

Metodología de clasificación de entes territoriales para la toma de decisiones en respuesta a la emergencia por el covid-19

Camila Ramirez¹, Ciro Alberto Amaya¹ [0000-0003-1537-1616], Andrea Herrera¹ [0000-0003-2898-7161], Catalina González-Uribe² [0000-0002-3322-5017], Estefania Hernandez³ [0000-0002-3689-8209], Johanna, Trujillo-Diaz² [0000-0001-6243-6393] y Nubia Velasco³ [0002-5717-4998]

{¹ Facultad de Ingeniería, ² Facultad de Medicina, ³ Facultad de Administración} Universidad de los Andes, Bogotá COLOMBIA
{c.ramirez, ca.amaya, a-herrer, cgonzalez, e.hernandez12, j.trujillo2, nvelasco}@uniandes.edu.co

Abstract. La pandemia por el covid-19 ha afectado en gran medida la salud de las personas y la economía en todo el mundo. En muchos países, los sistemas de salud han presentado problemas de ocupación y falta de recursos, debido a la gran cantidad de personas infectadas que no tenían previsto atender y a la escasez mundial de vacunas. Este artículo propone una metodología que permite clasificar las regiones de Colombia, en diferentes niveles de alerta de contagio SARS-Co2 con el fin de tomar decisiones de forma oportuna en términos de cuarentenas, y de despliegue de recursos de vacunación covid-19 que permitan disminuir las tasas de contagios y sus efectos. La metodología utiliza información proveniente de variables sobre el contagio y del avance de la vacunación. Dicha metodología puede ser empleada por los tomadores de decisiones de forma prospectiva para establecer estrategias en la prevención, y controlar el impacto de la pandemia en el país.

Keywords: Covid 19, Modelo de clasificación, Estratificación, clusterización

1 Introducción

La enfermedad del coronavirus (covid-19) se convirtió en una emergencia de salud pública a nivel mundial, que se propaga aceleradamente, ocasiona efectos que pueden conducir a la muerte, lo que conlleva a que los hospitales superen la ocupación de sus unidades de cuidados intensivos (UCI) y no logren atender adecuadamente la población infectada en estado grave, lo cual genera una gran cantidad de personas fallecidas e inequidades [1], [2]. El proceso de vacunación ha tomado un rol muy importante para mitigar el efecto del virus y es fundamental llevarlo a cabo a partir de estrategias que conlleven a una mejor planificación y ejecución de los recursos existentes[1], [3], [4]. Para lo cual Ministerio de Salud y de Protección Social (MSPS) en Colombia, diseño el Plan Nacional de Vacunación [5] contra el covid-19, que busca evitar la mayor cantidad de muertes y lograr una reactivación económica del país lo antes posible.

Dada la complejidad del PNV, es fundamental contar con metodologías de ayuda a la toma de decisiones, que permitan dar recomendaciones y alertas que ayuden al sector de salud pública y a los diferentes entes gubernamentales al despliegue efectivo de recursos del proceso de vacunación y de acciones de mitigación ante el contagio [6]. Esta investigación fue desarrollada con el objetivo de generar una metodología de clasificación de entes territoriales, con el objeto de dirigir acciones de mitigación del contagio por el covid-19, teniendo en cuenta variables relacionadas con *capacidad, cobertura, contagio y vacunación* que ayuden al sector de salud pública y a los diferentes gobernantes a tomar decisiones sobre el proceso de vacunación.

2 Metodología propuesta

Para la recolección de los datos, el proyecto de investigación cuenta con una herramienta de visualización de indicadores para la toma de decisiones sobre la vacunación y la mitigación del contagio en Colombia. Este estudio propone una metodología para clasificar entes territoriales, utilizando datos observados del avance de la vacunación y del estado de contagio. La metodología se basa en análisis multicriterio [7], en el cual se construyen índices teniendo en cuenta modelos explicativos de naturaleza aditiva. En dichos modelos los

índices son explicados a partir de las mediciones existentes o de datos observados, en este caso, del avance de la enfermedad y de la vacunación. La metodología general que se propone se divide en cinco pasos: (1) selección, recopilación y procesamiento de los datos iniciales disponibles, (2) construcción de índices, (3) clusterización, (4) graficación y análisis, y (5) acciones de mitigación.

