ReHuSo
Revista de Ciencias Humanísticas y Sociales

Predicción de evolución post covid19 en pacientes, usando herramientas de la big data

Prediction of post covid19 evolution in patients, using big data tools

Gilberto Horacio Fernández Cedeño [1] 0009-0008-4508-9957

Marely del Rosario Cruz Felipe [2] 0000-0003-1937-1568

Ermenson Ricardo Ordóñez-Ávila[3] 0000-0003-2583-2076

José Gabriel Moreira Vélez[4] 0000-0002-1743-322X

[1]*Universidad Técnica de Manabí, Ecuador*.gfernandez5134@utm.edu.ec

[2]*Universidad Técnica de Manabí, Ecuador*.marely.cruz@utm.edu.ec

[3]*Universidad Técnica de Manabí, Ecuador*.emerson.ordonez@utm.edu.ec

[4]*Universidad Técnica de Manabí, Ecuador*.jose.moreira@utm.edu.ec

Recepción: 22 de Febrero del 2023 / Aceptación: 22 de Mayo del 2023 / Publicación: 05 de Julio del 2023

Citación/como citar este artículo: Fernández, G., Cruz, M., Ordóñez, E. y Moreira, J. (2023). Predicción de evolución post covid19 en pacientes, usando herramientas de la big data. *ReHuSo, 8*(2), 125-136. https://doi.org/10.33936/rehuso.v8i2.5911

ReHuSo: Revista de Ciencias Humanísticas y Sociales
e-ISSN 2550-6587
https://revistas.utm.edu.ec/index.php/Rehuso/index
Vol. 8 Núm. 2 (125-136): Julio – Diciembre 2023
rehuso@utm.edu.ec
Universidad Técnica de Manabí
DOI: https://doi.org/10.33936/rehuso.v8i2.5911

Resumen

Esta investigación tiene como objetivo predecir la evolución post COVID-19 en pacientes del Hospital General de Portoviejo (IESS), identificando patrones similares en la propagación de casos futuros de esta enfermedad. Como metodología se efectuó un estudio de tipo descriptivo, retrospectivo, con enfoque cuantitativo y empleo del método de análisis documental, donde se realizó la extracción de información desde la base de datos del referido hospital, en el período 2020-2022. Para el análisis de los datos se utilizó el software Orange Data Mining, que es una herramienta de código abierto con una amplia gama de métodos de análisis de datos y aprendizaje automático. Del total de 18316 pacientes, se trabajó con una muestra intencional de 3678, por contar estos con los datos requeridos para el análisis. Entre los principales resultados se destaca que las personas más propensas a tener Covid, están en el rango de edades entre los 63 y 70 años; el sexo más expuesto es el masculino; los síntomas más comunes por los afectados son la insuficiencia respiratoria y enfermedad renal crónica, cuestiones que ayudan a predecir cuáles serán los pacientes que pudieran ser más propensos a contraer la enfermedad. A modo de conclusión se resalta que la aplicación de herramientas de minería de datos facilita la predicción y evolución futura de enfermedades como la analizada, facilitando la toma de decisiones en materia de prevención y control de la pandemia a las autoridades sanitarias.

Palabras clave

COVID-19, predicción, minería de datos, orange data mining, prevención.

Abstract

This research aims to predict the post COVID-19 evolution in patients at the Portoviejo General Hospital (IESS), identifying similar patterns in the spread of future cases of this disease. As a methodology, a descriptive, retrospective study was carried out, with a quantitative approach and use of the documentary analysis method, where information was extracted from the database of the aforementioned hospital, in the period 2020-2022. For the data analysis, the Orange Data Mining software was used, which is an open-source tool with a wide range of data analysis and machine learning methods. Of the total of 18,316 patients, an intentional sample of 3,678 was used, since they had the data required for analysis. Among the main results, it stands out that the people most likely to have Covid are in the age range between 63 and 70 years; the most exposed sex is the male; The most common symptoms for those affected are respiratory failure and chronic kidney disease, issues that help predict which patients may be more likely to contract the disease. In conclusion, it is highlighted that the application of data mining tools facilitates the prediction and future evolution of diseases such as the one analyzed, facilitating decision-making on the prevention and control of the pandemic for health authorities.

Keywords

COVID-19, prediction, data mining, orange data mining, prevention.

Introducción

En la historia reciente de la humanidad, siempre ha existido una exitosa simbiosis entre la tecnología y la medicina, desde simples algoritmos para determinar soluciones médicas, hasta los más complejos artefactos tecnológicos para dar una mejor determinación de la enfermedad (tales como resonadores, tomógrafos, etc.). La pandemia del COVID-19 ha sido un evento sin precedentes en la historia moderna que ha afectado a la economía, la sociedad y la salud en todo el mundo (Inca Ruiz y Inca León, 2020). En este contexto, la aplicación de la tecnología y, en particular, la aplicación de la "big data" han tenido un papel crucial en la gestión de la pandemia.

