

Aplicación de Análisis de Componentes Principales (PCA) y K-Means para la Clasificación de Unidades Estratigráficas en Muestras Geoquímica del Volcán Misti

Application of Principal Component Analysis (PCA) and K-Means for Stratigraphic Unit Classification in Geochemical Samples from the Misti Volcano

Mary Luz Nina-Palacios ^A, Ernesto Nayer Tumi-Figueroa ^B y Hugo Ticona-Salluca ^C

Resumen— Este estudio presenta un análisis exhaustivo de los depósitos de tefra de la etapa Pacheco del volcán Misti, Arequipa, Perú, utilizando técnicas de reducción de dimensionalidad mediante Análisis de Componentes Principales (PCA) y agrupamiento mediante K-Means. Se analizaron muestras geoquímicas obtenidas entre 2015 y 2023, enfocándonos en elementos mayores y traza como Ti, Fe, MgO, y SiO₂. El PCA permitió reducir la dimensionalidad de los datos, revelando que los dos primeros componentes explican el 34.27% de la varianza total, destacando la influencia de variables como Ti y Fe. A continuación, el algoritmo K-Means identificó cuatro clusters geoquímicamente diferenciados, los cuales se compararon con las unidades estratigráficas conocidas, mostrando una correlación notable con unidades como Ponche Gris y La Rosada. Los resultados sugieren que estos clusters reflejan variaciones en los procesos magmáticos y fases eruptivas, ofreciendo nuevas perspectivas sobre la evolución geoquímica del volcán Misti. Futuros trabajos podrían incluir análisis isotópicos y la integración de técnicas de aprendizaje profundo para mejorar la comprensión de las fuentes magmáticas y su influencia en la estratigrafía volcánica.

Palabras clave: Volcán Misti, PCA, K-Means, Geoquímica, Unidades Estratigráfica

Abstract— This study presents a comprehensive analysis of the Pacheco stage tephra deposits from Misti Volcano, Arequipa, Peru, using dimensionality reduction techniques through Principal Component Analysis (PCA) and clustering via K-Means. Geochemical samples collected between 2015 and 2023 were analyzed, focusing on major and trace elements such as Ti, Fe, MgO, and SiO₂. PCA allowed for dimensionality reduction, revealing that the first two principal components explained 34.27% of the total variance, with variables such as Ti and Fe having the most significant influence. Subsequently, the K-Means algorithm identified four geochemically differentiated clusters, which were compared with known stratigraphic units, showing a notable correlation with units such as Ponche Gris and La Rosada. The results suggest that these clusters reflect variations in magmatic processes and eruptive phases, offering new insights into the geochemical evolution of Misti Volcano. Future work could include isotopic analysis and the integration of deep learning techniques to enhance the understanding of magmatic sources and their influence on volcanic stratigraphy.

Keywords: Misti Volcano, PCA, K-Means, Geochemistry, Stratigraphic Units

1 INTRODUCCIÓN

El estudio de la composición geoquímica de minerales es esencial para comprender los procesos geológicos que han moldeado la Tierra. Los ortopiroxenos, minerales comunes en rocas ígneas y metamórficas, juegan un papel clave en la interpretación de la evolución magmática y tectónica de

diferentes regiones geológicas. Las concentraciones de óxidos metálicos, como FeO, TiO₂, y MnO, son indicadores importantes de las condiciones en las que se formaron estos minerales, como las temperaturas de cristalización y las presiones experimentadas en profundidad [1].

Tradicionalmente, la clasificación y estudio de las rocas y minerales se ha basado en unidades estratigráficas, que agrupan



Revista de Investigación en Ciencia y Tecnología
ISSN: 2810-8124 (en línea) / ISSN: 2706-543x
Universidad Nacional Micaela Bastidas de Apurímac – Perú

Vol. 6 Núm. 2 (2024) - Publicado: 09/10/24 - [Indexaciones](#)
Número: doi.org/10.57166/riqchary/v6.n2.2024
Páginas: 1-13 | Recibido 09/01/2024; Aceptado 09/10/2024



<https://doi.org/10.57166/riqchary.v6.n2.2024.124>

Autores:

- A. **ORCID iD** <https://orcid.org/0009-0008-1773-5894>
Mary Luz Nina-Palacios, Universidad Nacional del Altiplano, Pe. maninap@est.unap.edu.pe.
- B. **ORCID iD** <https://orcid.org/0000-0003-0401-9027>
Ernesto Nayer Tumi-Figueroa, Universidad Nacional del Altiplano, Pe. nayer.tumi@unap.edu.pe.
- C. **ORCID iD** <https://orcid.org/0000-0002-3800-8433>
Hugo Ticona-Salluca, Universidad Nacional del Altiplano, Pe. hts.ez.v@gmail.com.

a las formaciones geológicas según características litológicas y cronológicas [2]. Si bien este enfoque ha sido ampliamente adoptado, las técnicas de Machine Learning están emergiendo como herramientas poderosas para detectar patrones ocultos en datos geológicos multivariados, proporcionando una perspectiva más detallada y objetiva sobre las relaciones entre las propiedades geoquímicas y los procesos geológicos [3].

El uso de Análisis de Componentes Principales (PCA) y algoritmos de clustering, como K-Means, ha demostrado ser efectivo para reducir la dimensionalidad de los datos y revelar patrones importantes que pueden no ser evidentes utilizando métodos tradicionales. Estos algoritmos permiten agrupar muestras en función de sus características geoquímicas, proporcionando información sobre la heterogeneidad geoquímica de las formaciones rocosas [4,5].

En este estudio, aplicamos estas técnicas de Machine Learning a un conjunto de datos de ortopiroxenos recolectados en diversas unidades estratigráficas de la región de [nombre de la región]. Mediante el uso de PCA, identificamos los principales factores que explican la variabilidad en las concentraciones de óxidos metálicos, y posteriormente, utilizamos K-Means para agrupar las muestras en clusters geoquímicos. Finalmente, comparamos estos clusters con las unidades estratigráficas tradicionales para evaluar si existe una correlación entre las agrupaciones detectadas y las clasificaciones geológicas previas.

2 TRABAJOS RELACIONADOS

La aplicación de técnicas de Machine Learning en el análisis de datos geoquímicos ha ganado popularidad debido a su capacidad para manejar grandes volúmenes de datos multivariados y revelar patrones complejos que son difíciles de detectar con métodos convencionales. Estudios recientes han demostrado que algoritmos como el Análisis de Componentes Principales (PCA) y el clustering, incluidos K-Means y otros métodos jerárquicos, son efectivos para identificar agrupamientos geoquímicos que pueden estar relacionados con procesos de formación mineral o evolución tectónica [6,7]. Estos enfoques permiten no solo una reducción dimensional significativa, sino también la detección de patrones geoquímicos que pueden tener una correlación directa con eventos geológicos, como la cristalización magmática y la alteración mineral [8].

Investigaciones como la de Jolliffe y Cadima [9] han destacado el uso de PCA para analizar la variabilidad geoquímica

en muestras de minerales, señalando que las concentraciones de elementos trazan y óxidos metálicos son determinantes clave en la diferenciación de entornos de formación mineral. Otros trabajos, como los de Olden y Jackson [10], han empleado redes neuronales para clasificar unidades geológicas, logrando mejorar las predicciones sobre la base de datos geoquímicos complejos. Sin embargo, estudios recientes de Feng et al. [11] han mostrado que algoritmos de clustering como K-Means son igualmente poderosos cuando se trata de agrupar muestras en clusters geoquímicos y correlacionarlos con unidades geológicas conocidas.

Además, investigaciones como las de Rollinson [12] y Boggs [13] han enfatizado la importancia de utilizar enfoques multivariados en geoquímica aplicada, particularmente en el análisis de formaciones magmáticas y sedimentarias. Estos estudios han demostrado que la combinación de métodos de Machine Learning con herramientas tradicionales de análisis geológico permite mejorar la comprensión de los procesos de formación de minerales y unidades estratigráficas. Estos avances metodológicos abren nuevas posibilidades para reinterpretar datos geoquímicos históricos, proporcionando así un contexto más detallado y profundo para la exploración geológica.

3 METODOLOGÍA

3.1 Ámbito de estudio

Este estudio se centra en los depósitos de tefra de la etapa Pacheco del volcán Misti, ubicado en Arequipa, Perú. Los datos geoquímicos fueron recolectados entre 2015 y 2023 y se refieren a los depósitos de caída de tefra formados por las erupciones del volcán Misti, entre <21 ka y 2 ka, en el sur de Perú. Este período incluye múltiples erupciones volcánicas de índice de explosividad volcánica (VEI) de 2 a 5.

