

# Big data para visualizar y analizar la tendencia del COVID-19 en el Perú, 2020-2022

1<sup>st</sup> Luis Enrique Avila Velasquez 2<sup>nd</sup> Alvaro Huaysara Jauregui 3<sup>rd</sup> Luis Alejandro Gabriel Valenzuela Segovia  
*Ingeniería de Sistemas* *Ingeniería de Sistemas* *Ingeniería de Sistemas*  
UNMSM UNMSM UNMSM  
Lima, Perú Lima, Perú Lima, Perú  
luis.avila8@unmsm.edu.pe alvaro.huaysara@unmsm.edu.pe luis.valenzuela3@unmsm.edu.pe

4<sup>th</sup> Gianfranco Moisés Céspedes Francia  
*Ingeniería de Sistemas*  
UNMSM  
Lima, Perú  
gianfranco.cespedes@unmsm.edu.pe

**Abstract**—This article shows some research projects related to the use of Big Data as solutions to different areas of interest, from health with the monitoring, detection, diagnosis and treatment of diseases to the analysis of data generated by the government, companies and citizens, useful for decision making. With this in mind, it is intended to make use of these techniques to process, visualize and analyze data on COVID-19 in the world and then focus this comparison between Peru and Chile, and thus know its trend for the following periods. For this purpose, there is open data on the cases registered from 2020 to the present.

**Index Terms**—Big Data, COVID-19, comparison, trends.

**Resumen**—En el presente artículo se muestran algunos proyectos de investigación relacionados con el empleo del Big Data como soluciones a distintas áreas de interés, desde salud con el monitoreo, detección, diagnóstico y tratamiento de las enfermedades hasta el análisis de datos generados por el gobierno, empresas y ciudadanos, útiles para la toma de decisiones. Con esto en mente, se pretende hacer uso de estas técnicas para procesar, visualizar y analizar datos sobre el COVID-19 en el mundo y luego centrar esta comparación entre Perú y Chile, y así conocer su tendencia para los siguientes periodos. Para este fin se cuenta con datos abiertos sobre los casos registrados desde el año 2020 hasta la actualidad.

**Palabras clave**—Big Data, COVID-19, comparación, tendencias.

## I. INTRODUCCIÓN

A finales del 2019, exactamente el 17 de noviembre, se haría pública la aparición del COVID-19 al detectarse el primer paciente con dicho virus en la ciudad de Wuhan, China. Este virus terminaría trayendo diversos problemas para los países asiáticos, que, en vista de no encontrar una vacuna a tiempo para contrarrestarlo, seguiría expandiéndose por otros países y a mayor escala, continentes. Para darnos una idea de la velocidad de propagación que tuvo el virus, a finales de febrero del siguiente año el presidente de turno en Perú, Martin Vizcarra, anunciaría de manera oficial el primer paciente contagiado del virus del país donde en ese momento el COVID-19 había

llegado a un total de 98.123 infectados, 3.385 fallecidos y en un total de 87 países.

Actualmente en cifras totales en todo el mundo el COVID-19 lleva un acumulado de 586,435,868 contagios y 6,423,686 decesos, que si bien se puede considerar una cifra alarmante la línea de tiempo sobre los contagios ha ido cambiando de manera positiva, debido a que por mucha diferencia el año 2021 fue partícipe de uno de los momentos más cruciales que tuvo el COVID-19, no solo en base a contagios / muertes, sino también en los diversos ámbitos que termina afectando, como lo social, salud o económico. La aparición de las vacunas, ya sea por parte de países como china, rusia u otros, fue un factor favorable para la reducción de contagios donde se pudo ver como varios países optaron por su uso en numerosas dosis con la finalidad de disminuir la preocupación de la población y de esta manera se pueda reactivar sectores económicos volviendo viable la reaparición de empresas pertenecientes a varios sectores, favoreciendo a la creación masiva de empleos para los ciudadanos.

Teniendo ya en consideración la evolución que tuvo el COVID-19 en estos casi tres años además de conocer la falta de políticas para contrarrestar este tipo de pandemia por parte de los gobiernos nacionales, en este trabajo se busca aprovechar la información recopilada por la Universidad de Oxford para analizar el impacto generado por el COVID-19 en múltiples economías del mundo para determinar mediante cifras a través del tiempo la evolución de los casos / muertes en cada territorio haciendo uso de herramientas de Big Data aprovechando herramientas como Power Bi para la visualización de información como también el lenguaje de programación Python y el entorno de ejecución Google Colab para el desarrollo de gráficas personalizadas destinado a un análisis comparativo entre distintos países.