La "big data" se refiere a grandes conjuntos de datos que son demasiado complejos para ser analizados con herramientas de procesamiento de datos tradicionales. La capacidad de procesar grandes cantidades de datos ha permitido a los expertos en salud pública hacer predicciones más precisas y tomar decisiones basadas en datos en tiempo real (García et al., 2016). En el contexto del COVID-19, la aplicación de la "big data" ha permitido a los expertos en salud pública monitorear la propagación del virus, identificar los patrones de propagación y predecir los posibles puntos críticos. Además, la capacidad de analizar grandes conjuntos de datos les ha facilitado identificar factores de riesgo y evaluar la efectividad de las medidas de prevención y control (Naeem et al., 2022).

Los modelos de predicción basados en la "big data" se han utilizado ampliamente para predecir la propagación del COVID-19 y la capacidad de los sistemas de salud para atender a los pacientes. Estos modelos utilizan datos históricos y en tiempo real para predecir la propagación del virus y proporcionar información útil para la toma de decisiones. En resumen, la aplicación de la "big data" ha sido esencial en la gestión de la pandemia del COVID-19 (Brownlee, 2016).

La capacidad de procesar grandes cantidades de datos ha permitido a los expertos en salud pública hacer predicciones más precisas y tomar decisiones basadas en datos en tiempo real. Con la pandemia actual, una vez más el mundo se une para mitigar y resolver de la manera más precisa, eficaz y rápida esta crisis, se acota que la tecnología se presta como herramienta para colaborar con la medicina contra el mortal virus (Naeem et al., 2022). La unión entre la tecnología y la medicina siempre ha tenido resultados satisfactorios para la humanidad, y continúa en evolución, siendo una aliada indispensable. La aparición de la pandemia COVID-19 reflejó que en el mundo se requiere la ciencia de datos, que tiene como objetivo descubrir el conocimiento de estos grandes datos a través de algoritmos de minería de datos, herramientas de aprendizaje automático, modelos matemáticos y estadísticos, análisis de datos y análisis visual (Leung et al., 2020). El análisis de datos mediante estos recursos ofrece un camino para entender mejor el evento y cómo las organizaciones pueden elaborar estrategias y responder en esta nueva era mediante el uso de varios grandes métodos analíticos de datos. Entre las herramientas usadas esta la gestión de minería, modelado de datos y la Big Data.

A propósito, el presente estudio tiene como objetivo predecir la evolución post Covid 19 en pacientes del Hospital General de Portoviejo (HGP-IESS), identificando patrones similares en la propagación de casos futuros de esta enfermedad.

ReHuSo: Revista de Ciencias Humanísticas y Sociales
e-ISSN 2550-6587
https://revistas.utm.edu.ec/index.php/Rehuso/index
Vol. 8 Núm. 2 (125-136): Julio – Diciembre 2023
rehuso@utm.edu.ec
Universidad Técnica de Manabí
DOI: https://doi.org/10.33936/rehuso.v8i2.5911

Metodología

Se realizó un estudio de tipo descriptivo, retrospectivo, con enfoque cuantitativo y empleo del método de análisis documental, donde se efectuó la extracción de información desde la base de datos del referido hospital, en el período 2020-2022. Del total de 18316 pacientes, se trabajó con una muestra intencional de 3678, por contar estos con los datos requeridos para el análisis.

Con el propósito de encontrar un modelo adecuado que permita la predicción de casos de COVID-19 dentro de la provincia de Manabí, se presenta una investigación utilizando herramientas de big data. En este sentido, se realizó un análisis exploratorio de datos, considerando el criterio de Hastie et al. (2009) quiénes plantean que antes de realizar cualquier predicción, es importante comprender los datos que se están utilizando. Para los autores, el análisis exploratorio de datos permite identificar patrones y relaciones en los datos, lo que puede ser útil para desarrollar modelos predictivos. Dentro de la investigación el análisis exploratorio de datos (AED) permitió agrupar patrones que pueden incidir en los casos de contagios por COVID, tendencias y relaciones en un conjunto de datos. El objetivo principal del AED es obtener una comprensión inicial de los datos antes de realizar cualquier análisis estadístico formal o construir modelos de predicción.

