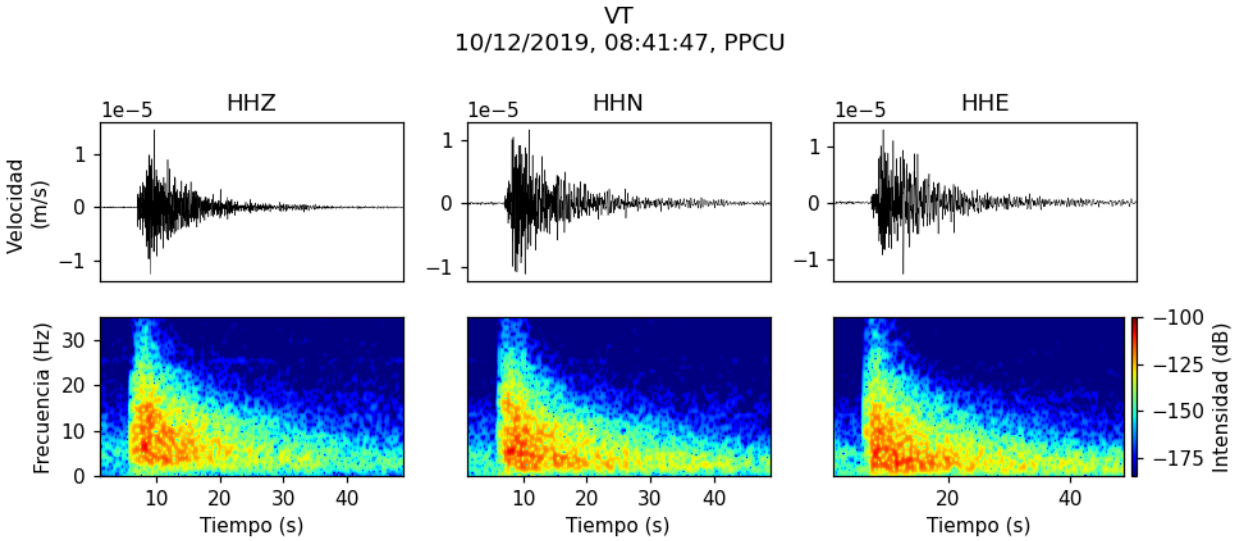

Figura 1.1
Evento VT del volcán Popocatépetl, se registró el día 10 de diciembre del 2019 a las 08:41:47 hora UTC. CENAPRED reporta una magnitud 2.3.

el pronóstico de erupciones porque sus hipocentros pueden delimitar la posición y/o migración del magma. Además, se pueden emplear para hacer tomografías de la estructura de velocidades sísmicas y de atenuación del interior del volcán; lo que permite la identificación de conductos, sistemas de fracturas, zonas en las que hay una acumulación de magma o fluidos y delimitar las propiedades térmicas de las rocas (de Lorenzo y col., 2001; Thurber, 1993). En la figura 1.1 se muestra un ejemplo de este tipo de señal registrada en el Popocatépetl. De acuerdo al reporte del Centro Nacional de Prevención de Desastres (CENAPRED) se trata de un sismo VT de magnitud 2.3.

Las señales del segundo grupo, en las que hay interacción entre sólidos y fluidos, son más diversas. En este se pueden incluir señales que se producen en el interior del volcán y también las que se generan sobre la superficie (por fenómenos como oleadas y flujos piroclásticos, lahares, etc.). Aquí nos enfocamos en las señales asociadas al movimiento de fluidos por conductos o fracturas en el interior del volcán. Se destacan 4 clases principales de eventos: sismos de periodo largo (LP), tremores, eventos de periodo muy largo (VLP, por sus siglas en inglés) y explosiones.

Los sismos LP se caracterizan por ser señales de corta duración –menor a un par de minutos– y la mayor parte de la energía de su contenido espectral está entre 0.5 y 5 Hz. Se ha demostrado que las fuentes de estos eventos son mecanismos de excitación que producen una perturbación de presión, seguida de la respuesta resonante de conductos o fracturas rellenas de fluidos (Chouet, 1996; Chouet & Matoza, 2013; Neuberg y col., 2000). Los procesos físicos responsables de la excitación pueden ser la descarga rápida de fluidos, el flujo no lineal y el colapso de burbujas, entre otros (Kawakatsu & Yamamoto, 2015). En la figura 1.2 se muestra un ejemplo de evento LP en el Popocatépetl.

Por otro lado, los tremores se caracterizan por un aumento de amplitud que se sostiene

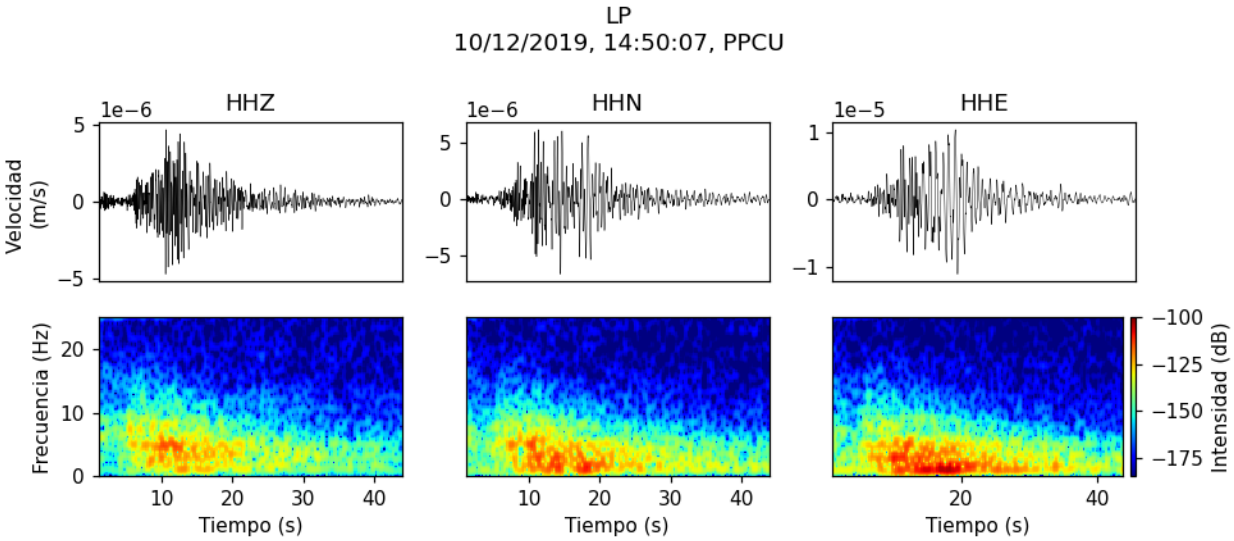

Figura 1.2
Evento LP del volcán Popocatépetl, se registró el día 10 de diciembre del 2019 a las 15:50:07 hora UTC.

durante periodos más largos de tiempo, sus duraciones pueden ir de unos cuantos minutos, hasta horas e incluso días (McNutt, 1992). Su contenido de frecuencias es similar al de los eventos LP y, en ciertos casos, su espectro se distingue por la presencia de una frecuencia dominante y sus sobretonos. Sus fuentes pueden ser las mismas que las de los eventos LP, pero es necesario que exista un mecanismo de refuerzo en la excitación que promueva ciclos de presión y descompresión en los fluidos (Neuberg y col., 2000). Además se han asociado al crecimiento de domos de lava, al movimiento de magma por conductos y, frecuentemente, se observan después de explosiones

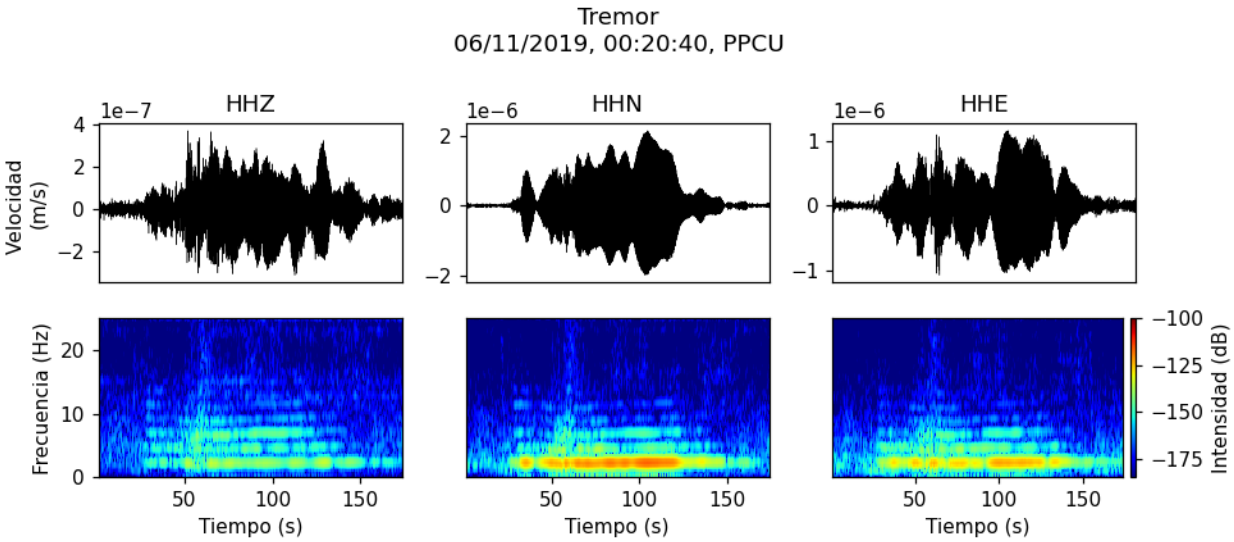

Figura 1.3
Señal de tremor armónico del volcán Popocatépetl, se registró el día 6 de noviembre del 2019 a las 00:20:40 en hora UTC.

(Arámbula-Mendoza y col., 2016; McNutt, 1992; McNutt & Nishimura, 2008). En la figura 1.3 se muestra un ejemplo de tremor armónico en el Popocatépetl.

Además de los eventos LP y los tremores, el movimiento de fluidos puede generar eventos de periodos muy largos o VLP. Estos se caracterizan por tener un ancho de banda entre los 0.01 y 0.5 Hz, no poseen características armónicas y se asocian a fuerzas inerciales producidas por las perturbaciones de presión en el magma y fluidos (Chouet y col., 2005).

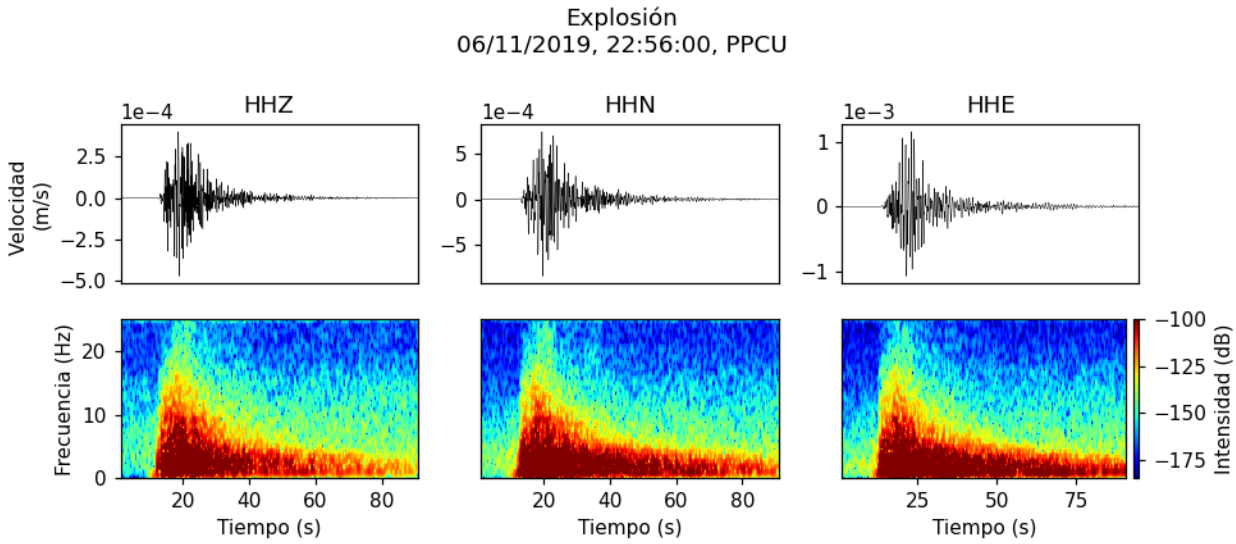

Figura 1.4
Señal asociada a una explosión del volcán Popocatépetl, registrada el día 6 de noviembre del 2019 a las 22:56:00 en hora UTC.

Por último, las explosiones se producen por excesos de presión en conductos cercanos a la superficie o en domos de lava. Resultan en la extrusión repentina de magma y en la emisión de gas y ceniza. Se caracterizan por un aumento brusco en la amplitud y la mayor parte de su energía espectral está concentrada en un ancho de banda de 1 a 5 Hz. Además, en los registros es posible observar una perturbación acoplada que corresponde a la onda acústica que se emite con la extrusión del material (R. S. Matoza y col., 2019; Tameguri y col., 2002). En la figura 1.4 se muestra un ejemplo de señal asociada a una explosión del Popocatépetl.

Por lo anterior, se subraya que la detección y clasificación de eventos sísmicos en volcanes es un paso necesario para entender los procesos físico-químicos que ocurren en su interior, obtener un modelo de su estructura y en la identificación de comportamientos que permitan pronosticar episodios que amenacen a las comunidades y actividades económicas en sus alrededores.

1.3. Aprendizaje automático y sismología volcánica

Los avances tecnológicos en sensores sísmicos y la disminución de su costo han facilitado la instrumentación de los volcanes. Esto ha resultado en un aumento considerable de la cantidad de datos disponibles, por lo que el análisis manual de los mismos es un proceso que consume mucho tiempo y está empezando a volverse poco eficiente (Malfante, Dalla Mura y col., 2018). La complejidad del sistema volcánico, aunado a la gran cantidad de datos que resultan de su monitoreo, hacen que incluso en etapas de quietud el uso de técnicas que permitan agilizar el manejo de información sea muy atractivo.

Este problema no es exclusivo de la sismología volcánica, sino que permea todas las áreas de las ciencias de la Tierra y, en realidad, aplica para cualquier problema en el que se recopilen datos por medio de tecnología actual. Desde finales del siglo pasado, el uso de técnicas de aprendizaje automático o machine learning se ha planteado como una de las soluciones más eficientes para lidiar con los retos del *big data*. Estas técnicas usan algoritmos o modelos estadísticos para analizar conjuntos de datos y, a partir de la identificación de patrones en ellos, hacer inferencias o predicciones.

En la sismología volcánica, la aplicación más frecuente de estos métodos es en la detección y clasificación automática de las señales sísmicas. Existen múltiples trabajos que demuestran la eficiencia de estas técnicas y los mejores resultados que se han publicado reportan una tasa de aciertos de alrededor del 94 % comparando con las detecciones y clasificaciones manuales (Bueno y col., 2019; Cortés y col., 2009; Falcin y col., 2021; Lara y col., 2020; Maggi y col., 2017; Malfante, Dalla Mura y col., 2018; Titos y col., 2018). Además de la tarea de clasificación de señales, las técnicas de machine learning se han implementado para disminuir el ruido en las señales (W. Zhu y col., 2019), en el picado automático de fases P y S (Chai y col., 2020; W. Zhu & Beroza, 2019), en técnicas tomográficas para obtener la estructura de velocidades (Araya-Polo y col., 2018; Bianco y col., 2019) y en el desarrollo de sistemas de alerta para el pronóstico de erupciones en tiempo real (Dempsey y col., 2020).

En resumen, el aprendizaje automático es una herramienta útil que puede usarse como alternativa y como complemento de las técnicas tradicionales. Su uso en la sismología volcánica y en el monitoreo volcánico ha aumentado de forma significativa en los últimos años y parece que esta tendencia va en aumento. Por lo que evaluar estos métodos e identificar los retos asociados a su aplicación es altamente relevante.

1.4. Objetivos y alcance

Los estudios volcano-sísmicos necesitan de una etapa de preprocesamiento en la que se realiza una selección de los eventos de interés –por ejemplo, si se realiza una tomografía de tiempos de viaje, la selección podría estar compuesta de eventos VT que se registran en al menos 4 estaciones y en los que se distinguen las fases P y S–. En la mayoría de los casos,

los eventos de interés poseen características específicas que cumplen con los modelos físicos involucrados en la creación y propagación de las ondas. Por lo que su búsqueda y selección puede ser una tarea compleja. Tradicionalmente, esta tarea se lleva a cabo de forma manual, usando métodos automáticos de detección (p.ej. STA/LTA) seguidos de una selección manual o empleando métodos automáticos basados en correlaciones cruzadas (p. ej. template matching, Gibbons y Ringdal, 2006; R. S. Matoza y col., 2015; Mendo-Pérez y col., 2021; Shelly y col., 2007; Stephens y Chouet, 2001). Sin embargo, todos estos métodos pueden consumir tiempo de forma considerable y son sensibles a la pérdida de información.

El objetivo principal de esta tesis fue obtener una serie de catálogos de los eventos sísmicos más frecuentes del volcán Popocatépetl. Estos se emplearán en estudios futuros para complementar las técnicas habituales de selección de eventos y permitirán que esta tarea se realice de forma más eficiente.

Los catálogos abarcan un periodo que va del 1 de noviembre del 2019 al 31 de julio del 2020. Para obtenerlos se exploraron dos algoritmos de machine learning (ML) –bosques aleatorios y máquinas de vectores de soporte– que permitieron la organización de los registros sísmicos en distintas clases.

El esquema de organización que se empleó está basado en la clasificación de eventos sísmicos que realiza el Centro Nacional de Prevención de Desastres (CENAPRED) –organismo encargado del monitoreo del volcán–. Las clases que se consideraron son: señales de periodo lardo (LP), tremores (TR), explosiones (EX), sismos volcano-tectónicos (VT), sismos regionales (RE) y ruido (NO).

Es importante aclarar que la clasificación está basada en aspectos generales y no se enfoca en aspectos específicos relacionados a los modelos físicos de los distintos eventos[†]. Por lo que este esquema no pretende ser una selección exhaustiva de eventos de interés para estudios futuros.

El CENAPRED publica reportes de sismicidad diarios que incluyen el número de explosiones y sismos VT, junto a la hora de su registro[‡], así como el número de exhalaciones y los minutos de tremor. Además, se solicitaron los catálogos internos de dos meses del periodo de estudio, en los que aparecen las horas de inicio y fin de eventos LP, tremores, explosiones y sismos VT. Estos datos sirvieron como referencia para evaluar el desempeño y reconocer las limitaciones de la metodología empleada y, en consecuencia, de los catálogos obtenidos.

[†]Por ejemplo, en el catálogo del CENAPRED aparecen eventos transitorios, con un contenido espectral menor a 10 Hz, que se clasifican como eventos LP. Sin embargo, este tipo de señales puede carecer de propiedades armónicas y, por lo tanto, no se pueden asociar a la respuesta resonante de una fractura rellena de fluido. Es decir, al modelo físico más robusto de los eventos LP, (Chouet, 1996). En el panel d) de la figura 5.10 se muestra un ejemplo de este tipo de señales.

[‡]En la detección de explosiones, el CENAPRED cuenta con la información de registros sísmicos, de infrasonido y con imágenes de telecámaras.

Capítulo 2

El volcán Popocatépetl

El periodo actual de actividad del volcán Popocatépetl ha durado 28 años y no hay indicios de que vaya a terminar pronto. Aún más, el rango de actividad que se ha presentado durante este periodo hace que sea el segundo volcán más activo de México –después del volcán de Colima– y el de mayor riesgo por la cercanía de regiones densamente pobladas. Por lo que su estudio es relevante para el desarrollo social y económico del país. (Ferres & Fonseca, 2017)

En este capítulo se presentan dos enfoques de estudio del Popocatépetl. En primer lugar, se habla sobre su contexto geológico y su historia eruptiva y en segundo lugar, se hace un recuento de los estudios que se han enfocado en caracterizar su sismicidad.

2.1. Contexto geológico

El Popocatépetl es un estratovolcán de composición andesítico-dacítica que se encuentra en la parte central del Cinturón Volcánico Transmexicano (CVTM), el cual tiene como origen la subducción de las placas oceánicas Cocos y Rivera por debajo la placa continental de Norteamérica (esquematizado en la figura 2.1). El CVTM atraviesa transversalmente a la Repúbilca Mexicana a la altura de los paralelos 19 y 20. En particular, el volcán se ubica en las coordenadas 19°01′23″N y 98°37′22″W y tiene una altura de 5419.43 msnm. Forma parte de la Sierra Nevada que es una alineación de volcanes orientada de norte a sur, que se encuentra a 40 km al oeste de la ciudad de Puebla y a 80 km al este de la Ciudad de México (Espinasa-Pereña & Martín-del Pozzo, 2006). La Sierra Nevada tiene una longitud de más de 80 km y está formada por los volcanes Tláloc, Telapón, Iztaccíhuatl y Popocatépetl. Este último es el más joven y el único activo. (Ferres & Fonseca, 2017).

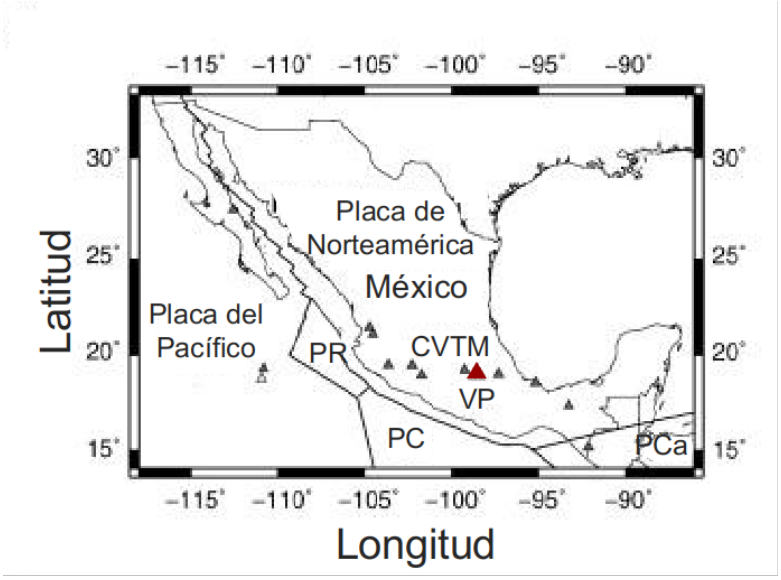

Figura 2.1
Ubicación del volcán Popocatépetl (VP) en el Cinturón Volcánico Transmexicano (CVTM). PC, PR y PCa son las placas de Cocos, Rivera y Caribe respectivamente. Los triángulos muestram los volcanes de México con actividad en tiempos históricos, el Popocatépetl se destaca con un triángulo rojo. **Fuente**: modificado de Arámbula-Mendoza y col., 2010

2.1.1. Historia eruptiva

La evolución temporal del Popocatépetl se caracteriza por ciclos de construcción de edificios seguidos de su destrucción parcial debida al colapso de sectores (Delgado Granados y col., 2017; Espinasa-Pereña & Martín-del Pozzo, 2006). Se tiene registro de la existencia de cuatro edificios volcánicos separados por sus respectivos colapsos. Sus nombres y edades –determinadas con métodos paleomagnéticos y dataciones radiométricas– se presentan en la tabla 2.1

Edificio volcánico	Edad
Volcán Tlamacas	$>$ 538 a $>$ 330 ka
Volcán Nexpayantla	ca. 330 a $>$ 98 ka
Volcán Ventorillo	ca. 98 a 23.5 ka
Volcán Popocatépetl	23.5 ka

Tabla 2.1: Edificios volcánicos en la evolución temporal del Popocatépetl y sus edades. **Fuente**: Gisbert y col., 2021

Todos los edificios antiguos han sido descritos a través de los depósitos de sus colapsos y remanentes de su actividad, generalmente mediante la caracterización de domos, coladas de lava o diques. En la mayoría de los casos, los colapsos producen calderas de más de 4 km

de diámetro –cuyas cicatrices aún se pueden identificar– y grandes avalanchas de escombros que se propagaron por decenas de kilómetros (Delgado Granados y col., 2017). Por ejemplo, el colapso del Ventorillo se caracterizó por la erupción pliniana conocida como "Pómez Blanca" o "Tochimilco" a la que se le asigna un VEI (Volcanic Explosivity Index, Newhall y Self, 1982) de 5. Se estima que el volumen de material juvenil extruido fue de 6.3 km^3, además de la movilización de alrededor de 10 km^3 de rocas preexistentes (Siebe y col., 2017). La reconstrucción de este evento indica que la erupción se originó por el colapso del sector SW. La rápida descompresión generó un blast lateral seguido de una columna eruptiva que, se estima, alcanzó alrededor de 33 km y se sostuvo durante varias horas. El colapso de la columna generó flujos de ceniza y pómez que alcanzaron distancias de 25 km y espesores máximos de 5 m. La fase final de la erupción se destacó por el flujo de lava Tochimilco que se emplazó hacia el sureste y alcanzó una longitud de 22 km, con espesores entre 20 y 200 m. En los años siguientes a la erupción, las lluvias temporales se encargaron de remover el material por medio de la generación de lahares que viajaron distancias de más de 100 km (Siebe y col., 2017).

Adicionalmente, la construcción del edificio actual está caracterizada por la recurrencia de erupciones plinianas. A través del estudio de sus depósitos, se han descrito de forma detallada cuatro erupciones de gran magnitud –la última ocurrió hace $\sim$ 1100 años–, y existe evidencia que sugiere al menos dos erupciones más. Los VEI estimados para estos eventos van desde 3 hasta 6. El tipo de actividad que se presentó durante estas etapas abarca desde pequeñas erupciones vulcanianas, eventos freatomagmáticos, flujos de bloques y cenizas, oleadas piroclásticas y hasta una columna eruptiva de 44 km de altura (Arana y col., 2017). Todo esto indica la gran capacidad explosiva del volcán y subraya la necesidad de continuar con los estudios geológicos y geofísicos, así como con un monitoreo adecuado y la constante actualización de planes de acción.

2.2. Actividad actual

El periodo de actividad actual comenzó, después de 70 años de reposo, con una serie de explosiones el 21 de diciembre de 1994. El tipo de actividad que ha presentado hasta la fecha está dominado por la emisión de gases y ceniza, acompañados por episodios efusivos que llevan a la construcción y subsecuente destrucción de domos de lava (Arciniega-Ceballos y col., 2008; Nieto Torres & Martin del Pozzo, 2017). Hasta la fecha, por medio de imágenes de telecámaras y sobrevuelos, se ha detectado la formación y destrucción de más de 80 domos de lava. La gran mayoría de las explosiones que se han presentado durante este periodo son de tamaño menor o moderado (figura 2.2) y alrededor del 2 % de las explosiones son grandes (CENAPRED, 2020; Chouet y col., 2005).

Durante estos 28 años la actividad ha sido más o menos regular, pero se han destacado algunos periodos de crisis. Por ejemplo, el 30 de junio de 1997 ocurrió una erupción con una columna de más de 8 km de altura sobre el nivel del cráter. La dispersión y caída de ceniza afectó varios estados del país y provocó el cierre de actividades del Aeropuerto Internacional de la Ciudad de México. Se estima que las pérdidas económicas de las aerolíneas fueron de alrededor de 5 millones de dólares (Guffanti y col., 2009; Nieto Torres & Martin del Pozzo, 2017). En

otro caso, durante el 2000 se registró un repunte en la actividad y en mayo se reportó un lahar provocado por lluvias intensas, para diciembre se recomendó la evacuación de poblaciones vulnerables ubicadas en un radio de seguridad de 13 km, ocasionando la movilización de más de 40,000 personas. Para enero de 2001 la actividad culminó con una explosión de VEI 3, cuya columna eruptiva llegó a los 13 km de altura. Los flujos de ceniza que resultaron de este episodio erosionaron el glaciar de la cima del volcán, generando flujos de lodo que viajaron más de 15 km (Capra y col., 2004; Nieto Torres & Martin del Pozzo, 2017).

