

Análisis de Datos y Aplicación de Clustering en Registros de Casos Confirmados por COVID-19 en la Provincia de Cañete

Data Analysis and Clustering Application in Case Records Confirmed by COVID-19 in Cañete Province

Mariano Anthony Cardenas-Quispe

Universidad Nacional de Cañete, Lima, Perú

E-mail: marianocardenas02021@gmail.com ORCID: <https://orcid.org/0000-0003-3033-7846>

Alex Pacheco

Universidad Nacional de Cañete, Lima, Perú

E-mail: apacheco@undc.edu.pe ORCID: <https://orcid.org/0000-0001-9721-0730>

Manuel Alberto Luis Manrique-Nugent

Universidad Nacional Autónoma de Huanta, Ayacucho, Perú

E-mail: manuelmanriquenu@gmail.com ORCID: <https://orcid.org/0000-0002-0816-2499>

Graciela Chela Quispe-Gonzales

Universidad Continental S.A.C., Junín, Perú

E-mail: chela.quispe@hotmail.com ORCID: <https://orcid.org/0000-0002-9394-331X>

Resumen

Tomar las medidas necesarias para la contención de la Pandemia por Covid19 resulta muy complejo. En la actualidad, los Países reconocen muy tarde las situaciones de riesgo de la propagación del virus provocando el colapso de camas UCI en los Hospitales. Por ello se busca identificar los niveles de alerta temprana para evitar una nueva ola de contagio y ubicar los distritos con mayor impacto del virus COVID-19. Esta investigación se desarrolló en 6 fases: La primera consistió en obtener y preparar los datos para el análisis, luego se cargaron los datos al entorno de trabajo Colab, después se seleccionaron los datos de reportes de contagio correspondientes a la provincia de Cañete, Perú; En cuarto lugar, se realizó un análisis utilizando las librerías pandas y Matplotlib, posterior a ello, se creó un nuevo DataFrame con los datos anteriores, finalmente se aplicó el algoritmo de agrupación para segmentar los datos del DataFrame. Obteniendo como resultado, cuatro niveles de alerta temprana, definidos de acuerdo con el total de casos reportados por día. Además, se obtuvo que los distritos con mayor impacto son Imperial y Mala, ya que presentaron mayor frecuencia de reportes y sus promedios de contagios por día son altos. Lo que contribuye a



Scientific Research Journal

Centro de Investigación y Desarrollo Intelectual CIDI

E-ISSN: 2789-2727 / Vol. 2, Núm. 3, 13-28, Julio 2022 / www.srjournalcidi.org

<https://doi.org/10.53942/srjcdi.v2i3.84>

determinar indicadores para plantear nuevas estrategias de prevención, vigilancia y contención de la pandemia en el País.

Palabras Clave

COVID 19, Python, Algoritmo de Clustering

Abstract

Taking the necessary measures to contain the Covid19 pandemic is very complex. At present, countries are very late in recognizing the risk of the spread of the virus, causing the collapse of ICU beds in hospitals. Therefore, the aim is to identify the early warning levels to avoid a new wave of contagion and to locate the districts with the greatest impact of the COVID-19 virus. This research was developed in 6 phases: The first consisted of obtaining and preparing the data for the analysis, then the data were loaded to the Colab work environment, then the data of contagion reports corresponding to the province of Cañete, Peru were selected; Fourthly, an analysis was performed using the pandas and Matplotlib libraries, after that, a new DataFrame was created with the previous data, finally the clustering algorithm was applied to segment the data of the DataFrame. As a result, four early warning levels were obtained, defined according to the total number of cases reported per day. In addition, it was found that the districts with the greatest impact are Imperial and Mala, since they had the highest frequency of reports and their average number of infections per day is high. This contributes to determine indicators to propose new strategies for prevention, surveillance and containment of the pandemic in the country.

Keywords

COVID 19, Python, Clustering Algorithm

Resumo

Tomar as medidas necessárias para conter a pandemia de Covid19 é muito complexo. Atualmente, os países estão muito atrasados em reconhecer o risco de propagação do vírus, causando o colapso dos leitos de UTI nos hospitais. Portanto, o objetivo é identificar os níveis de alerta precoce para evitar uma nova onda de contágio e localizar os distritos com o maior impacto do vírus COVID-19. Esta pesquisa foi realizada em 6 fases: O primeiro consistiu em obter e preparar os dados para a análise, depois os dados foram carregados no ambiente de trabalho Colab, depois foram selecionados os dados dos relatórios de contágio correspondentes à província de Cañete, Peru; quarto, uma análise foi realizada usando as bibliotecas de pandas e Matplotlib, depois disso, um novo DataFrame foi criado com os dados anteriores, finalmente o algoritmo de clustering foi aplicado para segmentar os dados do DataFrame. Como resultado, foram obtidos quatro níveis de alerta precoce, definidos de acordo com o número total de casos relatados por dia. Além disso, descobriu-se que os distritos com maior impacto são Imperial e Mala, pois tinham a maior frequência de relatórios e seu número médio de infecções por dia é alto. Isto ajuda a determinar indicadores para propor novas estratégias de prevenção, vigilância e contenção da pandemia no país.