## II. MARCO TEÓRICO

### A. Big data

Se entiende por Big Data como aquellos procedimientos computacionales que están dirigidos al análisis de una gran cantidad de datos con el objetivo de obtener información con valor para una determinada organización. [1]

La tendencia es la integración de big data con IoT o también conocido como internet de las cosas, los datos ya no solo se obtendrán de sensores sino también de sistemas biométricos, cámaras de monitoreo entre otros. Esta integración beneficia en particular al sector sanitario a través del uso de sensores y cámaras térmicas, así como también la identificación de focos infecciosos mediante los biosensores, el presente trabajo está enfocado a los datos de la reciente pandemia. [1]

El Big Data como tal está caracterizado principalmente por el volumen, variedad y velocidad, adicionalmente otros autores agregan la veracidad y el valor. Donde el término volumen se refiere a una gran cantidad de datos, variedad se refiere a las múltiples fuentes de donde se obtienen los datos, velocidad, a la rapidez con que se generan estos datos en periodos determinados, veracidad, a la autenticidad de los datos, es decir datos confiables y por último, el valor que se refiere a obtener información relevante para que pueda ayudar en la toma de decisiones. [1]

### B. COVID-19

La pandemia ocasionada por COVID-19 en el año 2020 tuvo como resultado un efecto devastador en la economía y salud en la población mundial, cuyas implicaciones sociales para los próximos años es aún incierta.

Sobre la transmisión del virus, en la revisión de 22 tipos de coronavirus, tanto el SARS-CoV, el MERS-CoV y otros tipos de coronavirus pueden persistir en superficies inanimadas como metal, vidrio o plástico por hasta nueve días, proporcionando evidencias sólidas de la supervivencia ambiental del patógeno. A estas evidencias se suma la contaminación del agua por heces de personas infectadas, ampliando otra posible vía de transmisión de esta enfermedad. [1]

La COVID-19 ha puesto en evidencia la fragilidad de la humanidad ante un virus que ha demostrado ser altamente resistente y peligroso, con un factor de propagación sin precedente alguno a escala global en tiempos modernos, exponiendo la falta de preparación del sector sanitario en atender una oleada creciente de infectados que asciende a millones y de víctimas fatales que suma cientos de miles.

Es normal que los virus cambien y evolucionen a medida que pasa el tiempo y se van propagando entre las personas. Cuando estos cambios hacen que un virus pase a ser sensiblemente distinto del virus original, se dice que el nuevo virus es una "variante". Para detectar variantes, los científicos analizan el material genético de los virus, proceso conocido como secuenciación, y luego buscan diferencias entre ellos para determinar si han cambiado. Dado que el virus SARS-CoV-2, causante de la COVID-19, se ha propagado por todo el planeta, han surgido y se han descrito variantes en muchos países del mundo. [1]

Entre las variantes que causaron mayor preocupación en el mundo tenemos a la variante Delta, la cual fue identificada en la India los primeros meses del 2020. Esta variante se propaga más fácilmente que otras variantes previamente registradas y las vacunas contra la covid-19 que se utilizaban en ese momento eran seguras y eficaces para prevenir una afección grave o la muerte a causa de la variante delta. Otra de las variantes que causó más preocupación y generó una gran cantidad de contagios es la variante Ómicron, esta variante fue identificada por primera vez en Sudáfrica en noviembre del 2021 y su propagación fue más rápida que cualquier otra variante anteriormente registrada, con un tiempo de duplicación de 2 a 3 días por lo que el riesgo general para esta variable fue de muy elevado.

## III. REVISIÓN DE LA LITERATURA

[2] En este trabajo mencionan el impacto del COVID-19 que generó un aumento en la cantidad de pacientes que tienen depresión ante esto los cuestionarios que se utilizaban para diagnosticar la depresión no son prácticos pues necesitan la participación activa del paciente; tienen como objetivo diagnosticar la depresión temprana para que puedan darle un tratamiento cuanto antes a través de un sistema de predicción de depresión automático y que recomiende el tratamiento adecuado. Como parte de su metodología toman las características depresivas multimodales según el Manual Diagnóstico y Estadístico de los Trastornos Mentales (DSM-5) además recopilan datos de los teléfonos inteligentes que van a una nube y el procesamiento que se le da está basado en big data.