Por su parte, Wickham & Grolemund (2017) mencionan que el AED se utiliza para realizar una exploración visual y estadística de los datos. Esto puede incluir en la identificación de valores atípicos, la comprobación de la normalidad de los datos, la búsqueda de patrones y tendencias, la identificación de relaciones entre variables, la identificación de valores faltantes y la exploración de distribuciones de frecuencia y estadísticas descriptivas. El AED es una técnica importante en el proceso de análisis de datos, ya que ayuda a los analistas a identificar cualquier problema con los datos antes de realizar cualquier análisis estadístico formal. También puede proporcionar información útil para la selección de variables en el análisis estadístico, la selección de modelos de predicción y la toma de decisiones.

Para el análisis de los datos se utilizó el software Orange Data Mining, que es una herramienta de código abierto con una amplia gama de métodos de análisis de datos y aprendizaje automático. Para utilizar Orange Data Mining, según Thange et al. (2021) en la predicción de casos de COVID-19, se requiere un conjunto de datos históricos que contenga información relevante, como el número de casos diarios, las características demográficas de la población, y otros factores relacionados con la propagación del virus, toda esta información fue extraídos de la base de datos del Hospital General de Portoviejo (IESS) entre 2020 y 2022. En este software se trabajó con cuatro modelos predictivos: Random Forest, Adaboost, Naive Bayes y Neural Network, mostrando mayor precisión el modelo Adaboost, cuyo algoritmo de aprendizaje automático permite la clasificación y regresión del conjunto de datos de estudio (Sujatha et al., 2022). Después de que se hayan generado los modelos de predicción, se pueden evaluar utilizando métricas adecuadas, como el error cuadrático medio o la precisión. Esto

ayuda a determinar la eficacia de los modelos y su capacidad para predecir con precisión los casos de COVID-19 (Ong et al.,2022).

El primer paso en el proceso de predicción es cargar y preparar los datos en Orange Data Mining. Esto implica la limpieza de los datos, la selección de características relevantes y la transformación de los datos en un formato adecuado para su procesamiento. Una vez que los datos están listos, se pueden aplicar diferentes técnicas de aprendizaje automático en Orange Data Mining para desarrollar modelos de predicción (Banluesapy & Jirapanthong, 2022).

Los datos correspondientes a las variables: sexo, edad, provincia, cantón, parroquia, síntomas y saturación de oxígeno, fueron introducidos en el software Orange Data Mining, arrojando diferentes resultados. Asimismo, se interrelacionaron las variables, realizando diferentes agrupaciones para identificar tendencias. Esta selección de variables puede ayudar a reducir el ruido en los datos y mejorar la precisión de las predicciones (Herrera et al.,2021).

A continuación, se muestran los resultados a partir de la aplicación de la herramienta.

Resultados

La primera variable con la que se trabajó fue el sexo. En la figura 1, se puede apreciar que la mayor cantidad de pacientes con COVID-19 corresponde al sexo masculino, representados por el color rojo (Covid identificado – Covid no identificado).