El monitoreo del volcán se lleva a cabo por el Centro Nacional de Prevención de Desastres (CENAPRED), en colaboración con la Universidad Nacional Autónoma de México (UNAM) y con reuniones regulares del Comité Científico Asesor de Volcán Popocatépetl para evaluar el estado de actividad eruptiva. La red de monitoreo cuenta con una red sísmica con tecnología de punta (véase capítulo 4 para una descripción detallada), red de cámaras permanentes para el monitoreo visual, así como un sistema que permite detectar anomalías térmicas. También se lleva un registro de las emisiones de SO_2 y se realizan campañas regulares de monitoreo hidrogeoquímico y sobrevuelos del cráter. También cuenta con una red de 5 GPS para la medición de deformaciones, medidores magnetométricos y sistemas de recolección de ceniza. De todos estos métodos, el que tiene mayor importancia debido al número de sensores, la continuidad de los datos y a la interpretación de los mismos, es el monitoreo sísmico. Al día de hoy, el semáforo de alerta volcánica se encuentra en amarillo fase 2 y se ejerce un radio de seguridad de 12 km que indica que la permanencia en esa área no está permitida.

Figura 2.2
Imágenes de una explosión registrada el 9 de enero del 2020. La altura de la columna alcanzó los 3 km. **Fuente**: Reporte del monitoreo de CENAPRED al volcán Popocatépetl, 2020.

2.3. Sismicidad en el Popocatépetl

Desde su reactivación en 1994, la sismicidad del volcán ha sido dominada por la presencia de eventos LP asociados a exhalaciones. Los eventos VT son mucho más escasos y se

han asociado a la intrusión de magma y a la distribución del campo de esfuerzos en la escala local y regional (Arámbula-Mendoza y col., 2010; Arciniega-Ceballos y col., 2003). Además, los episodios de tremor sostenido e intermitente son frecuentes y se han observado en varias de las fases eruptivas (Arámbula-Mendoza y col., 2016; Arciniega-Ceballos y col., 2000; Quezada-Reyes y col., 2013). A continuación se presenta una relación de algunos de los trabajos publicados sobre la sismología del Popocatépetl.

Chouet y col., 2005 usaron métodos de inversión de forma de onda para describir los mecanismos de fuente de señales VLP asociadas a explosiones moderadas. En particular, se analizaron las señales asociadas a dos eventos ocurridos en abril y mayo del 2000. Su solución indica que la fuente puntual de estos eventos se encuentra a 1.5 km por debajo del cráter. Además, el mecanismo de fuente que encontraron incluye componentes de fuerza simple y de tensor de momento. Estas últimas representan la intersección de dos fracturas: un sill con inclinación al este de 10° y un dique con inclinación mucho más pronunciada de 83° al noroeste, que interpretaron como el conducto principal. De acuerdo a sus resultados, ambas fracturas presentan un ciclo de presurización, despresurización y represurización en un intervalo de tiempo de 3 a 5 minutos. La historia temporal de las componentes volumétricas de su solución indica que el movimiento de masa comienza en el sill y, segundos después, genera una respuesta en el dique. Además, la componente de fuerza simple refleja procesos de advección en el magma que son respuesta a las perturbaciones de presión. Concluyen que su modelo es consistente con la expulsión de paquetes de gas presurizado originados en el sill que se forman por la cristalización del magma.

Por otro lado, Arciniega-Ceballos y col., 2008 realizaron un análisis de las señales de eventos LP y VLP registrados durante noviembre de 1999 a julio del 2000. Al igual que Chouet y col., 2005, usaron los datos de un experimento realizado con una red sísmica de banda ancha de 15 sismómetros que se instaló como parte de un programa de cooperación internacional entre el Centro de Geociencias en Potsdam, Alemania (GFZ-Potsdam), el Servicio Geológico de Estados Unidos (USGS) en Menlo Park y el Insituto de Geofísica de la UNAM (Arciniega-Ceballos, 2002, tesis doctoral). Observaron que las señales LP están asociadas a eventos de desgasificación o exhalación y reportan de decenas hasta cientos de eventos al día. El análisis de la forma de onda y las propiedades espectrales de los eventos LP permitió distinguir tres familias de eventos. Los eventos de Tipo-I (figura 2.3a) se distinguen porque la llegada de la onda P es impulsiva y por la presencia de una onda VLP con un periodo de alrededor de 30 s. Además, su forma de onda se caracteriza por ser armónica con frecuencias dominantes entre 0.5 y 0.7 Hz. Esta familia está asociada con eventos de desgasificación y tienen una duración aproximada de 1 a 3 minutos. Por otro lado, los eventos de la familia Tipo-II (figura 2.3b) son pares de pulsos energéticos que representan eventos LP muy cercanos en tiempo. Sus primeros arribos son emergentes y también contienen energía en la banda VLP de 30 s. Finalmente, la familia Tipo-III (figura 2.3c) está formada por los eventos más energéticos. Sus primeros arribos son emergentes, tienen una frecuencia dominante cercana a 0.9 Hz y no hay energía en la banda VLP. Debido a las similitudes entre las señales VLP descritas por las familias Tipo-I y Tipo-II con los eventos VLP descritos por Chouet y col., 2005 concluyen que los eventos LP se originan en la misma fuente (sill + dique) localizada 1.5 km por debajo del cráter. Por lo que se trata de un proceso de fuente no destructivo que puede estar relacionado con la inyección continua de

magma en niveles someros. Además, Arciniega-Ceballos y col., 2012, implementaron inversión de forma de onda para describir la fuente de los eventos Tipo-I. Su solución es consistente con una fractura horizontal, saturada de fluidos hidrotermales, localizada 250 m por debajo del cráter. Su modelo muestra una deflación inicial, provocada por la migración de los fluidos que salen de la fractura después de alcanzar un valor umbral de presión, seguida de ciclos de inflación y deflación asociados a la resonancia del sistema.

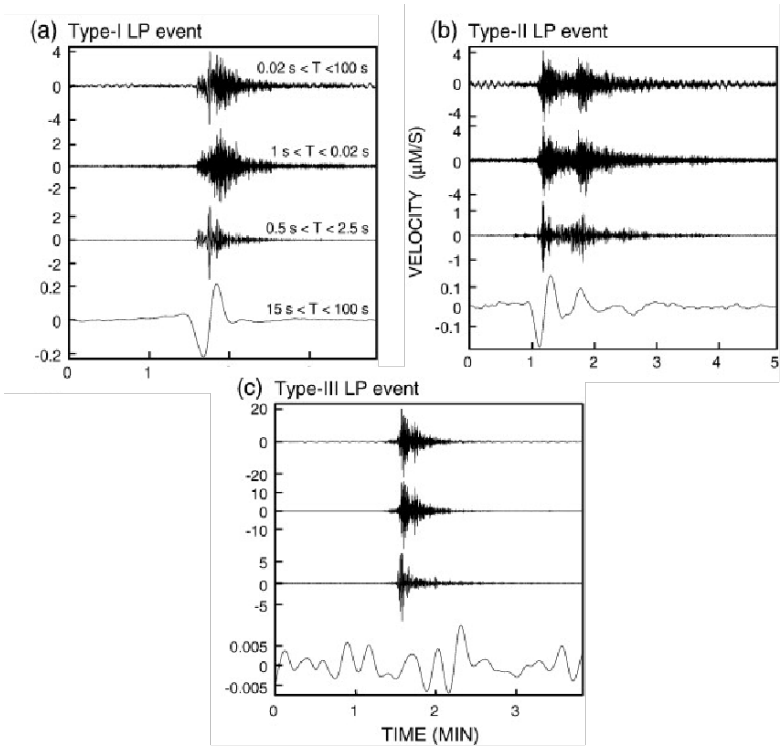

Figura 2.3
Registros de velocidad del suelo, componente vertical. (a) Ejemplo de evento LP Tipo-I. (b) Ejemplo de evento LP Tipo-II. (c) Ejemplo de eventos LP Tipo-III. Las propiedades características de cada familia se destacan al filtrarse en ciertos periodos. **Fuente:** Arciniega-Ceballos y col., 2008

Con respecto al tremor, este tipo de señal ha sido persistente desde el inicio del periodo de actividad actual y se ha estudiado por Arámbula-Mendoza y col., 2016; Arciniega-Ceballos y col., 2003; Arciniega-Ceballos y col., 2000. En términos generales, el tremor puede dividirse en tres tipos: armónico, espasmódico y pulsante. El tremor armónico (figuras 2.4 a y b) se caracteriza por la presencia de una frecuencia fundamental y sus sobretonos. Son los eventos con la amplitud más grande y su duración puede ser de varias horas. Este tipo de señal se asocia al movimiento de magma en el conducto y al emplazamiento de domos de lava en el interior del cráter. Además, reportan la existencia de señales VLP acopladas con estos eventos y se atribuyen a perturbaciones en el magma y gas durante su flujo. En cuanto al tremor espasmódico (figuras 2.4 c y d), se observa durante periodos de actividad explosiva y es muy común que acompañe periodos de emisión de ceniza. La energía se encuentra concentrada entre 1 y 5 Hz, sin embargo, no se puede distinguir, de forma clara, una frecuencia fundamental. Por último, el tremor pulsante (figuras 2.4 e y f) es el menos común y está formado por un tren

de eventos LP o explosiones pequeñas, plantean que la fuente más probable del tremor es el movimiento de fluidos a altas velocidades, acoplado con las paredes de la roca encajonante.

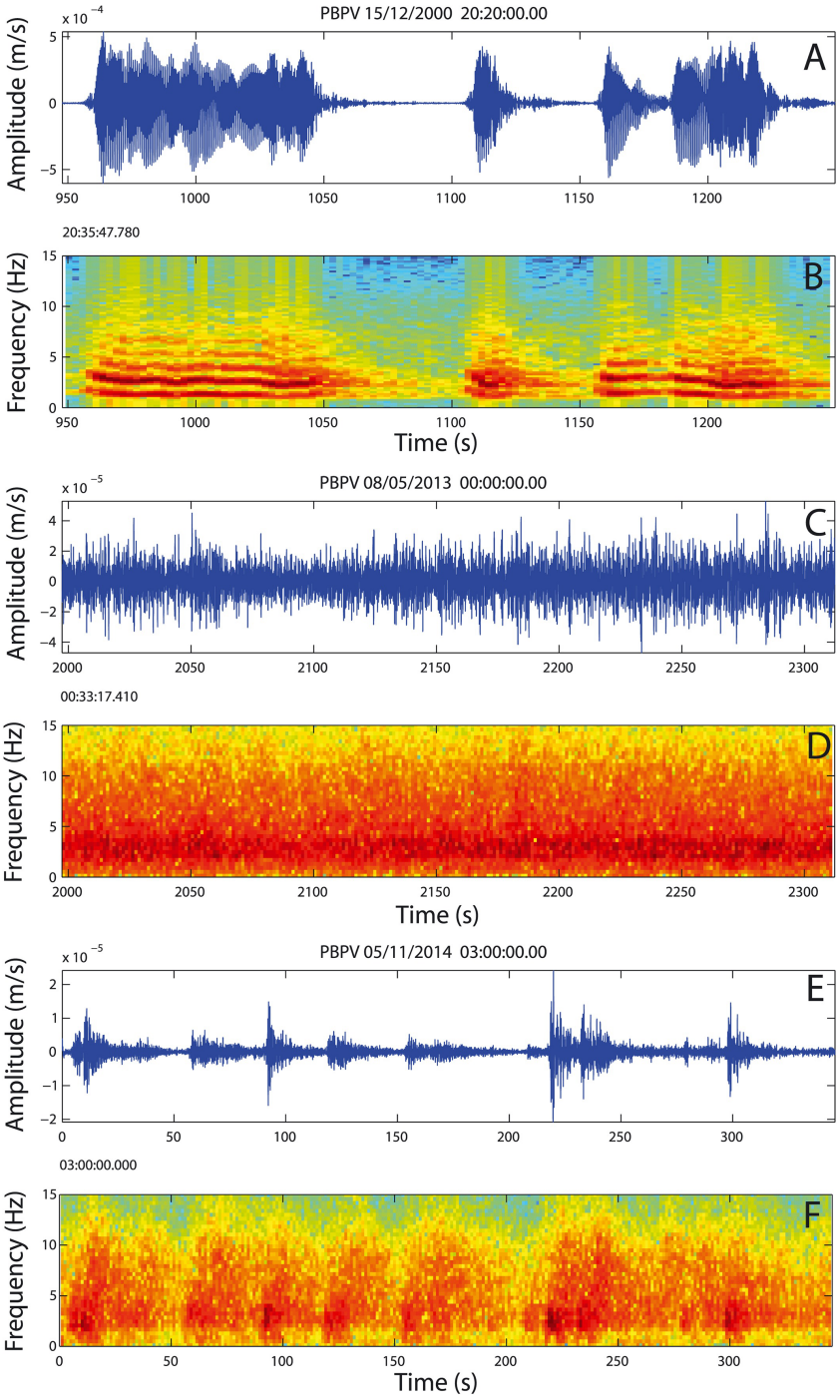

Figura 2.4
(a) Sismograma de un episodio de tremor armónico en diciembre 2000. (b) Espectrograma del evento a. (c) Sismograma de tremor espasmódico en mayo del 2013. (d) Espectrograma del evento c. (e) Tremor pulsante en noviembre del 2014. (f) Espectrograma del evento e. **Fuente:** Arámbula-Mendoza y col., 2016

En cuanto a la sismicidad VT, las observaciones indican que la ocurrencia de eventos es considerablemente menor que la sismicidad LP, con un promedio de un evento por día en periodos de actividad estable. Las magnitudes que se registran van de 1.4 a 3.6 (Arámbula-Mendoza y col., 2010; Arciniega-Ceballos y col., 2003; Quezada-Reyes y col., 2013). Por su parte, Arámbula-Mendoza y col., 2010 caracterizaron el estado de esfuerzos del volcán usando las soluciones de planos de falla de eventos VT ocurridos entre 1996 y 2003. En sus localizaciones encuentran que los eventos VTs están distribuidos en dos zonas principales: una por debajo del cráter, donde ocurren el 95 % de los sismos, y otra al sureste del volcán (véase figura 2.5). La orientación de los ejes de presión de la mayoría de los eventos que están por debajo del cráter es consistente con los esfuerzos regionales y sus soluciones tienen mecanismo normales con profundidades entre los 0 y 3 km s.n.m.. Sin embargo, observaron periodos en los que las soluciones consisten de mecanismos inversos que ocurren entre los -3 y 0 km s.n.m.. Concluyen que esta variabilidad se debe a la intrusión de magma en periodos explosivos o de crecimiento de domo. Pues las fuerzas asociadas al ascenso de magma producen un régimen de compresión en las paredes de las rocas por las que circula, resultando en mecanismos de fallas inversas. Por otro lado, los eventos al sureste tienen soluciones de corrimientos laterales a lo largo de dos posibles fallas en dirección N-S y W-E, que se activan debido a la intrusión de nuevo material.

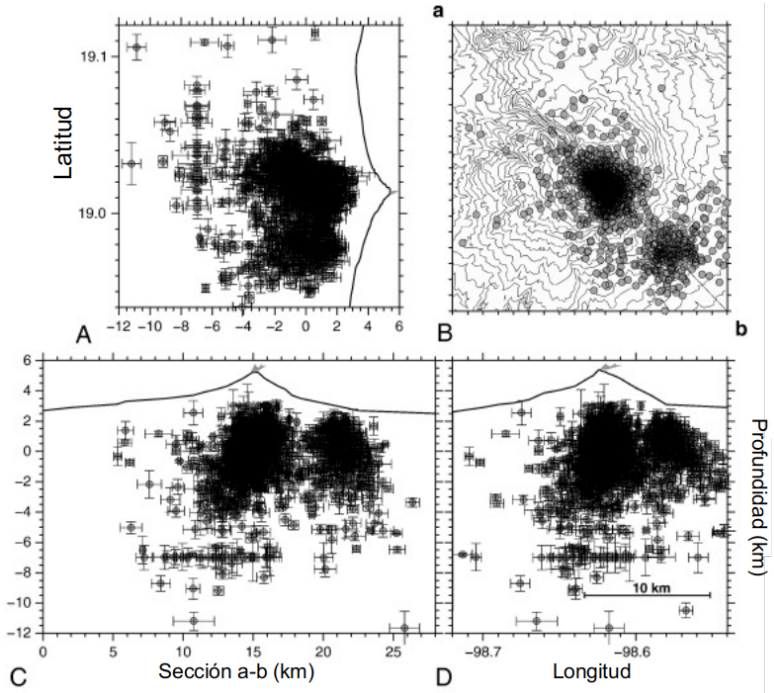

Figura 2.5
Distribución de eventos VT en el Popocatépetl de 1995 al 2003. Existen dos zonas principales de acumulación de VTs, una por debajo del cráter y otra al sureste del volcán. A) Vista de longitud. B) Vista epicentral. C) Sección a-b. D) Vista de longitud. **Fuente:** Arámbula-Mendoza y col., 2010

Finalmente, las señales sísmicas producidas por la actividad explosiva han sido estudiadas por varios autores (Arámbula-Mendoza y col., 2013; Arciniega-Ceballos y col., 2008; Arciniega-Ceballos y col., 1999; Chouet y col., 2005; Cruz-Atienza y col., 2001; R. Matoza y col., 2019; Mendo-Pérez y col., 2021; Zobin & Martínez, 2010). Algunos aspectos a destacar son: la

existencia de eventos VLP asociados a eventos energéticos (mencionados anteriormente), que los registros están formados por una fase de frecuencias bajas (0.1 - 10 Hz) seguida de un aumento de amplitud y frecuencias más altas (más de 10 Hz) y, por último, la presencia de ondas acústicas, que viajan por el aire, acopladas en los registros. De acuerdo al análisis realizado por Zobin y Martínez, 2010, la fase de frecuencias bajas de los registros se genera por el movimiento de magma fragmentado por el conducto, de modo que la duración de esta fase indica la profundidad a la que ocurre la fragmentación. Por otro lado, Arámbula-Mendoza y col., 2013, demuestran que la fase de altas frecuencias del registro se debe a la presencia de una onda acústica que se libera a la atmósfera durante la explosión. Además, esta fase incluye las vibraciones producidas por la caída de material balístico.

Capítulo 3

Aprendizaje automático

El aprendizaje automático o *machine learning* (ML) es un término genérico que se refiere a un conjunto de técnicas que permiten extraer información de un conjunto de datos a través de reglas de optimización bien definidas (Q. Kong y col., 2019). Estos métodos son atractivos porque nos permiten lidiar con grandes conjuntos de datos de forma rápida y automática, por su capacidad de detectar patrones no evidentes a simple vista o por técnicas tradicionales y por la versatilidad que tienen al permitir el análisis de datos de naturaleza diversa.

A modo general, los algoritmos de ML se pueden agrupar en dos esquemas: el aprendizaje supervisado y el aprendizaje no supervisado (figura 3.1). Donde el enfoque que se elije depende de que los datos estén o no "etiquetados" con una variable objetivo. En el aprendizaje supervisado los modelos que se obtienen son de carácter predictivo. Si las etiquetas son categorías o variables discretas, se usarán algoritmos de clasificación. Por el contrario, si las etiquetas son variables continuas, se usarán los algoritmos de regresión. En el caso del aprendizaje no supervisado, no se conocen las variables objetivo, por lo que el algoritmo debe de ser capaz de identificar patrones útiles en los datos. En los algoritmos de clustering se encuentran grupos de datos que guardan cierta semejanza. La reducción de dimensiones es otra tarea que se puede resolver con este enfoque, aquí se hace una proyección de los datos originales a un espacio de menor dimensión que preserva la mayoría de las propiedades originales.

3.1. Machine learning: flujo de trabajo

Las aplicaciones del machine learning pueden ser muy diversas, sin embargo tienden a seguir el mismo flujo de trabajo, el cual está esquematizado en la figura 3.2. A continuación se presentan los pasos que hay que seguir:

1. **Recolección de datos**: En este paso se crea el catálogo de datos que se va a utilizar. Un paso clave del ML es separar a los datos en 2 conjuntos: el conjunto de entrenamiento y el

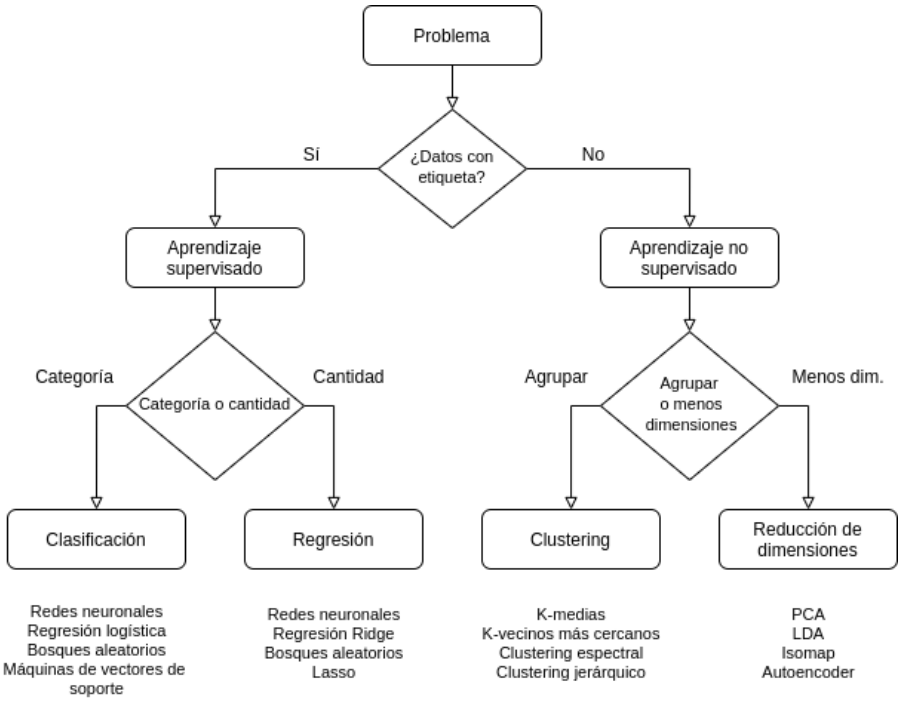

Figura 3.1
Tipos de algoritmos de machine learning. El aprendizaje supervisado funciona con conjuntos de datos etiquetados con el objetivo de crear modelos que predigan variables objetivo (categóricas o continuas). El aprendizaje no supervisado funciona con datos no etiquetados y tiene el objetivo de agrupar datos por semejanza o de reducir la dimensión de los datos de entrada. Además se nombran algunos algoritmos de cada categoría. **Fuente:** Q. Kong y col., 2019

conjunto de prueba. Con el primero se generará un modelo y con el segundo se evaluará el desempeño del modelo[†]. Si se va a utilizar un algoritmo de aprendizaje supervisado, en este paso se incluye la etapa de etiquetado de todos los datos.

2. **Preprocesamiento**: En este paso se incluyen todas las etapas necesarias para que los datos se puedan pasar como variables de entrada a los algoritmos de optimización. Se incluyen tareas como la limpieza del catálogo y, en caso de ser necesario, el filtrado de los datos. Además es posible aplicar métodos de aumento de datos en los que se extiende el número de datos agregando copias ligeramente modificadas de los datos original o creando datos sintéticos. Finalmente, es necesario que los datos pasen por un proceso de estandarización para que funcionen como variables de entrada en los algoritmos de ML, esta etapa se conoce como extracción de atributos. Los atributos son propiedades que

[†]Es común que se usen el 80 % de los datos en el entrenamiento y el 20 % en la prueba. Sin embargo, no existen reglas duras para hacer esta separación. El usuario debe tomar en cuenta: a) el tamaño del conjunto de datos. b) el hecho de que la generalización del modelo dependerá de qué tan representativa es la muestra de los datos que se usan en la etapa de aprendizaje. Por lo que es preferente que el conjunto de entrenamiento sea mayor que el de prueba. En la literatura también se encuentra un tercer conjunto de datos, el conjunto de validación. Este sirve para ajustar los hiperparámetros del modelo y su uso se recomienda cuando se tiene un conjunto de datos lo suficientemente grande. Como alternativa, se pueden usar técnicas de validación cruzada durante el entrenamiento (Géron, 2019).

permiten la caracterización o descripción de los datos. Es posible usar múltiples atributos, de modo que cada dato se representa por un vector de atributos.

3. **Entrenamiento**: En este paso los algoritmos formulan un modelo a partir de los vectores de atributos de los datos del conjunto de entrenamiento. El modelo se formula o se aprende por medio de la optimización de una función que depende del algoritmo que se elija (véase las secciones 3.2, 3.3 y el apéndice A para más detalle). Además en este paso es posible usar técnicas de validación cruzada para seleccionar los mejores hiperparámetros del modelo, que son variables que no se pueden aprender y que el usuario debe asignarles un valor de forma manual.

4. **Evaluación**: En este paso se usan los datos del conjunto de prueba para evaluar el desempeño del modelo. Existen distintas métricas que cuantifican el poder de generalización de un modelo (véase la sección 3.2.1).

5. **Aplicación**: El último paso consiste en la aplicación del modelo obtenido para la resolución del problema.

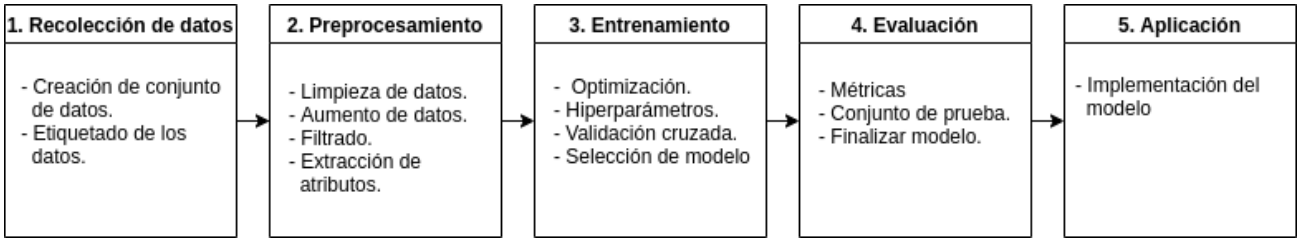

Figura 3.2
Flujo de trabajo genérico para las aplicaciones de machine learning: (1) Recolección de datos. (2) Preprocesamiento. (3) Entrenamiento. (4) Evaluación. (5) Aplicación. **Fuente:** Q. Kong y col., 2019

3.2. Algoritmos de clasificación

En este trabajo se plantea una aplicación de clasificación automática usando un esquema de aprendizaje supervisado. Por lo que en esta sección se presentan aspectos técnicos generales de los métodos de clasificación.

Un algoritmo de clasificación construye un modelo predictivo a partir de un conjunto de datos con etiquetas. Como se había mencionado, a estos datos se les llama conjunto de entrenamiento y formalmente se denotan como $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^{N}$, donde N es el número de instancias en el conjunto de entrenamiento. Cada $\mathbf{x}_i$ representa una instancia u observación de los datos y es un vector en el espacio de atributos, $\mathbf{x}_i \in \mathbb{R}^d$, donde la dimensión d es el número de atributos que se utilizan para describir las instancias. Cada una de estas instancias tiene asociada una etiqueta que describe la clase a la que pertenece. Al conjunto de etiquetas se les denota como

$Y = \{y_i\}_{i=1}^{N}$, con $y_i \in \mathbb{R}$.[‡] Entonces el modelo se puede definir como una función f que toma como entrada al vector $\mathbf{x}$ y devuelve un valor y.

$$f : \mathbf{X} \to Y$$
$$\mathbf{x} \mapsto y.$$

Es común pensar en la función f como una superficie en $\mathbb{R}^d$ que separa al espacio en distintas regiones. Usualmente a esta superficie se le llama *frontera de decisión.*

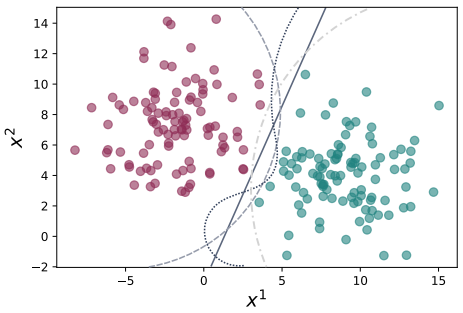

Figura 3.3
Existen múltiples fronteras de decisión para un problema de clasificación y cada algoritmo la define de formas distintas.