2.1 Recopilación y procesamiento de datos iniciales

Dos tipos de datos se utilizan dentro de la metodología: datos de contagio (ver Tabla 1) y datos de avance de vacunación (ver Tabla 2). En las siguientes tablas se muestran ejemplo de los tipos de datos y la fuente de recolección (ver en [8]).

Tabla 1. Información sobre el contagio

Información sobre contagio			
Información	Descripción	Siglas	Fuente
Casos activos	Cantidad de personas que se encuentran actualmente con covid-19.	CA	Datos Abiertos https://www.datos.gov.co/Salud-y-Protección-Social/Casos-positivos-de-COVID-19-en-Colombia/gt2j-8ykr/data
Casos de muertes	Cantidad de personas que fallecieron a causa de covid-19.	CM	Datos Abiertos https://www.datos.gov.co/Salud-y-Protección-Social/Casos-positivos-de-COVID-19-en-Colombia/gt2j-8ykr/data
Casos nuevos diarios	Cantidad de casos nuevos de covid-19 diariamente.	CN	Datos Abiertos https://www.datos.gov.co/Salud-y-Protección-Social/Casos-positivos-de-COVID-19-en-Colombia/gt2j-8ykr/data
Nuevos contagios semana i	Cantidad de casos nuevos en la semana i.	CNsem	Datos Abiertos https://www.datos.gov.co/Salud-y-Protección-Social/Casos-positivos-de-COVID-19-en-Colombia/gt2j-8ykr/data
Número de pruebas	Cantidad de pruebas realizadas diariamente para identificar covid-19, tanto de antígeno como PCR.	NumP	Datos Abiertos https://www.datos.gov.co/Salud-y-Protección-Social/Casos-positivos-de-COVID-19-en-Colombia/gt2j-8ykr/data

Tabla 2. Información sobre la vacunación

Información sobre vacunación			
Información	Descripción	Siglas	Fuente
Población total	Población total que tiene Colombia actualmente.	PobT	DANE (2018) https://www.dane.gov.co/index.php
Población por inmunizar	Cantidad de la población de Colombia que se debe inmunizar para cumplir con la inmunidad de rebaño, es decir el 70% de la población total.	PobInm	Plan Nacional de Vacunación https://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/VS/pnv-contra-covid-19.pdf
Puntos de vacunación habilitados	Cantidad de puntos de vacunación que se encuentran habilitados para la vacunación contra el COVID-19.	PVh	Ministerio de Salud https://www.minsalud.gov.co/salud/publica/Vacunacion/Paginas/puntos-de-vacunacion.aspx
Dosis a aplicar	Cantidad de dosis de vacunas que se deben aplicar para cumplir con la población por inmunizar.	DosisT	Plan Nacional de Vacunación https://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/VS/pnv-contra-covid-19.pdf
Inventario de vacunas	Cantidad de vacunas disponibles y que no han sido aplicadas.	InV	Ministerio de Salud https://www.minsalud.gov.co/Paginas/Norm_Resoluciones.aspx
Vacunas aplicadas	Cantidad total de dosis de vacunas que se han aplicado.	VacAp	Ministerio de Salud https://www.minsalud.gov.co/salud/publica/Vacunacion/Paginas/Vacunacion-covid-19.aspx
Tasa de vacunación diaria	Cantidad de dosis de vacunas que se están aplicando diariamente.	TVdiaria	Ministerio de Salud https://www.minsalud.gov.co/salud/publica/Vacunacion/Paginas/Vacunacion-covid-19.aspx