Palavras-chave

COVID 19, Python, Algoritmo de Aglomeração

1. Introdução

A inicios del año 2020, el virus COVID-19 se dispersaba en todo el mundo, originando en estos países muchos problemas de tipo salud pública, económico y social ocasionando una pandemia. La cual es un problema complejo, nuevo en el mundo y muy difícil de resolver, este problema se ve reflejado en varios países del mundo, y no distingue si el país es desarrollado o está en proceso de desarrollo, igualmente sus economías, que están compuestas por empresas y trabajadores, han sido muy afectadas (Cifuentes-Faura, 2020). Por otra parte en el sector salud se usan nuevas tecnologías como la inteligencia artificial que es una tendencia actualmente, además se observa que la IA, está ganando mucha popularidad en el área de la biomedicina, esto no sólo es causado por la constante optimización de esta ciencia, sino también por la utilidad que ofrece para resolver problemas complejos (Sancho-Escrivá et al., 2020). Una rama de esta tecnología es el machine Learning, que optimizo sus técnicas de aprendizaje en los últimos diez años, demostrando impresionantes resultados en varias áreas de la industria, donde también forma parte el manejo automático, la asistencia médica, fabricación y recolección de energía (Carleo et al., 2019). Por ejemplo, el popular algoritmo de agrupación denominado "K-means", es una técnica de aprendizaje no-supervisado, que trata de encontrar parecidos en los registros de la base de datos, es decir busca elementos que sean similares y los añaden a un grupo. (Sinaga & Yang, 2020)& (BBVA, 2019).

Es preciso señalar la importancia del análisis de datos, que consiste en evaluar un grupo de datos para obtener una conclusión, que permita tomar una decisión o solo aumentar nuestra información sobre diferentes contenidos (QuestiosPro, 2019), por consiguiente las librerías "Pandas" son un instrumento eficaz y fácil de usar para el análisis y manipulación de datos que son de código abierto y están desarrollados en el lenguaje de programación de Python (Pandas, 2021c), además las librerías de Matplotlib permiten crear



gráficos estáticos y dinámicos (Matplotlib, 2021) para una mejor comprensión de los datos analizados. Como menciona Juan Carlos Rodríguez (Accionista director y pionero de Qindice, empresa dedicada al estudio de mercado) que los países como Alemania y Corea del sur, usaron anticipadamente los datos y tuvieron un mayor éxito frenando la pandemia (Leal, 2020). En otras palabras, el análisis de datos ayuda a simplificar la problemática de analizar muchos registros de datos, y así poder conseguir una conclusión, esto es de utilidad en el estudio de los casos de infectados por covid-19.

El COVID-19 es parte de una inmensa familia de coronavirus que provoca enfermedades en humanos y animales, este nuevo coronavirus denominado COVID-19, que está afectando a todos los países del mundo, causando fiebre, tos seca, agotamiento, y otros síntomas poco comunes (Organización Mundial de la Salud, [OMS], 2020b). En consecuencia, la OMS propuso una estrategia sanitaria para combatir el COVID-19, que tiene como objetivo orientar y preparar a los países para tratar de forma correcta la propagación de este virus hasta que la situación sea mejor y estable (OMS, 2020a), como menciona (Alfonso-Sánchez & Fernández-Valdés, 2020), nuestra sociedad actual depende de dos factores importantes, que son la información y el conocimiento que guían a la persona a encontrar información. Debido a esto todas las naciones deben poner en acción las medidas necesarias para frenar la pandemia, de acuerdo con su facultad y contexto (OMS, 2020a).

En Colombia (Díaz-Pinzón, 2020) en su artículo titulado “Análisis de datos en el estudio del COVID-19 en Colombia”, determino las medidas de tendencia central para los casos de personas contagiadas, recuperadas y fallecidas, llegando a la conclusión que la cantidad de casos va en aumento y es necesario el estudio de brotes en el país. En un estudio del covid-19 en las naciones de todo el mundo, (Carrillo-Larco & Castillo-Cara, 2020) concluyeron que fue posible la agrupación de todos los países del mundo usando los datos de casos confirmados por Covid-19, sin embargo no fue posible la agrupación con datos de mortalidad de covid-19. Lo que significa que para estos tipos de estudio es recomendable utilizar los registros de casos confirmados. (Perdomo & Orjuela-Cañon, 2021) en su



investigación “Clustering Proposal Support for the COVID-19 Making Decision Process in a Data Demanding Scenario” concluyen que una estrategia semi-supervisada, da la posibilidad de contribuir en la toma de decisiones del manejo de reportes de Covid-19, en un ambiente con baja información. Además estos algoritmos de clustering son eficientes en encontrar semejanzas de acuerdo con sus características (Khmaissia et al., 2020). Es decir, existe un respaldo sobre la utilización de los algoritmos de agrupación y su eficacia para el análisis de datos.

En el Perú, el primer caso de COVID 19, fue confirmado el viernes 6 de marzo del año 2020, en el distrito de La Molina provincia de Lima, ante el suceso, el Ministerio de Salud (MINSU) pidió que la población mantuviera la calma y que ellos tomaran las acciones correspondientes para salvaguardar la salud de los ciudadanos del País (Ministerio de Salud, 2020). Nueve días después, el gobierno del Perú, con la intención de frenar el avance de la pandemia en el país, decreto una cuarentena general de 15 días calendarios (BBC, 2020). Cuarenta días después del decreto supremo, el Perú ya se había convertido en el segundo país con el mayor registro de infectados, en el ranking de América Latina (Noriega, 2020).