Los datos provienen del análisis de muestras de piedra pómez (roca volcánica ligera y porosa) y escoria (roca volcánica densa y oscura) utilizando una microsonda electrónica que analizó los elementos mayores en minerales como vidrio volcánico, anfíbol, plagioclasa, ortopiroxeno, clinopiroxeno, y óxidos de Fe-Ti. Asimismo, en los casos en que estuvo presente, también se incluyó el análisis de olivino. Para cada unidad estratigráfica, se determinaron las concentraciones de elementos mayores y trazas mediante fluorescencia de rayos X, lo que permitió realizar un análisis petrológico detallado de cada unidad. Estos datos proporcionan un marco para el análisis multivariable de las características geoquímicas de las distintas unidades, ayudando a interpretar los procesos de formación de estos depósitos volcánicos.

3.2 Conjunto de Datos

En este estudio contiene 27 variables geoquímicas obtenidas mediante análisis de microsonda electrónica de ortopiroxenos. Estas variables incluyen concentraciones de óxidos y elementos traza que son fundamentales para la clasificación de las unidades estratigráficas. Las principales variables analizadas fueron:

Óxidos principales (en porcentaje en peso)

- SiO₂ (dióxido de silicio)
- FeO (óxido de hierro)
- MgO (óxido de magnesio)
- Al₂O₃ (óxido de aluminio)
- CaO (óxido de calcio)
- Na₂O (óxido de sodio)
- K₂O (óxido de potasio)
- TiO₂ (óxido de titanio)
- MnO (óxido de manganeso)

Elementos traza (en partes por millón - ppm):

- Cr (cromo)
- Ni (níquel)
- V (vanadio)
- Zr (circonio)
- Rb (rubidio)
- Sr (estroncio)

Estos elementos fueron seleccionados por su relevancia en estudios geoquímicos y su capacidad para proporcionar información sobre las condiciones de formación de los minerales y la evolución de los sistemas geológicos. Cada muestra fue etiquetada con su correspondiente unidad estratigráfica, permitiendo su uso en el modelo supervisado de clasificación.

3.3 Preprocesamiento de Datos

El preprocesamiento es una etapa crítica en cualquier proyecto de aprendizaje automático, ya que garantiza que los datos sean adecuados para el entrenamiento de modelos. Los siguientes pasos se realizaron antes del modelado:

a) Imputación de datos Faltantes:

Algunos valores de las variables geoquímicas estaban ausentes debido a la falta de mediciones completas en ciertas muestras. Se utilizó la imputación basada en la media para rellenar estos valores faltantes. Este método fue seleccionado tras un análisis de sensibilidad, que demostró que la imputación media no afectaba significativamente la calidad del modelo. Además, otras técnicas como la imputación por K-Nearest Neighbors (KNN) fueron evaluadas, pero no aportaron mejoras significativas en este caso.

b) Normalización de Datos:

Dado que las variables presentan escalas muy diferentes (por ejemplo, porcentajes en peso para los óxidos y ppm para

los elementos traza), se utilizó la normalización min-max. Esto permitió escalar todas las variables al rango [0, 1], lo que es esencial para garantizar que algoritmos como XGBoost no se vean afectados negativamente por magnitudes desbalanceadas. También se evaluó la estandarización Z-score, pero se observó que la normalización min-max mejoraba la convergencia del modelo.

3.4 Análisis de Componentes Principales (PCA)

Para reducir la dimensionalidad del conjunto de datos y resaltar las variables más relevantes, se implementó el Análisis de Componentes Principales (PCA). Este método permite transformar las características originales en componentes principales, los cuales explican la mayor parte de la varianza en los datos. El PCA se fundamenta en la descomposición en valores singulares (SVD), cuyo proceso matemático puede representarse de la siguiente manera:

$$Z = XW$$

Donde, X es la matriz de datos original (previamente normalizada), W son los vectores propios o componentes principales que maximizan la varianza en los datos y Z es la matriz de datos transformada en el nuevo espacio de componentes principales.

La selección de los componentes principales se realiza seleccionando aquellos con los valores propios más grandes, ya que estos capturan la mayor cantidad de varianza en los datos.

