

Propuesta Inicial GRUPO 8

Análisis de muestreos geoquímicos de suelos superficiales usando métodos no supervisados de aprendizaje automático como estrategia para la prospección de yacimientos minerales en Perú.

Integrantes:

Diana Urbano

Edinson Fernandez

Daren Rodríguez

1. Resumen:

Entre el año 2000 al 2019 se realizaron muestreos geoquímicos en diferentes regiones del Perú, de los cuales se toma una base de datos que corresponde a 536 muestras geoquímicas de suelos superficiales en las que se determinaron concentraciones de 58 elementos químicos y 11 óxidos mayores.

Con este estudio se pretende usar herramientas de aprendizaje automático no supervisado, como análisis de componentes principales y conglomerados aglomerativos para generar agrupaciones de elementos indicadores “pathfinders” asociados a yacimientos minerales de interés como: Oro, Plata, Cobre entre otros. Estas agrupaciones facilitan la delimitación de nuevas zonas de prospección minera o nuevos “Target” que puedan ser visualizadas espacialmente y tener una idea de la abundancia, la posible distribución o migración de elementos asociados a minerales que facilitan la detección de yacimientos metálicos.

2. Introducción:

La geoquímica es una disciplina que trabaja en forma conjunta el conocimiento y la comprensión de los principios y procesos de la química con la geología y encuentra relaciones para resolver problemas teóricos y aplicados. (Siegel, 1992)

La base de datos a analizar corresponde a 536 muestras geoquímicas de suelos superficiales efectuados en Perú. Para cada muestra se determinaron concentraciones de 58 elementos químicos y 11 óxidos mayores. Con este estudio se pretende generar una delimitación por agrupaciones asociadas a estos elementos químicos y óxidos, usando métodos no supervisados como análisis de componentes principales y conglomerados aglomerativos de manera que se obtengan distribuciones espaciales que faciliten la visualización de áreas con minerales de interés. Esta información es útil en la prospección minera y permite tener una mejor comprensión del territorio

al reducir las áreas de exploración a sectores más puntuales, además históricamente han demostrado ser herramientas rápidas y económicas.

El análisis de los muestreos geoquímicos permite tener una idea de la abundancia, la posible distribución o migración de elementos asociados a minerales que facilitan la detección de yacimientos metálicos.

Muchos de estos elementos se conocen como “pathfinders” debido a que son elementos que comúnmente tienden a indicarnos la presencia de algún yacimiento como se observa en la Tabla 1.

Deposits of Interest	Type of Deposit	Main Pathfinder Minerals	Main Pathfinder Elements
Gold		Pyrite, chalcoppyrite, arsenopyrite, bismuthinite magnetite, tellurides, tetrahedrite, pyrite, sphalerite, muscovite, monazite, bastnäsité, quartz, scheelite, wolframite, cassiterite.	Fe, Mn, Cu, Co, Ni, Sb, Zn, As, Bi, Te, Sn, Se, Tl, Ag, Hg, Pb, Mo and W.
	Carbonatite rocks	Bastnäsité group, ancylite, monazite, (fluor)apatite, pyrochlore, xenotime, florencite.	Na, Mg, Fe, P, Ba, F, S, Sr, Ca, Nb, Th, U, Zr, Cu, Ta, Ti, V, Mn, Pb.
	Igneous rocks (including hydrothermal upgrade)	Bastnäsité group, aegirine, eudialyte, loparite, allanite, monazite, fergusonite, zircon, xenotime, fluorapatite, ancylite, gadolinite, euxenite, mosandrite.	Na, K, Fe, Al, Zr, Ti, Nb, Ta, Li, F, Cl, Si, Th, U, P, Cs, Rb, Sn, W, Mo, Be, Ga, Hf, Mn, B.
REE	Placers and palaeoplacers	Monazite, xenotime, allanite, euxenite.	Ti, Nb, Zr, Au, Sn, Th, U, Pb, F.
	Laterites	Monazite, apatite, pyrochlore, crandallite group, bastnäsité group, churchite, rhabdophane, plumbogummite, zircon, florencite, xenotime, cerianite.	Fe, Al, Nb, Zr, Ti, Sn, Mn, P, low Si, negative Ce anomaly.
	Ion-adsorption	Clay minerals (mainly kaolinite and halloysite).	High Si (>75%), low P.
	Iron oxide-associated (including IOCG) deposits	Bastnäsité, synchysite, monazite, xenotime, florencite, britholite.	Fe, Cu, U, Au, Ag, Ba, F, P, S.
	Seafloor deposits, such as manganese nodules, ferromanganese crust, phosphorite.	Vernadite, todorokite, Fe-oxyhydroxide, carbonate fluorapatite, francolite.	Mn, Fe, P, Cu, Ni, Co.
Cu-Ni-PGE		pentlandite, chalcoppyrite, pyrite, millerite, PGM, chromite, Cr-diopside, enstatite, olivine, Cr-andradite.	Ni, Cu, Pd, As, Cr, Co, S, PGE