En la provincia de Cañete, ha sido notorio la crisis en los sectores económico y salud a causa de la pandemia, (Jaramillo et al., 2021) mencionan, esta crisis ha originado el distanciamiento social que es un obstáculo para el avance de la economía, además del colapso de los sistemas de salud y las víctimas mortales. Como se menciona en la investigación de (Maguiña-Vargas, 2020), la nueva enfermedad ha resaltado la pobre disposición sanitaria en el Perú, donde se observa la escasez de materiales en los hospitales, los descuidos de los centros de salud y personal médico. Además del peligroso daño a la economía del país, creando una incertidumbre catastrófica, por ser la primera vez que la nación enfrenta una crisis sanitaria de nivel global (Herrera-Salino, 2020). Es decir, mientras siga aumentando el contagio y no se evalúen nuevas estrategias para mitigar la pandemia, estos sectores estarán en crisis más tiempo.

Por ello, la presente investigación, pretende identificar los niveles de alerta temprana para evitar una nueva ola de contagio y ubicar los distritos con mayor impacto



del virus. Lo que contribuye a determinar indicadores para plantear nuevas estrategias de prevención, vigilancia y contención de la pandemia en el País.

2. Estado del arte

2.1 COVID 19

La COVID-19 es provocado por un nuevo tipo de virus denominado “coronavirus”, el cual comenzó a propagarse en los últimos meses del año 2019, en la ciudad de Wuhan que se encuentra en el país de China (Perez-Abreu et al., 2020).

2.2 Análisis de datos cuantitativos

Según Tenelema-Cedeño en el 2018, afirma que el análisis de datos cuantitativos se refiere a la metodología de usar herramientas estadísticas para la descripción, resumen y comparación de los datos.

2.3 Python

Es un lenguaje de programación creado por Guido Van Rossum en el año 1989, el cual pertenece al tipo de lenguajes de programación interpretado, además de ser considerado como un lenguaje con una sintaxis simple y de fácil aprendizaje (Vidal-Silva et al., 2021).

2.2.1 DataFrame: Garcia-Sabater,2021, menciona que es una tabla de datos, donde en cada columna tiene el dato de una variable y cada fila contiene los datos de todas las variables.

2.2.2 Clustering: Es un método de aprendizaje no supervisado, que divide los datos en grupos, donde cada grupo contiene datos similares pero diferente a los demás grupos (Arias-Martinez & Zamora-Litardo, 2021).

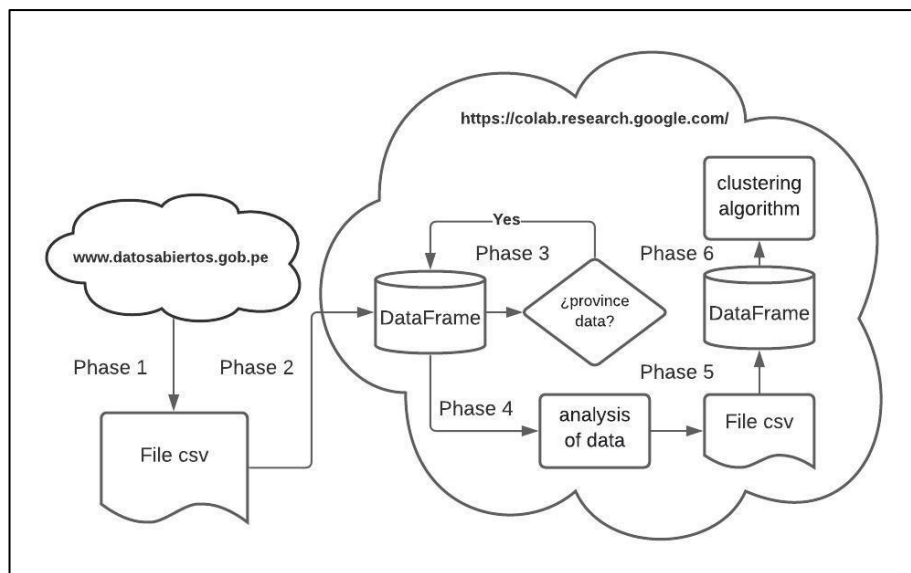
3. Materiales y métodos

El proyecto se desarrolló en 6 fases como muestra el diagrama de flujo de procesos (Fig.1), en el entorno de desarrollo “Colaboratory”, también denominado “Colab”, donde se programa en Python usando solo el navegador web, las líneas de código se ingresan y procesan en un entorno dinámico llamado “Notebook de Colab” (Colaboratory, 2021).

Figura 1.



Flujo de proceso de desarrollo



Fase 1, Obtención y preparación de los datos: Los datos fueron obtenidos de la página de datos abiertos del Perú, que se encuentra en el siguiente enlace: <https://www.datosabiertos.gob.pe/dataset/casos-positivos-por-covid-19-ministerio-de-salud-minsa>. Los datos recopilados corresponden a los registros de casos positivos por COVID-19 en el Perú, desde el 6 de marzo del 2020 hasta el 15 de enero del 2021

Los datos se descargaron en un archivo en formato “csv” que se caracterizan por tener filas de datos separados por comas que corresponde a la columna de una variable, y posteriormente se leyó los datos usando el software Excel, donde se realizó el cambio del formato de la fecha en: “Año,Mes,Día”, para una mejor manipulación de datos y después se ordenó en orden creciente; Además se eliminó la columna “Nombres”, ya que no formaba parte de la presente investigación.