[3] Este estudio tiene como objetivo tomar la gran cantidad de datos de datos clínicos de centros de investigación y hospitales para realizar análisis efectivos. La metodología empleada tiene como base identificar pacientes a través de big data y un procesamiento difuso donde calculan el Número de Identificación de la Enfermedad de Alzheimer (ADIN), esto permite calcular la estimación a corto plazo de la evolución de la enfermedad, así como también optar por un tratamiento adecuado. Sostienen que el ADIN brinda oportunidades para el manejo de esta enfermedad.

[4] Este estudio tiene como objetivo construir un modelo de machine learning para predecir de forma temprana el cáncer de próstata. La metodología utilizada se centra en el estudio de los cambios de frecuencia en 2916 pacientes con cáncer de próstata 3 años antes del diagnóstico con técnicas de machine learning para predecir el diagnóstico. Obtuvieron que el mejor modelo fue el modelo XGBoost ya que obtuvo la mejor precisión y tiene importancia en la detección temprana del cáncer de próstata.

[5] Este estudio tiene como objetivo mejorar el análisis de datos y la extracción de estos en el rubro de la energía eléctrica para así mejorar la capacidad y la calidad del servicio. Su metodología parte del uso de big data y la

computación en la nube para desarrollar un sistema de información para la gestión de empresas en el rubro de la energía eléctrica. El sistema de recopilación de información sobre el consumo de electricidad tiene una arquitectura distribuida, además establece una gran plataforma de datos en la nube y de ahí ayudan a tomar mejores decisiones a los gerentes.

[6] Este estudio trata sobre el análisis y comparativa entre el algoritmo tradicional de optimización de rutas y un algoritmo de optimización de rutas con big data. Los autores tienen como objetivo solucionar el problema de optimización de ruta tradicional pues toma mucho tiempo encontrar la ruta óptima y lo hacen mediante el diseño de un algoritmo de optimización de la ruta de distribución logística de comercio electrónico, usan el algoritmo NSGA-II para resolver problemas de optimización de múltiples objetivos. La función multiobjetivo considera cinco aspectos: índice de peso, índice de oportunidad, índice de importancia del cliente, índice de ventana de tiempo, índice de ruta total y por último establecen los pesos objetivo de distribución con la cual encuentran la mejor distribución. Los autores lograron demostrar que el algoritmo propuesto resultó ser mejor que el tradicional, lo cual resulta en reducir el tiempo de distribución logística en el campo de comercio electrónico.

[7] El autor propone la viabilidad de las herramientas de Big Data para las organizaciones sanitarias donde se busca medir el impacto. Este estudio concluye evidenciando resultados positivos sobre los sistemas de gestión basados en Big Data Analytics donde se pudo detectar la aparición nuevos modelos tanto para la medicina predictiva como para la personalizada mediante la explotación de dichas tecnologías.

[8] Considerando el constante avance de Big Data e Inteligencia Artificial el autor implementa estos avances, en apoyo con los macrodatos, en búsqueda de la detección de descargas epileptiformes por electroencefalografía, detección de convulsiones y pronóstico. Además del estudio de los algoritmos que implementó el autor, también se hace mención sobre la validez de las herramientas destinadas al manejo de datos buscando de esta forma mitigar el miedo sobre las nuevas tecnologías y apoyando la adopción de las mismas.

Según [9] para poder enfrentar el problema de la emisión de carbono se debe cambiar el enfoque trabajado ya que, al tomar como ciudad de estudio a Hangzhou se pudo ver que las investigaciones basadas en encuestas no presentan relevancia a gran escala debido al reducido panorama de la muestra, entonces, el autor plantea utilizar la información recopilada por el sistema taxis para poder acceder a una gran variedad de datos (velocidad instantánea, kilometraje, dirección) que nos garantiza precisión a la hora de determinar las emisiones de carbono generado por los taxis.

[10] El concepto de productos de Big Data es un panorama

poco estudiado frente a otras tecnologías tomando como referencia la producción de literatura. Todo lo referente a los productos de Big Data es explicado por el autor desde diversos ámbitos como la problemática de su difusión por internet hasta el escenario legal que conlleva su implementación.

[11] Se utilizó técnicas de big data y machine learning para evaluar las condiciones de salud de las personas en estudio y determinar la razón por la cual una persona muere después de ser contagiada con COVID-19. Se llegó a la conclusión de que los factores que agravan la condición de un paciente con COVID19 son: hipertensión arterial, edad avanzada y cualquier otra enfermedad. Esto permitió realizar un análisis predictivo para los grupos referentes a las provincias.