2.2 Construcción de índices

Para generar una clasificación de los entes territoriales se propone la construcción de dos índices: un índice de vacunación y un índice de contagio. A partir de los datos recopilados, se crean las variables correspondientes que van a definir cada uno de los índices. Cabe resaltar que cada variable se calcula para cada ente, con el fin de obtener el valor del índice departamental. Una vez se tienen definidas las variables correspondientes a cada uno de los índices, se prosigue con su cálculo. La fórmula general que se propone para el cálculo de cada índice es la siguiente:

$$\text{Índice}_i = \sum_{k=1}^{n_i} w_{ik} X_{ik}$$

Donde i hace referencia al índice que se está calculando, siendo 1 el índice de vacunación y 2 el índice de contagio. El cálculo del índice es una ponderación entre las variables que corresponden al índice (X_{ik}) y un peso (w_{ik}) que se ajusta a cada una de ellas.

2.2.1 Índice de vacunación

El índice de vacunación está compuesto por variables que representan lo relacionado con estado de avance de la vacunación. Dichas variables se calculan por ente territorial y se usaron los siguientes:

Tasa de vacunación de la última semana (TVus): Cantidad promedio de dosis aplicadas diariamente en la última semana. Se obtiene calculando el promedio de la tasa de vacunación diaria de los últimos siete días.

Días de inventario (DInv): Es la cantidad de tiempo, en días, que se va a contar con inventario existente de vacunas si se sigue el supuesto de que se está vacunando a la tasa promedio de la última semana.

Capacidad utilizada vs. Capacidad necesaria (ICap): La capacidad utilizada hace referencia a la capacidad de vacunación que se está utilizando a la tasa promedio de vacunación de la última semana, con base en la cantidad de puntos de vacunación habilitados por departamento. Por otro lado, la capacidad necesaria es la cantidad de dosis que se deberían aplicar diariamente por punto de vacunación habilitado para cumplir con la meta establecida por el PNV. Finalmente, la variable se define como la capacidad que se está utilizando sobre la capacidad que se necesitaría utilizar para alcanzar la meta nacional de vacunación.

Población por punto de vacunación (PobPV): Proporción entre la cantidad de la población que se va a inmunizar y los puntos de vacunación que se encuentran habilitados para covid-19.

Porcentaje de avance en vacunación (AV): Cantidad de vacunas que se han aplicado con respecto al total de vacunas que se deben aplicar para cumplir con la población por inmunizar.

2.2.2 Índice de contagio

El índice de contagio se calcula con base a las variables que representan el estado de contagio por la enfermedad. Dichas variables se calculan por ente territorial y son las siguientes:

Casos activos x10.000 habitantes (CAhab): Cantidad de personas que se encuentran actualmente con covid-19 por cada 10.000 habitantes.

Casos de muerte x10.000 habitantes (CMhab): Cantidad de personas que fallecieron a causa del covid-19 por cada 10.000 habitantes.

Razón de transmisibilidad (Rt): Proporción de los nuevos contagios en la semana i con los nuevos contagios en la semana $i-1$.

Tasa de incidencia x10.000 habitantes (TIn): Número de personas diagnosticadas como positivo para covid-19 por cada 10.000 habitantes.

Número de pruebas última semana (NumPsem): Promedio de la cantidad de pruebas que se realizaron en la última semana.

2.2.3 Asignación de pesos a las variables

En la construcción de los índices es necesario asignar un peso a cada variable. El peso asociado se calcula por medio de un ajuste a través de un modelo de regresión lineal, en el cual se establece la restricción de que los valores asociados a los betas sean entre 0 y 1, y la suma de todos ellos sea igual a 1. El modelo propuesto supone que cada variable de contagio puede ser explicada a partir de las variables de vacunación y cada variable de vacunación podría ser explicada a partir de las variables de contagio. A partir de lo anterior, se establecen las variables independientes como las variables explicativas y la variable dependiente cada una de las variables utilizadas. Para obtener los pesos de cada variable, se realizaron tantas regresiones lineales como variables, es decir, las variables que constituyen los índices 1 y 2. A partir de los resultados de las regresiones lineales se obtienen diferentes betas para cada variable, los cuales se promedian y así obtener un peso para cada una de ellas.