Fase 2, Cargar Datos: El archivo CSV denominado “datos”, fue cargado al drive personal. En el entorno Colaboraty, el archivo “datos” fue leído y guardado en un DataFrame.

Fase 3, Selección de datos de la provincia: Los datos del DataFrame contenían información de todos los departamentos del país, por ese motivo se seleccionaron solo los datos correspondientes a la provincia de Cañete, almacenándose en el mismo DataFrame.

Fase 4, Análisis de datos con pandas y Matplotlib: Se realizó un análisis sobre la cantidad de casos positivos por día en cada distrito. Después se mostraron los datos mediante el esquema de cuartiles (Pandas, 2021a), es decir se mostraron los datos en cuatro grupos denominados cuartiles, donde cada grupo represento el 25 % del total de los valores (Martínez et al., 2019), y también se usó el diagrama de línea que usa los valores de los



DataFrame como coordenadas, para hacer trazos de línea (Pandas, 2021b) y representar dichos resultados.

Fase 5: Creación de un nuevo DataFrame: Los resultados de la fase 5, se exportaron como archivo CSV, después se importó el archivo guardándolo en un nuevo DataFrame. Se asignó un código numérico a cada distrito (Fig. 2), y después se definió en cada fila la cantidad de días transcurridos desde la fecha del primer caso reportado en la provincia de Cañete.

Figura 2

Código asignado a cada distrito de la provincia de Cañete.

Codigo	Distrito	Codigo	Distrito	Codigo	Distrito	Codigo	Distrito
1	ASIA	5	COAYLLO	9	NUEVO IMPERIAL	13	SAN LUIS
2	CALANGO	6	IMPERIAL	10	PACARAN	14	SAN VICENTE
3	CERRO AZUL	7	LUNAHUNA	11	QUILMANA	15	SANTA CRUZ DE FLORES
4	CHILCA	8	MALA	12	SAN ANTONIO	16	ZUÑIGA

Fase 6: Análisis de datos mediante el algoritmo de agrupación: Previo a este análisis, se eliminó las columnas de distrito y fecha, ya que solo se necesitan valores numéricos para normalizar los datos y realizar el algoritmo de agrupación. Debido a la diferencia escalar de los datos de las variables, la normalización de datos busca que los valores de las variables tengan la misma escala (Isidro, 2017). Después se determinó la cantidad de grupos realizando la función del “codo” y finalmente se realizó el algoritmo de agrupación.

4. Resultados

4.1 Resultado 1

En este resultado se muestra la cantidad de casos por fecha, en la figura 3 se muestra que las cantidades son divididas en cuatro cuartiles, con cuatro valores atípicos mayores a 100; En la figura 4 se muestra las cantidades de casos confirmados de cada fecha. Comparando las dos figuras se deduce lo siguiente: a). El cuartil 1, con intervalos de 1 a 5 casos por día, representa un nivel de alerta moderado. b). El cuartil 2, con intervalos de 6 a 19 casos por día, representa un nivel de alerta alto. c). El cuartil 3, con intervalos de 20 a 47 casos por día, representa un nivel de alerta muy alto. d). El cuartil 4, con intervalos de 48 a 100 casos por día, representa un nivel de alerta crítico. e). Los puntos independientes representan los índices máximos de casos por día en la provincia.



Figura 3

Registros de casos reportados representados en cuartiles.

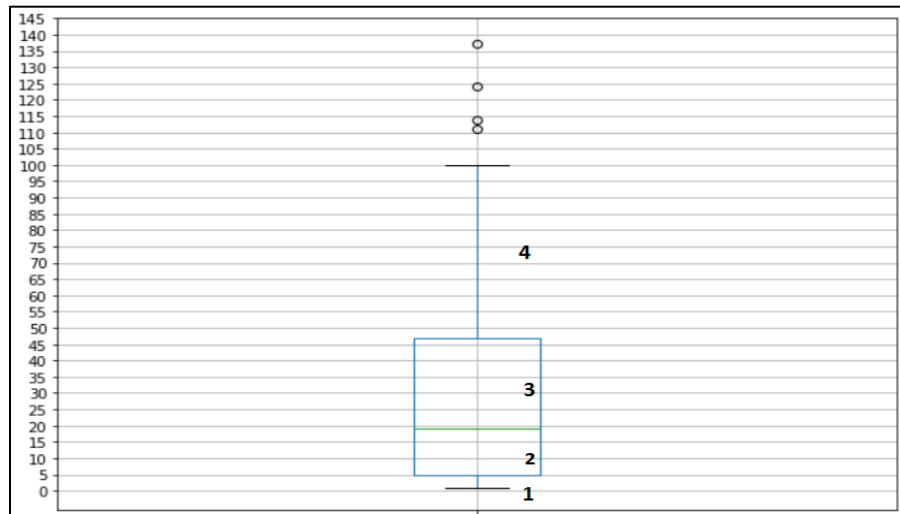
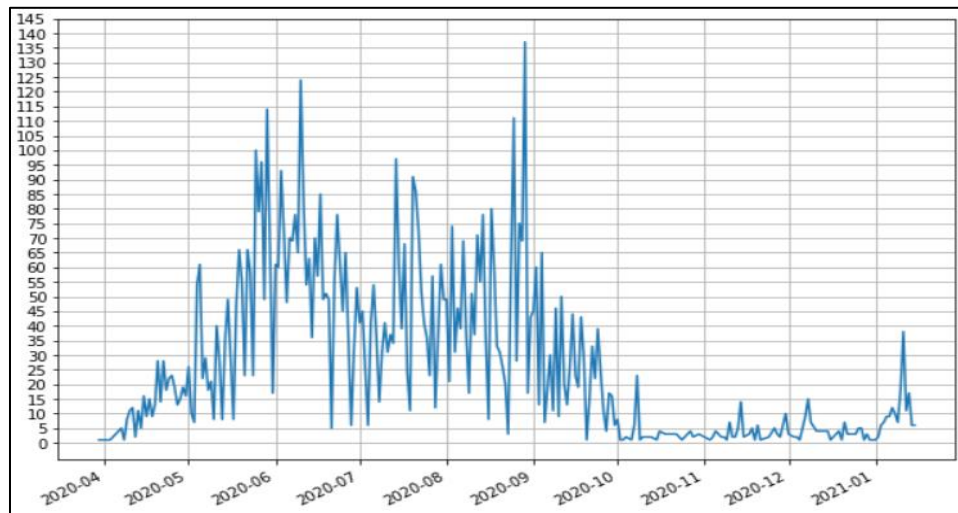