La figura 3.3 muestra un esquema de clasificación en dos dimensiones. Nótese que es posible definir más de una frontera de decisión y cada algoritmo la define con criterios distintos. La tarea de aprendizaje automático es básicamente un problema de optimización en el que se elije la mejor de las posibles fronteras de decisión de acuerdo a los criterios del algoritmo que se esté usando. Para esto se define una *función de costo*, $J_N(\theta)$, donde θ son los parámetros del modelo que se aprenden durante la etapa de entrenamiento. El trabajo del algoritmo es encontrar los parámetros θ que minimizan la función de costo.

Finalmente, existen numerosos algoritmos de aprendizaje orientados a la clasificación. Algunos ejemplos son la regresión logística (Menard, 2002), naive-bayes (Rish y col., 2001), modelos ocultos de Markov (Eddy, 2004), árboles de decisión (Quinlan, 1986), máquinas de vectores de soporte (Cortes & Vapnik, 1995) y redes neuronales convolucionales (Albawi y col., 2017). Más adelante, en la sección 3.3 se hace una descripción de los algoritmos que se utilizaron en este proyecto.

3.2.1. Evaluación del modelo

En secciones anteriores se ha discutido el concepto de la evaluación de un modelo. Sin embargo, hasta ahora no se ha presentado una metodología que permita cuantificarla. A continuación se exponen algunas ideas fundamentales para evaluar el desempeño de un modelo.

La única forma de saber qué tan generalizable es un modelo es probando su desempeño en nuevas instancias. Por lo que la evaluación se hace con las instancias del conjunto de prueba.

[‡]Las etiquetas no necesariamente son números reales. Para ilustrar, Dye y Morra, 2020 utilizan fotografías infrarrojas para describir el estado del Monte Erebus, consideraron las siguientes clases $Y = \{$*'En erupción', 'Sin erupción', 'Pre-eruptivo', 'Post-eruptivo*$\}$. En estos casos se hace un mapeo de las clases a un conjunto de números, por ejemplo los naturales $\mathbb{N}$. En el que '*En erupción*' $\mapsto 1$, '*Sin erupción*' $\mapsto 2$ y así sucesivamente. De este modo el formalismo matemático se sostiene.

La métrica más natural es la exactitud, que se define como la tasa de predicciones correctas sobre el total de predicciones (Malfante, 2018),

$$\text{Exactitud} = \frac{\#(\text{predicciones correctas})}{\#(\text{total de predicciones})}. \tag{3.1}$$

Además existen otras métricas que brindan información útil del modelo y que pueden medirse para cada clase (James y col., 2013). La figura 3.4 muestra un esquema del significado de estas métricas. La sensibilidad o exhaustividad indica cuántas de las instancias relevantes se identificaron,

$$\text{Sensibilidad} = \frac{\#(\text{predicciones correctas de la clase i})}{\#(\text{instancias de la clase i})}. \tag{3.2}$$

Mientras que la *precisión* indica la fracción de las instancias recuperadas que son relevantes.

$$\text{Precisión} = \frac{\#(\text{predicciones correctas de la clase i})}{\#(\text{predicciones de la clase i})}. \tag{3.3}$$

Puesto que ambas medidas brinda información útil del modelo, el valor F1 combina ambas en una sola métrica,

$$\text{Valor F1} = 2 \cdot \frac{\text{precisión} \cdot \text{sensibilidad}}{\text{precisión} + \text{sensibilidad}}. \tag{3.4}$$

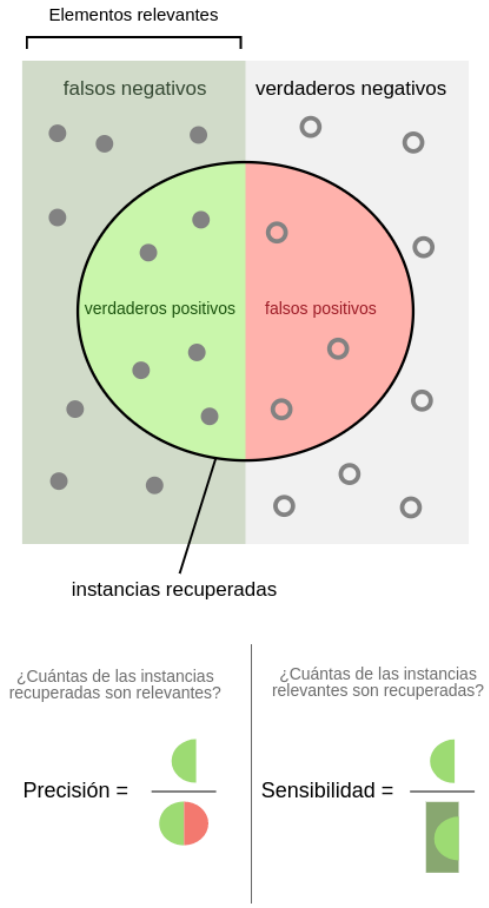

Figura 3.4
Métricas del desempeño de un modelo: sensibilidad y precisión. **Fuente**: Walber, Wikimedia Commons, 2014

Todas estas métricas se pueden calcular de la matriz de confusión que es una herramienta muy útil para analizar los resultados de un problema de clasificación. Se trata de una matriz cuadrada de $C \times C$, donde C es el número de clases que se consideran. Usualmente las columnas indican las predicciones del modelo y las filas las etiquetas reales de los datos. De modo que la diagonal de la matriz muestra las predicciones correctas y los elementos fuera de la diagonal permiten analizar los errores de clasificación.

Para concluir, hay que destacar que si se usan de forma aislada, ninguna de las herramientas que se presentan en esta sección es suficiente para validar o invalidar un proceso. Un análisis correcto del desempeño de un modelo debe incluir la integración de todas ellas.

3.3. Algoritmos utilizados

Hasta ahora se han expuesto aspectos generales del aprendizaje automático que son necesarios para comprender cómo funcionan los algoritmos. El objetivo de esta sección es presentar las técnicas específicas que se usaron en este estudio. En la sección 3.3.1 se introduce la técnica de *análisis de componentes principales* que se usó en la etapa de extracción de atributos. En seguida se discuten los dos algoritmos de clasificación que se emplearon: la sección 3.3.2 presenta el método de *bosques aleatorios* y, por último, las *máquinas de vectores de soporte* se describen en la sección 3.3.3. Intentando seguir un enfoque pedagógico estas secciones incluyen únicamente explicaciones intuitivas de las técnicas, en el apéndice A se presenta el formalismo matemático y algunos aspectos técnicos sobre ellas.

3.3.1. Análisis de componentes principales

El *análisis de componentes principales*, *Principal Component Analysis* o PCA es, probablemente, la técnica más popular para la reducción de dimensiones (Géron, 2019). Usa un enfoque de aprendizaje no supervisado ya que funciona únicamente con los vectores de atributos, sin necesitar de variables objetivo. La idea general del método es encontrar un número menor de atributo para describir los datos de la mejor manera posible. Los atributos nuevos van a ser combinaciones lineales de los originales. Para encontrarlos hay que hacer una proyección ortogonal a una superficie de dimensión menor a la que matemáticamente se le dice *hiperplano*. El método es particularmente útil cuando se tiene un conjunto de datos con atributos correlacionados, pues permite resumir las propiedades del conjunto original creando nuevos atributos independientes (James y col., 2013).

PCA: Intuición

Por simplicidad, considérese un problema en el que las instancias están descritas por dos atributos (x^1, x^2). En la figura 3.5.a se puede ver la distribución de los datos en el espacio de atributos. Sabemos que el PCA es una técnica para la reducción de dimensiones. En el ejemplo, las observaciones están en un plano por lo que reducir la dimensión implica proyectar los datos a una línea recta. Para ilustrar, en la figura 3.5.b se muestran las proyecciones de los datos con respecto a ambos ejes. ¿Cómo saber si una proyección es mejor que la otra y, en particular, cómo saber cuál es la mejor? El PCA contesta estas preguntas.

Naturalmente, la mejor proyección es la que más "parecido" tiene con la representación original, o usando otras palabras, la que minimiza la pérdida de información. En este tipo de problemas podemos pensar a la información como una medida que nos permite distinguir una observación del resto. Es importante notar que todo lo que se sabe de una observación esta descrito por su posición en el espacio de atributos. De modo que una forma de medir qué tanto se puede distinguir entre observaciones es por medio de la separación que hay entre los

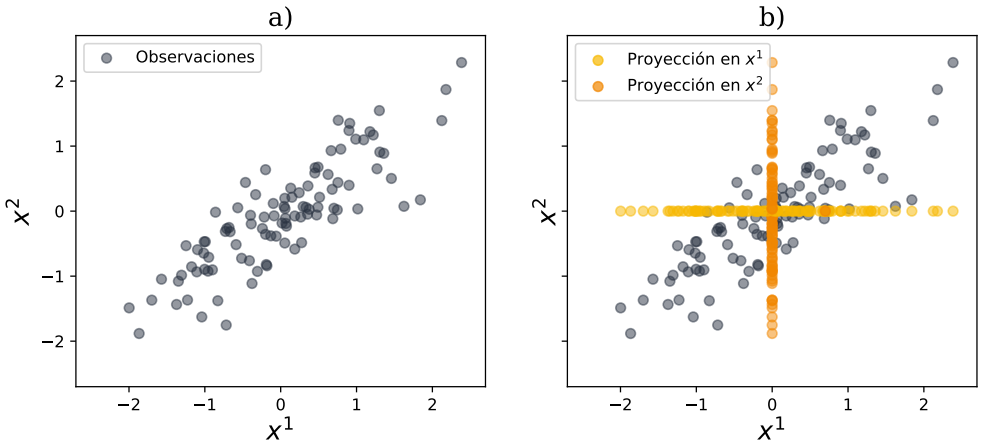

Figura 3.5
Intuición del análisis de componentes principales. En el panel a) se muestra la distribución de un conjunto de observaciones que se describen con los atributos x^1 y x^2. En el panel b) se muestra la proyección de los datos en los ejes x^1 y x^2. El objetivo del PCA es la elección óptima de la línea de proyección.

puntos. En la estadística este concepto es la varianza de una distribución. Así que el problema de optimización del PCA es encontrar una proyección ortogonal que maximice la varianza de los datos proyectados. Las componentes principales (CP) son los vectores que indican las direcciones que cumplen con esta condición y son perpendiculares entre sí. Por ejemplo, en la figura 3.6.a se muestran las dos componentes principales de la distribución de datos. Por convención, la primera componente principal es la que tiene la mayor varianza, las siguientes siguen un orden descendiente. En la figura 3.6.b se muestran las proyecciones sobre la primera componente

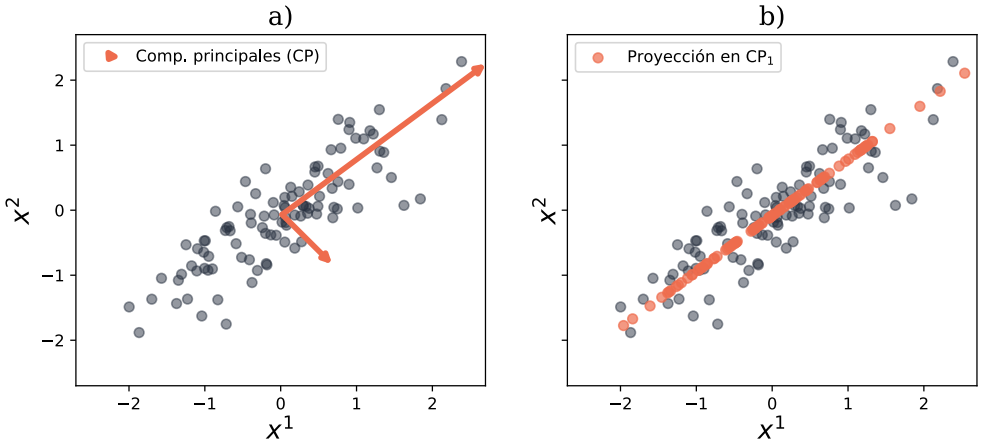

Figura 3.6
Las componentes principales son vectores ortogonales que definen las direcciones en las que la varianza de los datos proyectados es máxima. El panel a) muestra las componentes principales de las observaciones de la figura 3.5. El panel b) muestra las proyecciones de los datos sobre la primera componente principal.

principal.

El formalismo de la varianza máxima fue propuesto por Hotelling, 1933. Sin embargo, unas tres décadas antes, Pearson planteó el mismo problema y su respuesta tiene un enfoque más geométrico. Él propuso que la mejor proyección es aquella que minimiza la distancia entre los puntos originales y la línea sobre la que se proyectan (Pearson, 1901). Este resultado es familiar para cualquier científico, pues este es el mecanismo detrás de los mínimos cuadrados que se usan en una regresión lineal.

Ambos enfoques (varianza máxima y distancia mínima) son completamente equivalentes y fue hasta el desarrollo de las computadoras que se apreció su potencial para tratar grandes conjuntos de datos; más tarde, con la explosión del big data, se reafirmó el interés por esta técnica (Jolliffe, 2005).

3.3.2. Bosques aleatorios

El algoritmo de *bosques aleatorios*, *Random Forest* o *RF* es un método de ML que usa un enfoque de aprendizaje por ensamble o combinación de modelos, en el que se entrenan varios modelos y el resultado final es una media de los resultados individuales. En la literatura, es común encontrar el término *comité* para referirse a este tipo de combinaciones (Bishop, 2006). Por ejemplo, en un problema de clasificación, la clase que se asigna a una instancia es aquella que recibió el mayor número de "votos". Donde cada voto es la clase asignada por uno de los modelos individuales.

Como su nombre lo indica, los bosques aleatorios resultan de la unión de decenas o cientos de *árboles de decisión* (Quinlan, 1986). La idea del método se introdujo por Ho, 1995 pero el esquema que se usa actualmente fue desarrollado por Breiman, 2001. Desde entonces se ha aplicado con éxito en diversas áreas y es uno de los algoritmos de aprendizaje más usados debido a la facilidad con la que se pueden interpretar sus resultados.

RF: Intuición

Los árboles de decisión dividen el espacio de atributos en regiones cuboides, con bordes paralelos a los ejes, que delimiten instancias de la misma clase (Bishop, 2006). Para esto se usa un proceso recursivo; en el primer paso se elige un criterio que divida el espacio en dos regiones, en el segundo paso se eligen nuevos criterios para separar estas regiones en subregiones y así sucesivamente. Naturalmente, el objetivo es elegir criterios de partición que permitan que instancias de una misma clase estén dentro de la misma región.

Para ilustrar su comportamiento se presenta el ejemplo de la figura 3.7. Supongamos que se tiene una base de datos de los destinos vacacionales de un grupo de mexicanos. En el panel A se muestra la distribución de los turistas según su presupuesto y la duración del viaje. En el panel B se presenta el diagrama del modelo o árbol que se obtiene para su clasificación. Los

puntos amarillos son individuos que viajaron dentro de México, los morados fueron a Estados Unidos, los azules a Francia y los naranjas a Japón. En el primer paso o *nodo* del panel B, el espacio se divide entre los viajeros que pagaron menos de $\$10,000$ pesos y los que gastaron más. En la región de lado izquierdo solo quedan instancias de la misma clase, así que no tiene sentido hacer más divisiones. Por contrario, la región del lado derecho es heterogénea y se crea un nuevo nodo entre aquellos que gastaron menos de $\$30,000$ pesos y los que gastaron más. Finalmente, la subregión de más de $\$30,000$ se separa usando un criterio sobre la duración del viaje. Nótese que cualquier instancia nueva va a pertenecer a una de estas regiones y, por lo tanto, se le puede asignar una clase.

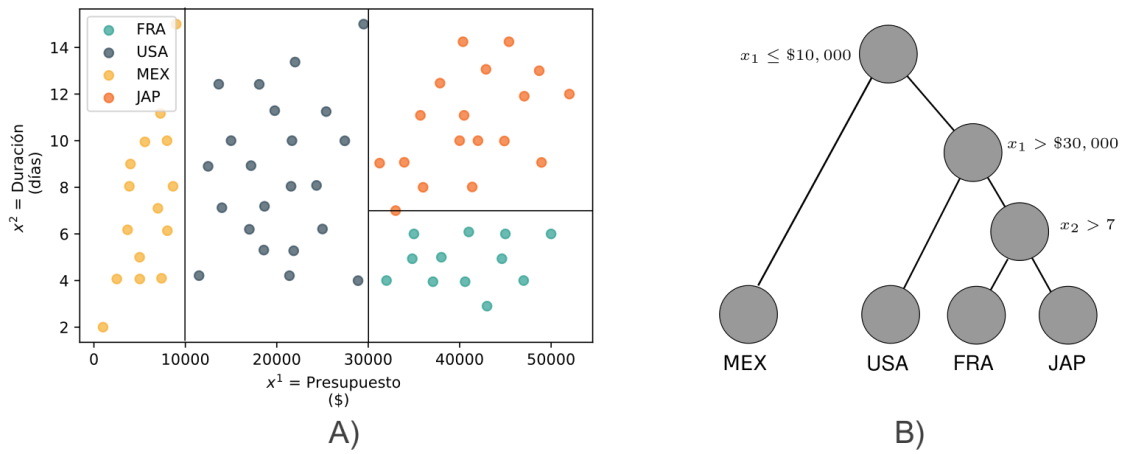

Figura 3.7
Ejemplo de aplicación de un árbol de decisión. A) Distribución de los datos en el espacio de atributos. B) Diagrama del modelo de un árbol de decisión para los datos del panel A.

El problema de optimización está en determinar la estructura del árbol. Para esto es necesario establecer dos parámetros por cada i nodo: El atributo x_i^j que se usará en el criterio y el valor umbral θ_i para hacer la partición. En este sentido se introduce el concepto de impureza, que es una medida de la heterogeneidad de las clases asociadas a los puntos que están dentro de una región. Por lo tanto, el objetivo es encontrar el par (x_i^j, θ_i) que minimice la impureza de las dos regiones definidas en cada nodo.

El modelo que se presenta en la figura 3.7.B es un claro ejemplo de la facilidad con la que se pueden interpretar los resultados de un árbol de decisión. Esta propiedad hizo que el algoritmo ganara mucha popularidad en las últimas décadas del siglo pasado. Sin embargo, en la práctica se encontró que los modelos tipo árbol son altamente sensibles al conjunto de datos; cambios pequeños producen modelos muy distintos. Ciertamente, valores altos de varianza afectan el poder de generalización del modelo y, en consecuencia, la relevancia del método es cuestionable.

La solución si bien es sencilla, también es muy ingeniosa. Breiman propone que es posible reducir la varianza si se usa la media de una colección de árboles (Breiman, 2001; Hastie y col., 2017). El detalle está en que para garantizar la reducción es necesario que no haya

correlación entre los árboles. La justificación es inmediata, el ruido de los modelos individuales se balancea cuando se considera un promedio. Para garantizar la descorrelación, es necesario incluir una componente aleatoria en la construcción de los clasificadores, específicamente se modifican dos aspectos: 1) Cada árbol se construye con un subconjunto de los datos de entrenamiento que se elige de forma aleatoria. 2) El criterio de selección de cada nodo se hace con un subconjunto aleatorio del total de atributos. El nombre natural de método que incluye estas modificaciones es *bosques aleatorios*.

3.3.3. Máquinas de vectores de soporte

Las *máquinas de vectores de soporte*, *Support Vector Machine* o SVM son modelos de ML capaces de realizar tareas de clasificación con fronteras de decisión lineales y no lineales, también pueden aplicarse en problemas de regresión. Siguen un enfoque de aprendizaje supervisado por lo que es necesario tener un conjunto de datos etiquetado.

Vapnik y Chervonenkis desarrollaron las primeras ideas de SVM en 1962 (Chervonenkis, 2013). Desde entonces, el método ha experimentado varias mejoras, la versión general más aceptada fue propuesta por Cortes y Vapnik, 1995. Se trata de uno de los algoritmos más usados en la industria e investigación y se ha aplicado con éxito en múltiples disciplinas. Dentro de las ciencias de la Tierra se ha implementado en las áreas de vulcanología (Malfante, Dalla Mura y col., 2018), sismología (J. Zhu y col., 2021), geofísica de exploración (Abedi y col., 2012), geología (Han y col., 2021), ciencias de la atmósfera (Chen y col., 2010) y ciencias del mar (X. Kong y col., 2018), entre otras.

SVM: Intuición

Pensemos en un problema de clasificación binario, es decir, que solo se consideran dos clases. El método de SVM busca separarlas por medio de una frontera lineal. Entonces cuando los datos se describen usando de dos dimensiones, la frontera lineal es una línea recta; en el caso de tres dimensiones se trata de un plano y cuando se trabaja con más de tres dimensiones se busca un hiperplano.

En la figura 3.8.a se aprecia que existen varias rectas que pueden dividir a las dos clases. Sin embargo, hay que recordar que los algoritmos de ML buscan fronteras de decisión que permitan que la generalización de un modelo sea buena. La forma en la que el SVM hacen esto es por medio de un margen, que se define como la distancia entre el hiperplano y el punto más cercano. La idea principal es que el mejor hiperplano es aquel que maximiza el margen, así se maximiza el espacio entre clases y se promueve la generalización. En la figura 3.8.b, la frontera es la línea continua y el margen de ambos lados se marca con una línea discontinua.

Una consideración importante es que no es necesario usar todos los datos para encontrar la frontera de decisión. Pues, considerando la definición del margen, basta con usar los puntos más cercanos al hiperplano. Estos se conocen como los vectores de soporte (Boser y col., 1992)

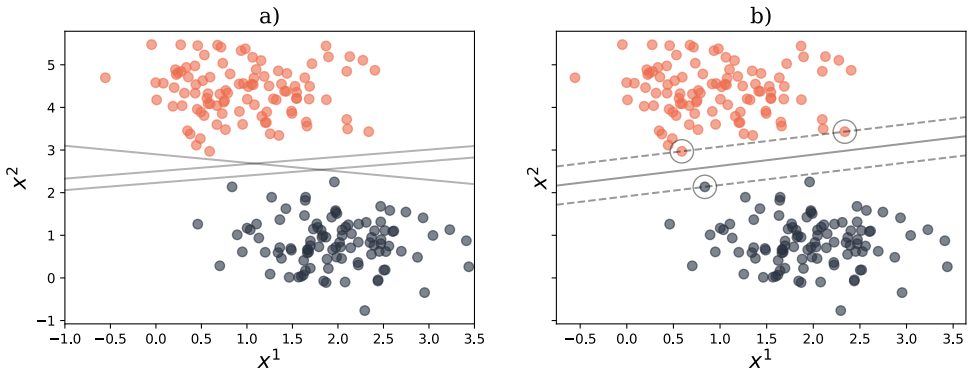

Figura 3.8
Frontera de decisión en las máquinas de vectores de soporte. En el panel a) se muestran varias distintas propuesta de fronteras de decisión lineales. La solución de la SVM se presenta en el panel b). Las líneas discontinuas son los márgenes y los vectores de soporte se marcan con círculos grises.

y su uso supone una simplificación en los cálculos y mejoras en el tiempo de cómputo. En la figura 3.8.b es fácil reconocer porqué los vectores de soporte son los únicos puntos necesarios para calcular la frontera de decisión, estos se destacan con un círculo gris.

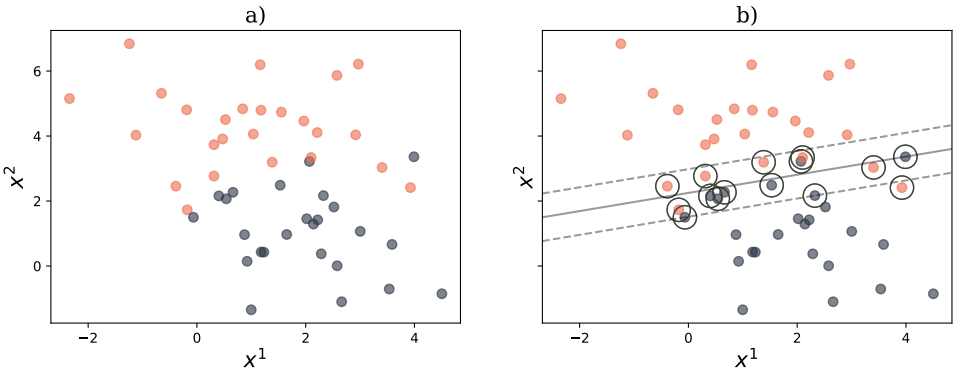

Figura 3.9
Clases con regiones de empalme. En el panel a) se muestra la distribución de las observaciones. La frontera calculada con un margen suave se presenta en el panel b). Los vectores de soporte se destacan con un círculo gris, los márgenes son las líneas discontinuas y la frontera de decisión es la línea continua.

Hasta aquí el planteamiento no está completo, pues hay dos situaciones que se pueden presentar con facilidad en aplicaciones reales. Por ejemplo, en el panel a) de la figura 3.9 se muestra un caso en el que la separación entre las dos clases es difusa, pues no es posible definir una línea recta que separe perfectamente todas las instancias de las dos clases. Por otro lado, también se puede presentar el caso de la figura 3.10.a en el que es imposible encontrar una frontera lineal que separe las clases de forma satisfactoria. Para lidiar con estos problemas se plantearon las siguientes consideraciones.

La solución al primer caso fue propuesta por Cortes y Vapnik, 1995. En ella proponen un margen suave que permite que el modelo cometa algunos errores de clasificación en los datos

el conjunto de entrenamiento. Por lo que es posible trabajar con regiones de empalme y clases con valores atípicos sin sufrir una pérdida de generalización. En la figura 3.9.b se muestra la frontera de decisión cuando se usa un margen suave. De nuevo, todos los vectores de soporte se destacan con un círculo gris, los márgenes son las líneas discontinuas y la frontera es la línea continua. Aquí se puede ver que sí hay errores en la clasificación pero se tiene un modelo que probablemente funcione mejor con instancias nuevas.

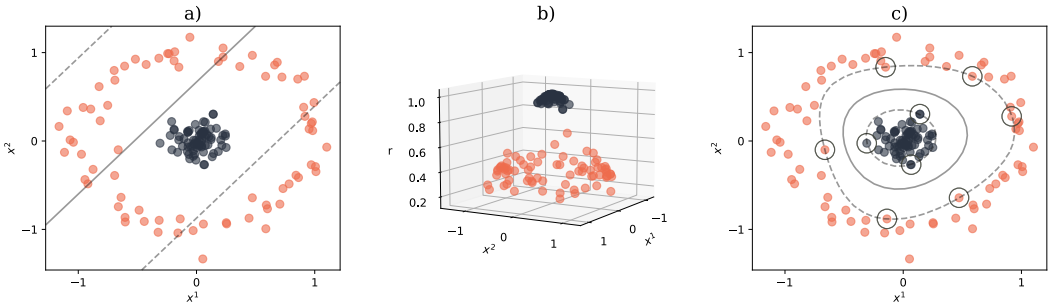

Figura 3.10
Clasificación no lineal con SVM. En el panel a) se presenta una distribución que no se puede separar con una frontera lineal. Las líneas grises muestran la frontera y los márgenes que calcula una SVM 'clásica'. Claramente los resultados no son satisfactorios. Las observaciones se pueden llevar a un espacio de más dimensiones en el que sí sean linealmente separables (panel b). Finalmente, en el panel c) se muestran la frontera, los márgenes y los vectores de soporte cuando se hace la transformación inversa.

Por otro lado, para resolver el problema de las clases que no se pueden separar linealmente hay que hacer un truco matemático. Formalmente, este método se conoce como el truco del kernel y consiste en aplicar una transformación no lineal que lleve a las observaciones a un espacio en el que sí se puedan separar con un hiperplano (Boser y col., 1992). Un ejemplo de esto se muestra en el panel b) de la figura 3.10. En el que se aplicó una función de base radial que lleva a las observaciones a un espacio tridimensional en el que sí es posible separar las clases con una frontera lineal. Una vez que se encuentra la frontera se aplica una transformación inversa, tanto a los datos como a la frontera, para regresar al espacio original. En el panel c) de la figura 3.10 se muestra la frontera, los márgenes y los vectores de soporte una vez que se transforman con la función de base radial inversa.