2.3 Clusterización

Luego de construir los índices de vacunación y contagio para cada ente territorial, se procede a realizar la clusterización de los departamentos. Existen diversos métodos [9] y debido a la característica de los índices, índices de distancias euclidianas, se propone la utilización del método de clusterización k-medias, para este es necesario establecer un número para k, el cual representa la cantidad de clústeres que se van a realizar [10]. Una estrategia simple es propuesta para el cálculo de k y es a través de *kmeans()* del paquete *stats* en R, a partir de los datos, generar un gráfico para diferentes valores de k, estableciendo la ubicación de una curva (codo) en el dibujo obtenido, ese se considera como un indicador del número apropiado de clústeres óptimo a utilizar [11].

Una vez se tiene identificado el número óptimo de clústeres, se procede con la agrupación por k-medias. En primer lugar, se debe establecer una medida, que hace referencia a la varianza dentro del clúster y es lo que se pretende minimizar dentro del modelo. Para efectos de esta metodología, se propone utilizar la distancia euclíadiana al cuadrado. Luego, se resuelve el algoritmo de clusterización y se encuentran los diferentes agrupamientos por departamentos. Finalmente, se grafican los resultados y de acuerdo con los clústeres generados se realiza un análisis para la toma de decisiones en el contexto de la situación por la pandemia del covid-19 en el país.

2.4 Graficación y análisis

Con los índices calculados para cada departamento (ente territorial), se realiza una gráfica 2D, en la cual el índice de vacunación (índice 1) representa la coordenada en el eje x, y el índice de contagio (índice 2), representa la coordenada en el eje y, como se puede ver en la Gráfica 1.

El rango de valores de cada eje está entre 0 y 1. Para cada eje, a mayor valor del indicador representa una mejor condición del ente territorial, quiere decir que, para un decisor, se busca que los entes territoriales sean graficados en la parte superior derecha del gráfico. Existen diferentes escenarios posibles dentro del gráfico, pues también puede pasar que la situación de un ente territorial en cuanto a contagio se encuentre controlada, pero la vacunación no tenga un avance significativo, o viceversa.

En la Gráfica 2, se propone la clusterización que permite agrupar a las entidades territoriales según su posición dentro del gráfico y, por tanto, plantear estrategias conjuntas para los diferentes grupos.

2.5 Acciones de mitigación

Como ya se ha mencionado el objetivo de los grupos es establecer acciones conjuntas para cada uno de los grupos identificados, esto permite un mejor uso de recursos. Estas acciones deben ser definidas por los expertos en mitigación de los efectos de la enfermedad.

3 Caso de aplicación, resultados y análisis

La metodología antes propuesta fue aplicada utilizando datos obtenidos del proceso de vacunación y del avance de la enfermedad de los datos abiertos [8] publicados para el 19 de febrero de 2021 al 21 de noviembre de 2021. A continuación, se presentan los resultados observados y la interpretación de las gráficas.

3.1 Índices

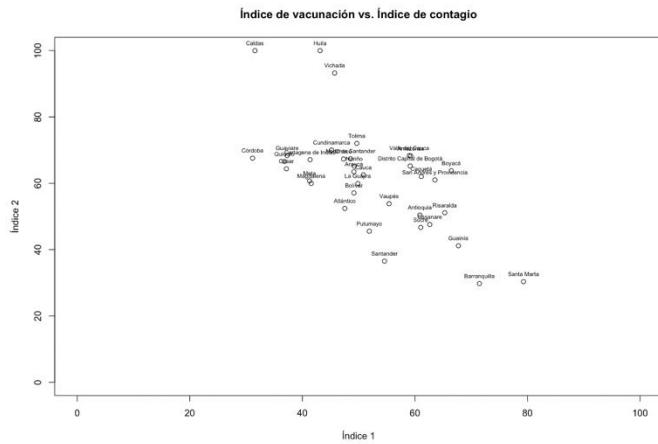
Los índices de vacunación y contagio se calculan departamentalmente a partir de las variables y los pesos asignados a cada una de ellas. Siguiendo la fórmula establecida para cada índice, se tiene que:

Índice de Vacunación:

$$\text{Índice 1} = x_1 * w_1 + x_2 * w_2 + x_3 * w_3 + x_4 * w_4 + x_5 * w_5$$

Índice de Contagio:

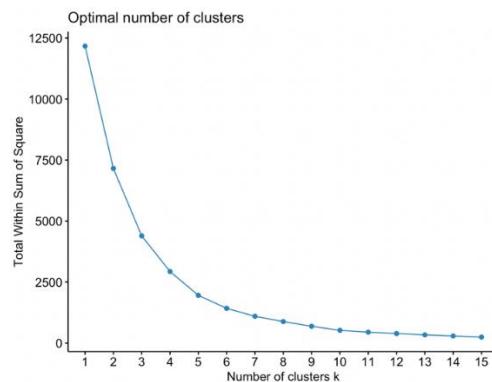
$$\text{Índice 2} = x_1 * w_1 + x_2 * w_2 + x_3 * w_3 + x_4 * w_4 + x_5 * w_5$$



Gráfica 1. Índice de vacunación vs. Índice de contagio

3.2 Clusterización, graficación y análisis

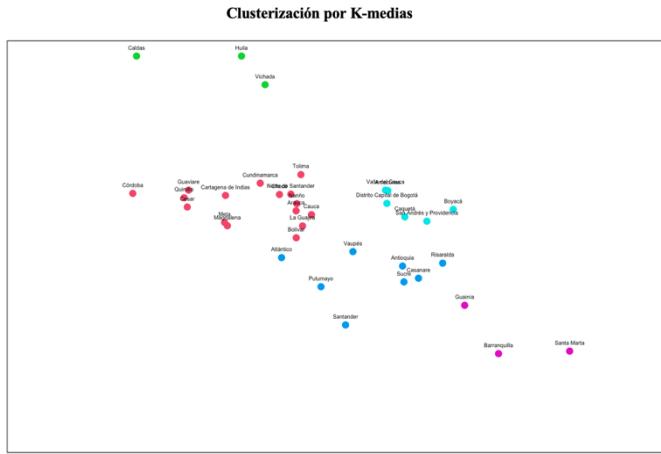
El objetivo de la clusterización es clasificar a los entes territoriales en diferentes grupos de tal manera que los entes pertenecientes a un grupo sean lo más similares entre sí como sea posible. Esto, con el fin de plantear diferentes estrategias y políticas de control basadas en el contagio y la vacunación. En la Gráfica 1 se muestra que el número de k-grupos que mejor se ajusta en la graficación de la suma de cuadrados internos de los grupos es 5. Para reconocer el número óptimo de clústeres, se debe observar el punto del gráfico en el cual, la suma total de cuadrados internos empieza a ser constante.



Gráfica 2. Número óptimo de grupos

La Gráfica 3, representa la clusterización por entidades territoriales, donde en el eje x, se observa el número de grupos y en el eje y, la suma total de cuadrados internos. La clusterización por k-medias tiene como objetivo minimizar la varianza interna de cada clúster, y para establecer el número óptimo de grupos que se van a generar,

se observa la reducción de la suma total de cuadrados internos. En la Gráfica 1 se puede ver que, a partir de cinco grupos, la reducción se estabiliza, lo que indica que utilizar un $k = 5$ es la mejor opción.



Gráfica 3. Clusterización por K-medias

4 Conclusiones y trabajo futuro

En este trabajo se propone la construcción de índices a partir del supuesto de aditividad de las variables usadas en la tabla 1 y 2. Como trabajo futuro es importante considerar efectos de interacción entre variables. La construcción de los índices, tanto de contagio como vacunación, funcionan como métricas que representan la situación actual por la que está pasando cada región geográfica.