Las técnicas de big data son útiles para el análisis de textos a gran escala apoyando a la investigación científica y periodística. [12] En este trabajo se presenta cuál es la infraestructura necesaria para el análisis de big data a través del despliegue de centros de cómputo distribuido y se valora el uso de las principales herramientas para la obtención de información a través de software comerciales y de paquetes de programación como Python o R, permitiendo aplicar técnicas como el web scraping para la recolección de información.

[13] El autor hace un análisis de las principales características, posibilidades, disciplinas técnicas y tecnológicas que contempla el big data, enfocando la utilidad de la información generada para la toma de decisiones, lo que se puede conseguir empleando técnicas de clasificación, regresión, detección de anomalías, clustering, etc. Una utilidad de estas técnicas que se presenta es la Analítica web, utilizada para conocer el comportamiento de las personas que interactúan con distintos sitios web y que es de interés para empresas dedicadas al comercio electrónico.

[14] El autor habla sobre el uso de los grandes volúmenes de datos a disposición de las administraciones públicas, de la ciudadanía y de las empresas que está generando un nuevo modelo de gestión pública, lo que nombra gobernanza inteligente. El uso combinado de la información generada por la administración pública con las técnicas de minería de datos, big data y las tecnologías de procesamiento del lenguaje natural puede facilitar la toma de decisiones públicas lo que generaría mejores políticas públicas. El paradigma de la gobernanza inteligente persigue dar respuesta a este nuevo escenario, donde los datos toman un rol muy importante.

[15] Los autores mencionan el impacto del big data en la interpretación de los medios sociales. Los social media, caracterizados por su rápida expansión y por la variedad de interacciones y de contenidos que en ellos circulan, se han incorporado al estudio del big data al convertirse en fuente de datos útiles para investigadores, entidades y empresas. La combinación de algoritmos que permitan el análisis de tendencias, y la aplicación de bases de datos cuyas

propiedades encajen en el paradigma big data, permitirán identificar nodos conectados y diseñar estrategias para la difusión selectiva de contenidos.

[16] Los autores de este artículo utilizan la información de los estudiantes como evaluación académica, actividades estudiantiles, evaluación psicológica, el ambiente estudiantil, trabajos prácticos, tareas, proyectos y número de ausencias para tener una visión de la calidad de los alumnos. Luego, aplicando métodos de inteligencia artificial y minería de datos como KNN y SVM predicen el éxito académico de los estudiantes.

[17] El artículo se centra en realizar un sistema de servicio inteligente de la salud física para estudiantes. Recopila datos masivos de los exámenes físicos de los estudiantes y los unifica. Esto no solo busca mejorar la conciencia de los estudiantes universitarios sobre el estado de la salud y el entusiasmo por participar sino que también mejora significativamente la eficiencia del trabajo de las instituciones de gestión de la salud.

[18] El autor propone mejoras en el proceso de educación profesional universitaria mediante el uso de big data y algoritmos de predicción. Lo primero que hace es identificar y analizar los problemas existentes en este ámbito de la educación universitaria, luego elige qué método de minería de datos es el más apropiado para realizar el procedimiento. Después, somete los datos a los métodos elegidos y compara los resultados con datos de verificación.

[19] El objetivo de este artículo es realizar un análisis del desarrollo de industrias emergentes de estudiantes universitarios haciendo uso de big data e inteligencia artificial. Para ello analiza los grandes datos dinámicos que corresponden a las tendencias en la industria, el mercado y la dirección de las industrias emergentes local y su relación con el espíritu empresarial universitario. Esto ayudará a identificar que apoyo teórico necesitan los estudiantes y así proporcionarles una base para el diseño de industrias emergentes y estratégicas.

#### IV. METODOLOGÍA

La metodología del presente artículo inicia desde la obtención de la data, el preprocesamiento, la limpieza de datos y selección de atributos relevantes para la visualización y comparación de gráficas resultantes para dos países, Perú y Chile. Y finaliza con la explicación de la relación e impacto de cada uno con respecto al tiempo.

##### A. Dataset

[20] El conjunto de datos para el presente trabajo se obtuvo gracias a la Universidad Oxford y Our World In-Data, este último es una publicación en línea que presenta datos y resultados empíricos que muestran el cambio en las

condiciones de vida en todo mundo, además los resultados se muestran mediante gráficos y mapas interactivos que muestran las tendencias de cambio por país y región. En conjunto desarrollaron un repositorio confiable de datos sobre temas enfocados a los grandes problemas que se tienen en el mundo. Desde el inicio de la pandemia originada por el brote de COVID-19 muchos investigadores se enfocaron en recopilar datos de todos los países sobre múltiples indicadores para que de ese modo se puedan tomar mejores decisiones; además este conjunto de datos se actualiza constantemente. Los atributos más relevantes de este conjunto de datos se puede visualizar en las siguientes tablas, Tabla. I y Tabla. II.