En resumen, las tres propiedades clave del método SVM son: 1) Maximizar el margen, 2) Permitir errores de clasificación durante el entrenamiento y 3) En caso necesario, llevar a las observaciones a un espacio en el que sean separables linealmente.

3.4. Clasificación de señales volcano-sísmicas.

Tradicionalmente la detección y clasificación de señales volcano-sísmicas se realiza de forma manual en los observatorios vulcanológicos. Esta es una tarea que requiere de grandes cantidades de tiempo y, debido a la manipulación por diferentes operadores, es común que sufra

por la falta de criterios estandarizados que puede resultar en catálogos heterogéneos (Cortés y col., 2021). Aún más, la clasificación manual no puede llevarse a cabo lo suficientemente rápido cuando hay periodos en los que la actividad incrementa de forma significativa, de modo que el análisis de comportamientos y la posible detección de precursores de actividad eruptiva se hace a posteriori.

El uso de técnicas de machine learning para abordar estas dificultades se ha investigado desde finales del siglo pasado (Falsaperla y col., 1992). Desde entonces se han implementado una gran variedad de algoritmos usando enfoques de aprendizaje supervisado (p.ej Langer y col., 2006; Maggi y col., 2017; Malfante, 2018; Ohrnberger, 2001) y también de aprendizaje no supervisado (p.ej. Carniel y col., 2013; Köhler y col., 2010). El desempeño de los primeros ha alcanzado rangos de $80 - 90\,\%$ en exactitud, mientras que los segundos difícilmente superan el $70\,\%$. En la tabla 3.1 se reporta un resumen de algunos de los estudios realizados en los últimos años. La experiencia adquirida del uso de estas técnicas ha hecho posible identificar cuatro grandes retos (Cortés y col., 2021):

1. El entrenamiento del modelo necesita de grandes catálogos de eventos previamente etiquetados. La tabla 3.1 muestra algunos ejemplos del número de eventos necesario.

2. Con frecuencia, los modelos que se obtienen no son robustos. Pues es común que el diseño de modelos se haga para un volcán y una estación específica. Así que su desempeño disminuye cuando hay cambios en el sensor o en el tipo de actividad.

3. La aplicación de los modelos al análisis de señales continuas se complica porque no solo hay que clasificar las señales, sino que el modelo debe de ser capaz de detectar los eventos. Aún más, en la gran mayoría de los casos, el entrenamiento se aplica sobre las señales de los eventos aislados; mientras que en el análisis continuo es frecuente que la ventana de análisis contenga solo un fragmento del evento o que esté "contaminada" por más de un evento.

4. La falta de software con interfaces intuitivas y sencillas de usar hace que la instalación de esquemas de aprendizaje automático en los observatorios requiera de personal experto en programación. Esto no solo resulta en un aumento de costo, sino que entorpece la integración entre protocolos estándar y los servicios de ML.

En los últimos años se han desarrollado estrategias que permiten abordar los retos que se mencionaron. Estas pueden ir desde soluciones sencillas hasta la integración de técnicas más complejas que complementan el flujo de trabajo. Por ejemplo, Malfante, Dalla Mura y col., 2018, agregan la clase de "Ruido" a su entrenamiento para que la detección de eventos en el análisis de señales continuas se haga de forma inmediata. Por otro lado, Bueno y col., 2019, utilizan los avances en deep learning para que su modelo funcione en la clasificación continua de 3 estaciones del volcán Bezymianny y 3 estaciones del monte Santa Helena. Por su parte, Cortés y col., 2021, usaron modelos ocultos de markov para desarrollar un sistema de reconocimiento que puede usarse en distintos volcanes y que contiene una interfaz de usuario gráfica que promete ser fácil de usar.

Estudio	Eventos	Clases	Algoritmo	Exactitud	A. continuo
Bueno y col., 2019	91,496	3	NN + TL	94.60 %	Sí
Cortés y col., 2009	6,788	7	HMM	85.48 %	Sí
Lara y col., 2020	49,675	5	EMD + PCA + SVM	90.50 %	No
Falcin y col., 2021	7,149	5	RF	83 % ± 2 %	No
Maggi y col., 2017	71,930	8	RF	93.56 % ± 1.65 %	No
Malfante, Dalla Mura y col., 2018	109,609	6	SVM	93.5 % ± 0.5 %	Sí
Titos y col., 2018	9,332	7	NN	94.32 % ± 0.66 %	No

Tabla 3.1: Resumen de algunos estudios en los que se ha aplicado técnicas de machine learning para la clasificación de señales sismo-volcánicas. Se muestra el número de eventos en el catálogo, el número de clases consideradas, los algoritmos empleados, la exactitud reportada y si el modelo se implementó en el análisis de señales continuas. Se usan las abreviaciones estándar de los algoritmos: NN redes neuronales, TL transfer learning, HMM modelos ocultos de markov, EMD descomposición empírica de modos, PCA análisis de componentes principales, SVM máquinas de soporte de vectores y RF bosques aleatorios

Por último, se destaca el estudio realizado por Cortés y col., 2009 en el que se emplean modelos ocultos de markov para la clasificación de señales sísmicas del Popocatépetl y del volcán de Colima. En su conjunto de entrenamiento incluyen 4687 eventos detectados en el volcán de Colima y 2101 eventos del Popocatépetl, por lo que el modelo que obtienen sirve en el análisis de ambos volcanes. Las clases que consideran son: ruido (NO), explosiones(EX), eventos LP, sismos regionales (RE), tremor pulsante (TR1), tremor espasmódico (TR2), tremor armónico (TR3), colapsos (CO), lahares (LA) y sismos VT. En la figura 3.11 se muestran los resultados obtenidos para el análisis continuo en señales desconocidas para el modelo. Los resultados para el Popocatépetl se pueden comparar con los que se obtienen en este trabajo y se retomarán en la discusión del capítulo 5.

Class accuracy (%)										
	NO	EX	LP	RE	TR1	TR2	TR3	CO	LA	VT
Popo	55.18		77.32	62.88	58.21	13.87				86.21
Colima	85.17	90.04	88.02	91.58	58.09	41.67	70.39	78.29	64.97	86.98
Joint	76.11	93.12	83.66	84.01	53.6	19.19	69.44	77.27	59.41	88.98

	Popo	Colima	Joint	Mean
Mean class acc (%)	58.95	75.52	70.48	68.32
Event acc (%)	72.42	72.42	81.71	77.47

Figura 3.11
Exactitud por clase para el análisis continuo en señales de prueba en el estudio hecho por Cortés y col., 2009. Las clases que se consideran son ruido (NO), explosiones(EX), eventos LP, sismos regionales (RE), tremor pulsante (TR1), tremor espasmódico (TR2), tremor armónico (TR3), colapsos (CO), lahares (LA) y sismos VT.

Capítulo 4

Metodología

La metodología que se siguió es la que se presenta en el flujo de trabajo de la sección 3.1 y que está formada por 5 pasos: recolección de datos, preprocesamiento, entrenamiento, evaluación y aplicación. En las secciones de este capítulo se describe la forma en la que se implementó cada uno de ellos. Para aplicar los últimos 4 pasos se usó el código desarollado por Malfante, Mars y col., 2018 y que lleva el nombre de *Arquitectura de Análisis Automático (AAA)*. Se trata de un esquema desarrollado en el lenguaje de programación Python que permite la detección y clasificación de eventos, para implementar los algoritmos de machine learning usa el paquete Scikit-learn (Pedregosa y col., 2011). La AAA puede descargarse a través de GitHub y es un software libre, lo que permitió hacer algunas modificaciones y aportaciones que se destacan a lo largo del capítulo. La AAA fue diseñada para aplicarse con cualquier tipo de señal digital, sin embargo se ha implementado con éxito en la clasificación de señales volcano-sísmicas por Falcin y col., 2021; Lara y col., 2020; Malfante, Dalla Mura y col., 2018.

4.1. Recolección de datos

El material con el que inició este proyecto fueron las señales sísmicas crudas que se obtienen de varias estaciones del volcán Popocatépetl. Tomando esto en cuenta, la recolección de datos se llevó a cabo en dos etapas: en la primera se realizó la conversión de los archivos a formatos típicos de forma de onda en el análisis sísmico (SAC y MiniSeed) y posteriormente se organizaron en una base de datos estandarizada, en la segunda etapa se creó el catálogo de eventos que se usó para entrenar y evaluar el modelo clasificador.

4.1.1. Base de datos de señales diarias

El Popocatépetl cuenta con un conjunto de hasta 19 estaciones sísmicas de banda ancha que se muestran en la figura 4.1. 11 de estas estaciones – marcadas con puntos azules–

son gestionadas por el CENAPRED y cuentan con teletransmisión para un monitoreo en tiempo real. Las 8 estaciones restantes –marcadas con puntos amarillos– están a cargo del doctor Marco Calò del Instituto de Geofísica de la UNAM y cuentan con un sistema de almacenamiento *in situ*. En la tabla B.1 del apéndice B se presentan algunos aspectos técnicos sobre las estaciones.

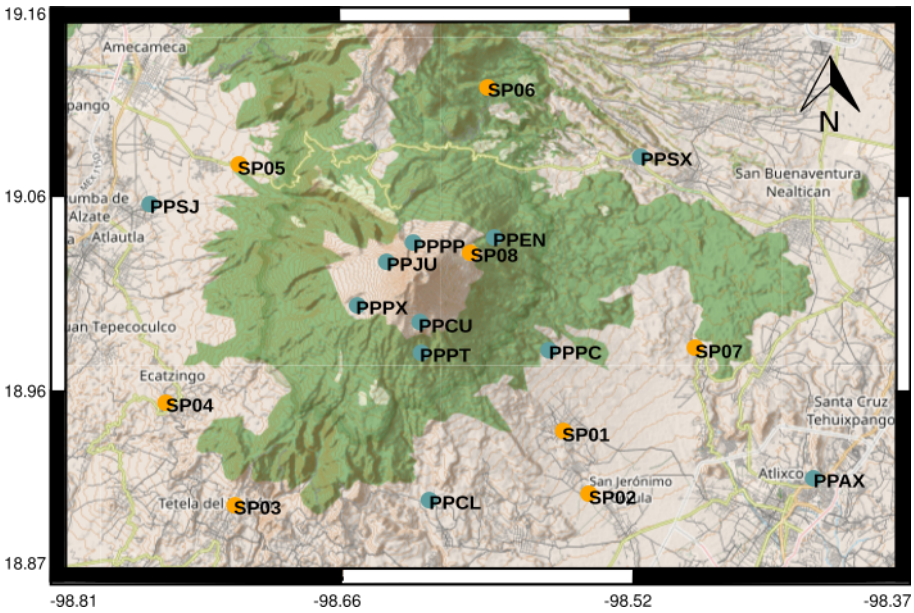

Figura 4.1
Red de estaciones sísmicas del volcán Popocatépetl. Puntos azules indican las estaciones gestionadas por Cenapred y los naranjas las gestionadas por el departamento de Vulcanología del Instituto de Geofísica, UNAM.

Debido a que el monitoreo del volcán se realiza de forma continua la red sísmica genera grandes cantidades de datos. Para que el acceso a estos datos fuera eficiente se planteó un esquema de organización que permitió estandarizar su manejo. La base de datos está almacenada en un clúster de computadoras llamado GAIA del Instituto de Geofísica de la UNAM. Cuenta con archivos diarios del registro en cada estación. Todos los archivos están disponibles en dos formatos: Seed y SAC. Además, usando los archivos de formato SAC se realizó la deconvolución de las señales para remover la respuesta instrumental. En la figura 4.2 se esquematiza la estructura de la base de datos. Cada formato (Seed, SAC y SAC sin respuesta intrumental) está almacenado en una carpeta que sigue el mismo esquema: en el primer nivel se tienen carpetas organizadas por año, en el segundo nivel se tienen carpetas organizadas por día juliano y en el tercer nivel se tienen los archivos de formas de onda. El nombre de los archivos también está estandarizado. Por ejemplo:

MC.SP08.HHZ.2019.353.000000.sac

Los primero dos caracteres indican el gestor de la estación (CN para CENAPRED y MC para el Dr. Marco Calò), los siguientes cuatro caracteres son el código de la estación, los siguientes tres la componente (HHE, HHN, HHZ), los siguiente 4 el año, después se usan tres caracteres

para el día juliano y finalmente se usan 6 caracteres para indicar la hora de inicio del archivo en el formato HH:MM:SS.

La base cuenta con los datos desde el 2017 para las estaciones gestionadas por CENAPRED y desde el 2018 para las estaciones gestionadas por el Dr. Marco Calò y se actualiza frecuentemente ya que ambas redes siguen funcionando. Contar con los archivos organizados de esta forma facilitó el análisis de todas las etapas de este proyecto y seguirá siendo útil en estudios futuros. Cabe aclarar que debido al volumen de datos la conversión y organización de los archivos fue un trabajo en equipo, la M. en C. Leonarda Isabel Esquivel Mendiola –también parte del Posgrado en Ciencias de la Tierra– estuvo a cargo de las estaciones gestionadas por el CENAPRED y yo me encargué de las estaciones gestionadas por el Dr. Marco Calò.

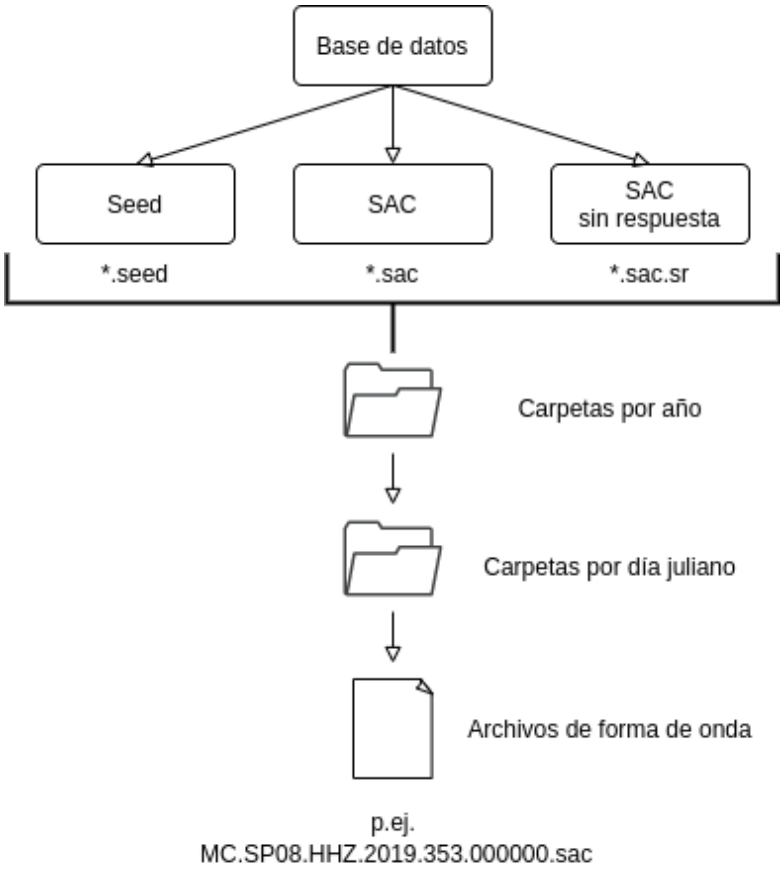

Figura 4.2
Base de datos de señales diarias. Cada formato (Seed, SAC y SAC sin respuesta intrumental) está almacenado en una carpeta que sigue el mismo esquema: en el primer nivel se tienen carpetas organizadas por año, en el segundo nivel se tienen carpetas organizadas por día juliano, en el tercer nivel se tienen los archivos de formas de onda.

4.1.2. Catálogo de eventos

Una vez que se organizó la base de datos con las señales diarias se creó un catálogo con eventos clasificados manualmente. El catálogo cuenta con 507 eventos ocurridos entre septiembre del 2019 y abril del 2021. El número de eventos por clase se muestra en la figura 4.3: 106 eventos de ruido (NO), 105 eventos LP, 91 tremores (TR), 85 explosiones (EX), 82 sismos VT y 38 sismos regionales (RE).

La identificación de estos eventos, en particular de las explosiones y los VT, se realizó con ayuda de los reportes diarios publicados por el CENAPRED. Para la mayoría de los eventos se consideró la señal registrada en la estación disponible más cercana al cráter (PPPP, PPPJU o PPCU). Sin embargo, para algunos eventos VT se consideró la estación más cercana al evento. En todos los casos se usó la componente vertical.

El tamaño del catálogo se podría considerar modesto, en particular si se compara con los catálogos de los estudios que se resumen en la tabla 3.1 del capítulo anterior. Esta aparente limitación se debe a cuestiones de tiempo y se compensa con una estrategia de aumento de datos que se describe en la siguiente sección.

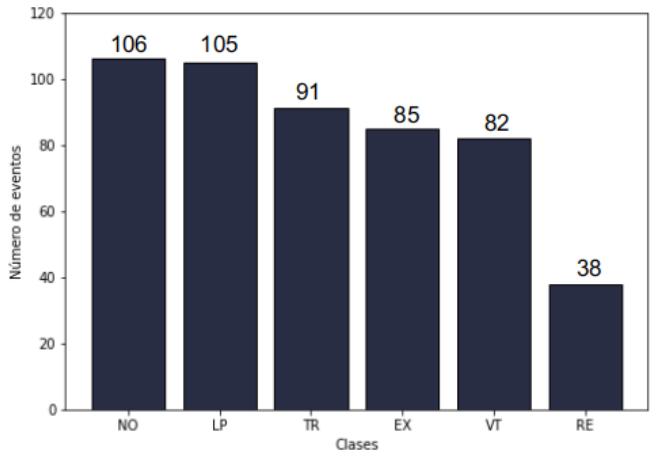

Figura 4.3
Catálogo de eventos clasificados manualmente. Se consideraron 106 eventos de ruido (NO), 105 eventos LP, 91 tremores (TR), 85 explosiones (EX), 82 sismos VT y 38 sismos regionales (RE).

4.2. Preprocesamiento

El preprocesamiento consistió en cuatro etapas: aumento de datos, filtrado de las señales, extracción de atributos y separación de los conjuntos de entrenamiento y prueba. Las etapas de aumento de datos y extracción de atributos tienen algunas sutilezas y se describen en las siguientes subsecciones. Por otro lado, el filtrado de los datos se realizó con un filtro de tipo butterworth de orden 4 con un ancho de banda de 0.66 a 45 Hz. Finalmente, la separación de los

conjuntos de entrenamiento y prueba se hizo con un 85 % y 15 % de señales, respectivamente.

Vale la pena mencionar que el ancho de banda considerado no fue una elección arbitraria, ya que se hicieron ejercicios variando la frecuencia mínima entre 0.01, 0.1, 0.3 y 0.66 Hz, los resultados no variaron para la clasificación aislada pero sí hubo una diferencia en la clasificación continua, donde el mejor resultado se obtuvo con 0.66 Hz.

4.2.1. Estrategia de aumento de datos

Debido a que el fundamento teórico de los métodos de ML es la estadística, es intuitivo que el desempeño de los modelos incrementa con el número de datos. Por esto existe un área de la ciencia de datos que se dedica al desarrollo de técnicas para aumentar el número de datos. En términos generales, las técnicas y estrategias de aumento de datos se pueden dividir en dos: aquellas que generan datos completamente sintéticos y las que que incluyen copias ligeramente modificadas de los datos originales.

Tomando como inspiración a los métodos en los que se añaden datos ligeramente modificados, se decidió aprovechar la información brindada por los distintos sensores triaxiales. En este sentido, se logró extender el catálogo original con dos estrategias: 1) se agregaron las señales de las componentes horizontales y 2) se agregaron algunas señales de los eventos del catálogo original registradas en otras estaciones cercanas.

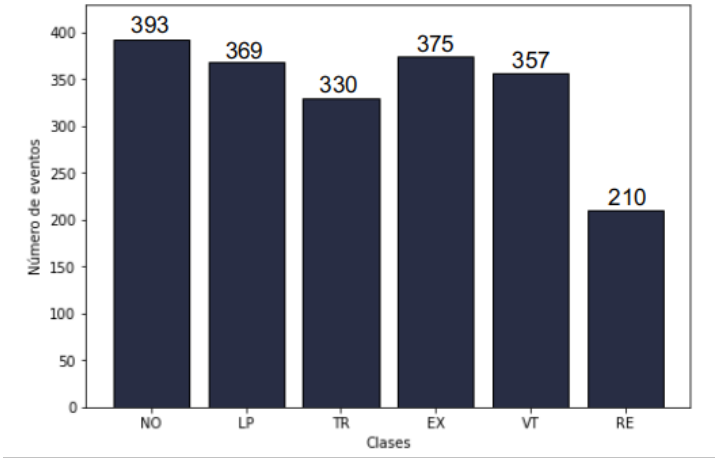

Figura 4.4
Catálogo extendido que incluye las señales de las tres componentes y los registros de otras estaciones. Se consideraron 393 señales de ruido (NO), 369 señales LP, 330 señales de tremor (TR), 375 señales de explosiones (EX), 357 señales de sismos VT y 210 señales de sismos regionales (RE).

El catálogo extendido cuenta con 2034 señales y la distribución por clase se muestra en la figura 4.4. Se consideraron 393 señales de ruido (NO), 369 señales LP, 330 señales de tremor (TR), 375 señales de explosiones (EX), 357 señales de sismos VT y 210 señales de sismos regionales (RE). Por simplicidad, en el resto del manuscrito se hará referencia a este catálogo como el catálogo 2 y al original como el catálogo 1. En el capítulo 5 se comparan los resultados

obtenidos para ambos catálogos. La etapa de aumento de datos no se incluye en la AAA y se considera una aportación original.

Vale la pena aclarar que, en la separación del catálogo extendido en los conjuntos de entrenamiento y prueba, se cuidó que todos las señales asociadas a un mismo evento estuvieran dentro del mismo conjunto. De esta forma la evaluación se hace con eventos completamente desconocidos para el modelo.

4.2.2. Extracción de atributos

La AAA incluye un método para la extracción de atributos que fue desarrollado como parte de la tesis doctoral de Malfante, 2018 y es la característica distintiva de la arquitectura. Este considera a las señales en tres dominios distintos que permiten resaltar algunas de sus características inherentes:

1. *Dominio temporal*: resalta las propiedades de la forma de onda, $z[t]$.

2. *Dominio espectral*: permite destacar el contenido de frecuencias de las señales. Se calcula con la transformada de Fourier, $\mathcal{F}\{\cdot\}$, de la señal en el dominio de tiempo, $Z[f] = \mathcal{F}\{z[t]\}$.

3. *Dominio cepstral*: se usa frecuentemente en el procesamiento digital de voz, se calcula aplicando la transformada de Fourier dos veces sobre la señal y destaca sus propiedades armónicas, $Z[q] = \mathcal{F}\{|Z[f]|\}$.

En cada uno de estos dominios se calculan 34 atributos que describen la forma, distribución estadística y entropía de las señales. Estos se presentan en la tabla 4.1, donde se incluye su descripción matemática y un número de referencia que permite distinguirlos en el análisis que se describe en el capítulo 5.

Todos los atributos de la tabla 4.1 se calculan en los tres dominios. Por lo que cada señal del catálogo se representa por medio de un vector de atributos que tiene 102 componentes.

Ahora bien, tomando en cuenta que la representación propuesta por Malfante, 2018 es una descripción genérica de las señales y no está vinculada –al menos de forma directa– con la física de las señales volcano-sísmicas, se decidió aplicar la técnica de PCA para obtener una representación alterna. Esto también es un aporte original a la metodología y en el capítulo 5 se justifica su uso y se comparan los resultados de las dos representaciones. Para distinguir entre ambas, me referiré como representación AAA a la que se obtiene de los atributos de la tabla 4.1 y como representación PCA a la que se obtiene del PCA.

Atributo	Fórmula	Ref
Longitud	n	1
Media	$\mu = \frac{\sum_i z_i}{n}$	2
Desviación estándar	$\sigma = \sqrt{\frac{1}{n-1}\sum_i (z_i - \mu)^2}$	3
Asimetría	$\frac{1}{n} \cdot \sum_i \left(\frac{z_i - \mu}{\sigma}\right)^3$	4
Curtosis	$\frac{1}{n} \cdot \sum_i \left(\frac{z_i - \mu}{\sigma}\right)^4$	5
Centroide	$\bar{i} = \frac{1}{E}\sum_i i \cdot E_i$	6
Ancho de banda RMS	$RMS_i = \sqrt{\sum_i \frac{i^2 \cdot E_i}{E} - \bar{i}^2}$	7
Asimetría media	$\left(\frac{\sum_i (i-\bar{i})^3 E_i}{E \cdot RMS_i^3}\right)^{1/2}$	8
Curtosis media	$\left(\frac{\sum_i (i-\bar{i})^4 E_i}{E \cdot RMS_i^4}\right)^{1/2}$	9
Entropía de Shannon	$-\sum_i p(z_i)\log_2(p(z_i))$	10 - 12
Entropía de Rényi	$\frac{1}{1-\alpha}\log_2\left(\sum_i p(z_i)^\alpha\right)$	13 - 18
Tasa de ataque	$\max_i \left(\frac{z_i - z_{i-1}}{n}\right)$	19
Tasa de descenso	$\min_i \left(\frac{z_i - z_{i+1}}{n}\right)$	20
Mínimo sobre media	$\min z_i/\mu$	21
Máximo sobre media	$\max z_i/\mu$	22
Energía	$E = \sum_i z_i^2$	23
Máximo de energía	$\max(z_i^2)$	24
Energía promedio	$\mu_E = \frac{\sum_i z_i^2}{n}$	25
Desviación estándar de energía	$\sigma_E = \sqrt{\frac{1}{n-1}\sum_i (E_i - \mu_E)^2}$	26
Asimetría de energía	$\frac{1}{n} \cdot \sum_i \left(\frac{E_i - \mu_E}{\sigma_E}\right)^3$	27
Curtosis de energía	$\frac{1}{n} \cdot \sum_i \left(\frac{E_i - \mu_E}{\sigma_E}\right)^4$	28
Mínimo de la señal	$\min_i(z_i)$	29
Máximo de la señal	$\max_i(z_i)$	30
i de mínimo	$\text{argmin}_i(z_i)$	31
i de máximo	$\text{argmax}_i(z_i)$	32
Tasa de cruce por umbral	$\frac{\#\text{cruces por umbral}}{n}$	33
Tasa de silencio	$\frac{\#z_i \text{ donde } z < \text{umbral}}{n}$	34

Tabla 4.1: Atributos que se utilizan para caracterizar a las señales. Estos se calculan sobre la señal $z[i]_{i=1}^n$, donde i se refiere a la muestra temporal, espectral o cepstral. Todos los atributo se calculan en los tres dominios. $E_i = z[i]^2$, es la energía en la muestra i. Los números de referencia de la tercera columna se usan en la discusión de resultados para distinguir entre atributos. La entropía mide el contenido promedio de información de la señal. La probabilidad, $p(z_i)$, se calcula a partir del histograma de los valores de amplitud de la señal en cualquiera de los dominios. Siguiendo la metodología de Malfante, Dalla Mura y col., 2018, para la entropía de Shannon se usaron n =5, 30 y 500 bins en el histograma, para la entropía de Rényi también se usaron n =5, 30 y 500 bins en histograma y $\alpha = 2$ e ínf. **Fuente:** Tabla modificada de Malfante, 2018

4.3. Entrenamiento y evaluación

La AAA está diseñada para implementar los algoritmos de RF y SVM, usando sus respectivas funciones en el paqueta de Scikit-learn. Se decidió implementar ambos algoritmos para seleccionar el modelo con mejor desempeño. Considerando los catálogos 1 y 2, las representaciones AAA y PCA, y ambos algoritmos de clasificación, se consideraron 8 modelos distintos. Cada uno de ellos se entrenó y se evaluó con los mismos conjuntos de entrenamiento y prueba.

