

Application of Association Rules for Discovering Patterns between Respiratory Diseases and Environmental Variables in the Municipality of Copiapó, Chile*

Andrés ALFARO, Servando CAMPILLAY,
Carlo TRONCOS and Luis ESPEJO

Universidad de Atacama. Copiapó, Chile

Correspondence should be addressed to: Andrés ALFARO, andres.alfaro@uda.cl

* Presented at the 44th IBIMA International Conference, 27-28 November 2024 Granada, Spain

Abstract

The study focuses on identifying patterns between respiratory diseases and environmental variables using data mining techniques and association rules. Information was collected from various data sources between January 2017 and December 2021, categorized to facilitate analysis. Although literature exists linking environmental variables to respiratory diseases, similar studies had not been conducted in Copiapó, highlighting the relevance of this work. A data mining model was employed to extract association rules that reveal relationships between respiratory diseases, environmental conditions, and pollution levels.

The results show that, in autumn, acute respiratory infections (ARI) are commonly associated with influenza and bronchitis, while in spring, climatic conditions and pollution are the main factors influencing bronchitis, rhinitis, pharyngitis, and tonsillitis. The study's conclusions also suggest that levels of air pollution (PM2.5 and PM10) are related to the number of consultations for respiratory diseases, affecting individuals of all ages. Furthermore, it is recommended to include COVID-19 in future studies, as it was not considered in this analysis due to its unexpected emergence during the five-year data collection period. This study provides evidence that air pollution is a public health issue in Copiapó and demonstrates that data mining can be a novel tool for identifying patterns and associations between environmental variables and respiratory diseases.

Keywords: Descriptive model, Relationship between variables, Apriori algorithm.

Introducción

Las enfermedades respiratorias representan una de las principales causas de mortalidad a nivel mundial (Rivera et al., 2016), afectando mayormente a las poblaciones más vulnerables como niños y adultos mayores. Estas enfermedades, incluyen desde infecciones agudas hasta condiciones crónicas como el asma y la enfermedad pulmonar obstructiva crónica (World Health Organization, 2015), siendo predominantes en Chile donde constituyen la tercera causa principal de muerte (Instituto Nacional de Estadísticas (INE, 2016). La Estrategia Nacional de Salud para el período 2011-2020 (Gobierno de Chile, 2010) destacó la necesidad de mejorar el acceso a servicios de salud de calidad y a tiempo, en un contexto donde los esfuerzos actuales no han sido suficientes para reducir el impacto de estas enfermedades. Esta situación es aún más crítica en regiones como Copiapó, donde los factores ambientales, incluidos los contaminantes atmosféricos, juegan un papel significativo en el desarrollo y gravedad de este tipo de patologías.

Cite this Article as: Andrés ALFARO, Servando CAMPILLAY, Carlo TRONCOS and Luis ESPEJO, Vol. 2024 (22) "Application of Association Rules for Discovering Patterns between Respiratory Diseases and Environmental Variables in the Municipality of Copiapó, Chile " Communications of International Proceedings, Vol. 2024 (22), Article ID 4439624, <https://doi.org/10.5171/2024.4439624>

La revolución digital en el sector salud ha generado una gran cantidad de datos que, si se utilizan adecuadamente, pueden ofrecer nuevas oportunidades para entender y mitigar los impactos de las enfermedades (Moreno, 2014). En este contexto, las tecnologías emergentes como la minería de datos y el aprendizaje automático presentan herramientas potenciales para analizar grandes volúmenes de datos y descubrir patrones que podrían ser significativos para la prevención y el tratamiento de estas enfermedades (Siebes, 2000; Vega et al., 2020).

Este trabajo se enfoca en la utilización de modelos de minería de datos para investigar las interacciones entre las enfermedades respiratorias y las variables ambientales en Copiapó, utilizando datos recopilados entre enero de 2017 y diciembre de 2021. En otras palabras, detectar cuáles factores medioambientales (climatológicos, contaminantes) afectan o influyen en la reacción de este tipo de enfermedades en la comuna de Copiapó. Para ello se propone una metodología de ocho etapas, basada en KDD (Knowledge Discovery in Databases) y CRISP-DM (Cross Industry Standard Process for Data Mining).

Principios Teóricos Y Estrategias De Análisis.