Fuente: Elaboración propia del autor

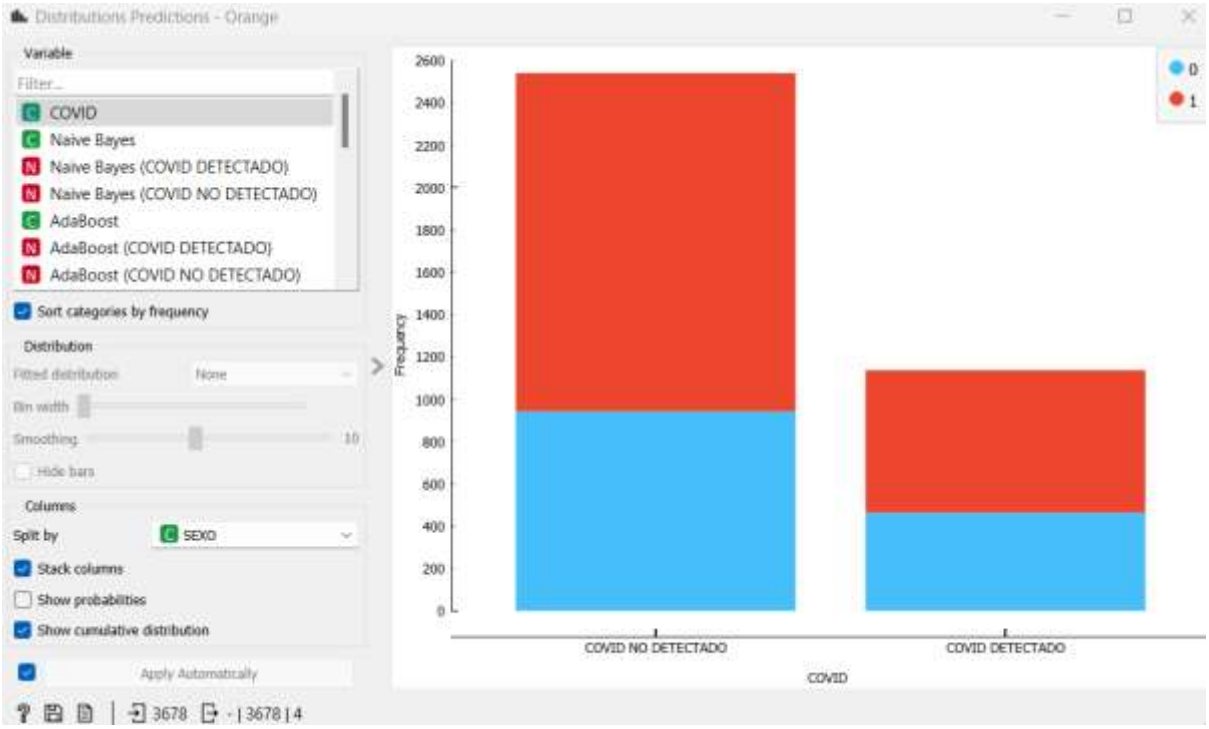

Fig. 1: *Casos COVID-19 agrupados por género femenino y masculino*

ReHuSo: Revista de Ciencias Humanísticas y Sociales
e-ISSN 2550-6587
https://revistas.utm.edu.ec/index.php/Rehuso/index
Vol. 8 Núm. 2 (125-136): Julio – Diciembre 2023
rehuso@utm.edu.ec
Universidad Técnica de Manabí
DOI: https://doi.org/10.33936/rehuso.v8i2.5911

Con la intención de identificar la cantidad de contagiados por sexo y edades, se interrelacionaron estas variables, cuyo resultado se muestra en la figura 2. Se puede observar que la media de las edades es de 63 años, la moda 70 años y la mediana 65 años, con una dispersión del 0.26%, tomando en cuenta edades comprendidas entre 0 y 98 años, siendo el sexo masculino los más afectadas por el virus del COVID-19 (color rojo).

Fuente: Elaboración propia del autor

Fig. 2: *Casos COVID-19 agrupados por edades*

La ubicación geográfica también juega un papel importante al momento de realizar predicciones mediante software de minería de datos, por ello, se tomaron los datos correspondientes a las provincias, nótese que la provincia de Manabí es la que aparece con la mayor cantidad de casos:

Fuente: Elaboración propia del autor

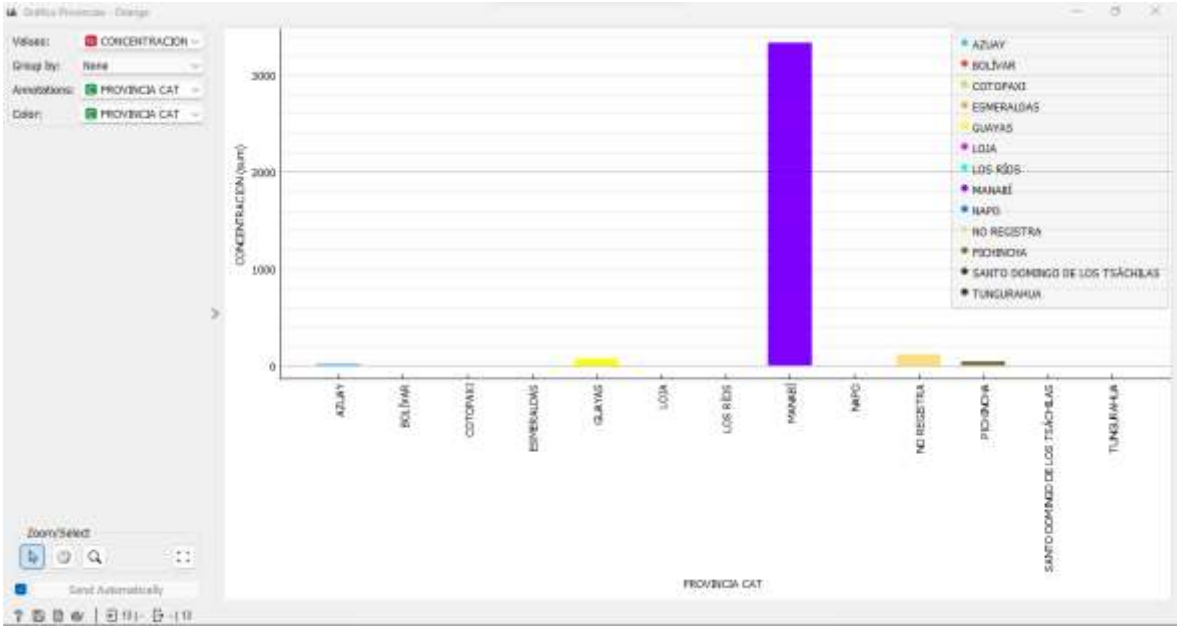

Fig. 3: *Casos COVID-19 registrados en el HGP-IESS agrupados por provincias*

Al identificar que la provincia con mayor cantidad de casos fue Manabí, se decidió revisar el comportamiento por cantones, en donde se puede observar que la mayor concentración de casos COVID-19 se encuentra en el cantón Portoviejo, registrando más de 1600 casos. La figura 4, muestra este análisis:

Fuente: Elaboración propia del autor

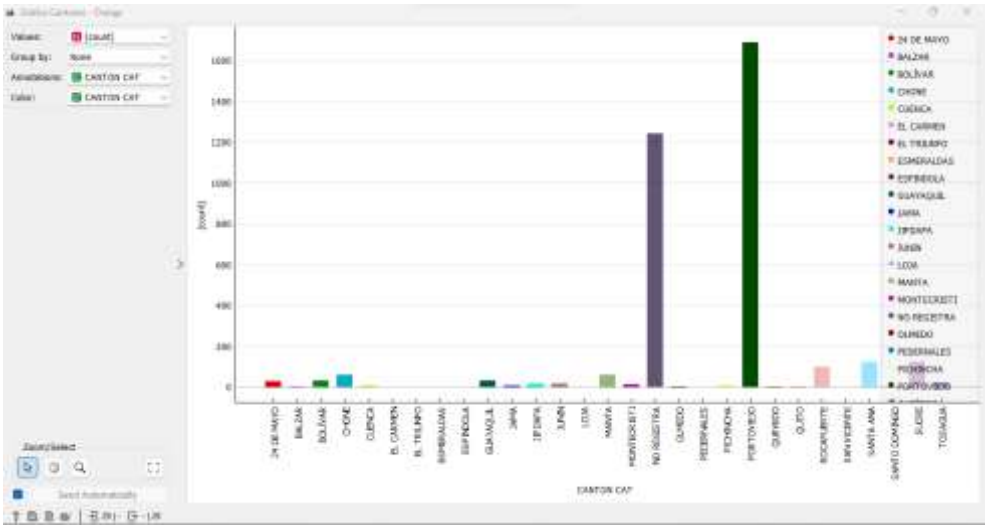

Fig. 4: Casos COVID registrados en el HGP-IESS agrupados por cantones

Al respecto, se consideró pertinente realizar un análisis por parroquias, factor importante al momento de realizar el modelo predictivo, para conocer en las que hubo mayor concentración de casos COVID-19. La figura 5 expone estos resultados, observándose que en la Parroquia de Portoviejo (color azul) está la mayor concentración de casos en comparación con otras parroquias. El color fucsia no registra de qué parroquia proceden los pacientes, por lo que estos datos no son relevantes para el análisis.