En el entrenamiento se incluyó una etapa de validación cruzada para seleccionar los hiperparámetros óptimos de los modelos. Para esto se hizo una búsqueda de mallado en la que se escoge la mejor combinación de hiperparámetros de una serie de valores que se establecen con anterioridad. La tabla B.2 del apéndice B presenta los hiperparámetros y los valores que se consideraron en la búsqueda. La validación cruzada se llevó a cabo con 10 iteraciones, en cada una el conjunto de entrenamiento se divide aleatoriamente en dos partes; el 80 % de los datos se usan para entrenar el modelo y el 20 % restante para hacer una validación. La elección de los hiperparámetros óptimos se hace con respecto a la exactitud media de las 10 iteraciones. En la figura 4.5 se muestran los hiperparámetros óptimos que se encontraron para los 8 modelos considerados. El significado de los hiperparámetros de cada algoritmo se discute en el apéndice B.2.

	Hiperparámetros óptimos	Exactitud media	Hiperparámetros óptimos	Exactitud media
	RF		*PCA + RF*	
Catálogo 1	N_estimators: 100 Criterion: gini Max_depth: 30 Max_features: log2 Bootstrap: False	80 ± 10%	N_estimators: 200 Criterion: entropy Max_depth: 30 Max_features: log2 Bootstrap: False	75 ± 5%
Catálogo 2	N_estimators: 200 Criterion: entropy Max_depth: 30 Max_features: log2 Bootstrap: False	88 ± 2%	N_estimators: 500 Criterion: entropy Max_depth: 30 Max_features: log2 Bootstrap: False	87 ± 5%
	SVM		*PCA + SVM*	
Catálogo 1	C: 10 Gamma: 0.01 Kernel: rbf	72 ± 11%	C: 100 Gamma: 0.001 Kernel: rbf	74 ± 5%
Catálogo 2	C: 1000 Gamma: 0.001 Kernel: rbf	85 ± 3%	C: 1000 Gamma: 0.001 Kernel: rbf	87 ± 4%

Figura 4.5
Conjuntos de hiperparámetros óptimos encontrados por medio de la validación cruzada. También se muestra la exactitud media asociada.

Después de la selección de los hiperparámetros óptimos se entrenó el modelo usando todas las señales del conjunto de entrenamiento. Posteriormente, se usaron las métricas de exactitud, precisión, sensibilidad y valor-F1 para evaluar el desempeño del modelos en la clasificación de las señales del conjunto de prueba. Al finalizar se eligió el mejor modelo.

4.4. Aplicación

En el capítulo 5 se muestran los resultados de la evaluación de los 8 modelos que se consideraron. Con base en estos se eligió al mejor modelo, que es el que se empleó en el análisis de señales continuas. El periodo de análisis va de noviembre del 2019 a julio del 2020. Se usó el registro de la componente norte de la estación disponible más cercana al cráter: PPJU o PPCU (en este periodo PPPP experimentó problemas de alimentación eléctrica y solo funcionaba algunas horas al día, por lo que se descartó).

Se decidió trabajar con las señales de la componente norte porque es común que la amplitud de la forma de onda sea mayor en las componentes horizontales. Esto se ejemplifica en los eventos que se muestran en las figuras 1.1 a 1.4. La elección de la componente norte, sobre la componente este, se hizo de forma arbitraria; pues no se encontró que la amplitud de una de las componentes dominara sobre la otra de forma consistente.

El análisis de señales continuas está incluido en la AAA. Se usa una ventana que se desliza sobre la señal de forma iterativa. En cada iteración se calcula el vector de atributos asociado a la ventana y, usando el modelo, se asocia una probabilidad de pertenencia a cada clase. La clase que se asigna corresponde al tipo de evento con mayor probabilidad, siempre y cuando se cumpla con un valor mínimo que se impone como umbral –en este caso se eligió el 60 %–, de otra forma la ventana se clasifica como *evento desconocido*. En la siguiente iteración, la ventana avanza sobre la señal y el ciclo continúa hasta que se haya clasificado todo el registro.

La probabilidad de pertenencia a cada clase se calcula mediante el método *decision_function* de los algoritmos de clasificación en Scikit-learn. La probabilidad se obtiene con el método propuesto por Wu y col., 2003, en el que se considera una regresión logística de las clases asignadas en distintas validaciones cruzadas. Los detalles del método quedan fuera del alcance de este texto.

Para elegir el valor umbral de probabilidad se hicieron pruebas con 40 % y 60 %. El número de errores de clasificación en el primer caso es mucho más abundante que en el segundo. Por otro lado, el número de eventos desconocidos es mucho mayor para el 60 %. Se decidió trabajar con un umbral del 60 % porque queda la posibilidad de que en el futuro se reentrene al modelo con más eventos y se analice de nuevo a las ventanas que no pudieron clasificarse.

Considerando que la longitud de los eventos del catálogo varía de forma considerable, se decidió trabajar con dos tamaños de ventana. En cada una de las iteraciones se empieza con una ventana de 40 segundos de longitud; sin embargo, si la máxima probabilidad calculada es menor a 60 % entonces la ventana crece al doble de tamaño (80 s) y se hace una nueva predicción. La ventana de menor tamaño se diseñó para detectar eventos de corta duración como sismos volcanotectónicos y eventos LP de poca energía. Por el contrario, la ventana de 80 segundos se consideró para la detección de explosiones, tremores y eventos LP de gran amplitud. Por lo general, los eventos regionales duran más de 80 s, lo cual impacta en la capacidad de detección de esta clase. La posibilidad de tener ventanas con longitud variable también es una propiedad que se agregó a la AAA.

El resultado del análisis continuo se puede visualizar de dos formas. La primera es una ligera modificación de la visualización que se incluye con la AAA, en la que se crean imágenes que muestran la forma de onda, el espectrograma, la clase asignada y la ventana de análisis de las tres componentes (véase figura 4.6). Esta forma de visualización tiene la ventaja de permitir un análisis rápido de la clasificación. La visualización propuesta por la AAA es muy similar pero solo tiene al espectrograma de la componente que se está analizando. La segunda forma de visualización es mediante el software SWARM (desarrollado por el USGS), en donde se crean archivos de etiquetas por clase que se pueden visualizar. La figura 4.7 muestra un ejemplo, los puntos amarillos sobre el helicorder muestran las ventanas clasificadas como LP por el modelo. La ventaja de esta visualización es que se puede interactuar con ella y permite ver las señales de otras estaciones. Esta es otra aportación novedosa a la AAA.

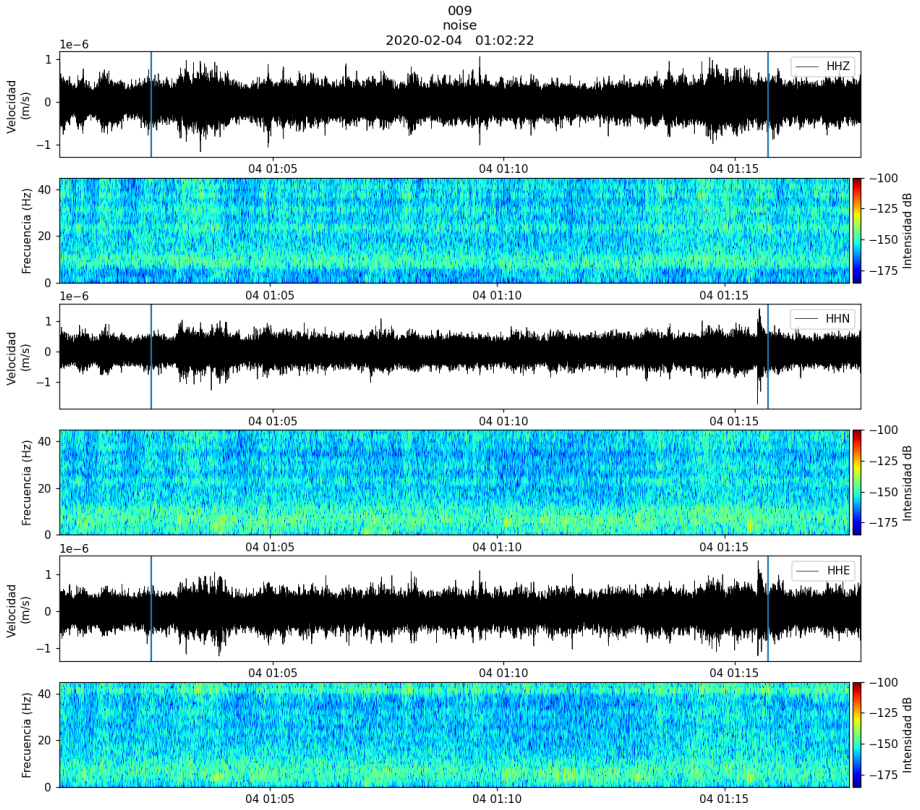

Figura 4.6
Ejemplo de visualización de la clasificación continua con imágenes de la forma de onda y el espectrograma de las tres componentes. Las líneas azules indican la ventana de análisis.

Por último, para tener un análisis cuantitativo de la clasificación continua se solicitó a CENAPRED el catálogo interno de eventos del mes de febrero del 2020 para comparar la clasificación manual con la del modelo. Se eligió este mes porque según los reportes diarios del CENAPRED es el mes con más explosiones en el periodo estudiado. En específico se compararon

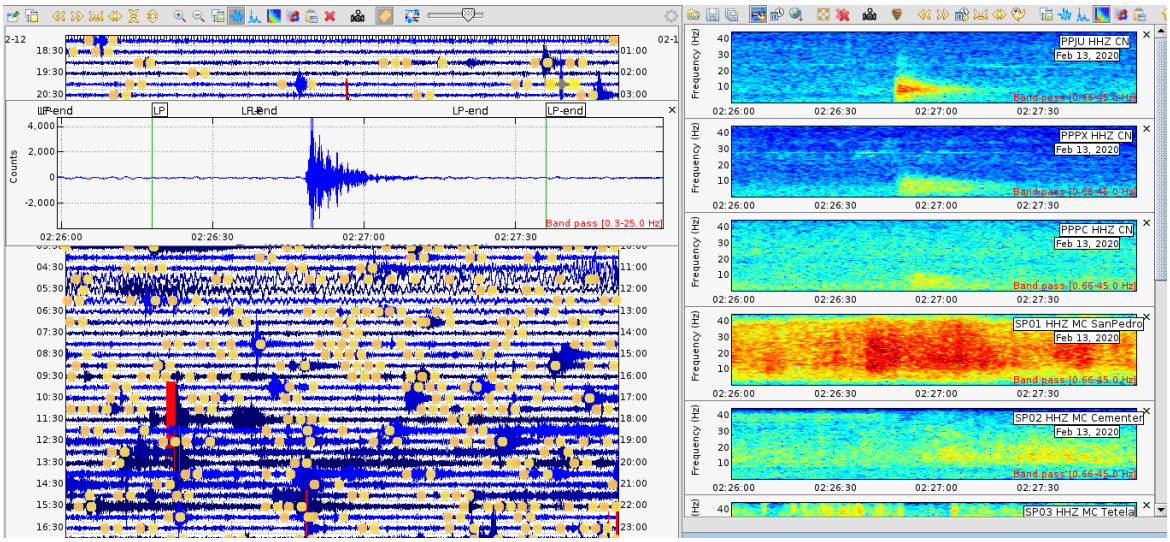

Figura 4.7
Ejemplo de visualización de la clasificación continua en el software SWARM. Los puntos amarillos muestran las ventanas clasificadas como LP por el modelo.

2141 eventos, que corresponde a los eventos identificados manualmente en 10 días de este mes. El catálogo de CENAPRED solo incluye eventos LP, explosiones, tremores y VTs por lo que son las únicas clases que se pueden comparar.

4.5. Resumen

La figura 4.8 resume las etapas de la metodología que se aplicaron en cada paso.

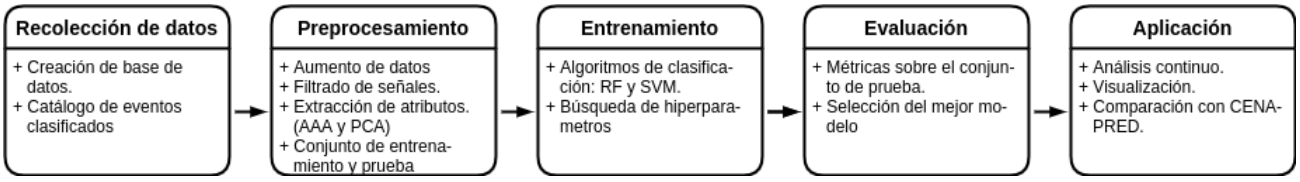

Figura 4.8
Resumen de las etapas que se aplicaron en cada paso de la metodología. Las aportaciones que se hicieron a la AAA se destacan en el texto.

Las aportaciones originales a la metodología son:

- Estrategia de aumento de datos.
- Implementación de la técnica de PCA durante la etapa de extracción de atributos.
- Búsqueda de hiperparámetros óptimos mediante la validación cruzada.

- Visualización de los resultados del análisis continuo usando el software SWARM e imágenes que incluyen las tres componentes con su forma de onda y espectrogramas.

Capítulo 5

Resultados y discusión

A continuación se presentan los resultados obtenidos y se hace una discusión sobre los mismos. La exposición se divide en tres secciones principales: en la primera se presenta un análisis de correlación de los atributos de la representación AAA para justificar el uso de una representación alterna, en la segunda se compara el desempeño de los modelos que se obtuvieron y se resaltan las propiedades del mejor modelo y, en la última sección, se evalúa el desempeño del mejor modelo en el análisis de señales continuas de noviembre 2019 a julio 2020.

5.1. Correlación de los atributos

Se sabe que el espacio de atributos propuesto por la AAA funciona en la clasificación de señales volcano-sísmicas porque se probó en el volcán Ubinas, Perú (Lara y col., 2020; Malfante, Dalla Mura y col., 2018) y en La Soufrière, Guadalupe (Falcin y col., 2021). Sin embargo, cuando se trabaja con representaciones genéricas –no diseñadas para el conjunto de datos con el que se cuenta– es común que haya atributos que aporten poca o incluso nada de información al modelo. También se puede presentar el caso en el que más de un atributo aporta información similar al modelo, cayendo en redundancias que pueden afectar las propiedades de generalización del mismo e incrementar los tiempos de cómputo. Tomando en cuenta estas consideraciones, se calculó la correlación entre pares de atributos de las señales del conjunto de entrenamiento. Se usó el coeficiente de correlación de Spearman –pues los atributos no siguen una distribución normal–, el cual evalúa qué tan bien se puede describir la relación entre dos atributos (X y Y) usando una función monótona. Formalmente, se define como:

$$r_s = \frac{\mathrm{cov}(R(X), R(Y))}{\sigma_{R(X)}\sigma_{R(Y)}},$$

donde $R(X)$ y $R(Y)$ son los rangos o *rankings*[†] de los atributos, cov es la covarianza y σ es la desviación estándar. El coeficiente de correlación de Spearman puede tomar valores entre -1 y 1, su comportamiento se ilustra en la figura 5.1. El signo describe la dirección de la relación, es decir, los valores negativos indican que cuando uno de los atributos aumenta, el otro disminuye. Cuando el coeficiente es 0, implica que los atributos no pueden describirse por una función monótona.

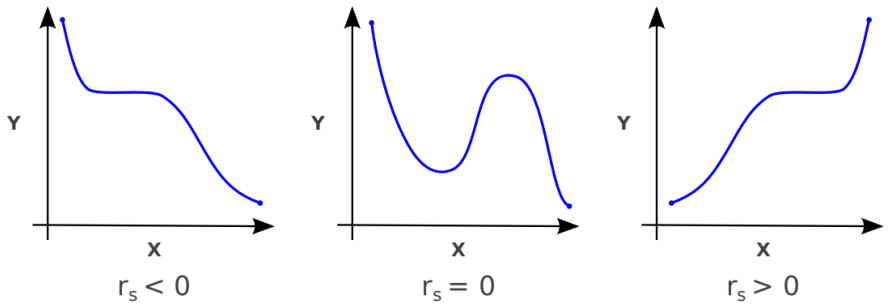

Figura 5.1
Comportamiento del coeficiente de correlación de Spearman (r_s). La correlación se calcula entre dos atributos cualesquiera X y Y. **Fuente**: Figura modificada de O. Alexandrov, 2007. Wikimedia Commons.

En la representación AAA, las señales se describen con un vector de 102 atributos, de modo que la correlación entre todos los pares de atributos se puede expresar con una matriz de 102×102. Para facilitar su consulta, la tabla de atributos que se presenta en la metodología (tabla 4.1) se replica en esta sección (tabla 5.1). En la figura 5.2 se muestra la matriz de correlación calculada sobre las señales del conjunto de entrenamiento. Las etiquetas usan el número de referencia de la tabla 5.1 y la primera letra indica el dominio: t para el temporal, s para el espectral y c para el cepstral; los colores azules indican una correlación positiva y los rojos una negativa. Se puede observar claramente que existe un grado de correlación significativo entre algunos atributos. Por lo que la representación AAA contiene información redundante que, en principio, puede expresarse en menos dimensiones. Este resultado fue el que motivó el uso de la técnica de PCA.

Hay algunos aspectos que vale la pena discutir sobre la matriz de correlación. Por ejemplo, las zonas de mayor correlación están asociadas a las entropías de Shannon y de Rényi en los tres dominios (atributos t10 - t18, s10 - s18 y c10 - c18) y se pueden identificar como bloques azules que pasan sobre la diagonal. Como se puede ver en la tabla 5.1, en el cálculo de entropía se asume que el histograma de las amplitudes de la señal se asocia a la función de densidad de probabilidad (fdp). De modo que las zonas de alta correlación entre atributos de entropía, para un mismo dominio, indican que el comportamiento general de la fdp se mantiene, aún cuando se varía la construcción del histograma (número de bins que se considera). Un aspecto

[†]En la estadística, el rango o *ranking* se refiere a una transformación de los datos, en la que el valor de los atributos se reemplaza por el lugar que ocupan en una lista ordenada de forma ascendente. Por ejemplo, supongamos que el conjunto de entrenamiento está compuesto de tres señales. El valor del atributo X de cada señal es: 4.5, 7.8 y 2.3, respectivamente. Al ordenar estos valores de forma ascendente, los rangos del atributos X son: 2, 3 y 1, respectivamente. (Spearman, 1904)

Atributo	Fórmula	Ref
Longitud	n	1
Media	$\mu = \frac{\sum_i z_i}{n}$	2
Desviación estándar	$\sigma = \sqrt{\frac{1}{n-1}\sum_i (z_i - \mu)^2}$	3
Asimetría	$\frac{1}{n} \cdot \sum_i \left(\frac{z_i - \mu}{\sigma}\right)^3$	4
Curtosis	$\frac{1}{n} \cdot \sum_i \left(\frac{z_i - \mu}{\sigma}\right)^4$	5
Centroide	$\bar{i} = \frac{1}{E}\sum_i i \cdot E_i$	6
Ancho de banda RMS	$RMS_i = \sqrt{\sum_i \frac{i^2 \cdot E_i}{E} - \bar{i}^2}$	7
Asimetría media	$\left(\frac{\sum_i (i-\bar{i})^3 E_i}{E \cdot RMS_i^3}\right)^{1/2}$	8
Curtosis media	$\left(\frac{\sum_i (i-\bar{i})^4 E_i}{E \cdot RMS_i^4}\right)^{1/2}$	9
Entropía de Shannon	$-\sum_i p(z_i)\log_2(p(z_i))$	10 - 12
Entropía de Rényi	$\frac{1}{1-\alpha}\log_2\left(\sum_i p(z_i)^\alpha\right)$	13 - 18
Tasa de ataque	$\max_i \left(\frac{z_i - z_{i-1}}{n}\right)$	19
Tasa de descenso	$\min_i \left(\frac{z_i - z_{i+1}}{n}\right)$	20
Mínimo sobre media	$\min z_i/\mu$	21
Máximo sobre media	$\max z_i/\mu$	22
Energía	$E = \sum_i z_i^2$	23
Máximo de energía	$\max(z_i^2)$	24
Energía promedio	$\mu_E = \frac{\sum_i z_i^2}{n}$	25
Desviación estándar de energía	$\sigma_E = \sqrt{\frac{1}{n-1}\sum_i (E_i - \mu_E)^2}$	26
Asimetría de energía	$\frac{1}{n} \cdot \sum_i \left(\frac{E_i - \mu_E}{\sigma_E}\right)^3$	27
Curtosis de energía	$\frac{1}{n} \cdot \sum_i \left(\frac{E_i - \mu_E}{\sigma_E}\right)^4$	28
Mínimo de la señal	$\min_i(z_i)$	29
Máximo de la señal	$\max_i(z_i)$	30
i de mínimo	$\text{argmin}_i(z_i)$	31
i de máximo	$\text{argmax}_i(z_i)$	32
Tasa de cruce por umbral	$\frac{\#\text{cruces por umbral}}{n}$	33
Tasa de silencio	$\frac{\#z_i \text{ donde } z < \text{umbral}}{n}$	34

Tabla 5.1: Atributos que se utilizan para caracterizar a las señales. Estos se calculan sobre la señal $z[i]_{i=1}^n$, donde i se refiere a la muestra temporal, espectral o cepstral. Todos los atributo se calculan en los tres dominios. $E_i = z[i]^2$, es la energía en la muestra i. Los números de referencia de la tercera columna se usan en la discusión de resultados para distinguir entre atributos. La entropía mide el contenido promedio de información de la señal. La probabilidad, $p(z_i)$, se calcula a partir del histograma de los valores de amplitud de la señal en cualquiera de los dominios. Siguiendo la metodología de Malfante, Dalla Mura y col., 2018, para la entropía de Shannon se usaron n =5, 30 y 500 bins en el histograma, para la entropía de Rényi también se usaron n =5, 30 y 500 bins en histograma y $\alpha = 2$ e ínf. **Fuente:** Tabla modificada de Malfante, 2018

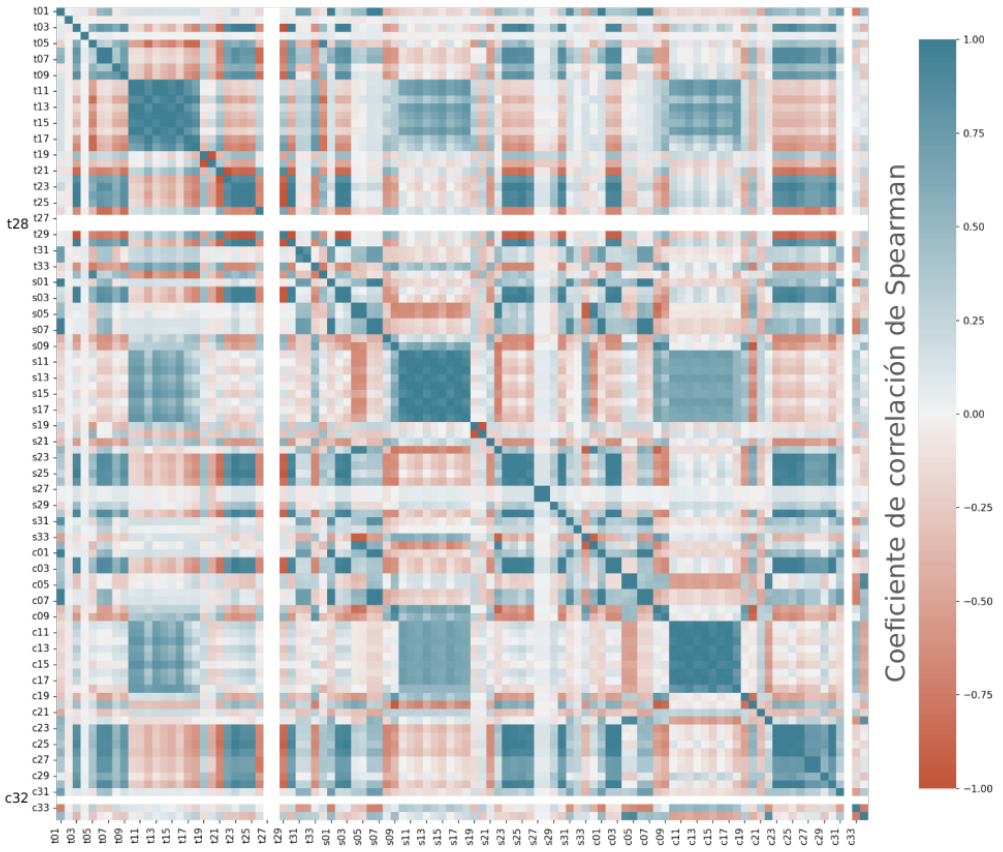

Figura 5.2
Matriz de correlación de los atributos de las señales del conjunto de entrenamiento. Se usa el coeficiente de correlación de Spearman. Únicamente se muestras las etiquetas impares y se usan los números de referencia que se presentan en la tabla 4.1 para distinguir entre atributos. La letra de prefijo indica el dominio de la señal: t para temporal, s para espectral y c para el dominio cepstral. Los 34 atributos de la tabla 5.1 se calculan en los tres dominios, por lo que las señales se representan con 102 atributos.

más interesante, es la correlación que existe entre los atributos de entropía para los distintos dominios (bloques de azul menos intenso que están fuera de la diagonal), lo que sugiere que la entropía de una señal es un atributo que exhibe cierto grado de invarianza ante la transformada de Fourier. Profundizar en su eficiencia para la clasificación de señales volcano-sísmicas y su posible relación con los procesos físicos que las generan puede ser de interés en estudios futuros.

Por otro lado, la matriz de correlación también indica que hay tres atributos, la asimetría de energía en el tiempo (t27), la curtosis de energía en el tiempo (t28) y la posición del máximo en el cepstral (c32), que no están aportando información al modelo. El valor de estos atributos es cero para todas las señales y por eso aparecen como líneas completamente blancas en la matriz. En el caso de t27 y t28, los valores son tan pequeños que se consideran cero. En el caso de c32, la aparición de valores máximos del cepstrum en el cero parece ser una propiedad de ciertas señales y la discusión al respecto queda fuera del alcance de este texto (Fraile & Godino-Llorente, 2014).

Estas consideraciones ilustran la importancia de un análisis exploratorio de datos cuando se usa una representación que no fue diseñada específicamente para las señales con las que se trabaja. Es pertinente mencionar que en los estudios de Malfante, Dalla Mura y col., 2018 y Falcin y col., 2021 no se presenta un análisis de esta naturaleza, sin embargo en ambos estudios se hace una selección de los mejores atributos usando métodos específicos para el algoritmo de RF. En este caso, se decidió implementar el PCA para evitar sesgos cuando el algoritmo clasificador es el SVM.

5.2. Modelo clasificador

Una vez establecida la representación de las señales, se procedió al cálculo del modelo clasificador. En los resultados que se presentan a continuación se comparan los modelos que se obtuvieron usando los dos catálogos propuestos: el de los eventos registrados en una sola estación y una componente (catálogo 1) y el catálogo extendido que incluye los eventos registrados en las tres componentes y en más de una estación (catálogo 2). Además, para cada uno de ellos se comparan los resultados que se obtienen de las dos representaciones –AAA y PCA– y los dos algoritmos de clasificación que se usaron –RF y SVM–. En los modelos que usan la representación PCA se consideraron 90 componentes principales (CP) que explican alrededor del 99 % de la varianza de los datos. Se hicieron pruebas considerando 20 y 40 componentes –equivalentes aproximadamente al 90 % y 95 % de la varianza– pero los mejores resultados se obtuvieron con 90 CP. Originalmente, las señales se representan con 102 atributos, por lo que se logró una disminución de 12 dimensiones. Aún más, la implementación del PCA permite disminuir alrededor del 13 % del tiempo de cómputo de ambos algoritmos –RF y SVM– en las pruebas aisladas.

En la tabla 5.2 se presentan los resultados de la evaluación de los 8 modelos. Se muestran los valores de exactitud calculados en el conjunto de prueba y se hace un desglose del valor F1 que se obtuvo para cada clase. Si nos enfocamos en la exactitud, todos los modelos del catálogo 2 tienen un mejor desempeño que los modelos entrenados con los eventos del catálogo 1. En particular, para los modelos de SVM, se obtiene una mejora del 6 % cuando se usa el catálogo extendido. Aún más, el uso de la técnica de PCA también tiene un resultado favorable. En el caso de RF, la combinación del catálogo 2 y el PCA producen un aumento del 7 % en la exactitud obtenida para el conjunto de prueba. En el caso de SVM, el aumento es del 8 %.