Las enfermedades respiratorias, definidas por la Organización Mundial de la Salud (OMS) como infecciones agudas y condiciones crónicas que afectan las vías respiratorias, son una causa significativa de morbilidad y mortalidad a nivel global, afectando principalmente a niños y adultos mayores (World Health Organization, 2015). En Chile, representan la tercera causa de muerte, componiendo un 9.5% del total de defunciones (Instituto Nacional de Estadísticas (INE), 2016). La lucha por controlar y atenuar este tipo de patologías ha sido compleja e insatisfactoria, como lo expone en (González et al., 2017), que relatan las reiteradas crisis nacionales provocadas por este tipo de enfermedades. Actualmente, se implementan numerosas estrategias de prevención, como la campaña de invierno, planes de vacunación, reforzamiento de la red sanitaria y vigilancia epidemiológica (Gobierno de Chile, 2024). Todo esto con el objetivo de prevenir, disminuir y evitar el brote de estas enfermedades (García Pérez & Aguilar, 2013). En este contexto, las variables ambientales que miden el estado de la atmósfera en un momento y lugar determinado, como la temperatura, la humedad y la presión atmosférica (Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM, 2019; Rodríguez Jiménez et al., 2004), y las variables contaminantes que miden los niveles de contaminación existentes en el aire, como el ozono (O₃), el dióxido de nitrógeno (NO₂) y el dióxido de azufre, así como el material particulado de MP10 y MP2.5 (World Health Organization, 2015), son fundamentales para comprender las dinámicas de estas enfermedades en entornos urbanos. Más cuando la prevalencia y complejidad de las enfermedades respiratorias dependen directamente de estos factores, según lo expone (Barros Monge, 2005; Asociación Latinoamericana de Tórax, 2017; Cuesta Cambra et al., 2019; Bellinger et al., 2017). Estas conexiones permiten analizar cómo los factores ambientales y contaminantes contribuyen a la incidencia y gravedad de este tipo de patologías, proporcionando una base para la aplicación de técnicas de minería de datos en el análisis de estos factores (Riquelme Santos et al., 2006; Bellinger et al., 2017). La minería de datos se presenta como una técnica para analizar grandes volúmenes de datos y extraer patrones significativos, permitiendo identificar asociaciones y tendencias que de otro modo serían difíciles de detectar (Linoff & Berry, 1997). En este contexto, los algoritmos de reglas de asociación, como Apriori, Eclat y FP-Growth, son herramientas para identificar patrones y relaciones entre las variables estudiadas. Estos algoritmos facilitan el desarrollo de modelos predictivos y descriptivos útiles en salud pública (Sidow Osman, 2019).

Reglas De Asociación Y Algoritmo Apriori

Las reglas de asociación se utilizan para identificar relaciones interesantes en grandes conjuntos de datos, representadas en la forma $\{X\} \rightarrow \{Y\}$ (Moya Amaris & Rodríguez Rodríguez, 2003). Los conceptos clave son: Soporte (*Support*), que indica la frecuencia con la que un conjunto de ítems aparece en la base de datos; Confianza (*Confidence*), que mide la probabilidad de que el conjunto $\{Y\}$ aparezca en las transacciones que contienen $\{X\}$; y *Lift*, que relaciona la confianza de una regla con la frecuencia de aparición del consecuente $\{Y\}$, indicando la asociación entre $\{X\}$ e $\{Y\}$ (Zaki & Wong, 2004; Romero, 2003; Silverstein et al., 1998). Dentro de los diferentes algoritmos de reglas de asociación, se encuentra el algoritmo Apriori, utilizado en minería de datos para encontrar conjuntos de ítems frecuentes, los cuales sirven para generar reglas de asociación. Su funcionamiento se resume en dos pasos: primero, genera todos los itemsets que contienen un solo elemento y luego combina estos conjuntos en pares, eliminando los que no cumplen con las métricas mínimas de soporte. En segundo lugar, genera reglas a partir de los conjuntos que cumplen con el mínimo de confianza, descartando los conjuntos cuyos subconjuntos no cumplen con las condiciones de soporte y confianza (Moya Amaris & Rodríguez Rodríguez, 2003).

Metodología

Basado en CRISP-DM y KDD (Moine et al., 2011; Mejía, 2019), se propone una metodología de 8 pasos o etapas: 1. Selección del subconjunto de datos, 2. Pre-procesamiento de los datos, 3. Revisión del conocimiento de dominio, 4. Creación del dataset, 5. Diseño e implementación del modelo, 6. Aplicación del modelo, 7. Validación del modelo, 8. Análisis y evaluación de los resultados. En la Figura 1 se puede apreciar cada una de las etapas que componen el proceso.

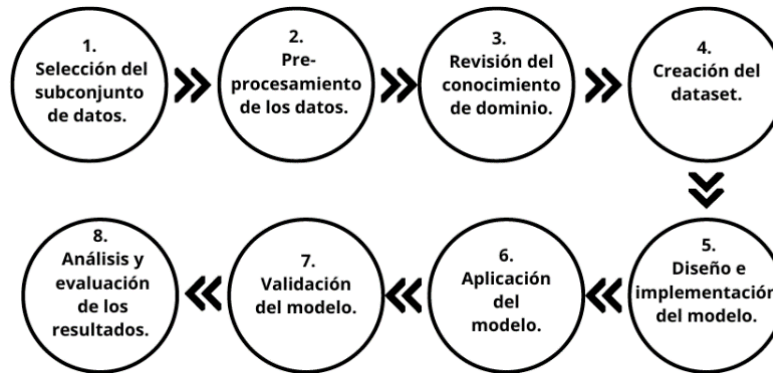


Fig. 1. Metodología propuesta (Fuente: Elaboración propia)