La Gráfica 3 es una herramienta para la toma de decisiones que permite observar cómo se encuentra una región geográfica respecto a otras en términos del contagio y la vacunación. Los grupos obtenidos permiten a los tomadores de decisiones plantear estrategias de control por conjuntos que impidan el colapso de los sistemas de salud y el avance en la vacunación en el país.

Otro aspecto a considerar en trabajos futuros es el análisis dinámico y la evolución de los grupos, que, junto a la actualización constante de la información, puede generar dinámicas interesantes para los tomadores de decisión.

Dentro de las limitaciones, sería importante observar la variabilidad de las betas dentro del modelo de regresión lineal y validar con un especialista la relevancia de las variables que se están utilizando para la construcción de los índices, con el fin de obtener un modelo más robusto. Por otro lado, sería interesante poder utilizar el modelo planteado en otro contexto o situación que requiera información departamentalmente para el planteamiento de estrategias y políticas de control frente al problema.

5 Reconocimientos

Este trabajo se desarrolló en el contexto de un estudio de métodos mixtos sobre el diseño de estrategias basadas en inteligencia artificial y ciencia de datos para informar las respuestas de salud pública al COVID-19 en diferentes ecosistemas de salud locales dentro de Colombia (COLEV), proyecto financiado por el Centro Internacional de Investigaciones para el Desarrollo (IDRC), y la Agencia Sueca de Cooperación Internacional para el Desarrollo (Asdi).

6 Referencias

- [1] T. Burki, “Global COVID-19 vaccine inequity,” *Lancet Infect. Dis.*, vol. 21, no. 7, pp. 922–923, 2021.
- [2] G. España, Z. M. Cucunubá, J. Cuervo Rojas, H. Díaz, M. González Mayorga, and J. D. Ramírez, “The impact of vaccination strategies for COVID-19 in the context of emerging variants and increasing social mixing in Bogotá, Colombia: a mathematical modelling study,” 2021.
- [3] L. Arregoces Castillo *et al.*, “Effectiveness of COVID-19 Vaccines in Older Adults in Colombia: First Report of the Esperanza Cohort. A Matched-Pair, National Study,” *A Matched-Pair, Natl. Study*.
- [4] C. Gomez-Restrepo and Z. M. Cucunubá, “cinco propuestas para optimizar la vacunación en Colombia,” 2021.
- [5] M. de S. y P. S. MSPS, “Plan Nacional de Vacunación contra el COVID-19. Documento técnico, v2,” 2021.
- [6] R. Prieto Curiel and H. González Ramírez, “Vaccination strategies against COVID-19 and the diffusion of anti-vaccination views,” *Sci. Rep.*, vol. 11, no. 1, pp. 1–13, 2021.
- [7] Y. Siskos and A. Spyridakos, “Intelligent multicriteria decision support: Overview and perspectives,” *Eur. J. Oper. Res.*, vol. 113, no. 2, pp. 236–246, 1999.
- [8] I. N. de S. INS, “Conjunto de datos abiertos: Casos positivos de Covid-19 Colombia,” 2021.
- [9] L. A. García-Escudero, A. Gordaliza, C. Matrán, and A. Mayo-Iscar, “A review of robust clustering methods,” *Adv. Data Anal. Classif.*, vol. 4, no. 2, pp. 89–109, 2010.
- [10] D. T. Pham, S. S. Dimov, and C. D. Nguyen, “Selection of K in K-means clustering,” *Proc. Inst. Mech. Eng. Part C J. Mech. Eng. Sci.*, vol. 219, no. 1, pp. 103–119, 2005.
- [11] C. G. Martínez, “Métodos de Clustering,” 2018. [Online]. Available: https://rpubs.com/Cristina_Gil/Clustering.