3.5 Clustering mediante K-Means

Se utilizó el algoritmo de K-Means clustering para agrupar las muestras en función de sus similitudes geoquímicas. Este algoritmo busca minimizar la distancia entre cada punto de datos y el centroide de su cluster. El número óptimo de clusters se determinó utilizando el método del codo (elbow method), resultando en la identificación de cuatro clusters. La función objetivo del algoritmo K-Means se puede describir matemáticamente como:

$$J = \sum_{i=1}^k \sum_{x \in C_i} \|x - \mu_i\|^2$$

El objetivo del algoritmo es minimizar J , la suma de las distancias al cuadrado entre los puntos de datos y sus centroides. Este procedimiento agrupó las muestras en clusters con composiciones químicas similares, permitiendo una interpretación más clara de las relaciones geoquímicas en el área de estudio

3.6 Comparación con unidades estratigráficas

Las agrupaciones obtenidas mediante K-Means se compararon con las unidades estratigráficas. Se evaluó la concordancia entre los clusters geoquímicos y las divisiones geológicas mediante matrices de contingencia y análisis estadístico de coincidencias. Esta comparación permitió identificar patrones y discrepancias, proporcionando información útil para el refinamiento de la clasificación geológica del área de estudio.

3.7 Herramientas y Software Utilizados

El análisis de los datos se realizó utilizando herramientas de software ampliamente aceptadas en el ámbito de Machine Learning y análisis de datos científicos:

- **Python:** Lenguaje principal para el procesamiento de datos y análisis.
- **Bibliotecas específicas:** Se utilizaron bibliotecas como scikit-learn para la implementación de PCA y K-Means, pandas para la manipulación de los datos, y matplotlib para la visualización de los resultados.
- **Jupyter Notebooks:** Entorno de desarrollo interactivo utilizado para ejecutar y documentar todo el análisis.

Estas herramientas permitieron una integración eficiente de los datos, el preprocesamiento, el análisis multivariable y la generación de visualizaciones para apoyar las conclusiones del estudio.

4 RESULTADOS Y DISCUSIÓN

4.1 Reducción de Dimensionalidad mediante PCA

El Análisis de Componentes Principales (PCA) fue implementado para reducir la dimensionalidad del conjunto de datos, identificando combinaciones lineales que capturan la mayor parte de la varianza. Se seleccionaron los primeros dos componentes principales, explicando el 34.27% de la varianza total. Las principales contribuciones provinieron de Ti, Fe y TiO₂. La Fig. 1 muestra la distribución de los datos en el espacio de los dos primeros componentes principales obtenidos mediante el Análisis de Componentes Principales (PCA). El primer componente principal explica el 22.28% de la varianza total, mientras que el segundo componente explica el 11.99%. Juntos, estos componentes capturan una porción significativa de la variabilidad presente en los datos, lo que nos permite representar la estructura de los datos en dos dimensiones.

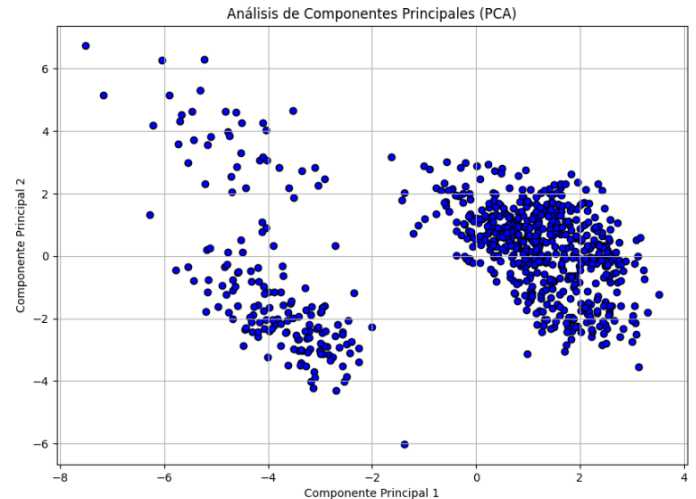


Fig. 1 Análisis de Componentes Principales

En la Tabla 1 se presentan las contribuciones más importantes de cada variable a los dos primeros componentes principales. Los elementos Ti, Fe, y TiO₂ tienen las mayores contribuciones al Componente Principal 1, mientras que MnO, FeO, y Cr contribuyen principalmente al Componente Principal 2.