TABLE I  
DESCRIPCIÓN DE ATRIBUTOS 1

Nº	Atributo	Descripción	Valor
1	iso_code	Identificador único internacional del país continente	Cadena
2	continent	Continente donde se ubica el país	Cadena
3	location	Nombre del país	Cadena
4	date	Fecha en que se realizó el registro	Date
5	total_cases	Casos totales	Entero
6	new_cases	Casos nuevos	Entero
7	total_deaths	Muertes totales	Entero
8	new_deaths	Nuevas muertes	Entero
9	total_cases_per_million	Total casos por millón	Entero
10	new_cases_per_million	Nuevos casos por millón	Entero
11	total_deaths_per_million	Total muertes por millón	Entero
12	new_deaths_per_million	Nuevas muertes por millón	Entero
13	reproduction_rate	Tasa de reproducción	Entero
14	icu_patients	Pacientes uci	Entero
15	hosp_patients	Pacientes hospitalizados	Entero
16	hosp_patients_per_million	Pacientes hospitalizados por millón	Entero
17	weekly_icu_admissions	Admisiones semanales uci	Entero

#### V. RESULTADOS

Luego de procesar el conjunto de datos, enfoque que se aborda para la visualización y el correspondiente análisis es desde los datos globales a los casos particulares en los cuales se realiza la comparación entre Chile y Perú.

Una herramienta que provee la página de donde se obtuvo el conjunto de datos nos muestra el porcentaje de cada variante de COVID-19, el caso particular con los dos países escogidos para este trabajo se muestra en la Fig. 2. Se tiene que la variante predominante en Chile es Omicron (BA.4) con un 50.47% seguido de Omicron (BA.5) con un 43.93%. Para

TABLE II  
DESCRIPCIÓN DE ATRIBUTOS 2

N°	Atributo	Descripción	Valor
18	weekly_icu_admissions_per_million	Admisiones semanales uci por millón	Entero
19	weekly_hosp_admissions	Admisiones semanales hosp	Entero
20	weekly_hosp_admissions_per_million	Admisiones semanales hosp por millón	Entero
21	new_tests	Nuevas pruebas	Entero
22	total_tests	Total pruebas	Entero
23	total_tests_per_thousand	Pruebas totales por mil	Entero
24	new_tests_per_thousand	Nuevas pruebas por mil	Entero
25	new_tests_smoothed	Nuevas pruebas por mil	Entero
26	positive_rate	Tasa positiva	Entero
27	tests_per_case	Pruebas por caso	Entero
28	tests_units	Unidades de prueba	Entero
29	total_vaccinations	Vacunas totales	Entero
30	people_vaccinated	Personas vacunadas	Entero
31	people_fully_vaccinated	Personas totalmente vacunadas	Entero
32	total_boosters	Potenciadores totales	Entero
33	new_vaccinations	Nuevas vacunas	Entero
34	total_vaccinations_per_hundred	Total Vacunaciones por cien	Entero

el caso de Perú se tiene que la variante predominante es Omicron(BA.5) con un 60.58% seguido de Omicron (BA.4) con un 21.50%.

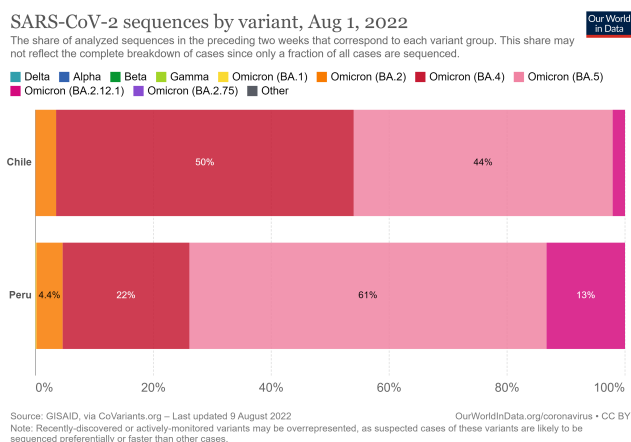


Fig. 1. Porcentaje de las variantes de COVID-19 actualmente

En la Fig. 2 se puede visualizar las 4 ondas que afectaron al mundo entero, empezando por la variante original, luego la variante beta, seguido de la variante delta y por último la