El valor F1 permite evaluar el desempeño de los modelos en cada clase. El mejor desempeño se marca con negritas en la tabla 5.2. El comportamiento más interesante se presenta en la clase LP. Pues, en todos los modelos, se tienen desempeños más pobres cuando se usa el catálogo 2. Este resultado es consecuencia del uso de la estrategia de aumento de datos. Recordemos que el catálogo 2 incluye las señales de un mismo evento registradas en distintas estaciones. Al agregar estas señales es necesario tomar en cuenta los efectos de atenuación. Pues, su atenuación puede influir en la clasificación del modelo. Por ejemplo, en el panel a) de la figura 5.3, se muestra un VT de magnitud 2.3 registrado en la estación PPCL –que es la estación más cercana al evento y se encuentra a 12 km del cráter, en dirección sur–. Se trata de un registro

	RF		**PCA + RF**		**SVM**		**PCA + SVM**	
	Cat 1	*Cat 2*	*Cat 1*	*Cat 2*	*Cat 1*	*Cat 2*	*Cat 1*	*Cat 2*
Exactitud (%)	62	64	61	69	64	70	66	**72**
	Valor F1 (%) (conjunto de prueba)							
EX	**92**	71	75	82	73	78	82	81
LP	**72**	60	61	57	56	53	58	52
RE	67	55	25	61	50	68	60	**76**
TR	50	67	47	70	60	73	55	**74**
VT	69	69	64	64	77	70	70	**78**
RU	21	57	67	75	74	80	**81**	76

Tabla 5.2: Evaluación de los modelos clasificadores. Se muestra la exactitud para el conjunto de prueba y un desglose del valor F1 obtenido en el conjunto de prueba para cada clase. EX: explosiones, LP: periodo largo, RE: regionales, TR: tremor, VT: volcanotectónicos, NO: ruido.

energético con contenido espectral por arriba de los 20 Hz. En cambio, en el panel b) de la misma figura, se muestra el mismo evento registrado en la estación PPCU –ubicada a 2.6 km del cráter, también en dirección sur–, donde el contenido espectral de la señal se ha atenuado considerablemente y, en efecto, posee características similares a las señales de la clase LP. Si bien este ejemplo es extremo –la distancia entre ambas estaciones es de más de 9 km–, resulta útil para destacar una limitación importante de la estrategia de aumento de datos. Para evitar estos inconvenientes, se recomienda tener un control más estricto al agregar la información de otras estaciones.

Sin embargo, en la tabla 5.2 se puede apreciar que para el resto de las clases, el desempeño del modelo se beneficia del catálogo extendido y del uso del PCA. Por ejemplo, para la clase de tremor, el aumento del desempeño es del 20 % cuando se usa RF y del 14 % cuando se usa SVM.

Tomando en cuenta la exactitud de todos los modelos y el desempeño general por clase, se considera que el mejor modelo obtenido es el que usa el catálogo 2 y los algoritmos de PCA + SVM. Este tiene una exactitud del 72 % y, a excepción de la clase LP, el desempeño promedio de las clases está por arriba del 70 %. De este resultado, se concluye que las aportaciones que se hicieron a la metodología están justificadas y se recomiendan para trabajos futuros.

Para ilustrar, de forma más completa, el comportamiento del mejor modelo obtenido, se presenta la matriz de confusión obtenida con el conjunto de prueba (tabla 5.3). La clase con mejor desempeño es la de las explosiones que tiene sensibilidad y precisión alrededor del 80 %. Por otra parte, la mejor métrica que se obtuvo fue en la precisión de los sismos regionales, con un valor del 94 %. Sin embargo la exactitud de esta clase fue del 64 % por lo que el modelo tiene problema en identificar algunas de las instancias de esta clase. Esto se debe al desbalance que hay en el número de instancias de la clase RE, tiene alrededor de 1/3 menos que el resto de las clases (véase figura 4.4). En cuanto a tremores, sismos VT y ruido, el comportamiento del

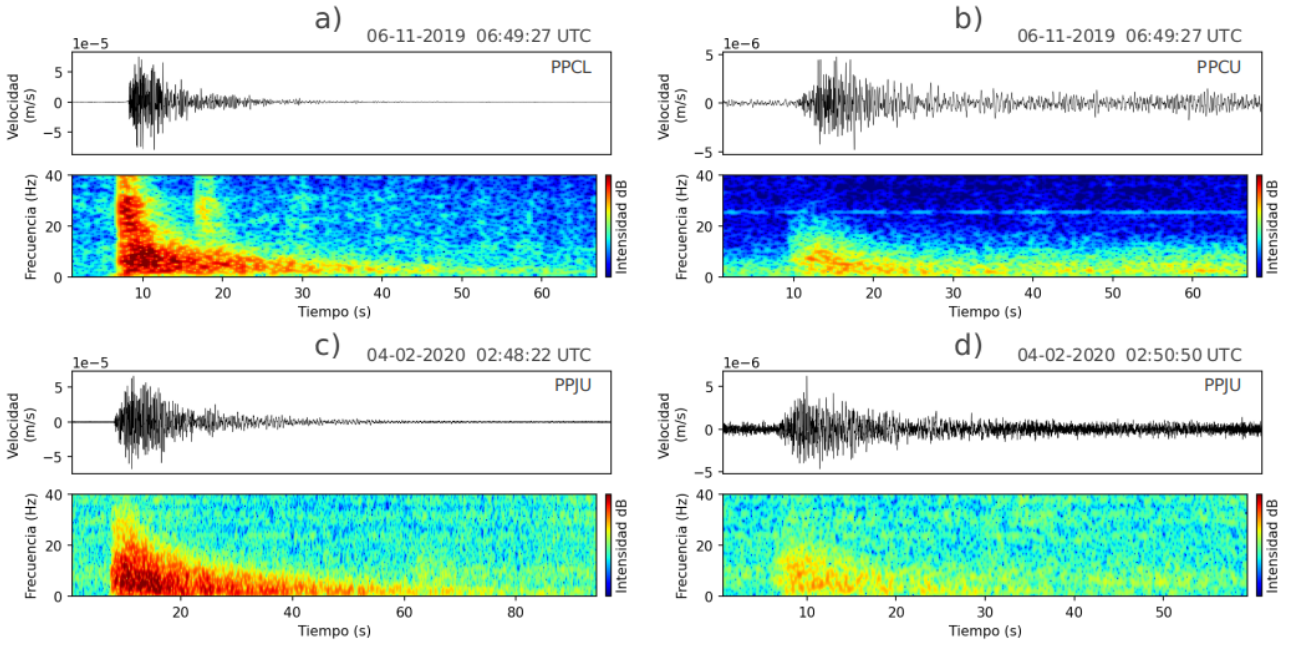

Figura 5.3
Señales asociadas a sismos VT. Los paneles a) y b) son señales del mismo evento registrado el 06/11/2019 de magnitud 2.3 de acuerdo al reporte del CENAPRED. En el panel a) se muestra la señal registrada por la estación PPCL. En el panel b) se tiene el registro en la estación PPCU. En el panel d) se muestra un evento registrado el 04/02/2020 en la estación PPJU de magnitud 2.7, según el reporte de CENAPRED. En el panel e) se tiene un evento registrado el 04/02/2020 en la estación PPJU, magnitud reportada de 1.7.

modelo es bastante estable con valores de sensibilidad ligeramente menores que los de precisión pero todos por arriba del 70 %.

	Predicciones							
	EX	LP	RE	TR	VT	NO	Sensibilidad	Valor F1
Explosión	41	5		1	3	1	**80 %**	**81 %**
LP	6	31		2	3	6	65 %	52 %
Regional		8	29	6	2		64 %	76 %
Tremor	3	14	1	46		2	70 %	74 %
VT		10			41		**80 %**	78 %
Ruido		3	1	4	5	35	73 %	76 %
Precisión	**82 %**	44 %	**94 %**	78 %	76 %	**80 %**		
				Exactitud			72 %	

Tabla 5.3: Matriz de confusión para el modelo PCA + SVM, usando el catálogo 2. Se consideran 90 componentes principales. Este fue el mejor modelo clasificador que se obtuvo. Los hiperparámetros utilizados son $C_{RBF} = 1000$ y $\gamma = 0.001$.

En particular, se observa que la métrica que presenta mayores dificultades es la precisión de la clase LP, en la que se obtiene un valor del 44 %. Esto indica que un número importante

de los eventos que el modelo clasifica como LP, no pertenecen a esta clase. Lo cual es consistente con la discusión previa, referente al uso del catálogo 2. Los errores más frecuentes son con las clases TR y VT. La confusión con los tremores se justifica porque los procesos físicos que generan ambas clases están asociados al movimiento de fluidos, de modo que sus señales comparten algunas características, tanto en la forma de onda como en el contenido de frecuencias. En lo referente a la confusión con la clase VT, la baja precisión puede deberse a dos causas y para ilustrarlas se usan los eventos de la figura 5.3. El caso de las señales a) y b) se debe a efectos de atenuación y ya se discutió. Por otro lado, en los paneles c) y d) se muestran sismos VT que ocurrieron el mismo día, en ambos casos, la primera estación en registrarlos es PPJU. De acuerdo al reporte de CENAPRED, el primero tiene una magnitud de 2.7 (panel c) y el segundo de 1.7 (panel d). Se puede apreciar que ambos eventos tienen grandes diferencias tanto en la forma de onda como en el contenido espectral. Si se compara al de menor magnitud con algunas señales de clase LP de la figura 5.10 se pueden observar varias semejanzas, incluso considerando que el contenido espectral del VT tiene algo de energía cerca de los 20 Hz. Por lo tanto, el tamaño del evento VT puede influir en su clasificación. A pesar de los errores de clasificación responsables de la precisión de la clase LP, su sensibilidad es del 64 %, lo que nos indica que el modelo es capaz de detectar la mayoría de los señales. Finalmente, la sensibilidad de los tremores y los sismos VT tienen un buen desempeño, 70 % y 80 % respectivamente, por lo que los errores de clasificación de la clase LP no afectan al modelo de forma considerable.

Hasta ahora, se ha discutido que algunas señales de eventos VT poseen propiedades similares a los eventos LP debido a procesos asociados a la atenuación geométrica y la dispersión (scattering). Sin embargo, un estudio reciente de Clarke y col., 2021 demuestra, por medio de simulaciones numéricas, que la presencia de zonas de alta atenuación intrínseca pueden transformar las señales de VTs en señales que parecen eventos LP. Este resultado es interesante en el contexto del Popocatépetl porque Novelo-Casanova y Martínez-Bringas, 2005 y Shapiro y col., 2000 han detectado zonas anómalas de atenuación debajo del volcán. Por otro lado, la actividad del Popocatépetl se caracteriza por tener bajos niveles de sismicidad VT en comparación con los altos niveles de producción de eventos LP (Arciniega-Ceballos y col., 2003; Chouet y col., 2005). La correlación de ambos fenómenos a la luz de los resultados de Clarke y col., 2021 puede indicar la presencia de procesos anelásticos que no se han considerado hasta la fecha.

5.3. Análisis continuo

En esta sección se muestran los resultados de aplicar el modelo PCA + SVM al análisis de señales continuas. El periodo de estudio va del 1 de noviembre del 2019 al 31 de julio del 2020. Se usaron las señales registradas en la estación disponible más cercana al cráter; para los meses de diciembre, noviembre y enero se utilizó la estación PPCU y para los 6 meses restantes se utilizó la estación PPJU. La exposición de los resultados tiene el siguiente orden: en primera instancia, se muestran y discuten algunos ejemplos del análisis continuo; posteriormente, se muestra el número de eventos detectados por día para cada clase y se hace una comparación con el número de eventos reportados por el CENAPRED en los boletines diarios; finalmente, se muestra una comparación cuantitativa de 2141 eventos de febrero 2020, entre la clasificación

manual realizada por personal del CENAPRED y las predicciones del modelo.

Ejemplos de clasificación continua

Para discutir aspectos generales de la clasificación continua, se usan los ejemplos de la figura 5.4. Se usan líneas verticales para delimitar las ventanas de análisis. La primera característica que hay que destacar es que el empalme entre las ventanas no es constante. Esto es consecuencia de la propiedad de la ventana, de crecer al doble de tamaño, cuando no se identifica un clase con su longitud original.

En la señal del panel a) el modelo predice la siguiente secuencia de clases: ruido, desconocido y tremor. En el espectrograma se puede observar que dentro de la ventana de ruido, entre las 12:59:00 y las 13:01:00, hay un tremor de baja amplitud que el modelo no reconoce; este tipo de comportamiento es muy común. En seguida se presenta una secuencia típica en la que el inicio del tremor no se logra clasificar, pero una vez que la ventana avanza, el evento se identifica y se clasifica de forma correcta. Es probable que este resultado sea consecuencia del entrenamiento del modelo, en donde se utiliza la señal del evento completo. Una recomendación importante para trabajos a futuro, es que el entrenamiento se haga con segmentos secuenciales del evento con la misma longitud que se usará en la ventana del análisis continuo. De esta forma, el modelo aprenderá a clasificar con la información que tendrá al momento de su aplicación.

En la señal del panel b) se tiene la secuencia: ruido, tremor, ruido y LP. De nuevo, la parte central del tremor sí se identifica, pero sus extremos no. También se pueden ver pequeños eventos dentro de la segunda ventana de ruido que el modelo no identifica. La ventana LP es un claro ejemplo de evento corto que se detecta en su totalidad y se clasifica correctamente.

En el panel c) la secuencia identificada por el modelo es: ruido, desconocido, VT, ruido, regional y VT. En la ventana de desconocido no se logra clasificar el evento de clase LP, que a simple vista se identifica con facilidad. Este tipo de errores se pueden corregir ampliando el número de eventos en el catálogo. Por otro lado, la predicción de evento regional es un claro ejemplo de error en la clasificación, pues en realidad se trata de un evento VT. Sin embargo, los VTs y los sismos regionales pueden compartir características espectrales, por lo que la clasificación del modelo no es completamente descabellada. Los errores en la clase RE son muy frecuentes, se deben al tamaño de la ventana de análisis y, como ya se mencionó, al desequilibrio de las clases en el catálogo. Finalmente, un aspecto positivo a resaltar es que ambos eventos VT tienen propiedades distintas y el modelo los puede identificar.

En el panel d) la secuencia es: tremor, desconocido, LP, ruido, LP y desconocido. De nuevo se muestran problemas con la identificación del inicio de los tremores. La segunda ventana de clase LP es un ejemplo de otro escenario común, en el que la ventana clasifica de forma correcta a más de un evento. En este panel, las ventanas de eventos desconocidos se justifican porque incluyen secciones de distintos eventos.

Por último, la secuencia del panel e) es: tremor, LP, explosión, LP, VT, desconocido

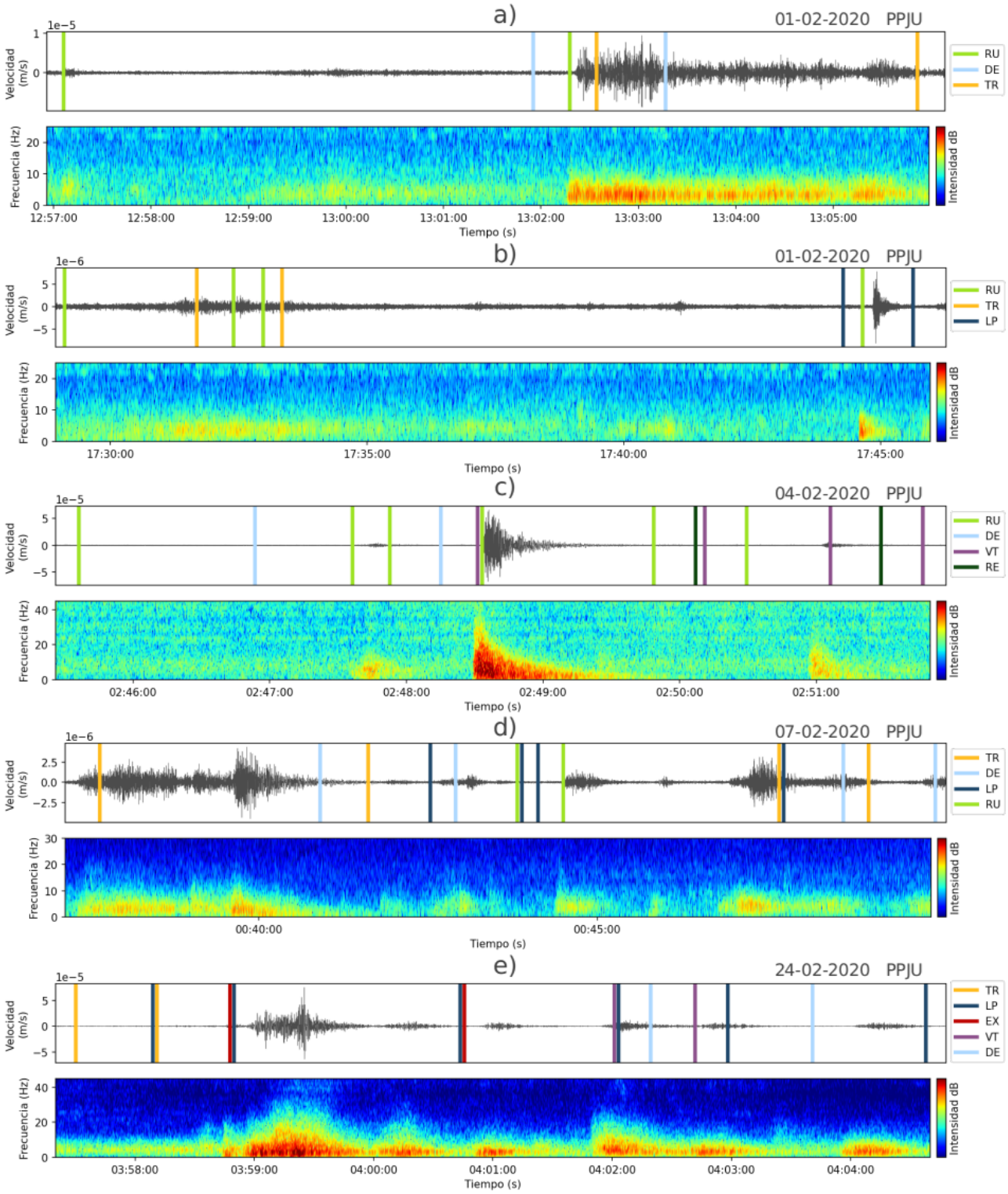

Figura 5.4
Ejemplos del desempeño del modelo PCA + SVM en el análisis continuo. Las predicciones del modelo se muestran con las lineas verticales de colores que identifican el inicio y el final de la ventana temporal seleccionada. NO es ruido, DE es desconocido, TR es tremor, LP es periodo largo, VT es volcano-tectónico, RE es regional y EX es explosión. Para todos los casos se muestra la señal registrada en la componente norte de la estación PPJU.

y LP. La explosión, el tremor y los LPs se identifican con éxito. La ventana VT es un error de clasificación, a pesar de que el contenido de frecuencias es superior a los 10 Hz, no se puede identificar una fase S en el registro.

Número de eventos diarios

Ahora se presenta la comparación de eventos por día entre los resultados del modelo y los reportes diarios del CENAPRED. Los boletines del CENAPRED solo reportan explosiones, VTs, tremores y exhalaciones. Múltiples estudios han asociado a los eventos LP con las exhalaciones (p.ej. Arciniega-Ceballos y col., 2008), por lo que el número de LPs se compara con el número de exhalaciones reportado. Una sutileza que hay que comentar es que los reportes de CENAPRED se hacen con hora local y las señales usan hora UTC; además los reportes abarcan periodos de 24 h que empiezan a las 10 de mañana, por lo que incluyen eventos que ocurren en días distintos. La comparación se muestra en la figura 5.5 y la etiqueta de los eventos reportados por CENAPRED incluyen en sufijo *_CN*. Los tremores no se cuentan por evento, sino que se comparan los minutos de tremor al día.

Figura 5.5
Comparación entre el número de eventos por día detectados por el modelo y reportados por CENAPRED.

En cuanto a la comparación del número de eventos LP y el número de minutos de tremor, las predicciones del modelo siguen las tendencias de los reportes, en particular con la clase de tremores. Lo cual sugiere que el modelo es eficiente en la detección de estas dos clases. Por otro lado, en todas las clases, el modelo sobrestima el número de eventos con factores que van de 1 a 2 órdenes de magnitud. En donde destacan algunos días atípicos, en los que el

modelo detecta más de 200 explosiones o sismos VT. Particularmente en estas clases, la mayoría de las clasificaciones son espurio y necesitan una revisión manual profunda. Sin embargo, los resultados de la evaluación del modelo en la sección anterior (tabla 5.3) y los ejemplos de la figura 5.4 sugieren que el modelo es capaz de detectar la mayoría de los eventos verdaderos.

Aún más, es importante resaltar que los resultados del modelo pueden complementar los reportes de CENAPRED. Por ejemplo, en la figura 5.6, el panel a) es una explosión reportada por CENAPRED, el modelo también asigna la clase EX. La señal del panel b) no se reporta como explosión por el CENAPRED, pero el modelo sí asigna la clase EX. Ambas señales ocurren el mismo día y se registran en la misma estación. Nótese como ambas tienen un contenido de frecuencias muy similar, tienen la misma amplitud y, fuera de un pequeño evento LP al inicio del evento b), sus formas de onda tienen varias semejanzas. Este ejemplo ayuda a resaltar la utilidad de los métodos de ML para obtener clasificaciones consistentes, donde es posible disminuir los errores asociados al operador. De igual forma, en el panel c) se muestra la señal de una explosión reportada por CENAPRED y detectada por el modelo. La señal del panel d) no se reporta como explosión pero así la clasifica el modelo. Nuevamente, ambas señales se registran el mismo día por la misma estación. En este caso hay más diferencias en las formas de onda y en el contenido de frecuencias. Sin embargo, el evento d) ocurre durante un episodio de tremor muy energético y su amplitud es mayor que la del evento en c). En el reporte de CENAPRED para ese día, se destaca la falta de visibilidad del volcán, por lo que sabemos que no es posible identificar el evento d) en las cámaras. Por otro lado, el aumento de amplitud que ocurre en el evento d) a los 28 segundos de su inicio, se podría asociar con la onda de sonido que viaja por el aire. De modo que, en efecto, se podría tratar de una explosión.

Por otro lado, en la figura 5.7 se muestra el número de eventos por día para las clases: ruido (NO), sismos regionales (RE) y eventos desconocidos (DE). La identificación de sismos regionales se ve seriamente afectada por la longitud de la ventana; pues la duración promedio de estos eventos, en el conjunto de entrenamiento, es de 195 segundos (p.ej. sismo de la figura 5.8.a). En consecuencia, la gran mayoría de los eventos de esta clase son espurio. En la figura 5.8.b se muestra un ejemplo clásico de esta situación. Donde el modelo asigna la clase RE, basándose en la forma de la señal, a un evento que no corresponde a un sismo regional.

También vale la pena discutir sobre los eventos desconocidos (clase DE). En la figura 5.7 se observa que, en promedio, el modelo es incapaz de identificar 200 ventanas al día; que equivalen a más de dos horas de registro. Dentro de esta clase se incluyen ventanas "contaminadas" por segmentos de más de un evento y ventanas en las que predomina un solo evento. En la figura 5.9 se muestran algunos ejemplos del segundo caso. Una alternativa para mejorar el desempeño del modelo es agregar estas señales al catálogo 2 y reentrenar. El evento de la figura 5.9.c es interesante porque son tres ventanas consecutivas de clase DE que, en realidad, corresponden a un sismo regional.

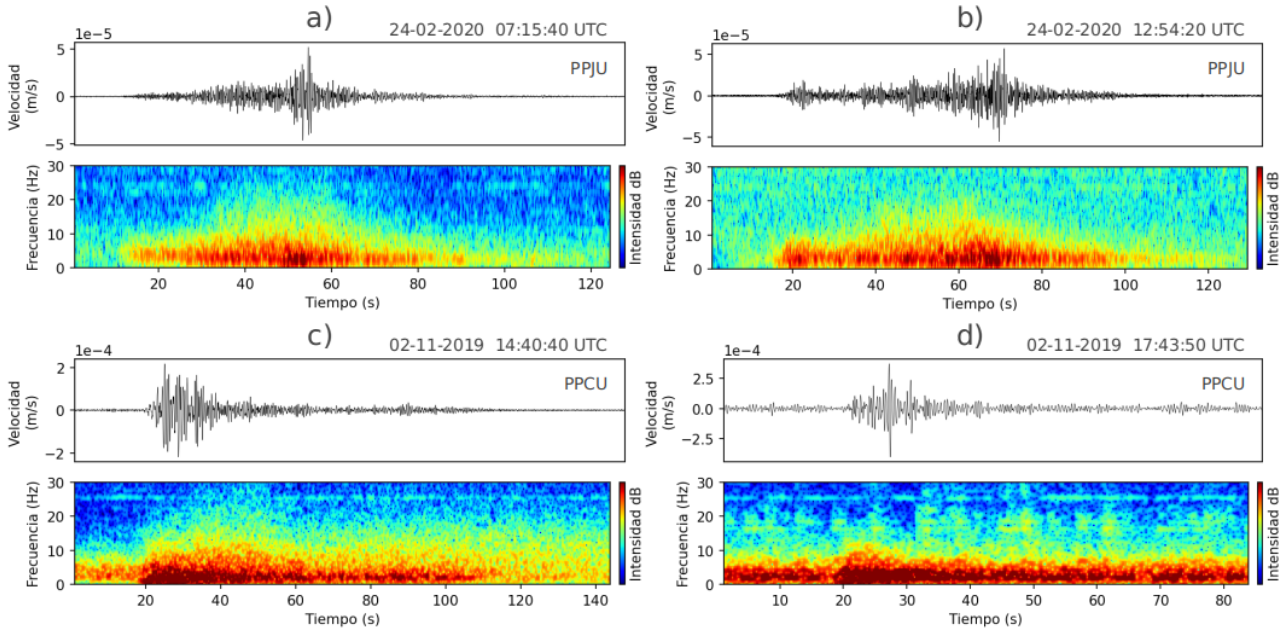

Figura 5.6
a) Explosión reportada por CENAPRED y detectada por el modelo. b) Explosión no reportada por CENAPRED, pero detectada por el modelo. Nótese la semejanza con el evento del panel a), ambos eventos se registran el mismo día por la misma estación (PPJU). c) Explosión reportada por CENAPRED y detectada por el modelo. d) Explosión no reportada por CENAPRED, pero detectada por el modelo. Los eventos en c) y d) ocurren el mismo día y se registran por la misma estación (PPCU).

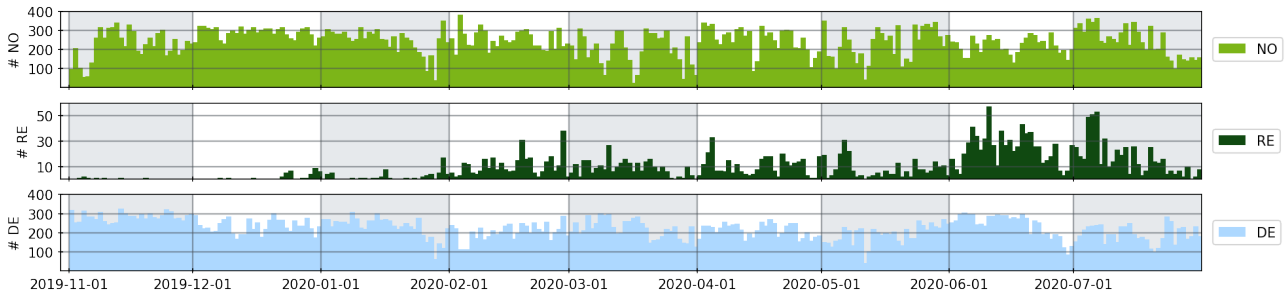

Figura 5.7
Número de eventos por día para las clases: ruido (NO), sismos regionales (RE) y eventos desconocidos (DE).