Volcanogenic massive sulphide (VMS) deposits (Cu, Pb, Zn, Ag, Au)	Galena, sphalerite, chalcoppyrite, pyrrhotite, gold, pyrite, gahnite, staurolite, cassiterite, spessartine, sillimanite, andalusite, beudantite, jarosite, barite, tourmaline, hogcomite, nigerite.	Cu, Zn, Pb, Ag, Mo, Sn, Ba, As, Sb, In, Te, Bi, and Tl
W-Mo-Bi, and Sn-Zn-In deposits	Cassiterite, wolframite, molybdenite, topaz, chalcoppyrite, galena, sphalerite, arsenopyrite, pyrite, loellingite, beudantite, anglesite, plumboferrite, plumbogummite.	Ag, As, Cd, Cu, Pb, Re, Te, Tl
Li	Spodumene, petalite, amblygonite, quartz, K-feldspar, albite, or montebrasite, lepidolite, zinnwaldite, eucryptite, cassiterite, lithiophilite, holmquistite, triphylite, quartz, muscovite, apatite, tourmaline tantalite-columbite, beryl.	K, Ca, Rb, Sr, Y, Nb, Sn, Cs, Ta, Sb, W, Bi, As, Ga, Tl, and the REE
Kimberlite-hosted diamonds	Cr-pyrope, Cr-diopside, eclogitic garnet, Mg-ilmenite, chromite, olivine, diamond.	C
U	Uraninite (pitchblende), thoriantite, tourmaline, sulphides, monazite, allanite, zircon, baddelyite, niccolite, U-Th anatase, U-Th rutile, brannerite, magnetite.	Cu, Ag, As, Cr, Pb, Zn, Ni, Co, Re, Be, P, Mo, Mn, REE and radiogenic Pb isotopes

Tabla 1. Elementos indicadores de la presencia de yacimientos metálicos según los tipos de depósitos. (Balaram y Sawant, 2021)

Algunos estudios han demostrado que puede haber relaciones significativas entre el contenido de ciertos elementos y metales de interés, como lo relaciona Nude, y otros, (2012), encontrando

asociaciones de oro con otros elementos, usando metodos como el analisis de factores y agrupamiento jerarquico, llegando a muy buenos resultados de agrupamiento.

Levitan, y otros, (2015) realizan un estudio similar al anterior pero con el uranio como elemento de interes y usando analisis de componentes principales y conglomerados jerarquicos.

La motivación en este tipo de análisis usando métodos no supervisados como análisis de componentes principales y métodos de conglomerados aglomerativos, nace de buscar las posibles relaciones entre elementos o conjuntos de elementos que nos puedan indicar anomalías, o tendencias en un área o zona de estudio, lo que nos permite tener un área mucho más focalizada siendo este un primer insumo para delimitar las primeras zonas con mayores probabilidades de estar cerca o en un yacimiento metálico y donde se puedan realizar estudios adicionales, como: sondeos, métodos geofísicos o campañas de cartografía geológica entre otras.