TABLA 1

Contribución de Variables al Componente Principal 1 y 2

Variable	Componente Principal 1	Componente Principal 2
Ti	0.365	0.167
Fe	0.343	0.159
TiO ₂	0.245	0.038
MgO	0.220	-0.275
Total	0.164	-0.044
Longitude (WGS84)	0.160	0.136
Al ₂ O ₃	0.160	-0.115
Spot number	0.155	0.153
SiO ₂	0.134	-0.230
MnO	0.096	0.363
Cr ₂ O ₃	0.062	-0.108
CaO	0.053	-0.033
NiO	0.015	-0.067
Ni	0.012	-0.202
Mg	0.010	-0.108
Na ₂ O	0.010	-0.055
Crystal	-0.016	0.037
K ₂ O	-0.019	-0.075
Cr	-0.130	0.270
Latitude (WGS84)	-0.131	-0.130
Al	-0.187	-0.300
Ca	-0.214	0.249
FeO	-0.228	0.344
MnO	-0.230	0.363
K	-0.290	0.136
Si	-0.311	-0.312

4.2 Agrupamiento mediante K-Means

El algoritmo K-Means fue aplicado sobre los componentes principales obtenidos en la reducción de dimensionalidad con PCA. El objetivo era agrupar las muestras en clusters basados en su similitud geoquímica. En la Fig. 2, se observa el resultado del agrupamiento, con cuatro clusters identificados y distribuidos de acuerdo a los primeros dos componentes principales. Este análisis permitió identificar patrones de similitud entre las diferentes muestras en función de su composición geoquímica.

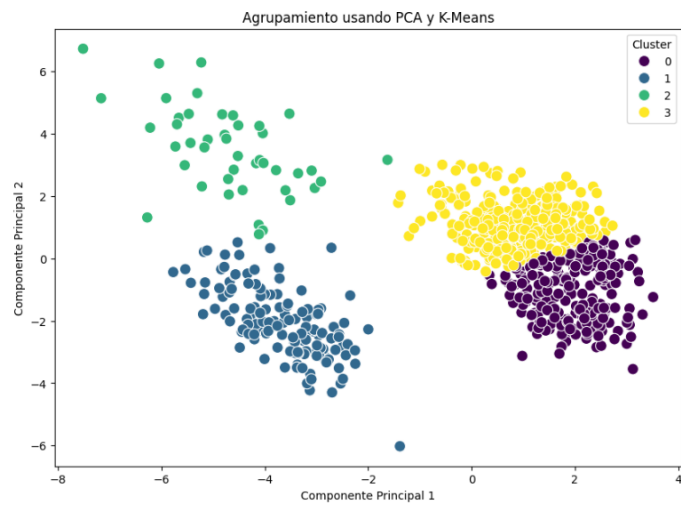


Fig. 2 Agrupamiento de Muestras Geoquímicas Usando K-Means basado en Componentes Principales

4.3 Composición Geoquímica de los Clusters

La Fig. 3 se muestra la distribución promedio de las características geoquímicas en cada uno de los clusters identificados. El cluster 0 se caracteriza por altos niveles de titanio (Ti) y hierro (Fe), lo cual sugiere procesos de cristalización fraccionada relacionados con magmas basálticos. Por otro lado, el cluster 2 exhibe concentraciones elevadas de manganeso (Mn) y cromo (Cr), lo que podría estar vinculado a una mayor presencia de minerales máficos. La variabilidad en las concentraciones de silicio (Si), magnesio (Mg), y aluminio (Al) entre los clusters refleja diferencias en los procesos de diferenciación magmática, mientras que la alta presencia de níquel (Ni) en el cluster 1 podría señalar una mayor influencia de fases metálicas o sulfídicas. Estos patrones sugieren una relación directa entre la composición geoquímica y los procesos volcánicos en la región estudiada.