Fuente: Elaboración propia del autor

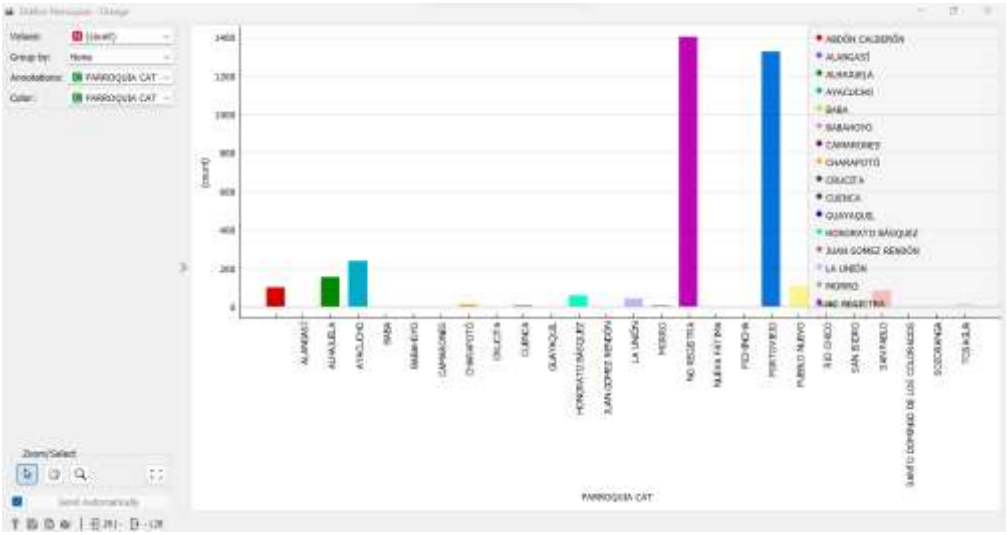

Fig. 5: Casos COVID registrados en el HGP-IESS agrupados por parroquias

ReHuSo: Revista de Ciencias Humanísticas y Sociales
e-ISSN 2550-6587
https://revistas.utm.edu.ec/index.php/Rehuso/index
Vol. 8 Núm. 2 (125-136): Julio – Diciembre 2023
rehuso@utm.edu.ec
Universidad Técnica de Manabí
DOI: https://doi.org/10.33936/rehuso.v8i2.5911

Referente a los síntomas, cuya variable es la más importante y relevante para obtener resultados en los modelos predictivos, se puede observar en la figura 6, el agrupamiento de los síntomas más comunes que los pacientes con COVID tenían al momento de ingresar al HGP, entre los síntomas más comunes se tiene insuficiencia respiratoria, enfermedades renales, hipertensión, infecciones en las vías urinarias.

Fuente: Elaboración propia del autor

Fig. 6: Síntomas más comunes que compartían los pacientes con COVID-19

Los resultados obtenidos a través de la aplicación de Orange Data Mining mostraron que los casos de COVID-19 en Manabí han tenido una evolución variable a lo largo del tiempo, entre el 2020 y 2022. En cuanto a las variables más relevantes que influyen en la propagación del virus en la provincia de Manabí, los resultados indicaron que la densidad poblacional, la ubicación geográfica, el porcentaje de población con edades de 63 a 70 años con ciertas patologías comunes son variables significativas para explicar la evolución de los casos de COVID-19 en la provincia.

En cuanto a la predicción de la evolución de la pandemia en Manabí, los resultados obtenidos a través de la aplicación de modelos de aprendizaje automático mostraron que es posible realizar predicciones precisas sobre el número de casos en la provincia en el corto y mediano plazo, siendo el Modelo Adaboost el que muestra mayor precisión con un 71.5%; sin embargo, es importante destacar que estas predicciones están sujetas a incertidumbres y dependen mucho de las variables utilizadas y, al mismo tiempo, de ciertos factores como la calidad de datos de entrenamiento, tamaño del conjunto de datos, elección del algoritmo de aprendizaje, cambio en datos de entrada, sesgo y variación.

Discusión

El análisis realizado sobre los casos de COVID-19 en la provincia de Manabí, la utilización de herramientas de minería de datos como Orange Data Mining y modelos de aprendizaje automático como Naive Bayes, Redes Neuronales, AdaBoost y Random Forest, demostraron una aplicación efectiva de la tecnología en la lucha contra la pandemia, así también como lo afirman Medel-Ramírez & Medel-López (2020) quiénes alegan que con el uso de minería de datos se puede encontrar un algoritmo que permita identificar pacientes con COVID-19.

Se coincide con Pérez-Milena et al. (2022) que es importante destacar la identificación de las variables más relevantes que influyen en la propagación del virus en la provincia de Manabí, como la densidad poblacional, la ubicación geográfica y el porcentaje de población con edades y patologías específicas.

Los resultados obtenidos dentro de la investigación concuerdan con otras similares, tal es el caso de Crespo (2019) en su artículo Análisis de la encuesta de Salud Nacional y Examen de Nutrición de Estados Unidos (NHANES) usando machine learning en el que se puede destacar que el mejor modelo se obtiene con AdaBoost al tener una exactitud de 76.33, aunque los otros modelos muestran resultados aceptables, el modelo indicado es el que menor cantidad de falsos negativos muestra al momento de compararlos.

Como describen Raftarai et al. (2021) los métodos y algoritmos basados en el aprendizaje automático y la extracción de datos han resultado exitosos para la predicción de la tasa de reingreso, ellos proponen un nuevo modelo predictivo de la tasa de reingreso basado en el clasificador de conjuntos AdaBoost mejorado. El modelo propuesto se basa en técnicas de aprendizaje automático y combina de forma inteligente tres clasificadores en un clasificador de conjunto. Los resultados obtenidos han sido evaluados por precisión 91,61%, sensibilidad 95,80% y valores predictivos positivos (VPP) 90,25% y valores predictivos negativos (VPN) 89,31% y también comparados con clasificadores básicos.