Figura 4

Registro de casos reportados en la provincia de Cañete.



4.2 Resultado 2

Después de aplicar el algoritmo clustering, la figura 5 muestra los promedios de días transcurridos después del primer caso de contagio registrado y la cantidad de casos por día de cada distrito, estos datos fueron clasificados en tres grupos: a). En el grupo de etiqueta 0, los promedios del total de casos registrados por distrito son menores a 7 casos de contagio, y los promedios de días transcurridos son mayores a 100 días. b). En el grupo de



etiqueta 1, los promedios del total de casos registrados por distrito son entre 30 y 6, y los promedios de días transcurridos son menores son entre 85 y 26. c). En el grupo de etiqueta 2, los promedios del total de casos registrados por distrito son menores a 5 y los promedios de días transcurridos son menores a 28.

Figura 5

Promedios de número de días y cantidades de casos reportados de cada distrito.

Cod	NumD	CANT	ETIQUETA
6.0	134.021127	6.176056	0
8.0	145.562130	5.822485	0
9.0	118.752212	4.371681	0
4.0	112.978417	4.251799	0
11.0	119.517241	4.241379	0
13.0	132.623932	3.333333	0
14.0	174.492424	3.000000	0
3.0	124.566038	2.811321	0
1.0	133.037500	2.775000	0
15.0	153.107143	2.357143	0
12.0	135.528736	1.988506	0
7.0	103.659091	1.954545	0
10.0	144.913043	1.739130	0
5.0	148.187500	1.500000	0
2.0	141.333333	1.000000	0
16.0	288.500000	1.000000	0

Cod	NumD	CANT	ETIQUETA
4.0	84.666667	29.166667	1
13.0	94.333333	22.000000	1
3.0	81.000000	20.666667	1
14.0	59.758621	20.517241	1
6.0	62.114286	20.371429	1
9.0	70.000000	19.666667	1
11.0	78.800000	16.800000	1
8.0	59.923077	15.538462	1
1.0	47.000000	10.000000	1
5.0	26.000000	6.000000	1

Cod	NumD	CANT	ETIQUETA
14.0	27.382353	4.941176	2
9.0	18.444444	3.444444	2
6.0	14.666667	2.222222	2
8.0	19.000000	2.200000	2
11.0	28.000000	2.125000	2
13.0	31.300000	2.100000	2
12.0	29.500000	1.750000	2
7.0	15.000000	1.000000	2
15.0	23.000000	1.000000	2

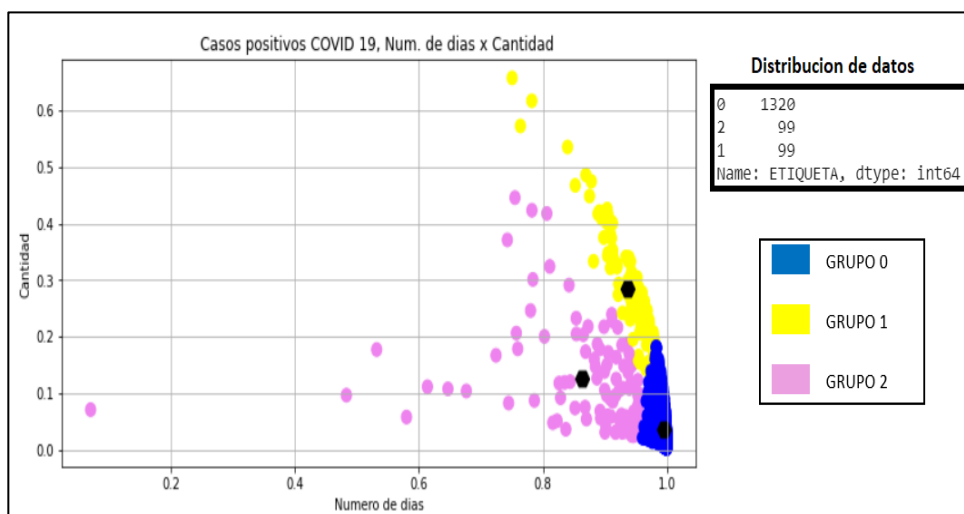
4.3 Resultado 3

Después de aplicar el algoritmo clustering, La figura 6 muestra un gráfico de los datos normalizados en tres grupos, representan una distribución en las dimensiones del número de días y cantidad de casos registrados: a). El grupo 0 contiene el 87% de los datos con alta similitud, que representan una cantidad baja de casos en un mayor número de días. b). El grupo 1 contiene el 6.5% con una similitud media, que representan cantidades altas de casos en un alto número de días. c). El grupo 2 representa el 6.5% de los datos con baja similitud, que representan diferentes cantidades de casos en un intervalo de tiempo.