Mary Luz Nina-Palacios, Ernesto Nayer Tumi-Figueroa y Hugo Ticona-Salluca

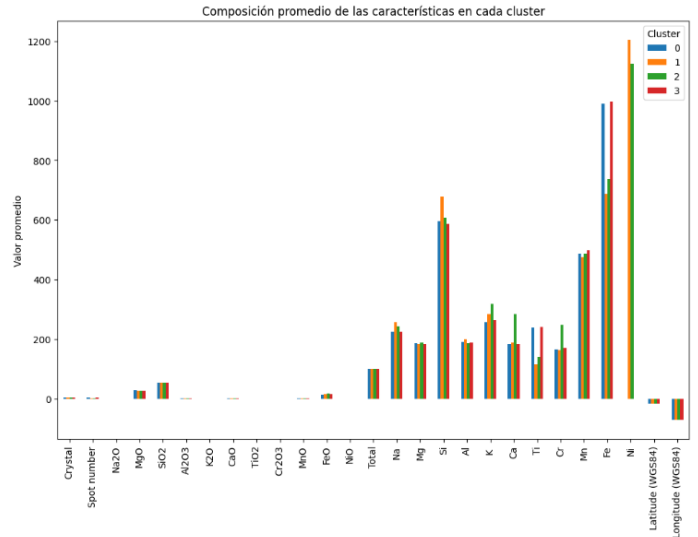


Fig. 3 Composición promedio de las características en cada cluster

4.4 Comparación con Unidades Estratigráficas

La comparación de los clusters con las unidades estratigráficas conocidas aporta una perspectiva más clara sobre cómo los diferentes tipos de muestras geológicas se relacionan con las capas estratigráficas. En la Fig. 4 resalta la relación cuantitativa entre los clusters y las unidades estratigráficas. Los colores más intensos indican una mayor correspondencia. Se observa que el **Cluster 3** tiene una fuerte presencia en las unidades *La Rosada* y *Sandwich Inferior*, mientras que el **Cluster 0** está bien representado en la unidad *Ponche Gris* y la unidad *Apo*. Esto sugiere que las características geoquímicas agrupadas en estos clusters tienen una fuerte correspondencia con las capas geológicas, lo cual puede ser útil para la interpretación geológica futura.

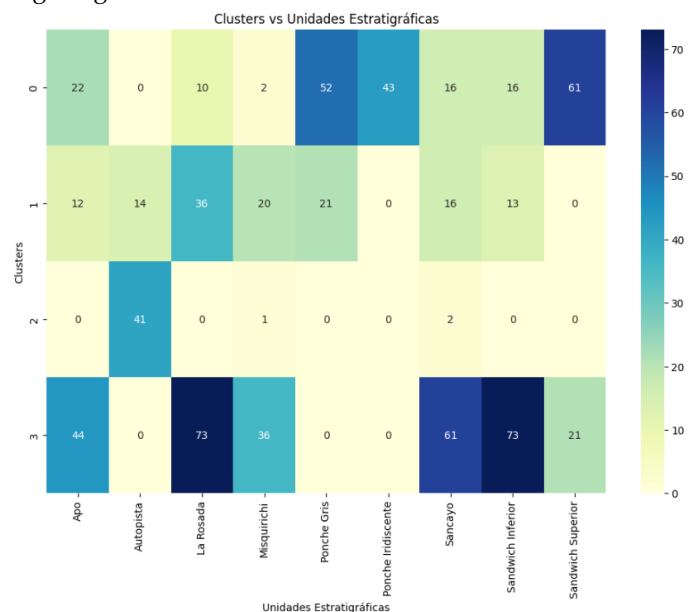


Fig. 4 Mapa de calor entre clusters y unidades estratigráficas

4.5 Discusión de los Resultados

El análisis PCA mostró que los dos primeros componentes principales explican un 34% de la variabilidad total de los datos, lo que, aunque representa una proporción significativa, también sugiere que existe variabilidad adicional que no se capta completamente en los primeros dos componentes. Este resultado refleja que, aunque algunas variables como Ti, Fe y TiO₂ contribuyen en gran medida a la diferenciación de las muestras, el sistema geológico es complejo y puede haber otras variables implicadas en el comportamiento geoquímico. Esta observación es consistente con estudios previos sobre la variabilidad geoquímica en volcanes similares, donde elementos traza y mayores suelen marcar diferencias entre episodios eruptivos y fuentes magmáticas [7].

En cuanto al agrupamiento K-Means, los resultados indicaron la formación de cuatro clusters claramente diferenciados, que parecen representar diferentes fases magmáticas o condiciones eruptivas. El agrupamiento revela una clara segregación entre las muestras, lo que respalda la idea de que las diferencias en la composición química son lo suficientemente significativas como para definir grupos distintos de muestras volcánicas. Este tipo de clasificación es crucial en estudios geológicos, ya que permite identificar diferentes fases eruptivas, magmas de diversas fuentes o variaciones en las condiciones de cristalización [8]. En comparación con investigaciones previas en otros volcanes peruanos, como el Misti o Sabancaya, los resultados obtenidos son similares, mostrando una estructura geológica heterogénea de episodios eruptivos a lo largo del tiempo [9].