Desde el punto de vista de Byeon (2021) el modelo AdaBoost confirmó que el nivel de educación, el conocimiento de la respuesta COVID-19 de vecinos/colegas, la edad, el género, y el estrés subjetivo fueron cinco variables clave con alto peso en la predicción de la ansiedad inducida por COVID-19 para adultos que viven en comunidades de Corea del Sur. Para una mujer adulta mayor que sentía mucho estrés subjetivo, no asistía a una escuela secundaria, tenía 70.6 años y pensaba que los vecinos y los colegas respondieron a COVID-19 de manera adecuada (exactitud de clasificación = 0,812, precisión = 0,761, recuerdo = 0,812, AUC = 0,688 y puntaje F-1 = 0,740). Como se puede comprobar, la edad de los pacientes tomados en esta investigación tiene una media de 70 años que coincide también con la investigación realizada en este artículo, además se utilizó el modelo de Adaboost con variables de interés para poder realizar la predicción, y obtener resultados más precisos en comparación a los otros modelos.

La validación cruzada utilizada en la investigación, también es una herramienta importante para garantizar la precisión y confiabilidad de los modelos de aprendizaje automático utilizados para predecir la evolución de la pandemia (Villena et al., 2021). La combinación de estas

ReHuSo: Revista de Ciencias Humanísticas y Sociales
e-ISSN 2550-6587
https://revistas.utm.edu.ec/index.php/Rehuso/index
Vol. 8 Núm. 2 (125-136): Julio – Diciembre 2023
rehuso@utm.edu.ec
Universidad Técnica de Manabí
DOI: https://doi.org/10.33936/rehuso.v8i2.5911

técnicas puede ser una herramienta valiosa para las autoridades sanitarias al momento de planificar y ejecutar acciones específicas para controlar y prevenir la propagación del virus en la provincia y otras zonas afectadas.

En cuanto a la implementación de medidas y la toma de decisiones para la prevención y control de la enfermedad, hay que estar atentos a personas con las sintomatologías más comunes: insuficiencia respiratoria, hipertensión esencial primaria, infecciones en las vías urinarias y enfermedad renal crónica.

Conclusiones

El estudio realizado pone de manifiesto que en la actualidad hay una fuerte tendencia a la utilización de la minería de datos para estimar hechos y eventos. En esta investigación, La herramienta de minería de datos Orange Data Mining puede ser valiosa para desarrollar modelos de predicción que ayuden a comprender y anticipar la propagación del virus COVID-19. Sin embargo, es esencial tener en cuenta las limitaciones y la incertidumbre asociada con la predicción de una pandemia en constante evolución. Estos resultados pueden ser útiles para las autoridades sanitarias y para la toma de decisiones en materia de prevención y control de la pandemia en la región.

Quedó evidenciado a partir de la utilización de las herramientas de minería de datos Orange Data Mining, que la evolución de la pandemia de COVID-19 en la provincia de Manabí sigue patrones asociados a: las sintomatologías detectadas en los pacientes, la densidad poblacional, la ubicación geográfica, el porcentaje de población con edades 63 y 70 años, lo que puede ser útil para predicciones precisas sobre el número de enfermos en el corto y mediano plazo y la toma de decisiones en futuras apariciones de casos.

Referencias bibliográficas

Banluesapy S. & Jirapanthong, W. (2022, del 10 al 11 de noviembre). A Prediction Model for Screening COVID-19 Patients [conference]. *2022 6th International Conference on Information Technology (InCIT)*, Nonthaburi, Thailand. https://10.1109/InCIT56086.2022.10067446.

Byeon, H. (2021). Predicting high-risk groups for COVID-19 anxiety using adaboost and nomogram: Findings from nationwide survey in South Korea. *Applied Sciences*, *11*(21), 1-15. https://doi.org/10.3390/app11219865

Brownlee, J. (2016). *Machine Learning Mastery with Python: Understand Your Data, Create Accurate Models, and Work Projects End-to-End*. Machine Learning Mastery. https://n9.cl/lnk4v

Herrera, C. E., Lage, D., Betancourt, J., Barreto, E., Sánchez, L., y Crombet, T. (2021). Nomograma de predicción para la estratificación del riesgo en pacientes con COVID-19. *European Journal of Health Research:(EJHR)*, *7*(2), 1-19. https://doi.org/10.32457/ejhr.v7i2.1592

Crespo, M. (2019). *Análisis de la Encuesta de Salud Nacional y Examen de Nutrición de Estados Unidos (NHANES) usando machine learning*. [Tesis de maestría, Universidad Oberta de Catalunya].UOC. https://hdl.handle.net/10609/99127

García, S., Ramírez-Gallego, S., Luengo, J., & Herrera, F. (2016). Big Data: Preprocesamiento y calidad de datos. *Big Data monografía*, (237), 17-23. https://n9.cl/spesd