Figura 6



Distribución y agrupación de datos realizado por el algoritmo de clustering



4.4 Resultado 4

Después de aplicar el algoritmo clustering, la figura 7 muestra la frecuencia de cada distrito en tres grupos: a). El grupo 0 está compuesto por los 16 distritos en total de la provincia, donde los distritos con código 8 y 6 (Fig. 2), que corresponden a los distritos de Mala e Imperial, son los distritos con mayor frecuencia. b). El grupo 1 está compuesto por 10 de los 16 distritos en total de la provincia, donde los distritos con código 6 y 14 (Fig. 2), que corresponden a los distritos de Imperial y San Vicente, son los distritos con mayor frecuencia. c). El grupo 2 está compuesto por 9 de los 16 distritos en total de la provincia, donde los distritos con código 14 y 9 (Fig. 2), que corresponden a los distritos de San Vicente y Nuevo Imperial, son los distritos con mayor frecuencia.

Figura 7

Agrupaciones con frecuencia de reportes por cada distrito

Group 0		Group 1		Group 2	
Code	Frequency	Code	Frequency	Code	Frequency
8.0	169	6.0	35	14.0	34
6.0	142	14.0	29	9.0	18
4.0	139	8.0	13	13.0	10
14.0	132	4.0	6	8.0	10
13.0	117	11.0	5	6.0	9
11.0	116	3.0	3	11.0	8
9.0	113	13.0	3	12.0	8
3.0	106	9.0	3	15.0	1
12.0	87	1.0	1	7.0	1
1.0	80	5.0	1		
7.0	44				
15.0	28				
10.0	23				
5.0	16				
2.0	6				
16.0	2				

Name: Cod, dtype: int64



5. Discusión

El resultado 1, determina cuatro niveles de alerta, donde el nivel uno y dos representan una situación controlable, en cambio desde el nivel tres que inicia con una cantidad de 20 casos por día, representa un riesgo de rebrote muy alto y se debe ejecutar protocolos de contención de emergencia. Esto concuerda con (Instituto Nacional de Vigilancia de Medicamentos y Alimentos, 2021), donde menciona que ante un indicio de riesgo latente para la salud de los ciudadanos, se deberá desarrollar actividades competentes y urgentes de salud pública. Además (Guillot et al., 2017) menciona que los sistemas de alerta temprana son fundamentales para disminuir los riesgos de desastres.

En el resultado 2, del total de casos registrados en el grupo 0, se muestra que los distritos de Imperial y Mala han tenido en promedio 6 casos por día después de 100 días en promedio, de registrarse el primer caso de contagio en la provincia de Cañete; Del total de casos registrados en el grupo 1 se muestra que los distritos de Chilca y San Luis han tenido en promedio 29 y 22 casos por día después de 84 días en promedio, de registrarse el primer caso de contagio en la provincia de Cañete; Del total de casos registrados en el grupo 2 se muestra que los distritos de San Vicente y Nuevo Imperial han tenido en promedio 5 y 3 casos por día después de 18 días en promedio, de registrarse el primer caso de contagio en la provincia de Cañete. Con estos resultados fue posible medir el impacto de contagio por Covid-19 y realizar un análisis comparativo con los demás distritos de la provincia de Cañete a través del tiempo desde el primer día en el que se registró un caso de Covid-19. En ese sentido, (Biele, 2020) menciona que es necesario realizar una normalización de los datos de la investigación, para poder comparar grupos de datos de manera justa e independiente de la cantidad de registros que contengan.

El resultado 3, muestra como el número de casos reportados se han vuelto más similares a través del tiempo, donde el grupo 0 contiene el 87% de estos registros más comunes y frecuentes. Como menciona en sus conclusiones (Biele, 2020), los algoritmos de agrupación permiten obtener grupos con características similares y únicas, como el agrupar países con situaciones similares.

El resultado 4, muestra que los distritos de Imperial y Mala tienen un mayor impacto de contagio por Covid-19, porque presentan 169 y 142 reportes en total en el grupo 0, a comparación con los demás grupos que contienen como máximo 35 reportes, además que este grupo contiene los registros más frecuentes y similares de acuerdo con el resultado 3. Como menciona (Aguilar-Cortes, 2018), conocer las medidas métricas de cada entorno posibilita tomar decisiones que optimicen una estrategia, para ello es importante plantearse objetivos de análisis de datos.



6. Conclusiones

Se concluye que el límite máximo de contagios que debe permitirse es hasta 19 casos por día (nivel de alerta alto), conociendo este límite de contagio se permitirá tomar las acciones necesarias en cada situación.

Después de analizar los tres resultados del algoritmo clustering, se deduce que los primeros epicentros virales corresponden a los distritos de Nuevo Imperial y San Vicente, después estos epicentros virales se ubicaron en Chilca y San Luis, por último, se convirtieron en los nuevos epicentros virales, los distritos de Imperial y Mala, también estos distritos mencionados al último pertenecen al grupo con más reportes de contagio y por consiguiente mayor impacto viral.