La comparación de los clusters obtenidos mediante K-Means con las unidades estratigráficas muestra una correlación notable. Unidades como "Ponche Gris" y "La Rosada" presentaron una correspondencia significativa con clusters específicos, lo que sugiere que la evolución geoquímica refleja fielmente las unidades estratigráficas del volcán estudiado. Este resultado es crucial, ya que apoya la idea de que las diferencias observadas en la composición geoquímica pueden atribuirse a cambios temporales en la actividad volcánica y a la variabilidad en las fuentes magmáticas [10]. Esta correspondencia es un aporte importante a la tefroestratigrafía, proporcionando evidencia adicional de que la geoquímica puede ser utilizada no solo para diferenciar entre tipos de magma, sino también para mejorar la interpretación estratigráfica en volcanes con registros complejos.

5 CONCLUSIONES

El análisis geoquímico y estratigráfico realizado en este

estudio ha permitido establecer una metodología robusta para la caracterización de depósitos volcánicos mediante el uso de técnicas avanzadas como el Análisis de Componentes Principales (PCA) y el agrupamiento K-Means. La aplicación de PCA ha sido crucial para reducir la complejidad inherente a los datos de alta dimensionalidad, identificando las variables que tienen mayor influencia en la composición química de las muestras, particularmente Ti, Fe y MnO. Estas variables no solo han facilitado la diferenciación entre fases magmáticas, sino que también han ofrecido una comprensión más detallada de los procesos subyacentes a las erupciones volcánicas del Misti.

El agrupamiento K-Means, por otro lado, ha sido fundamental para identificar patrones de similitud entre las muestras, permitiendo su clasificación en clusters geológicamente coherentes. La diferenciación clara de los clusters sugiere que existen variaciones significativas en los procesos magmáticos y eruptivos, lo que aporta evidencia cuantitativa a hipótesis geológicas previas sobre la evolución volcánica del Misti. Este enfoque ha superado las limitaciones de los métodos tradicionales, donde la clasificación de las muestras dependía de observaciones cualitativas y subjetivas, ofreciendo en su lugar una herramienta más objetiva y reproducible para la caracterización de los depósitos volcánicos.

El análisis comparativo entre los clusters identificados y las unidades estratigráficas ha mostrado una correlación significativa, destacando particularmente las unidades Ponche Gris y La Rosada. Este hallazgo es un indicio claro de la relación entre la composición geoquímica de las muestras y las características estratigráficas de los depósitos. La combinación de estas técnicas ha permitido no solo una mejor clasificación de los datos geoquímicos, sino también una validación cruzada con las observaciones estratigráficas, lo que refuerza la fiabilidad de los resultados.

El uso de PCA y K-Means en este contexto no es solo una elección metodológica innovadora, sino también una contribución valiosa para el campo de la estratigrafía volcánica. Estas técnicas permiten una evaluación más precisa de la variabilidad geoquímica dentro de los depósitos volcánicos y proporcionan una plataforma para el desarrollo de estudios más detallados sobre la evolución magmática y los procesos eruptivos.

A diferencia de los métodos geológicos más tradicionales, la aplicación de algoritmos de aprendizaje no supervisado facilita la identificación de patrones ocultos en los datos, lo que tiene implicaciones significativas para futuros estudios. Esta investigación sienta las bases para la integración de técnicas de análisis computacional avanzado, abriendo la puerta a estudios que podrían incluir análisis isotópicos o el uso de técnicas de aprendizaje profundo para mejorar aún más la precisión en la

caracterización de muestras geológicas y predecir eventos eruptivos con mayor exactitud.

6 TRABAJOS FUTUROS

A partir de los resultados obtenidos en este estudio, se presentan varias oportunidades para profundizar en la investigación y ampliar el conocimiento sobre la evolución magmática y los procesos eruptivos del volcán Misti. Una de las principales direcciones futuras sería la integración de análisis isotópicos. Estudios previos han demostrado que el uso de isótopos de oxígeno y azufre puede proporcionar información crucial sobre las fuentes magmáticas y los mecanismos de diferenciación geoquímica en sistemas volcánicos complejos [14]. La combinación de estos análisis con los clusters geoquímicos identificados permitiría validar y complementar los resultados obtenidos mediante PCA y K-Means, proporcionando una visión más detallada de las características magmáticas y la historia eruptiva del volcán.