Debido a que la base de datos corresponde a información recopilada en Perú, el éxito de estos análisis puede ser de interés para Instituto Geológico, Minero y Metalúrgico del Perú, así como el gobierno y autoridades mineras que otorgan títulos mineros de acuerdo a las características geológicas y prospectivas de cada territorio. Por otra parte, las metodologías pueden ser replicadas en otros contextos geológicos y sirven de herramientas para las universidades, gobiernos, así como empresas privadas sin importar el país de origen, debido a que el éxito o fracaso de estas metodologías depende en gran medida de las condiciones geológicas de cada zona de estudio particular y la ocurrencia natural de yacimientos metálicos.

3. Revisión preliminar de antecedentes en la literatura:

“La exploración minera es un esfuerzo de equipo multidisciplinario que involucra a expertos de diferentes áreas, como la geología, la geofísica, la geoquímica, la petrología y la ingeniería, donde las técnicas geoquímicas, en particular, han contribuido significativamente a los descubrimientos de varios yacimientos minerales. Los datos estadísticos de China revelan que el 71% del total de los depósitos minerales fueron descubiertos por métodos geoquímicos durante el período 1981–2000” (Balaram y Sawant, 2021)

Las anomalías geoquímicas del suelo se pueden utilizar para identificar “Pathfinders” en la exploración de depósitos minerales y apoyarse en métodos no supervisados para facilitar la búsqueda de yacimientos minerales como lo sugieren las siguientes investigaciones:

Levitan, y otros, (2015) realizan el análisis de datos de composición de suelos con métodos estadísticos multivariados usando conglomerados jerárquicos y análisis de componentes principales, para analizar los datos geoquímicos del suelo recopilados del depósito de uranio de Coles Hill, Virginia, EE. UU., para identificar los “Pathfinders” asociados con este depósito.

Un trabajo similar al anterior se realizó por Nude, y otros, (2012) con un análisis estadístico multivariante en datos geoquímicos del suelo de múltiples elementos de los prospectos de oro Koda Hill-Bulenga en el cinturón de oro de Wa-Lawra, al noroeste de Ghana. Los objetivos del estudio fueron definir las relaciones del oro con otros elementos traza para determinar posibles elementos

“Pathfinders” del oro a partir de los datos geoquímicos del suelo. El estudio se centró en siete elementos, a saber, Au, Fe, Pb, Mn, Ag, As y Cu. Se realizaron análisis factoriales y análisis de conglomerados jerárquicos en las muestras analizadas.

4. Descripción detallada de los datos:

4.1 Descarga de los datos

La base de datos corresponde a información de diferentes campañas de campo recopilada en el programa de Prospección Geoquímica realizado en Perú entre los años 2000 y 2019, de la cual se realizará el análisis de los muestreos realizados en suelo.

La información geoquímica es de carácter regional y permite conocer las concentraciones de los elementos mayores y los principales elementos traza, entre los que destacan el Au, Ag, Cu, Mo, Pb, Zn, Cd y Hg, entre otros.

Cada muestra de suelo fue analizada en el laboratorio de INGEMMET y se determinaron concentraciones de 58 elementos químicos y 11 óxidos mayores. Las siglas ppm corresponden a partes por millón y pct corresponde al porcentaje.

La información se encuentra disponible para libre descarga por el Instituto Geológico, Minero y Metalúrgico de Perú en su plataforma GEOCATMIN y pertenecen a la serie B Prospección Geoquímica del Perú.

El archivo **2_BD_Geoquímica de Suelos.rar** corresponde a 4 documentos de Excel con muestras de suelo superficial y suelo profundo en Perú. Para los fines de este trabajo se unificaron los Excel de suelos superficiales en un solo documento y usando solamente la información del laboratorio INGEMMET.

Enlaces de descarga:

<https://www.gob.pe/institucion/ingemmet/informes-publicaciones/1423546-base-de-datos-de-geoquimica>

4.2 Limpieza de datos

La base de datos unificada de suelos superficiales contiene inicialmente 540 registros que posteriormente luego de la limpieza se reducen a 536 registros de muestras Figura 1.