7. Recomendaciones

Se recomienda, que, si el límite máximo de contagio es superado en tres ocasiones, la situación sanitaria en Cañete estaría en un riesgo muy alto, y se debería tomar acción con protocolos de contención de emergencia, para prevenir y no alcanzar el nivel de alerta crítico. También se recomienda una vigilancia sanitaria en los últimos epicentros virales, que son los distritos de Imperial y Mala.

Referencias

- Aguilar-Cortes, D. C. (2018). La Importancia del Contenido de Valor y su Medicion Mejorar la Estrategia Digital de los Medios de Comunicacion: Caso Revista Semana. In *Fundación Universitaria los Libertadores* (Issue Marzo).
https://repository.libertadores.edu.co/bitstream/handle/11371/2963/Cortés_Dini_2018.pdf?sequence=1&isAllowed=y
- Alfonso-Sánchez, I. R., & Fernández-Valdés, M. de las M. (2020). Comportamiento Informacional, Infodemia y Desinformación Durante la Pandemia de COVID-19. *Academia de Ciencias de Cuba*, 10(2), 19. <http://orcid.org/0000-0003-2296-5041MaríadelasMercedesFernándezValdés2>, <http://orcid.org/0000-0002-9551-7437>
- Arias-Martinez, B. J., & Zamora-Litardo, B. S. (2021-03). Algoritmo de Clustering dinámico para trayectoria GPS. [Tesis, *Universidad de Guayaquil*]. Recuperado a partir de <http://repositorio.ug.edu.ec/handle/redug/52739>
- BBC. (2020). *Coronavirus: Perú decreta cuarentena general en el país y el cierre de fronteras durante 15 días ante la pandemia de covid-19 - BBC News Mundo*. BBC. <https://www.bbc.com/mundo/noticias-america-latina-51902989>
- BBVA. (2019). *Te contamos qué es el “machine learning” y cómo funciona. “Machine Learning”: ¿qué Es y Cómo Funciona?* <https://www.bbva.com/es/machine-learning->



- que-es-y-como-funciona/
Biele, R. (2020). *Un análisis de agrupación de aprendizaje automático de la serie temporal de COVID-19. | Hacia la ciencia de datos*. Towards Data Science.
<https://towardsdatascience.com/which-countries-are-affected-the-most-by-covid-19-4d4570852e31>
- Carleo, G., Cirac, I., Cranmer, K., Daudet, L., Schuld, M., Tishby, N., Vogt-Maranto, L., & Zdeborová, L. (2019). Machine learning and the physical sciences. *Reviews of Modern Physics*, 91(4), 45002. <https://doi.org/10.1103/RevModPhys.91.045002>
- Carrillo-Larco, R. M., & Castillo-Cara, M. (2020). Using country-level variables to classify countries according to the number of confirmed COVID-19 cases: An unsupervised machine learning approach. *Wellcome Open Research*, 5, 56.
<https://doi.org/10.12688/wellcomeopenres.15819.3>
- Cifuentes-Faura, J. (2020). Crisis del coronavirus: impacto y medidas económicas en Europa y en el mundo. *Espaço e Economia*, 18.
<https://doi.org/10.4000/espacoeconomia.12874>
- Colaboratory. (2021). *Te damos la bienvenida a Colaboratory - Colaboratory*. Te Damos La Bienvenida a Colaboratory Te.
<https://colab.research.google.com/notebooks/intro.ipynb#recent=true>
- Díaz-Pinzón, J. E. (2020). Análisis de Datos en el Estudio del COVID-19 en Colombia. *Revista Repertorio de Medicina y Cirugía*, 29, 103–107.
<https://doi.org/10.31260/repertmedcir.01217372.1121>
- García-Sabater, Jose P. (2021) Programación Matemática en Python con PULP RIUNET Repositorio UPV <http://hdl.handle.net/10251/158416>
- Guillot, J. D., Robles, C. A., & Callejas, J. D. (2017). Adquisición de Señales Ambientales para un Sistema de Alerta Temprana Adquisición de Señales Ambientales para un Sistema de Alerta Temprana Environmental Signals Acquisition for an Early Warning System. *Información Tecnológica*, 28(5), 45–54. <https://doi.org/10.4067/S0718-07642017000500007>
- Herrera-Salino, L. (2020). *El COVID – 19: las consecuencias en la economía peruana y mundial - Facultad de Derecho*. USIL- Facultad de Derecho.
<https://facultades.usil.edu.pe/derecho/carrera-de-relaciones-internacionales/el-covid-19-las-consecuencias-en-la-economia-peruana-y-mundial/>
- Instituto Nacional de Vigilancia de Medicamentos y Alimentos. (2021). *Alertas Sanitarias e Informes de seguridad*. [INVIMA]. <https://www.invima.gov.co/Alertas-Sanitarias-E-Informes-De-Seguridad>
- Isidro, G. O. (2017). *Algoritmos de clustering y búsqueda de asociaciones*. http://rstudio-pubs-static.s3.amazonaws.com/285462_c6f85d1b9c784c17b0b87dc7e018ae3c.html
- Jaramillo, M., Kristian, B., & Vargas, L. (2021). Políticas para combatir la pandemia del COVID-19 GRADE Informe preliminar. *Grupo de Analisis Para El Desarrollo - GRADE*.
- Khmaissia, F., Haghighi, P. S., Jayaprakash, A., Wu, Z., Papadopoulos, S., Lai, Y., & Nguyen, F. T. (2020). *An Unsupervised Machine Learning Approach to Assess the ZIP Code Level Impact of COVID-19 in NYC*. <https://www.citibikenyc.com/system-data>