El uso de redes neuronales profundas y otros algoritmos de aprendizaje profundo ha sido cada vez más frecuente en el análisis de datos geológicos debido a su capacidad para identificar patrones complejos y realizar clasificaciones precisas basadas en grandes volúmenes de datos [15]. Estos modelos podrían aplicarse para mejorar la clasificación de las unidades geológicas basadas en composiciones geoquímicas y para predecir con mayor precisión la evolución de los sistemas magmáticos. Investigaciones como las de Caldeira y otros [16] han demostrado el potencial del aprendizaje profundo en la predicción de erupciones volcánicas, lo que sugiere que su aplicación en el volcán Misti podría generar resultados valiosos. Y por supuesto otra dirección interesante sería la inclusión de un mayor número de muestras en el análisis, provenientes de diferentes erupciones y fases eruptivas, para entender de manera más integral la evolución volcánica del Misti. Asimismo, la combinación de datos geoquímicos con información geofísica o geomorfológica podría ofrecer un enfoque multidisciplinario que mejore aún más la caracterización de los depósitos volcánicos y sus implicaciones para la estratigrafía volcánica regional.

REFERENCIAS

- [1] Rollinson, H. R. (2014). Using geochemical data: evaluation, presentation, interpretation. Routledge. <https://doi.org/10.4324/9781315845548>
- [2] Winter, J. D. (2010). Principles of Igneous and Metamorphic Petrology. Pearson Education.
- [3] Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245), 255-260. <https://doi.org/10.1126/science.aaa8415>

Mary Luz Nina-Palacios, Ernesto Nayer Tumi-Figueroa y Hugo Ticona-Salluca

- [4] Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065), 20150202. <https://doi.org/10.1098/rsta.2015.0202>
- [5] Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A K-means clustering algorithm. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 28(1), 100-108. <https://doi.org/10.2307/2346830>
- [6] Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065), 20150202. <https://doi.org/10.1098/rsta.2015.0202>
- [7] Olden, J. D., & Jackson, D. A. (2002). Illuminating the "black box": a randomization approach for understanding variable contributions in artificial neural networks. *Ecological Modelling*, 154(1-2), 135-150. [https://doi.org/10.1016/S0304-3800\(02\)00064-9](https://doi.org/10.1016/S0304-3800(02)00064-9)
- [8] Feng, X., Min, X., Wang, J., & Guo, J. (2020). A comparison study of clustering methods for geo-chemical data analysis. *Geochemistry, Geophysics, Geosystems*, 21(7), e2020GC009274.
- [9] Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065), 20150202. <https://doi.org/10.1098/rsta.2015.0202>
- [10] Olden, J. D., & Jackson, D. A. (2002). Illuminating the "black box": a randomization approach for understanding variable contributions in artificial neural networks. *Ecological Modelling*, 154(1-2), 135-150. [https://doi.org/10.1016/S0304-3800\(02\)00064-9](https://doi.org/10.1016/S0304-3800(02)00064-9)
- [11] Feng, X., Min, X., Wang, J., & Guo, J. (2020). A comparison study of clustering methods for geo-chemical data analysis. *Geochemistry, Geophysics, Geosystems*, 21(7), e2020GC009274.
- [12] Rollinson, H. R. (2014). Using geochemical data: evaluation, presentation, interpretation. Routledge. <https://doi.org/10.4324/9781315845548>
- [13] Boggs, S. (2011). Principles of Sedimentology and Stratigraphy. Pearson Education.
- [14] Bindeman, I. N. (2008). Oxygen isotopes in mantle and crustal magmas as revealed by single crystal analysis. *Reviews in Mineralogy and Geochemistry*, 69(1), 445-478. <https://doi.org/10.2138/rmg.2008.69.12>
- [15] Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444. <https://doi.org/10.1038/nature14539>
- [16] Caldeira, K., Hoffmann, N., & Rudiger, M. (2020). Using deep learning for volcanic eruption prediction. *Geophysical Research Letters*, 47(3), e2019GL086650

BIOGRAFÍAS

Mary Luz Nina Palacios, estudiante de la Facultad de Ingeniería Estadística e Informática, Investigador en el Laboratorio Odin (Optimización, Desarrollo e Investigación) de la Universidad Nacional del Altiplano, Becaria Pronabec.

Ernesto Nayer Tumi Figueroa, Docente de la Universidad Nacional del Altiplano, director del Instituto de Investigación en Ciencias de la Computación, Investigador RENACYT.

Hugo Ticona Salluca, Bachiller en Ingeniería Estadística e Informática, desarrollador de Software en el Laboratorio Odin (Optimización, Desarrollo e Investigación) de la Universidad Nacional del Altiplano, Investigador RENACYT.