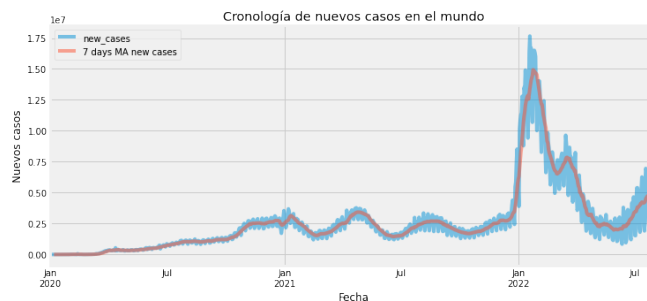


Fig. 2. Nuevos casos a nivel mundial

variante omicron. Esta variante actual se ha convertido en la dominante por mucho, su pico fue más de 4 veces más alto que el pico anterior que corresponde a la variante de onda beta, este fenómeno ocurre principalmente porque, como dice el Dr. Tom Frieden: es tan infeccioso como el sarampión que se propaga en una población no inmune, con un tiempo de incubación mucho más corto por lo tanto un tiempo de duplicación mucho más rápido. Esto se puede contrastar con la Fig. 1 en la cual se puede apreciar que la variante predominante es la Omicron.

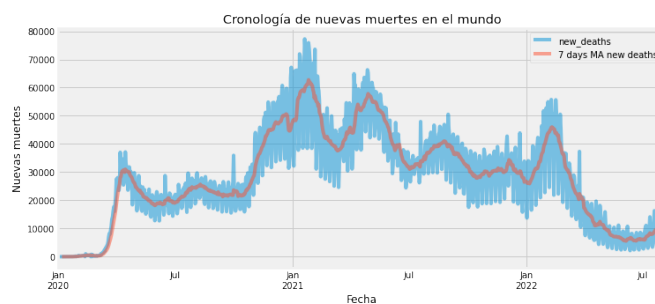


Fig. 3. Nuevas muertes a nivel mundial

Al comienzo de la ola actual ocasionada por la variante Omicron vimos un aumento relativamente pequeño en la curva de nuevas muertes lo que ha hecho que los especialistas formulen una enorme cantidad de hipótesis pendientes: ¿Omicron es menos grave que las variantes anteriores?, sin importar lo que pudieran decir las publicaciones del mundo las autoridades sanitarias están avalando seguir las recomendaciones para evitar contagiarse porque todos estos experimentos aún están en desarrollo. No obstante en un fragmente de una publicación de BBC New que data del 14 de enero del 2022 dice: "Ahora hay mucha confianza en que Omicron es relativamente leve para la mayoría de los adultos, dice la Agencia de Seguridad Sanitaria del Reino Unido en su última evaluación de riesgos. Esto se debe principalmente a que las vacunas brindan una alta protección contra enfermedades graves causadas por esta última variante de COVID-19 que está infectando a muchos."

No es nada nuevo que el objetivo principal de las vacunas es exactamente ese, pero todos sabemos que algunos países todavía tienen problemas para acceder a las vacunas, es por eso que en este trabajo hemos seleccionado dos países, Chile como pionero en la lucha contra el COVID-19 y Perú para

comparar los resultados; por lo que el comportamiento de la pandemia en Chile podría hacer de sus estrategias un modelo a seguir por otros países para que mejoren sus políticas públicas sobre el COVID-19.

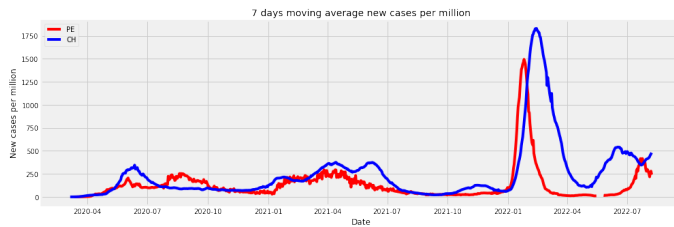


Fig. 4. Comparativa de nuevos casos por millón entre Chile y Perú

En la Fig. 4 se puede apreciar que las dos curvas tienen patrones un poco similares, pero desplazados en el tiempo, en el periodo de julio del 2021 hasta enero del 2022 tanto Chile como Perú disminuyeron los contagios hasta que llegó la variante Omicrón y es donde se disparan los nuevos casos de contagio por COVID-19, pero Chile a pesar de ser uno de los países con mayor porcentaje de vacunados tuvo un pico más alto de contagios a comparación de Perú en donde el pico de contagios fue menor y se redujo rápidamente.