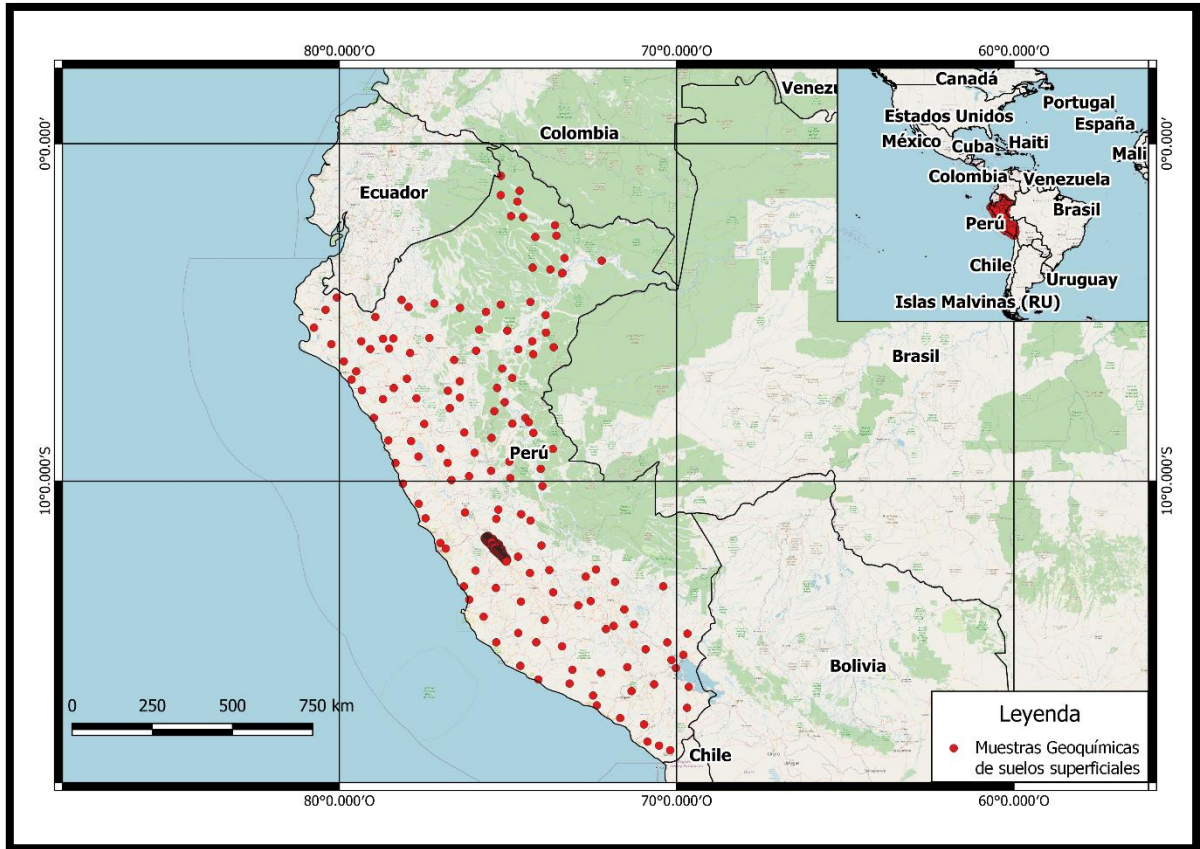


Figura 1. Distribución de las 536 muestras geoquímicas de suelos superficiales en Perú. Fuente Autores.

Para verificar la ubicación espacial de las muestras se realiza una visualización inicial usando el software libre QGIS y con este se verifica que 3 muestras se encuentran por fuera del territorio de Perú y 1 muestra cae sobre el mar, por lo cual se eliminan estos 4 registros.

Algunos símbolos como el “<” se refieren al mínimo de detección del instrumento por ejemplo <0.005 para el oro quiere decir que el instrumento no puede detectar valores de este elemento inferiores a 0.005 partes por millón por lo cual para este y los demás elementos se decide reemplazar estos valores por cero y de esta manera poder usar los datos de forma numérica para los análisis.