Comparación cuantitativa

Para realizar una comparación cuantitativa, se analizaron 2141 eventos clasificados por personal experto del CENAPRED. Estos corresponden a los eventos identificados los días 1, 4, 7, 10, 13, 16, 19, 22, 25 y 28 de febrero del 2020. En la tabla 5.4 se presentan los resultados. En este caso, la clasificación se considera correcta, incluso, si el modelo asigna la clase reportada por CENAPRED a un segmento del evento completo. Con esta condición, se obtiene una exactitud total del 78 %. El mejor desempeño del modelo es en la clase LP, con una precisión del 92 % y

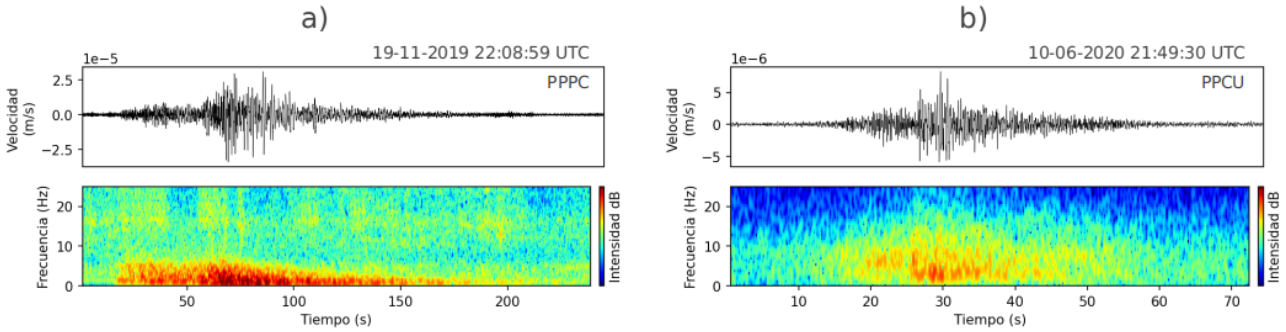

Figura 5.8
a) Sismo regional (RE) registrado el 19 de noviembre del 2019 a las 22:08:59 (UTC), en la estación PPPC. b) Ejemplo de error de clasificación en la clase RE durante el análisis continuo.

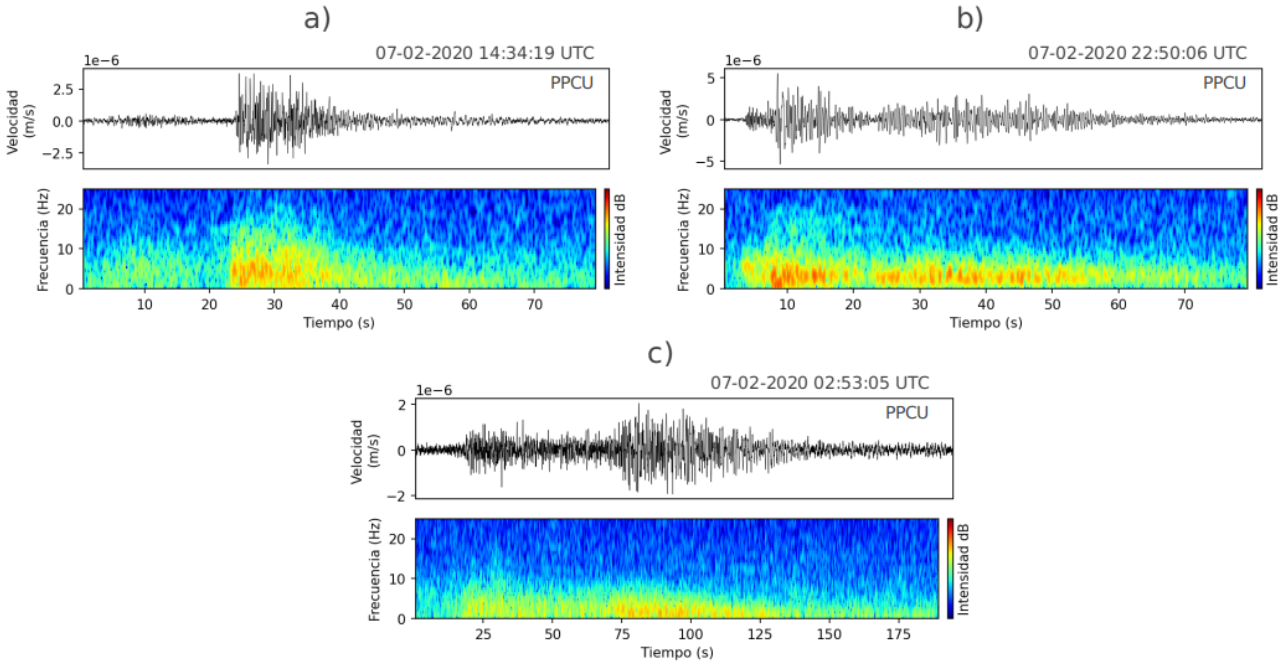

Figura 5.9
Ejemplos de eventos desconocidos que se podrían clasificar para aumentar el número de señales del catálogo y mejorar el desempeño del modelo.

sensibilidad del 81 %. El resto de las clases tiene una sensibilidad promedio del 66 %, que nos indica que el modelo es capaz de detectar la mayoría de los eventos reportados por CENAPRED.

Por otro lado, los valores de precisión reflejan el comportamiento que, cualitativamente, se aprecia en la comparación de eventos por día (figura 5.5). Es decir, para las clases LP y TR, la mayoría de las predicciones son verdaderos positivos (92 % y 86 %, respectivamente). La situación es opuesta en las clases EX y VT, donde únicamente el 35 % y 24 % de las predicciones son verdaderos positivos.

	Predicciones								
	LP	EX	TR	VT	NO	DE	RE	**Sensibilidad**	**Valor F1**
LP	1253	12	63	18	86	106	4	**81 %**	86 %
Explosión	4	9				1		64 %	45 %
Tremor	109	5	396	1	15	50		69 %	77 %
VT				6	1	2		67 %	35 %
Precisión	**92 %**	35 %	**86 %**	24 %					
				Exactitud				78 %	

Tabla 5.4: Comparación entre los eventos reportados en el catálogo de CENAPRED y los resultados del modelo PCA+SVM en el análisis de señales continuas. Se comparan 2141 eventos detectados en 10 días del mes de febrero del 2020. Se consideran las clases: LP, Explosión (EX), Tremor (TR), VT, Ruido (NO), Desconocido (DE) y Regional (RE). Es importante notar que CENAPRED no reporta ruido o sismos regionales.

5.4. Recomendaciones

Es importante notar las diferencias que existen entre los resultados del análisis de señales aisladas y los del análisis continuo (tablas 5.3 y 5.4). Por ejemplo, en la clasificación aislada, el peor desempeño se obtuvo en la clase LP; mientras que en la clasificación continua, la clase LP es la de mejor desempeño. Otra diferencia notable ocurre con las clases EX y VT, donde los valores F1 en la clasificación aislada son 81 % y 78 %, respectivamente; mientras que el la clasificación continua, los valores son 45 % y 35 %. Este comportamiento es un indicio de *overfitting* del modelo; pues sus capacidades de generalización son pobres. Para superar estas dificultades se hacen las siguientes recomendaciones:

- Modificar el código de la AAA para que el entrenamiento se haga con segmentos secuenciales de los eventos. Estos deberán ser de la longitud de la ventana que se usará en análisis continuo. De esta forma, el modelo aprenderá a clasificar con la información que tendrá al momento de su aplicación. Esta estrategia facilitará la comparación entre el análisis aislado y el continuo, contribuirá a la clasificación de eventos largos (p.ej. sismos regionales y tremores) en el análisis continuo y reduciría la necesidad de contar con ventanas de análisis más flexibles.

- Aumentar el número de eventos del catálogo 2 y reentrenar el modelo. Se recomienda incluir: i) eventos de la clase DE y ii) eventos que se clasificaron de forma correcta en el análisis continuo. Cuando se añadan eventos se debe mantener un balance entre las clases. Además, al incluir señales de un mismo evento, registradas en estaciones distintas, hay que cuidar que los efectos de atenuación sean menores. Con esta estrategia se pretende mejorar las propiedades de generalización del modelo.

- En la literatura se aprecia que los modelos construidos con redes neuronales tienen un

mejor desempeño que los que se obtienen con algoritmos de aprendizaje estadístico (p.ej. RF y SVM). Por lo que se recomienda explorar y comparar ambos enfoques.

Las primeras dos recomendaciones se pueden aplicar de forma recursiva y, de esta forma, mejorar el desempeño del modelo y la calidad de los catálogos. En ese sentido, los catálogos que se obtuvieron en este proyecto no son estáticos. Más aún, su uso requiere de una revisión de los mismos, por lo que el reentrenamiento se puede hacer de forma natural.

5.5. Trabajo futuro

En las secciones anteriores se discuten las bondades y limitaciones de los catálogos obtenidos. A continuación, se hacen algunas reflexiones sobre su utilidad en trabajos futuros.

Para comenzar, hay que remarcar que los catálogos son un marco de organización de los datos sísmicos. En ese sentido, se usarán, como punto de partida, en la selección de eventos de interés de estudios específicos. En particular, durante mis estudios de doctorado, trabajaré en la formulación de modelos 3D de la estructura del Popocatépetl. Por consiguiente, se empleará el catálogo de la clase VT para seleccionar los eventos que se utilizarán en las inversiones de velocidad y de atenuación. Además, en el proyecto doctoral, se plantea el uso de registros de tremor y eventos LP para complementar los modelos de atenuación; así que los catálogos de las clases TR y LP también serán relevantes.

Por otro lado, es claro que el desempeño actual del modelo sería poco eficiente en las tareas de monitoreo del volcán. Sin embargo, el potencial de los métodos de ML, en esta área, se ha justificado en múltiples estudios[§] y vale la pena continuar con su exploración para que, en el futuro, su implementación sea rigurosa. En ese sentido, las recomendaciones que se hacen en la sección anterior son importantes. Además, hay que destacar que el uso de estos métodos tiene beneficios, aún cuando los modelos no son lo suficientemente robustos para su uso en tiempo real. Por ejemplo, su aptitud para disminuir los errores debidos al operador y, en general, su uso para la estandarización de los criterios de clasificación, es una ventaja que, con frecuencia, es infravalorada.

Finalmente, algo que hay que tomar en cuenta, es que las técnicas de ML no son un sustituto de los métodos basados en modelos físicos, ni de los operadores que realizan las tareas de forma manual. Más bien, deben entenderse como herramientas que complementan a las técnicas tradicionales. Si bien es cierto que tiene limitaciones importantes, el modelado basado

[§]Por ejemplo, Dempsey y col., 2020, entrenaron un modelo de bosques aleatorios para desarrollar un prototipo de alerta temprana en el volcán Whakaari de Nueva Zelanda. Su modelo trabaja con registros de tremor y, en caso de identificar precursores, emite una alerta. Usando los datos de la erupción del 2 de diciembre del 2019, donde murieron 21 personas, encontraron que su prototipo es capaz de emitir una alerta 4 horas antes del evento. Naturalmente, la implementación de su prototipo en el observatorio del volcán implica grandes retos; no obstante, sus resultados exhortan a continuar con la evaluación de estas técnicas en el monitoreo volcánico.

en datos es relevante en una disciplina en la que la disponibilidad de datos está creciendo de forma exponencial (Q. Kong y col., 2019).

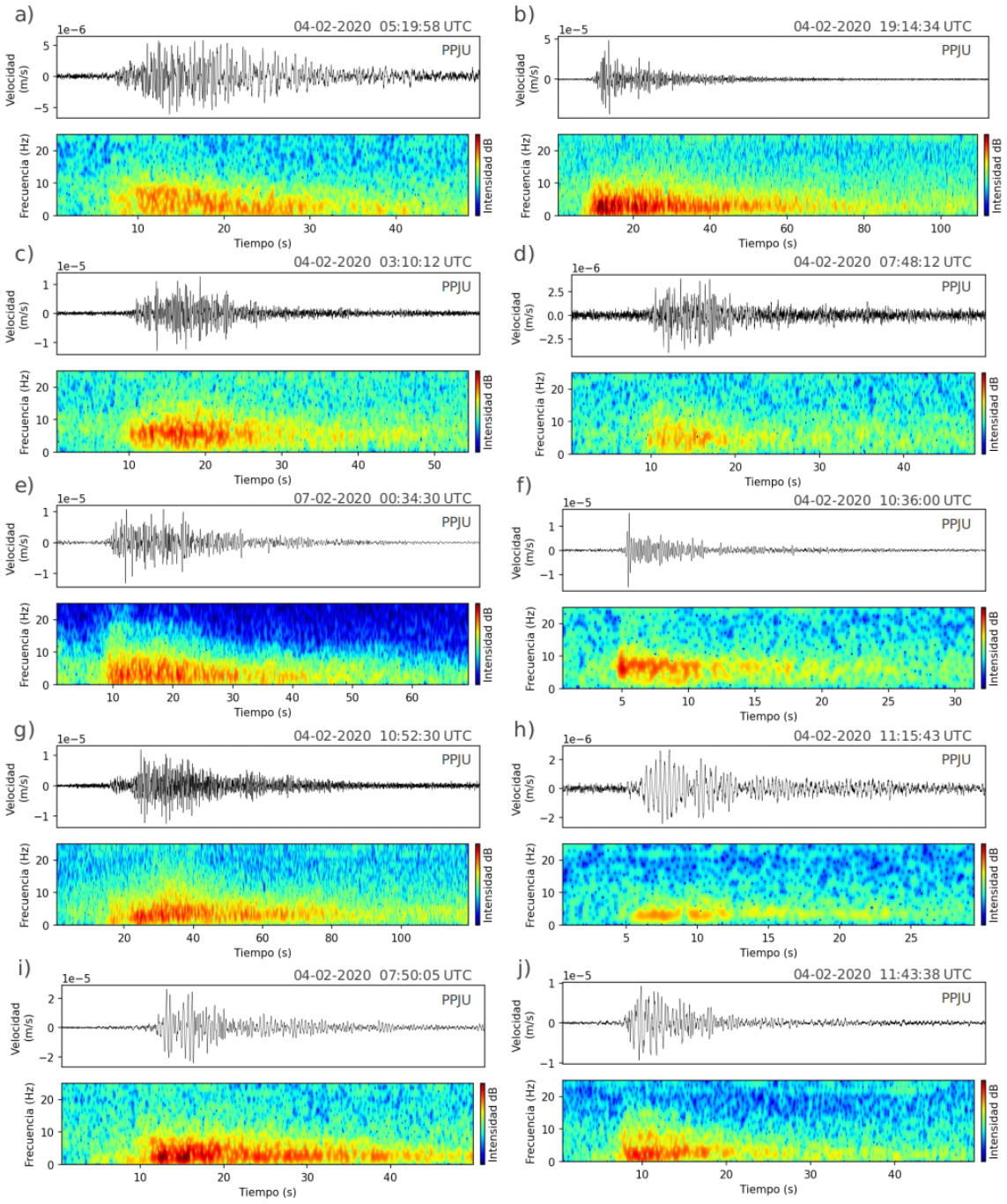

Figura 5.10
Señales de la clase LP. Nótese la diversidad de las señales en cuanto a forma de onda, amplitud, duración y contenido de frecuencia. Todas las señales están clasificadas como eventos LP en el catálogo interno de CENAPRED. El modelo clasificador también les asigna esta clase.

Capítulo 6

Conclusiones

En este proyecto se implementaron métodos de machine learning para obtener una serie de catálogos de las señales sísmicas más frecuentes del volcán Popocatépetl. Los cuales serán útiles en la etapa de selección de eventos de interés de estudios futuros. Para su obtención se usó un esquema de aprendizaje supervisado; en el que un conjunto de señales, previamente etiquetadas, le enseñó al modelo a distinguir entre distintas clases. Se usó un esquema de clasificación general, que incluye las clases: periodo largo (LP), explosiones (EX), tremor (TR), volcano-tectónicos (VT), sismos regionales (RE) y ruido (NO). Para realizar la clasificación se implementó un programa, de software libre, desarrollado por Malfante, Mars y col., 2018. Las principales aportaciones que se hicieron al la metodología fueron: i) el uso de una estrategia de aumento de datos para mejorar el desempeño del modelo, en ella se incluyen las señales de las tres componentes, de una misma estación, y los registros del mismo evento en distintas estaciones; ii) la implementación de la técnica de PCA para disminuir las dimensiones del espacio de atributos propuesto por Malfante, Dalla Mura y col., 2018. Ambas modificaciones produjeron un aumento del 8 % en el desempeño del modelo y agregaron la capacidad de clasificar registros de más de una estación. El mejor modelo que se obtuvo (PCA + SVM) alcanzó una exactitud del 72 % en la clasificación aislada de eventos desconocidos. Posteriormente, se usó este modelo para el análisis de señales continuas que abarcan un periodo de 9 meses entre noviembre del 2019 y julio del 2020. La comparación de los resultados del análisis continuo con los reportes diarios del CENAPRED indican que el modelo es particularmente bueno en la detección de las clases LP y TR. Por otro lado, la limitación más importante del modelo es su tendencia a cometer errores de clasificación en las clases VT y EX; sobrevaluando, de forma considerable, el número de eventos de ambas clases.

Apéndice A

Formalismos de los algoritmos utilizados

A.1. Análisis de componentes principales (PCA)

A continuación, se presenta únicamente el formalismo del enfoque de la varianza máxima. Si el lector está interesado en la formulación de la distancia mínima, se recomienda ampliamente la lectura del capítulo 12 de Bishop, 2006.

Además, para facilitar los cálculos se considera que los datos se han centrado en el origen, es decir, que la media de todos los atributos es cero. Cualquier conjunto de datos se puede transformar de esta forma así que no hay pérdida de generalidad.

Para encontrar la primera componente principal se busca una dirección definida por el vector $\mathbf{u}_1$ en la que la varianza sea máxima. Además, no nos interesa la magnitud del vector por lo que se impone una norma unitaria, $\mathbf{u}_1^T\mathbf{u}_1 = 1$. Sea la matriz $\mathbf{X}$ aquella cuyas filas son los vectores asociados a cada observación. La proyección de los datos en la primera componente principal es $\mathbf{X}\mathbf{u}_1$ y, como los datos tienen media cero, la varianza de la proyección es:

$$\frac{1}{N}\left(\mathbf{X}\mathbf{u}_1\right)^T \cdot \mathbf{X}\mathbf{u}_1 = \mathbf{u}_1^T \cdot \left(\frac{1}{N}\mathbf{X}^T\mathbf{X}\right) \cdot \mathbf{u}_1 = \mathbf{u}_1^T\mathbf{C}\mathbf{u}_1, \tag{A.1}$$

donde $\mathbf{C}$ es la matriz de covarianza.

Entonces buscamos $\mathbf{u}_1$ tal que $\mathbf{u}_1^T\mathbf{C}\mathbf{u}_1$ sea máximo y se cumpla la restricción $\mathbf{u}_1^T\mathbf{u}_1 = 1$. Usando el método de multiplicadores de Lagrange, se puede definir un problema de optimización sin restricciones en el que la función de costo es:

$$J = \mathbf{u}_1^T\mathbf{C}\mathbf{u}_1 + \lambda_1(1 - \mathbf{u}_1^T\mathbf{u}_1). \tag{A.2}$$

Donde λ es el multiplicador de Lagrange, Para maximizar se deriva con respecto a $\mathbf{u}_1$ y se iguala a cero. La expresión tiene un punto estacionario cuando

$$\mathbf{C}\mathbf{u}_1 = \lambda_1\mathbf{u}_1, \tag{A.3}$$

Que nos dice que $\mathbf{u}_1$ debe de ser un *eigenvector* o vector propio de $\mathbf{C}$. Si se multiplica por $\mathbf{u}_1^T$ por la izquierda y se hace uso de su norma unitaria se obtiene que la varianza está dada por

$$\mathbf{u}_1^T \mathbf{C} \mathbf{u}_1 = \lambda_1, \tag{A.4}$$

que tendrá un valor máximo cuando la primera primera componente principal, $\mathbf{u}_1$, sea el vector propio asociado al *eigenvalor* o valor propio más alto.

El resto de las componentes principales se encuentran escogiendo nuevas direcciones que maximicen la varianza de la proyección y, además, sean ortogonal a las componentes que ya se hayan definido. En resumen, si se desea encontrar una representación de dimensión $M < D$, con D la dimensión original, la proyección lineal óptima que maximiza la varianza de los datos proyectados está definida por M vectores propios $\mathbf{u}_1, ..., \mathbf{u}_M$ de la matriz de covarianza $\mathbf{C}$ que corresponden a los M valores propios más grandes $\lambda_1, ..., \lambda_M$ (Bishop, 2006; James y col., 2013; Jolliffe, 2005).

A.2. Bosques aleatorios (RF)

La implementación del método que se describe en la sección 3.3.2 se puede hacer con un algoritmo sencillo que se presenta más adelante, sin embargo es necesario saber cómo medir la impureza (Q) de una región. En la literatura se usan dos métricas distintas. Por lo general, ambas dan resultados similares y la elección entre una u otra es uno de los hiperparámetros del modelo. Si se considera $p_{R_\tau k}$ la proporción de datos de la región R_τ a los que se le asigna la clase k, donde $k = 1, ..., K$ y $\tau = 1, 2$ (cada nodo define dos regiones), entonces las dos opciones son la *entropía cruzada*

$$Q(R_\tau) = -\sum_{k=1}^{K} p_{R_\tau k} \ln p_{R_\tau k} \tag{A.5}$$

y el *índice de Gini*

$$Q_(R_\tau) = \sum_{k=1}^{K} p_{R_\tau k}(1 - p_{R_\tau k}). \tag{A.6}$$

Bishop, 2006 destaca que ambas métricas dan cero cuando $p_{\tau k} = 0$ y $p_{\tau k} = 1$ y toman un valor máximo cuando $p_{\tau k} = 0.5$, que es el comportamiento que se espera.

La función de costo de cada nodo debe considerar la impureza de las dos regiones por crear (R_1 y R_2) y el número de instancias en cada una de ellas (m_1 y m_2), luego

$$J(\theta) = \frac{m_1}{N}\ Q(R_1) + \frac{m_1}{N}\ Q(R_1), \tag{A.7}$$

donde N es el número total de instancias.

Algoritmo 1: Bosques aleatorios

Sean $\mathcal{D} = \{(\mathbf{x}_1, y_1), ..., (\mathbf{x}_N, y_N)\}$ el conjunto de entrenamiento y $\mathcal{P} = \{x^1, ..., x^p\}$ el conjunto de atributos;

for $\forall$ árbol **do**

Definir un conjunto aleatorio de instancias:

$$D_{rand} = \text{random}(\mathcal{D})$$

while *criterio de parada = false* **do**

for $\forall$ $x^j \in random(\mathcal{P})$ **do**

for $\forall$ θ_i, punto de separación **do**

Definir las regiones R_1 y R_2:

$$R_1(x^j, \theta_i) = \{\mathbf{x}_i \mid x^j \leq \theta_i\}, \quad x_i \in \mathcal{D}_{rand}$$

$$R_2(x^j, \theta_i) = \{\mathbf{x}_i \mid x^j > \theta_i\}, \quad x_i \in \mathcal{D}_{rand}$$

Calcular la función de costo del nodo:

$$J(x^j, \theta_i)$$

end

Seleccionar el mejor punto de separación:

$$\theta_i = \arg\min_{\theta_i}(J(x^j, \theta_i))$$

end

Seleccionar el mejor atributo:

$$x^j = \arg\min_{x^j}(J(x^j, \theta_i))$$

end

end

Otra factor importante que hay que tomar en cuenta es el criterio para detener el crecimiento de los árboles (criterio de parada). En principio la subdivisión de cualquier región se detiene cuando su impureza es nula, sin embargo, también es posible definir un valor umbral que frene la subdivisión aún cuando la impureza sea distinta de cero (Géron, 2019). Además, es común que se establezca una profundidad máxima que le indica al algoritmo el número máximo de iteraciones. La profundidad se refiere a los conectores o *ramas* que hay en un diagrama como el de la figura 3.7.B. Nótese que estos valores son hiperparámetros del modelo.

Un hiperparámetro adicional es el número de árboles que tiene el bosque. La forma de

determinar el valor óptimo es por medio de un proceso de validación. Por lo general se consideran entre decenas o cientos de árboles (Géron, 2019).

Una ventaja importante de este método es que ofrece una métrica de la importancia de los atributos. La relevancia de un atributo en particular se calcula promediando la reducción de impureza de los todos los nodos en los que se usa. De forma más precisa, se hace un promedio pesado donde los pesos son el número de instancias de las regiones definidas por los nodos, la métrica se conoce como *disminución media de la impureza* o MDI, por sus siglas en inglés (Breiman, 2001). Sin embargo, una sutileza que hay que destacar es que estos valores no necesariamente se refieren al poder de predicción de cada atributo, sino a su relevancia en el modelo clasificador.

En el recuadro Algoritmo 1 se presenta un esquema simplificado del algoritmo de RF, para más detalles sobre el método se invita al lector a consultar las fuentes Bishop, 2006; Breiman, 2001; Cutler y col., 2012; Hastie y col., 2017; Malfante, 2018.

A.3. Máquinas de vectores de soporte (SVM)

Teniendo en mente las ideas de la sección 3.3.3, el problema se puede plantear de la siguiente forma. Consideremos un problema binario en el que, sin pérdida de generalidad, la clases son -1 y 1. El conjunto de entrenamiento está formado por N instancias con sus respectivas variables objetivo,

$$(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_N, y_N), \quad y_i \in \{-1, 1\}. \tag{A.8}$$

Se dice que las instancias son linealmente separables cuando es posible encontrar un vector $\mathbf{w}$ y un escalar b que cumplen las siguientes expresiones:

$$\begin{aligned} \mathbf{w} \cdot \mathbf{x}_i + b \geq 1 \ \text{ if } \ y_i = 1, \\ \mathbf{w} \cdot \mathbf{x}_i + b \leq -1 \ \text{ if } \ y_i = -1. \end{aligned} \tag{A.9}$$

Esto es equivalente a decir que todos los puntos satisfacen la restricción

$$y_n(\mathbf{w}^T \mathbf{x}_n + b) \geq 1, \quad n = 1, ..., N \tag{A.10}$$

De modo que el hiperplano óptimo,

$$\mathbf{w}_o \cdot \mathbf{x} + b_o = 0, \tag{A.11}$$

es aquel que separa los datos de entrenamiento con un margen máximo. La distancia ρ de los puntos más cercanos al hiperplano es

$$\rho(\mathbf{w}_o, b_o) = \min_{\{\mathbf{x}: y=1\}} \frac{\mathbf{w} \cdot \mathbf{x}}{\mid \mathbf{w} \mid} - \max_{\{\mathbf{x}: y=-1\}} \frac{\mathbf{w} \cdot \mathbf{x}}{\mid \mathbf{w} \mid}, \tag{A.12}$$

donde se están considerando los vectores de soporte de la clase positiva y de la negativa. Entonces, dadas las ecuaciones A.9, los parámetros óptimos $(\mathbf{w}_o, b_o)$ son tales que el margen,

$$\rho(\mathbf{w}_o, b_o) = \frac{2}{\mid \mathbf{w}_o \mid} = \frac{2}{\sqrt{\mathbf{w}_o \cdot \mathbf{w}_o}}, \tag{A.13}$$

se maximiza. Por último, nótese que el problema de optimización es equivalente a minimizar $\mathbf{w} \cdot \mathbf{w}$ sujeto a las restricciones A.9.

Además esta formulación se sostiene aún cuando es necesario transformar las instancias a un espacio en el que sean linealmente separables,

$$\mathbf{x} \mapsto \phi(\mathbf{x}). \tag{A.14}$$

En la literatura es común encontrar que al espacio transformado se le llame espacio de atributos, pero para evitar confusión aquí se le llamará el *espacio del kernel* (Malfante, 2018). El truco del kernel se puede aplicar con cualquier transformación, pero con frecuencia se usan funciones polinomiales o gaussianas (de base radial). Nótese que los parámetros de estas funciones serán hiperparámetros del modelo, es decir, la transformación debe estar completamente definida por el usuario.