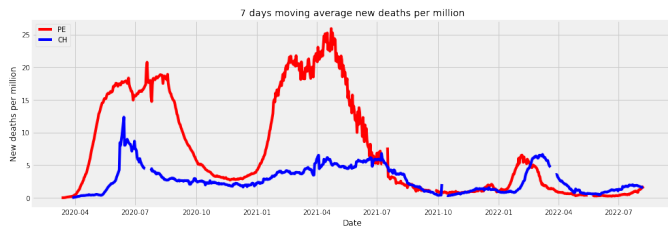


Fig. 5. Comparativa de muertes por millón entre Chile y Perú

Gracias a la Fig. 5 se puede apreciar que entre enero del 2021 y julio del 2021 hubo una diferencia extrema en las muertes por millón pues mientras Chile presenta una baja cantidad de muertes, Perú alcanzaba el pico de fallecidos y esto puede justificarse a través de la cantidad de personas vacunadas y las medidas que adoptaron cada país. A pesar de esta enorme diferencia en cuanto a fallecidos en el periodo mencionado anteriormente, desde finales de julio del 2021 hasta enero del 2022 ambos países tuvieron la misma tendencia de disminución en la cantidad de fallecidos, aunque a inicios del 2022 debido a la variante omicron la cantidad de fallecidos aumentó, pasado ese periodo, es decir desde abril hasta la actualidad la cantidad de fallecidos a disminuido y tiene una tendencia constante.

## VI. CONCLUSIÓN

Si bien los casos en el mundo así como las muertes están influenciadas por el COVID-19, específicamente por sus variantes, los países que tuvieron mejor plan de acción contra esta pandemia redujeron sus cifras, pero a pesar de eso las tendencias con cada variante tienen cierta similitud en el mundo. En el presente trabajo se tomó como caso

particular las comparaciones entre dos países de Sudamérica, uno que fue tomado como modelo gracias a sus políticas sobre la vacunación en la cual casi la totalidad de habitantes se vacunaron desde el inicio y otro que a pesar de no ser tomado como referencia muestra unas tendencias que no son opuesto, es decir no se fue al extremo de ser considerado un país que no tuvo medidas adecuadas para enfrentar la pandemia. Esto nuevamente nos lleva a pensar por qué en nuestro país no estamos preparados para este tipo de situaciones, y no nos referimos solo a una emergencia sanitaria sino por ejemplo frente a un desastre natural como un terremoto no estamos preparados para las consecuencias que trae consigo, por eso es de suma importancia los estudios realizados a través de herramientas o técnicas que nos ayuden a manejar modelos complejos de aprendizaje automático con el objetivo de mitigar estas pérdidas y también para ayudar a una mejor toma de decisiones para los encargados de dirigir el país; Big Data tiene muchas aplicaciones como se pudo observar en todos los estudios revisados y combinado con otras herramientas puede ser una de las soluciones en lo respecta a predicción de fenómenos o tendencias que podrían afectar a la población y la estabilidad de un Estado.