A continuación, se describen los métodos de limpieza que se realizaron en los datos inicialmente en la base de datos unificada en Excel ver Tabla 2:

Variables eliminadas sin valores registrados(N.R.)	Variables eliminadas solo un valor valido	Valores reemplazados por cero (corresponden al valor mínimo de detección del instrumento)	Símbolos eliminados	Registros completos eliminados (filas)
Au_ppb, Hg_ppm, Ag_ppb, Mn_ppm, P_ppm, Ti_ppm, B_ppm, Ge_ppm, Se_ppm, Sn_ppm, Te_ppm, S_pct, Re_ppm, F_ppm	Tl_2_ppm	<0.01, <0.005, <0.5, <2, <0.6, <0.1, <0.3, <0,15	* (se elimina el símbolo * de algunos valores altos de Hg_ppb)	Se eliminan 3 registros cuyas coordenadas no coinciden con el país de origen y una que cae en el mar.

Tabla 2. Limpieza de la base de datos.

La base de datos contiene 536 registros de muestras y 79 variables, de las cuales 10 son informativas y 69 variables corresponden a los elementos y óxidos de interés:

['Muestra', 'Codigo_estandar', 'Longitud', 'Latitud', 'UTM_E', 'UTM_N', 'Zona', 'Region', 'Region_Hidrografica', 'Cuenca', 'Au_ppm', 'Hg_ppb', 'Ag_ppm', 'Al_pct', 'As_ppm', 'Ba_ppm', 'Bi_ppm', 'Ca_pct', 'Cd_ppm', 'Co_ppm', 'Cr_ppm', 'Cu_ppm', 'Fe_pct', 'K_pct', 'La_ppm', 'Li_ppm', 'Mg_pct', 'Mn_pct', 'Mo_ppm', 'Na_pct', 'Ni_ppm', 'P_pct', 'Pb_ppm', 'Sb_ppm', 'Sc_ppm', 'Si_pct', 'Rb_ppm', 'Sr_ppm', 'Ti_pct', 'V_ppm', 'Y_ppm', 'Zn_ppm', 'Zr_ppm', 'Be_ppm', 'Ce_ppm', 'Cs_ppm', 'Dy_ppm', 'Er_ppm', 'Eu_ppm', 'Ga_ppm', 'Gd_ppm', 'Hf_ppm', 'Ho_ppm', 'In_ppm', 'Lu_ppm', 'Nb_ppm', 'Nd_ppm', 'Pr_ppm', 'Sm_ppm', 'Tl_1_ppm', 'Tb_ppm', 'Th_ppm', 'Ta_ppm', 'Tm_ppm', 'U_ppm', 'W_1_ppm', 'W_2_ppm', 'Yb_ppm', 'Al2O3_pct', 'CaO_pct', 'Fe2O3_pct', 'K2O_pct', 'MgO_pct', 'MnO_pct', 'Na2O_pct', 'P2O5_pct', 'SiO2_pct', 'TiO2_pct', 'LOI_pct']

A continuación, se describen las principales características estadísticas de algunos de los principales elementos (Oro (Au), Plata (Ag), Cobre (Cu), Zinc(Zn) y Molibdeno(Mo) Tabla 3:

	Au_ppm	Ag_ppm	Cu_ppm	Zn_ppm	Mo_ppm
count	536	536	536	536	536
mean	0.003165	1.764307e+08	1.394794e+08	1.141555e+08	116606.102612
std	0.008886	6.366175e+08	2.028466e+08	2.180378e+08	198725.586928
min	0.000000	0.000000e+00	3.000000e+00	1.300000e+01	0.000000
25%	0.000000	0.000000e+00	2.875000e+01	1.210000e+02	0.000000
50%	0.000000	0.000000e+00	5.650000e+01	2.330000e+02	2.000000
75%	0.005448	6.824524e-01	2.883085e+08	1.449035e+08	250222.500000
max	0.140045	5.231002e+09	9.586025e+08	9.966708e+08	823411.000000

Tabla 3. Descripción estadística general para los elementos (Oro (Au), Plata (Ag), Cobre (Cu), Zinc(Zn) y Molibdeno(Mo)).