- Leal, F. (2020). *El coronavirus realza el valor de la analítica de datos para gobiernos*. El Periodico.
<https://www.elperiodico.com/es/activos/innovadores/20200610/coronavirus-analisis-datos-7993823>
- Maguiña-Vargas, C. (2020). Reflexiones sobre el COVID-19, el Colegio Médico del Perú y la Salud Pública. *Acta Medica Peruana*, 37(1), 8–10.
<https://doi.org/10.35663/amp.2020.371.929>
- Martínez, P., (dir.), I., Alonso, R. A., Varela, M. C., Benzies, Y. C., Polo, F. X. F., García, L. G., Rúa, P. L., Rodríguez, Y. R., & Pérez, J. R. V. (2019). *Definición de: Cuartil | Dicenlen*. Diccionario Electrónico de Enseñanza y Aprendizaje de Lenguas.
<https://www.dicenlen.eu/es/diccionario/entradas/cuartil>
- Matplotlib. (2021). *Matplotlib: trazado de Python - documentación de Matplotlib 3.3.3*. Matplotlib. <https://matplotlib.org/index.html>
- Ministerio de Salud. (2020). *Coronavirus en el Perú: casos confirmados | Gobierno del Perú*. Ministerio de Salud. <https://www.gob.pe/8662-coronavirus-en-el-peru-casos-confirmados>
- Noriega, C. (2020). *Éxodo masivo en Perú: miles de personas retornan a pie a sus pueblos por hambre y falta de trabajo*. Nodal - Noticias de América Latina y El Caribe.
<https://www.nodal.am/2020/04/exodo-masivo-en-peru-miles-de-personas-retornan-a-pie-a-sus-pueblos-por-hambre-y-falta-de-trabajo/>
- OMS. (2020a). Actualización de la estrategia frente a la COVID-19. *Organización Mundial de La Salud*, 1–17. https://www.who.int/docs/default-source/coronaviruse/covid-strategy-update-14april2020_es.pdf?sfvrsn=86c0929d_10
- OMS. (2020b). *Preguntas y respuestas sobre la enfermedad por coronavirus (COVID- 19)*. Organización Mundial de La Salud.
<https://www.who.int/es/emergencias/diseases/novel-coronavirus-2019/advice-for-public/q-a-coronaviruses>
- Pandas. (2021a). *pandas.DataFrame.plot.box - documentación de pandas 1.2.1*. <https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.plot.box.html>
- Pandas. (2021b). *pandas.DataFrame.plot.line - documentación de pandas 1.2.1*. <https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.plot.line.html>
- Pandas. (2021c). *Pandas - Python Data Analysis Library*. <https://pandas.pydata.org/>
- Perdomo, O., & Orjuela-Cañon, A. D. (2021). Clustering Proposal Support for the COVID-19 Making Decision Process in a Data Demanding Scenario. *IEEE*.
<https://latamt.ieee9.org/index.php/transactions/article/view/4403/1028>
- Perez-Abreu, M. R., Gómez-Tejeda, J. J., & Dieguez-Guach, R. A. (2020). Características clínico-epidemiológicas de la COVID-19. *Revista Habanera de Ciencias Médicas*, 19(2). <http://www.revhabanera.sld.cu/index.php/rhab/article/view/3254/2505>
- QuestiosPro. (2019). *Análisis de Datos | QuestionPro*. QuestionPro.
<https://www.questionpro.com/es/analisis-de-datos.html>



- Sancho-Escrivá, J. V., Fanjul-Peyró, C., De la Iglesia-Vayá, M., Montell, J. A., & Escartí-Fabra, M. J. (2020). Aplicación de la Inteligencia Artificial con Procesamiento del Lenguaje Natural para textos de investigación cualitativa en la relación médico-paciente con enfermedad mental mediante el uso de tecnologías móviles. *Revista de Comunicación y Salud*, 10(1), 19–41. [https://doi.org/10.35669/rcys.2020.10\(1\).19-41](https://doi.org/10.35669/rcys.2020.10(1).19-41)
- Sinaga, K. P., & Yang, M. S. (2020). Unsupervised K-means clustering algorithm. *IEEE Access*, 8, 80716–80727. <https://doi.org/10.1109/ACCESS.2020.2988796>
- Tenelema-Cedeño, J. D. (2018). Análisis de las Estrategias Comerciales del Almacén Casa Wilson de la Ciudad de Portoviejo y su Influencia en el Comportamiento de las Ventas. [Tesis, UNIVERSIDAD SAN GREGORIO DE PORTOVIEJO]. <http://repositorio.sangregorio.edu.ec/handle/123456789/889>
- Vidal-Silva, C. L., Sánchez-Ortiz, A., Serrano, J., & Rubio, J. M. (2021). Experiencia académica en desarrollo rápido de sistemas de información web con Python y Django. *Formación universitaria*, 14(5), 85–94. <https://doi.org/10.4067/s0718-50062021000500085>