## REFERENCES

- [1] J. Márquez Díaz, "Inteligencia Artificial y Big Data como soluciones frente a la COVID-19," *Rev. Bioética y Derecho Perspect. Bioéticas*, vol. 37, no. c, pp. 315–331, 2020, [Online]. Available: <https://scielo.isciii.es/pdf/bioetica/n50/1886-5887-bioetica-50-00315.pdf>
- [2] J. Hong et al., "Depressive Symptoms Feature-Based Machine Learning Approach to Predicting Depression Using Smartphone," *Healthcare*, vol. 10, no. 7, p. 1189, 2022, doi: 10.3390/healthcare10071189.
- [3] K. Munir, A. de Ramón-Fernández, S. Iqbal, and N. Javaid, "Neuroscience patient identification using big data and fuzzy logic—An Alzheimer's disease case study," *Expert Syst. Appl.*, vol. 136, pp. 410–425, 2019, doi: 10.1016/j.eswa.2019.06.049.
- [4] J. Finkelstein, W. Cui, T. C. Martin, and R. Parsons, "Machine Learning Approaches for Early Prostate Cancer Prediction Based on Healthcare Utilization Patterns," *Stud. Health Technol. Inform.*, vol. 289, pp. 65–68, 2022, doi: 10.3233/SHTI210860.
- [5] X. Zhao, "Research on management informatization construction of electric power enterprise based on big data technology," *Energy Reports*, vol. 8, pp. 535–545, 2022, doi: 10.1016/j.egy.2022.05.124.
- [6] L. P. Qu and H. Li, "Analysis of distribution path optimization algorithm based on big data technology," *J. King Saud Univ. - Sci.*, vol. 34, no. 5, p. 102019, 2022, doi: 10.1016/j.jksus.2022.102019.
- [7] N. Cozzoli, F. P. Salvatore, N. Faccilongo, and M. Milone, "How can big data analytics be used for healthcare organization management? Literary framework and future research from a systematic review," *BMC Health Serv. Res.*, vol. 22, no. 1, 2022, doi: 10.1186/s12913-022-08167-z.
- [8] Y. G. Chung, Y. Jeon, S. Yoo, H. Kim, and H. Hwang, "Big data analysis and artificial intelligence in epilepsy – common data model analysis and machine learning-based seizure detection and forecasting," *Clin. Exp. Pediatr.*, vol. 65, no. 6, pp. 272–282, 2022, doi: 10.3345/cep.2021.00766.
- [9] S. Li et al., "A Big Data-Based Commuting Carbon Emissions Accounting Method—A Case of Hangzhou," *Land*, vol. 11, no. 6, p. 900, 2022, doi: 10.3390/land11060900.
- [10] Y. Ye, Y. Zhang, and Y. Zhu, "Exploring the form of big data products and the supporting systems," *J. Big Data*, vol. 9, no. 1, 2022, doi: 10.1186/s40537-022-00604-4.
- [11] C. Peña, F. Peralta, and R. Hurtado, "Subgroup classification model identifying the most influential factors in the mortality of patients with COVID-19 using data analysis; Subgroup classification model identifying the most influential factors in the mortality of patients with COVID-19 using data analysis," *2020 IEEE Int. Autumn Meet. Power, Electron. Comput.*, vol. 4, 2020, doi: 10.1109/ROPEC50909.2020.9258683.

- [12] J. Serrano Cobos, "Big data y analítica web. Estudiar las corrientes y pescar en un océano de datos," 2014. <https://revista.profesionaldelainformacion.com/index.php/EPI/article/view/epi.2014.nov.01/16929> (accessed Aug. 08, 2022).
- [13] S. Martínez Martínez and P. Lara Navarra, "El big data transforma la interpretación de los medios sociales," 2014.
- [14] A. Cerrillo Martínez, "Datos masivos y datos abiertos para una gobernanza inteligente," 2018. <https://revista.profesionaldelainformacion.com/index.php/EPI/article/view/epi.2018.sep.16/40860>.
- [15] C. Arcila Calderón, E. Barbosa Caro, and F. Cabezuelo Lorenzo, "Big data techniques: Large-scale text analysis for scientific and journalistic research," 2016, doi: 10.3145/epi.2016.jul.12.
- [16] F. Ouatik et al., "Predicting Student Success Using Big Data and Machine Learning Algorithms," *International Journal of Emerging Technologies in Learning*, vol. 17, no. 12, p. 236–251, 2022, doi: [doi.org/10.3991/ijet.v17i12.30259](https://doi.org/10.3991/ijet.v17i12.30259).
- [17] Y. Jifang et al., "Construction of Intelligent Service System for Adolescent Students' PE Based on Big Data Analysis," *Wireless Communications and Mobile Computing*, vol. 2022, no. 1, p. 8, 2022, doi: [doi.org/10.1155/2022/4922918](https://doi.org/10.1155/2022/4922918).
- [18] L. Cai et al., "Prediction and Influencing Factors of College Students' Career Planning Based on Big Data Mining," *Mathematical Problems in Engineering*, vol. 2022, no. 1, p. 11, 2022, doi: [10.1155/2022/5205371](https://doi.org/10.1155/2022/5205371).
- [19] Y. Pan et al., "Analysis of the Correlation between Emerging Industry Development and University Students' Entrepreneurship Based on Big Data," *Discrete Dynamics in Nature and Society*, vol. 2022, no. 1, p. 11, 2022, doi: [10.1155/2022/1098903](https://doi.org/10.1155/2022/1098903).
- [20] E. Mathieu, H. Ritchie, and E. Ortiz-Ospina, "Coronavirus (COVID-19) Vaccinations - Our World in Data," 2022. <https://ourworldindata.org/covid-vaccinations> (accessed Aug. 08, 2022).