Los histogramas de la mayoría de elementos muestran que la mayoría de valores se ubican en cero o muy cerca a cero con muy pocos valores altos Figura 2.

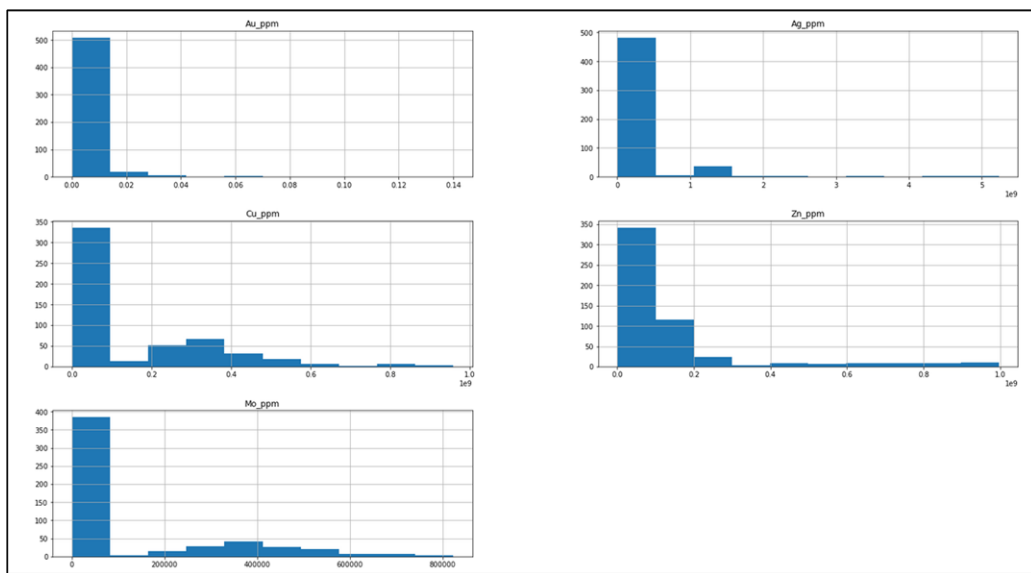


Figura 2. Histogramas para los elementos (Oro (Au), Plata (Ag), Cobre (Cu), Zinc(Zn) y Molibdeno(Mo)).

En la Tabla 4 se observa la correlación entre algunos elementos y óxidos con los elementos (Oro (Au), Plata (Ag), Cobre (Cu), Zinc(Zn) y Molibdeno(Mo)), destacándose el cobre con los valores mas altos de correlación con varios elementos y en menor medida la plata, el molibdeno y el zinc. El oro presenta valores muy bajos de correlación con otros elementos.

Correlación elementos	Cu_ppm	Correlación elementos	Ag_ppm	Correlación elementos	Au_ppm	Correlación elementos	Mo_ppm	Correlación elementos	Zn_ppm
Al_pct	80,64%	Cd_ppm	67,12%	Cd_ppm	20,43%	Si_pct	68,21%	Al_pct	62,29%
Fe_pct	76,67%	In_ppm	72,59%	Rb_ppm	-10,45%	Er_ppm	67,51%	Sc_ppm	-55,37%
Sc_ppm	-72,73%	W_1_ppm	67,05%	In_ppm	20,36%	Nb_ppm	68,22%	Si_pct	61,96%
Si_pct	74,43%							Pr_ppm	62,53%
Be_ppm	74,48%								
Ce_ppm	75,50%								
Dy_ppm	74,50%								
Er_ppm	74,93%								
Ga_ppm	75,06%								
Gd_ppm	71,96%								
Hf_ppm	73,48%								
Nb_ppm	73,68%								
Nd_ppm	75,75%								
Pr_ppm	76,08%								
Sm_ppm	72,89%								
Th_ppm	74,88%								
U_ppm	73,52%								
Fe2O3_pct	73,41%								
K2O_pct	70,26%								

Tabla 4. Correlación entre elementos y óxidos con algunos de los principales elementos de interés (Oro (Au), Plata (Ag), Cobre (Cu), Zinc(Zn) y Molibdeno(Mo))

5. Propuesta metodológica:

Los métodos no supervisados buscan patrones ocultos y aprender de la estructura subyacente en datos no etiquetados.