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer. https://n9.cl/yumt2

Leung, C. K., Chen, Y., Hoi, C. S.H., Shang, S., Wen, Y. & Cuzzocrea, A. (2020, del 07 al 11 de septiembre). Big Data Visualization and Visual Analytics of COVID-19 Data [conference]. *24th International Conference Information Visualisation (IV)*. Melbourne, Australia. https://doi.org/10.1109/IV51561.2020.00073

Medel-Ramírez, C. & Medel-López, H. (2020). *Data Mining for the Study of the Epidemic (SARS- CoV-2) COVID-19: Algorithm for the Identification of Patients (SARS-CoV-2) COVID 19 in Mexico*. SSRN. https://dx.doi.org/10.2139/ssrn.3619549

Naeem, M., Jamal, T., Diaz-Martínez, J., Butt, S. A., Montesano, N., Tariq, M. I., De-La-Hoz-Franco, E. & De-La-Hoz-Valdiris, E. (2022). Trends and future perspective challenges in big data. In J-S. Pan., V.E. Balas. y C. M. Chen. (Eds.), Advances in Intelligent Data Analysis and Applications. *Smart Innovation, Systems and Technologies* (pp. 309-325). Springer, Singapore. https://doi.org/10.1007/978-981-16-5036-9_30

Ong, A. K. S., Prasetyo, Y. T., Yuduang, N., Nadlifatin, R., Persada, S. F., Robas, K. P. E., Chuenyindee Thanatorn & Buaphiban, T. (2022). Utilization of random forest classifier and artificial neural network for predicting factors influencing the perceived usability of COVID-19 contact tracing "Morchana" in Thailand. *International Journal of*

ReHuSo: Revista de Ciencias Humanísticas y Sociales
e-ISSN 2550-6587
https://revistas.utm.edu.ec/index.php/Rehuso/index
Vol. 8 Núm. 2 (125-136): Julio – Diciembre 2023
rehuso@utm.edu.ec
Universidad Técnica de Manabí
DOI: https://doi.org/10.33936/rehuso.v8i2.5911

Environmental Research and Public Health, 19(13),2-28. https://doi.org/10.3390/ijerph19137979

Pérez-Milena, A., Leyva-Alarcón, A., Barquero-Padilla, R. M., Peña-Arredondo, M., Navarrete-Espinosa, C. & Rosa-Garrido, C. (2022). Valoración y seguimiento de los pacientes con sospecha de COVID-19 en la primera ola pandémica en una zona urbana de Andalucía. *Atención Primaria, 54*(1), 1-8. https://doi.org/10.1016/j.aprim.2021.102156

Raftarai, A., Mahounaki, R. R., Harouni, M., Karimi, M. & Olghoran, S. K. (2021). Predictive models of hospital readmission rate using the improved AdaBoost in COVID-19. In T. Saba. y A. Rehman. (Ed.), *Intelligent Computing Applications for COVID-19* (pp. 67-86). CRC Press. https://doi.org/10.1201/9781003141105

Inca Ruiz, G. P. y Inca León, A.C. (2020). Evolución de la enfermedad por coronavirus (COVID-19) en Ecuador. *La ciencia al servicio de la salud, 11*(1), 5-15. https://dx.doi.org/10.47244/cssn.Vol11.Iss1.441

Sujatha, R., Venkata Siva, B., Chatterjee, J. M., Rahul Naidu, P., Jhanjhi, N. Z., Charita, C., Mariya, E. & Baz, M. (2022). Prediction of Suitable Candidates for COVID-19 Vaccination. *Intelligent Automation & Soft Computing*, *32*(1) https://n9.cl/ktzk3

Thange, U., Shukla, V. K., Punhani, R. & Grobbelaar, W. (2021, del 19-21 de enero). Analyzing COVID-19 Dataset through Data Mining Tool "Orange" [conference]. In *2021 2nd International Conference on Computation, Automation and Knowledge Management (ICCAKM)*. Dubai, United Arab Emirates 10.1109/ICCAKM50778.2021.9357754

Villena-Ortiz, Y., Giralt, M., Castellote-Bellés, L., Lopez-Martínez, R. M., Martinez-Sanchez, L., García-Fernández, A. E., Ferrer-Costa, R., Rodriguez-Frias, F. & Casis, E. (2021). Estudio descriptivo y validación de un modelo predictivo de severidad en pacientes con infección por SARS-CoV-2. *Advances in Laboratory Medicine/Avances en Medicina de Laboratorio, 2*(3), 399-408. https://doi.org/10.1515/almed-2021-0006

Wickham, H., & Grolemund, G. (2017). *R for Data Science*. O'Reilly Media. https://n9.cl/etad3