En el espacio del kernel el hiperplano simplemente se escribe como

$$f(\phi(\mathbf{x})) = \mathbf{w}^T \phi(\mathbf{x}) + b. \tag{A.15}$$

Usando el método de multiplicadores de Lagrange, el problema de optimización es

$$L(\mathbf{w}, b, \mathbf{a}) = \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{n=1}^{N}\{y_n(\mathbf{w}^T\phi(\mathbf{x}_n) + b) - 1\}. \tag{A.16}$$

Donde el factor $1/2$ se agrega por conveniencia y $\mathbf{a} = (a_1, ..., a_N)^T$ es un vector de multiplicadores de Lagrange. El signo negativo antes del multiplicador es porque hay que minimizar con respecto a $\mathbf{w}$ y b, y maximizar con respecto a $\mathbf{a}$ $(a_n \geq 0)$ Bishop, 2006.

Para completar el formalismo hace falta considerar el margen suave. Para esto se introducen las *variables de holgura*, ξ_n, que funcionan como penalización cuando hay errores de clasificación. A cada instancia se le asocia una de estas variables, $\xi_n = 0$ cuando la instancia está bien clasificada y $\xi_n = |y_n - f(\mathbf{x}_n)|$ para los otros puntos. Por lo que las restricciones del problema (ec. A.10) se pueden remplazar con

$$y_n f(\phi(\mathbf{x}_n)) \geq 1 - \xi_n, \qquad n = 1, ..., N. \tag{A.17}$$

En este caso los objetivos del modelo son maximizar el margen y minimizar la penalización por los errores de clasificación. Por lo que el problema de optimización es minimizar

$$C\sum_{n=1}^{N}\xi_n + \frac{1}{2}\|\mathbf{w}\|^2, \tag{A.18}$$

sujeto a las restricciones A.17. Donde $C > 0$ es un hiperparámetro que permite hacer una compensación entre minimizar los errores de clasificación y la complejidad del modelo. Finalmente, usando multiplicadores de Lagrange, el modelo debe encontrar una solución para

$$L(\mathbf{w}, b, \mathbf{a}, \mu_i) = \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{n=1}^{N}\xi_n - \sum_{n=1}^{N}a_n\{y_n f(\phi(\mathbf{x}_n)) - 1 + \xi_n\} - \sum_{n=1}^{N}\mu_n\xi_n. \tag{A.19}$$

Más detalles sobre la solución se pueden encontrar en Bishop, 2006, Cortes y Vapnik, 1995 y Malfante, 2018.

Apéndice B

Aspectos técnicos de la metodología

B.1. Red sísmica del Popocatépetl

El Popocatépetl cuenta con una red de monitoreo que tiene 19 estaciones de banda ancha. 11 de estas estaciones –su nombre empieza con PP– son gestionadas por el CENAPRED y cuentan con teletransmisión para un monitoreo en tiempo real. Las 8 estaciones restantes –su nombre empieza con SP– están a cargo del doctor Marco Calò del Instituto de Geofísica de la UNAM y cuentan con un sistema de almacenamiento *in situ*.

B.2. Búsqueda de hiperparámetros

En la tabla B.2 se muestran los valores que se usaron en la búsqueda de mallado para la selección de los mejores hiperparámetros. Además se incluye una pequeña descripción de cada uno de ellos.

Estación	*Nombre*	*Distancia al cráter (km)*	*Instrumento*
PPPP	Canario	2.20	Trillium posthole 120 s
PPPX	Chipiquixtle	4.11	Trillium posthole 120 s
PPJU	Juncos	2.33	Trillium posthole 120 s
PPCU	Cuervos	2.65	Trillium posthole 120 s
PPPC	Colibrí	8.25	Guralp CMG-DM24 (hasta 08/07/2019) Guralp 6TD-T6W01 (hasta 23/01/2019) Guralp 6TD-T6V92
PPPT	Tetexcaloc	4.39	Trillium posthole 120 s
PPEN	Encinos	4.65	Trillium posthole 120 s
PPCL	Cuilotepec	12.69	Trillium posthole 120 s
PPAX	Atlixco	21.41	Trillium posthole 120 s
PPSJ	San Juan Tehuixtitlán	15.67	Trillium posthole 120 s
PPSX	Santiago Xalitzintla	13.89	Guralp 6TD-T6A26 (hasta 28/03/2019) Guralp 6TD-T6W12 (hasta 29/07/2020)
SP01	San Pedro	11.79	Trillium compact posthole 120 s
SP02	Cementerio	15.69	Trillium compact posthole 120 s
SP03	Tetela	16.58	Trillium compact posthole 120 s
SP04	Ecatzingo	15.80	Trillium compact posthole 120 s
SP05	Parque ecológico	12.08	Trillium compact posthole 120 s
SP06	Buenavista	11.48	Trillium compact posthole 120 s
SP07	Altimeyaya	15.62	Trillium compact posthole 120 s
SP08	Golondrinas	2.91	Trillium compact posthole 120 s

Tabla B.1: Especificaciones de la red de monitoreo sísmico del volcán Popocatépetl. Se muestra código, nombre, distancia al cráter e instrumento de cada estación.

Hiperparámetro	**Descripción**	**Valores considerados**
	Bosques aleatorios (RF)	
Número de estimadores	Número de árboles de decisión que se usan	100, 200, 500
Máximo de atributos	El número de atributos que se consideran en cada nodo. Se calculan sobre el total de atributos	'sqrt', 'log2'
Profundidad máxima	Profundidad máxima para cada árbol de decisión. Máximo número de nodos	5, 30, 60, None
Criterio de impureza	Función para medir la calidad del punto de separación	'gini', 'entropy'
Bootstrap	La selección de atributos en cada nodo se hace con reemplazo	True, False
	Máquinas de vectores de soporte (SVM)	
Kernel	Transformación del espacio de atributos. Lineal o función de base radial	'linear', **'rbf'**
C	Parámetro de regularización que permite errores en la clasificación. Valores pequeños permiten menos errores	0.1, 1, 10, 100, **1000**
Gamma	Radio de influencia de cada punto. Controla la curvatura de la frontera de decisión. Valores pequeños implican poca curvatura	0.0001, 0.001, **0.01**, 0.1, 1, 10, 100

Tabla B.2: Hiperparámetros que se consideraron en el aprendizaje del modelo. Para más detalle sobre su significado véase las secciones 3.3.2 y 3.3.3. La elección de los hiperparámetros óptimos se hace usando validación cruzada y comparando la exactitud media de todas las combinaciones posibles.

Referencias

Abedi, M., Norouzi, G.-H. & Bahroudi, A. (2012). Support vector machine for multi-classification of mineral prospectivity areas. *Computers & Geosciences*, *46*, 272-283. https://doi.org/https://doi.org/10.1016/j.cageo.2011.12.014

Albawi, S., Mohammed, T. A. & Al-Zawi, S. (2017). Understanding of a convolutional neural network. *2017 International Conference on Engineering and Technology (ICET)*, 1-6. https://doi.org/10.1109/ICEngTechnol.2017.8308186

Arámbula-Mendoza, R., Valdés-González, C. & Martínez-Bringas, A. (2010). Temporal and spatial variation of the stress state of Popocatépetl Volcano, Mexico. *Journal of Volcanology and Geothermal Research*, *196*(3), 156-168. https://doi.org/https://doi.org/10.1016/j.jvolgeores.2010.07.007

Arámbula-Mendoza, R., Valdés-González, C., Varley, N., Juárez-García, B., Alonso-Rivera, P. & Hernández-Joffre, V. (2013). Observation of vulcanian explosions with seismic and acoustic data at Popocatépetl volcano, Mexico. *Monitoring of Volcanic Activity: Methods and Results*, 13-33.

Arámbula-Mendoza, R., Valdés-González, C., Varley, N., Reyes-Pimentel, T. & Juárez-Garíca, B. (2016). Tremor and its duration-amplitude distribution at Popocatépetl volcano, Mexico. *Geophysical Research Letters*, *43*(17), 8994-9001.

Arana, L., Siebe, C. & Macías, J. L. (2017). Construcción del volcán Popocatépetl actual o moderno: una historia repetida de erupciones plinianas. En J. Yamamoto Victorio (Ed.), *Memoria técnica del mapa de peligros del volcán Popocatépetl*. Unidad de apoyo editorial, Instituto de Geofísica.

Araya-Polo, M., Jennings, J., Adler, A. & Dahlke, T. (2018). Deep-learning tomography. *The Leading Edge*, *37*(1), 58-66.

Arciniega-Ceballos, A. (2002). *Análisis de datos sísmicos de banda ancha registrados en el volcán Popocatépetl* (Tesis doctoral). Universidad Nacional Autónoma de México.

Arciniega-Ceballos, A., Chouet, B. & Dawson, P. (2003). Long-period events and tremor at Popocatepetl volcano (1994–2000) and their broadband characteristics. *Bulletin of Volcanology*, *65*(2), 124-135.

Arciniega-Ceballos, A., Chouet, B., Dawson, P. & Asch, G. (2008). Broadband seismic measurements of degassing activity associated with lava effusion at Popocatépetl Volcano, Mexico. *Journal of Volcanology and Geothermal Research*, *170*(1-2), 12-23.

Arciniega-Ceballos, A., Chouet, B. A. & Dawson, P. (1999). Very long-period signals associated with vulcanian explosions at Popocatépetl volcano, Mexico. *Geophysical Research Letters*, *26*(19), 3013-3016.

Arciniega-Ceballos, A., Dawson, P. & Chouet, B. A. (2012). Long period seismic source characterization at Popocatépetl volcano, Mexico. *Geophysical Research letters*, *39*(20).

Arciniega-Ceballos, A., Valdes-Gonzalez, C. & Dawson, P. (2000). Temporal and spectral characteristics of seismicity observed at Popocatépetl volcano, central Mexico. *Journal of volcanology and geothermal research*, *102*(3-4), 207-216.

Bianco, M., Gerstoft, P., Olsen, K. B. & Lin, F.-C. (2019). High-resolution seismic tomography of Long Beach, CA using machine learning. *Scientific reports*, *9*(1), 1-11.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer New York.

Boser, B. E., Guyon, I. M. & Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers. *Proceedings of the fifth annual workshop on Computational learning theory*, 144-152.

Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5-32.

Brown, S. K., Auker, M. & Sparks, R. (2015). Populations around Holocene volcanoes and development of a Population Exposure Index. *Global volcanic hazards and risk*, 223-232.

Bueno, A., Benitez, C., De Angelis, S., Moreno-Díaz, A. & Ibáñez, J. (2019). Volcano-seismic transfer learning and uncertainty quantification with Bayesian neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, *58*(2), 892-902.

Capra, L., Poblete, M. & Alvarado, R. (2004). The 1997 and 2001 lahars of Popocatépetl volcano (Central Mexico): textural and sedimentological constraints on their origin and hazards. *Journal of Volcanology and Geothermal Research*, *131*(3-4), 351-369.

Carniel, R., Jolly, A. D. & Barbui, L. (2013). Analysis of phreatic events at Ruapehu volcano, New Zealand using a new SOM approach. *Journal of volcanology and geothermal research*, *254*, 69-79.

CENAPRED. (2020). Clasificación, cuantificación y alertamiento de las explosiones del volcán Popocatépetl en tiempo cuasi-real, como apoyo en la etapa de prevención. *Subdirección de Riesgos Volcánicos*. https://www1.cenapred.unam.mx/DIR_INVESTIGACION/2021/1er_Trimestre/FRACCION_XLI/RV/Clasificacion_cuantificacion_y_alertamiento_de_las_explosiones_del_volcan_Popocatepetl_en_tiempo_cuasi-real_como_apoyo_en_la_etapa_de_prevencion.pdf

Chai, C., Maceira, M., Santos-Villalobos, H. J., Venkatakrishnan, S. V., Schoenball, M., Zhu, W., Beroza, G. C., Thurber, C. & Team, E. C. (2020). Using a deep neural network and transfer learning to bridge scales for seismic phase picking. *Geophysical Research Letters*, *47*(16), e2020GL088651.

Chen, H., Guo, J., Xiong, W., Guo, S. & Xu, C.-Y. (2010). Downscaling GCMs using the Smooth Support Vector Machine method to predict daily precipitation in the Hanjiang Basin. *Advances in Atmospheric Sciences*, *27*(2), 274-284.

Chervonenkis, A. Y. (2013). Early History of Support Vector Machines. En B. Schölkopf, Z. Luo & V. Vovk (Eds.), *Empirical Inference: Festschrift in Honor of Vladimir N. Vapnik* (pp. 13-20). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-41136-6_3

Chouet, B. (1996). Long-period volcano seismicity: its source and use in eruption forecasting. *Nature*, *380*(6572), 309-316.

Chouet, B., Dawson, P. & Arciniega-Ceballos, A. (2005). Source mechanism of Vulcanian degassing at Popocatépetl Volcano, Mexico, determined from waveform inversions of very long period signals. *Journal of Geophysical Research: Solid Earth*, *110*(B7).

Chouet, B. & Matoza, R. (2013). A multi-decadal view of seismic methods for detecting precursors of magma movement and eruption. *Journal of Volcanology and Geothermal Research*, *252*, 108-175.

Clarke, J., Adam, L. & van Wijk, K. (2021). LP or VT signals? How intrinsic attenuation influences volcano seismic signatures constrained by Whakaari volcano parameters. *Journal of Volcanology and Geothermal Research*, *418*, 107337.

Cortes, C. & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*(3), 273-297. https://doi.org/10.1007/BF00994018

Cortés, G., Arámbula, R., Gutiérrez, L., Benítez, C., Ibánez, J., Lesage, P., Alvarez, I. & Garcia, L. (2009). Evaluating robustness of a HMM-based classification system of volcano-seismic events at Colima and Popocatépetl volcanoes. *2009 IEEE International Geoscience and Remote Sensing Symposium*, *2*, II-1012.

Cortés, G., Carniel, R., Lesage, P., Mendoza, M. Á. & Della Lucia, I. (2021). Practical volcano-independent recognition of seismic events: VULCAN. ears project. *Frontiers in Earth Science*, *8*, 616676.

Cruz-Atienza, V., Pacheco, J., Singh, S., Shapiro, N., Valdés, C. & Iglesias, A. (2001). Size of Popocatépetl volcano explosions (1997–2001) from waveform inversion. *Geophysical research letters*, *28*(21), 4027-4030.

Cutler, A., Cutler, D. R. & Stevens, J. R. (2012). Random forests. *Ensemble machine learning* (pp. 157-175). Springer.

Delgado Granados, H., Cassatta, W., Gisbert Pinto, G. & Renee, P. (2017). Historia geológica y eruptiva del volcán Popocatépetl. En J. Yamamoto Victorio (Ed.), *Memoria técnica del mapa de peligros del volcán Popocatépetl*. Unidad de apoyo editorial, Instituto de Geofísica.

de Lorenzo, S., Zollo, A. & Mongelli, F. (2001). Source parameters and three-dimensional attenuation structure from the inversion of microearthquake pulse width data: Qp imaging and inferences on the thermal state of the Campi Flegrei caldera (southern Italy). *Journal of Geophysical Research: Solid Earth*, *106*(B8), 16265-16286.

Dempsey, D., Cronin, S. J., Mei, S. & Kempa-Liehr, A. W. (2020). Automatic precursor recognition and real-time forecasting of sudden explosive volcanic eruptions at Whakaari, New Zealand. *Nature communications*, *11*(1), 1-8.

Dye, B. C. & Morra, G. (2020). Machine learning as a detection method of Strombolian eruptions in infrared images from Mount Erebus, Antarctica. *Physics of the Earth and Planetary Interiors*, *305*, 106508. https://doi.org/https://doi.org/10.1016/j.pepi.2020.106508

Eddy, S. R. (2004). What is a hidden Markov model? *Nature biotechnology*, *22*(10), 1315-1316.

Espinasa-Pereña, R., Arámbula, R., Ramos, S., Sieron, K., Capra, L., Hernández-Oscoy, A., Alatorre, M. & Montiel, F. C. (2021). Monitoring volcanoes in Mexico. *Volcanica*, *4*(S1), 223-246.

Espinasa-Pereña, R. & Martín-del Pozzo, A. L. (2006). Morphostratigraphic evolution of Popocatépetl volcano, México. *Special Papers-GEOLOGICAL SOCIETY OF AMERICA*, *402*, 115.

Falcin, A., Métaxian, J.-P., Mars, J., Stutzmann, É., Komorowski, J.-C., Moretti, R., Malfante, M., Beauducel, F., Saurel, J.-M., Dessert, C. y col. (2021). A machine-learning approach for automatic classification of volcanic seismicity at La Soufrière Volcano, Guadeloupe. *Journal of Volcanology and Geothermal Research*, *411*, 107151.

Falsaperla, S., Fortuna, L., Graziani, S. & Nunnari, G. (1992). Automatic classification of seismic events by neural networks. *IGARSS'92; Proceedings of the 12th Annual International Geoscience and Remote Sensing Symposium*, *1*, 224-226.

Ferres, D. & Fonseca, R. (2017). Introducción y estudios geológicos. En J. Yamamoto Victorio (Ed.), *Memoria técnica del mapa de peligros del volcán Popocatépetl*. Unidad de apoyo editorial, Instituto de Geofísica.

Fraile, R. & Godino-Llorente, J. I. (2014). Cepstral peak prominence: A comprehensive analysis. *Biomedical Signal Processing and Control*, *14*, 42-54.

Géron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems* (2nd). O'Reilly Media, Inc.

Gibbons, S. J. & Ringdal, F. (2006). The detection of low magnitude seismic events using array-based waveform correlation. *Geophysical Journal International*, *165*(1), 149-166.

Gisbert, G., Delgado-Granados, H., Mangler, M., Prytulak, J., Espinasa-Pereña, R. & Petrone, C. M. (2021). Evolution of the Popocatépetl Volcanic Complex: constraints on periodic edifice construction and destruction by sector collapse. *Journal of the Geological Society*. https://doi.org/10.1144/jgs2021-022

Guffanti, M., Mayberry, G. C., Casadevall, T. J. & Wunderman, R. (2009). Volcanic hazards to airports. *Natural hazards*, *51*(2), 287-302.

Han, H., Shi, B. & Zhang, L. (2021). Prediction of landslide sharp increase displacement by SVM with considering hysteresis of groundwater change. *Engineering Geology*, *280*, 105876. https://doi.org/https://doi.org/10.1016/j.enggeo.2020.105876

Hastie, T., Tibshirani, R. & Friedman, J. (2017). *The elements of statistical learning: Data mining, inference, and prediction*. Springer.

Ho, T. K. (1995). Random decision forests. *Proceedings of 3rd international conference on document analysis and recognition*, *1*, 278-282.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, *24*(6), 417-441.

James, G., Witten, D., Hastie, T. & Tibshirani, R. (2013). *An Introduction to Statistical Learning: with Applications in R*. Springer.

Jolliffe, I. (2005). Principal Component Analysis. *Encyclopedia of Statistics in Behavioral Science*. John Wiley & Sons, Ltd. https://doi.org/https://doi.org/10.1002/0470013192.bsa501

Kawakatsu, H. & Yamamoto, M. (2015). Volcano seismology. *Earthquake Seismology*, 389-419.

Köhler, A., Ohrnberger, M. & Scherbaum, F. (2010). Unsupervised pattern recognition in continuous seismic wavefield records using self-organizing maps. *Geophysical Journal International*, *182*(3), 1619-1630.

Kong, Q., Trugman, D. T., Ross, Z. E., Bianco, M. J., Meade, B. J. & Gerstoft, P. (2019). Machine learning in seismology: Turning data into insights. *Seismological Research Letters*, *90*(1), 3-14.

Kong, X., Che, X., Su, R., Zhang, C., Yao, Q. & Shi, X. (2018). A new technique for rapid assessment of eutrophication status of coastal waters using a support vector machine. *Journal of Oceanology and Limnology*, *36*(2), 249-262.

Langer, H., Falsaperla, S., Powell, T. & Thompson, G. (2006). Automatic classification and a-posteriori analysis of seismic event identification at Soufriere Hills volcano, Montserrat. *Journal of volcanology and geothermal research*, *153*(1-2), 1-10.

Lara, P., Espinoza, E., Fernandes, C., Rolim, A., Inza, A., Mars, J., Métaxian, J.-P., Dalla Mura, M. & Malfante, M. (2020). Automatic multichannel volcano-seismic classification using machine learning and EMD. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *13*, 1322-1331.

Maggi, A., Ferrazzini, V., Hibert, C., Beauducel, F., Boissier, P. & Amemoutou, A. (2017). Implementation of a multistation approach for automated event classification at Piton de la Fournaise volcano. *Seismological Research Letters*, *88*(3), 878-891.

Malfante, M., Mars, J. & Dalla-Mura, M. (2018). malfante/AAA: v1.0.0. https://doi.org/10.5281/zenodo.1216028

Malfante, M. (2018). *Automatic classification of natural signals for environmental monitoring* (Theses 2018GREAU025). Université Grenoble Alpes. https://tel.archives-ouvertes.fr/tel-01944587

Malfante, M., Dalla Mura, M., Mars, J. I., Métaxian, J.-P., Macedo, O. & Inza, A. (2018). Automatic Classification of Volcano Seismic Signatures. *Journal of Geophysical Research: Solid Earth*, *123*(12), 10, 645-10, 658. https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018JB015470

Matoza, R., Arciniega-Ceballos, A., Sanderson, R. W., Mendo-Pérez, G., Rosado-Fuentes, A. & Chouet, B. A. (2019). High-Broadband Seismoacoustic Signature of Vulcanian Explosions at Popocatépetl Volcano, Mexico. *Geophysical Research Letters*, *46*(1), 148-157.

Matoza, R. S., Arciniega-Ceballos, A., Sanderson, R. W., Mendo-Pérez, G., Rosado-Fuentes, A. & Chouet, B. A. (2019). High-Broadband Seismoacoustic Signature of Vulcanian Explosions at Popocatépetl Volcano, Mexico. *Geophysical Research Letters*, *46*(1), 148-157. https://doi.org/https://doi.org/10.1029/2018GL080802

Matoza, R. S., Chouet, B. A., Dawson, P. B., Shearer, P. M., Haney, M. M., Waite, G. P., Moran, S. C. & Mikesell, T. D. (2015). Source mechanism of small long-period events at Mount St. Helens in July 2005 using template matching, phase-weighted stacking, and full-waveform inversion. *Journal of Geophysical Research: Solid Earth*, *120*(9), 6351-6364.

McNutt, S. (1992). Volcanic tremor. *Encyclopedia of earth system science*, *4*, 417-425.

McNutt, S. & Nishimura, T. (2008). Volcanic tremor during eruptions: temporal characteristics, scaling and constraints on conduit size and processes. *Journal of Volcanology and Geothermal Research*, *178*(1), 10-18.

Menard, S. (2002). *Applied logistic regression analysis* (Vol. 106). Sage.

Mendo-Pérez, G., Arciniega-Ceballos, A., Matoza, R. S., Rosado-Fuentes, A., Sanderson, R. W. & Chouet, B. A. (2021). Ground-coupled airwaves template match detection using broadband seismic records of explosive eruptions at Popocatépetl volcano, Mexico. *Journal of Volcanology and Geothermal Research*, *419*, 107378.

Neuberg, J., Luckett, R., Baptie, B. & Olsen, K. (2000). Models of tremor and low-frequency earthquake swarms on Montserrat. *Journal of Volcanology and Geothermal Research*, *101*(1-2), 83-104.

Newhall, C. & Self, S. (1982). The volcanic explosivity index (VEI) an estimate of explosive magnitude for historical volcanism. *Journal of Geophysical Research: Oceans*, *87*(C2), 1231-1238.

Nieto Torres, A. & Martin del Pozzo, A. L. (2017). Actividad reciente en el Popocatépetl 1993-2016. En J. Yamamoto Victorio (Ed.), *Memoria técnica del mapa de peligros del volcán Popocatépetl*. Unidad de apoyo editorial, Instituto de Geofísica.

Novelo-Casanova, D. & Martínez-Bringas, A. (2005). A seismic attenuation zone below Popocatépetl volcano inferred from coda waves of local earthquakes. *Geofísica internacional*, *44*(2), 177-186.

Ohrnberger, M. (2001). *Continuous automatic classification of seismic signals of volcanic origin at Mt. Merapi, Java, Indonesia* (Tesis doctoral). Potsdam, Univ., Diss., 2001.

Pearson, K. (1901). Principal components analysis. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, *6*(2), 559.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. & Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*, 2825-2830.

Quezada-Reyes, A., Lesage, P., Valdés-González, C. & Perrier, L. (2013). An analysis of the seismic activity of Popocatépetl volcano, Mexico, associated with the eruptive period of December 2002 to February 2003: looking for precursors. *Geological Society of America Special Papers*, *498*, 89-106.

Quinlan, J. R. (1986). Induction of decision trees. *Machine learning*, *1*(1), 81-106.

Rish, I. y col. (2001). An empirical study of the naive Bayes classifier. *IJCAI 2001 workshop on empirical methods in artificial intelligence*, *3*, 41-46.

Shapiro, N., Singh, S., Iglesias-Mendoza, A., Cruz-Atienza, V. & Pacheco, J. (2000). Evidence of low Q below Popocatépetl volcano, and its implication to seismic hazard in Mexico City. *Geophysical Research Letters*, *27*(17), 2753-2756.

Shelly, D. R., Beroza, G. C. & Ide, S. (2007). Non-volcanic tremor and low-frequency earthquake swarms. *Nature*, *446*(7133), 305-307.

Siebe, C., Arana, L. & Macías, J. L. (2017). Inicio de la construcción del Popocatépetl actual o joven: la erupción pliniana "Pómez Blanca"Tochimilco, la de mayor magnitud de los últimos 23,500 años A.P. En J. Yamamoto Victorio (Ed.), *Memoria técnica del mapa de peligros del volcán Popocatépetl*. Unidad de apoyo editorial, Instituto de Geofísica.

Spearman, C. (1904). The proof and measurement of association between two things. *The American journal of psychology*, *100*(3/4), 441-471.

Stephens, C. & Chouet, B. (2001). Evolution of the December 14, 1989 precursory long-period event swarm at Redoubt Volcano, Alaska. *Journal of Volcanology and Geothermal Research*, *109*(1-3), 133-148.

Tameguri, T., Iguchi, M. & Ishihara, K. (2002). Mechanism of explosive eruptions from moment tensor analyses of explosion earthquakes at Sakurajima volcano, Japan. *Bulletin of the Volcanological Society of Japan*, *47*(4), 197-215.

Thurber, C. (1993). Local earthquake tomography: velocities and V_p/V_s-Theory. *Seismic tomography: theory and practice*, 563-583.

Titos, M., Bueno, A., García, L., Benítez, M. C. & Ibañez, J. (2018). Detection and classification of continuous volcano-seismic signals with recurrent neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, *57*(4), 1936-1948.

Wu, T.-F., Lin, C.-J. & Weng, R. (2003). Probability estimates for multi-class classification by pairwise coupling. *Advances in Neural Information Processing Systems*, *16*.

Zhu, J., Li, S. & Song, J. (2021). Magnitude Estimation for Earthquake Early Warning with Multiple Parameter Inputs and a Support Vector Machine. *Seismological Research Letters*. https://doi.org/10.1785/0220210144

Zhu, W. & Beroza, G. (2019). PhaseNet: a deep-neural-network-based seismic arrival-time picking method. *Geophysical Journal International*, *216*(1), 261-273.

Zhu, W., Mousavi, M. & Beroza, G. (2019). Seismic signal denoising and decomposition using deep neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, *57*(11), 9476-9488.

Zobin, V. (2016). *Introduction to volcanic seismology* (3rd edition). Elsevier.

Zobin, V. & Martínez, A. (2010). Quantification of the 1998–1999 explosion sequence at Popocatépetl volcano, Mexico. *Journal of volcanology and geothermal research*, *194*(4), 165-173.