Inicialmente se usarán 2 métodos no supervisados usando aprendizaje automático con código en Python para generar las agrupaciones de interés en los datos de los muestreos geoquímicos de suelos superficiales.

El primer método corresponde al análisis de componentes principales que será el primer insumo al hacer una reducción lineal de la dimensionalidad para posteriormente aplicar el método de clustering aglomerativo. Estos dos métodos se usarán inicialmente debido a que históricamente y según los artículos revisados son métodos adecuados para análisis geoquímico de muestras de suelo asociadas a yacimientos minerales.

5.1 Método PCA (análisis de componentes principales)

Este método de análisis de componentes principales tiene como objetivo la reducción lineal de dimensionalidad. El algoritmo busca crear una representación reducida de los datos, mientras se conserva la mayor cantidad de información posible.

El método aborda la correlación entre las diferentes características. Si la correlación es muy alta entre un subconjunto de características, el método combina las características altamente correlacionadas y representa estos datos con un número menor de características linealmente no correlacionadas. El algoritmo sigue realizando esta reducción de correlación, encontrando las direcciones de máxima variación en los datos originales de alta dimensión y proyectándolos en un espacio dimensional más pequeño. Estos nuevos componentes generados se conocen como componentes principales (Patel, 2019).

5.2 Método clúster aglomerativo

El método hace parte del aprendizaje no supervisado usando el análisis de clúster, que se encarga de formar grupos diferentes dentro de los datos. Al usar algoritmos de agrupamiento se encuentra la estructura en los datos, de manera que los elementos del mismo clúster o agrupación sean más similares entre sí que con los otros grupos generados (Román, 2019).

El clúster aglomerativo es un tipo de clúster jerárquico en el cual el agrupamiento se inicia con todas las observaciones separadas, cada una formando un clúster individual. Los clústeres se van combinando a medida que la estructura crece hasta converger en uno solo (Amat, 2020).

6. Bibliografía

Amat, J. (2020). Clustering con Python. <https://www.cienciadedatos.net/documentos/py20-clustering-con-python.html>

Balaram, V.; Sawant, S.S. (2021). Indicator Minerals, Pathfinder Elements, and Portable Analytical Instruments in Mineral Exploration Studies. *Minerals* 2022, 12, 394.

Instituto Geológico, Minero y Metalúrgico de Perú. (2021). GEOCATMIN - Prospección Geoquímica del Perú, Serie B. Obtenido de <http://metadatos.ingemmet.gob.pe:8080/geonetwork/srv/spa/catalog.search#/metadata/b1cc5e47-88c8-4c5e-832f-f6dcdbd20211b>

Levitan, D.M.; Zipper, C.E.; Donovan, P.; Schreiber, M.E.; Seal, R.R.; Engle, M.A.; Chermak, J.A.; Bodnar, R.J.; Johnson, D.K.; Aylor, J.G.(2015). Statistical analysis of soil geochemical data to identify pathfinders associated with mineral deposits: An example from the Coles Hill uranium deposit, Virginia, USA. *Journal of Geochemical Exploration*, 154, 238–251.

Nude, P. M., Asigri, J. M., Yidana, S. M., Arhin, E., Foli, G., & Kutu, J. M. (2012). Identifying pathfinder elements for gold in multi-element soil geochemical data from the Wa-Lawra belt, northwest Ghana: A multivariate statistical approach. *International Journal of Geosciences*, 3, 62-70.

Patel, A. (2019). *Hands-On unsupervised learning using Python: How to build applied Machine Learning Solutions from unlabeled data*. O'Reilly Media.

Román, V. (12 de junio de 2019). Aprendizaje no supervisado en Machine Learning: agrupación. <https://medium.com/datos-y-ciencia/aprendizaje-no-supervisado-en-machine-learning-agrupaci%C3%B3n-bb8f25813edc>.

Siegel, F. (1992). Geoquímica aplicada, Departamento de geología, The George Washington University, D. C., Estados Unidos.