Selección del subconjunto de datos

La primera etapa del proceso metodológico consiste en la selección del subconjunto de datos necesarios para el análisis. En este estudio, los datos provienen de diversas fuentes como el Hospital Regional de Copiapó, la Dirección General de Aeronáutica Civil (DGAC), y el Sistema de Información Nacional de Calidad del Aire (SINCA). Estas fuentes proporcionan variables relacionadas con enfermedades respiratorias, variables ambientales y contaminantes respectivamente. Las enfermedades respiratorias son siete variables: Infección respiratoria aguda (IRA), Influenza, Neumonía, Bronquitis, Crisis obstructiva bronquial, Otra causa del sistema respiratorio, y Causas del sistema respiratorio. Los datos, son agrupados por año (2017-2021) y rango de edades (todas las edades, menores de 1 año, niños de 1 a 4 años, niños de 5 a 14 años, adultos de 15 a 64 años, adultos mayores de 65 años), se complementan con cinco variables medioambientales (humedad, rocío, temperatura, temperatura mínima y temperatura máxima) y dos variables contaminantes MP10 y MP2.5. Inicialmente, se consideraron las variables de COVID-19 confirmado en urgencia y hospitalizado, pero se eliminaron del análisis debido a su impacto limitado en el tiempo dentro del período de estudio de cinco años.

Pre-procesamiento de los datos

Una vez recopilados los datos, se procede a un pre-procesamiento que incluye la limpieza de datos, el manejo de datos faltantes y la transformación de datos (Shafique & Qaiser, 2014). Se emplean técnicas para reemplazar datos faltantes como el método de la media y el método del vecino más cercano para asegurar la integridad y preparar la base de datos para el análisis (Dagnino, 2014; Rosas & Verdejo, 2009; Allison, 2001). En algunas fuentes de datos se detectaron datos faltantes, específicamente en variables medioambientales, que fueron completados utilizando estos métodos para asegurar la integridad y minimizar el impacto en la varianza (Rosas & Verdejo, 2009). Posteriormente, se creó una estructura de datos que reúne todas las variables de interés en un solo lugar para facilitar su almacenamiento. Se creó un dataframe con 16 columnas y 261 filas, representando las semanas desde enero de 2017 hasta diciembre de 2021. En esta etapa, además se realizó un análisis estadístico descriptivo (Allison, 2001) que se centra en la utilización de técnicas como la media, la varianza, gráficas y un análisis de componentes principales (PCA), ver Figura 2, como parte del pre-procesamiento de los datos (Aroca et al., 2009; Chávez Chong et al., 2015).

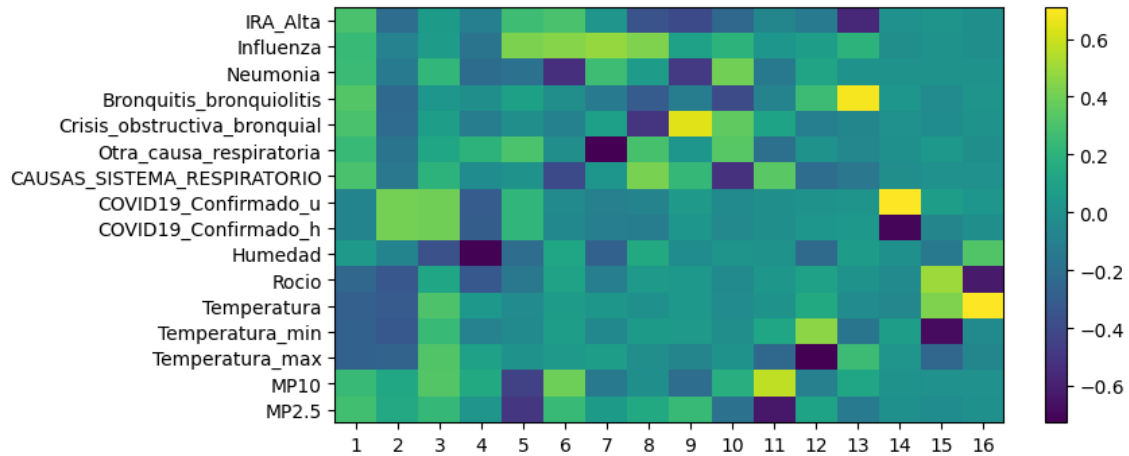


Fig. 2. Mapa de calor de todas las componentes principales

Estas técnicas permitieron calcular medidas de tendencia central y dispersión, describiendo así la distribución de los datos recopilados entre 2017 y 2021. Este análisis mostró, por ejemplo, que la media de la temperatura obtenida es aproximadamente 16.3 °C, lo cual coincide con trabajos externos que estiman la temperatura media entre 15 y 16 °C (Karamizadeh et al., 2013; Infante Amunátegui, 2015; Gómez Sarria, 2014). En el análisis de los niveles de contaminación (MP2.5 y MP10), se observa, ver Figura 3, que estos se mantienen dentro de los límites establecidos por las normas chilenas, aunque siguen siendo factores desencadenantes de enfermedades respiratorias.

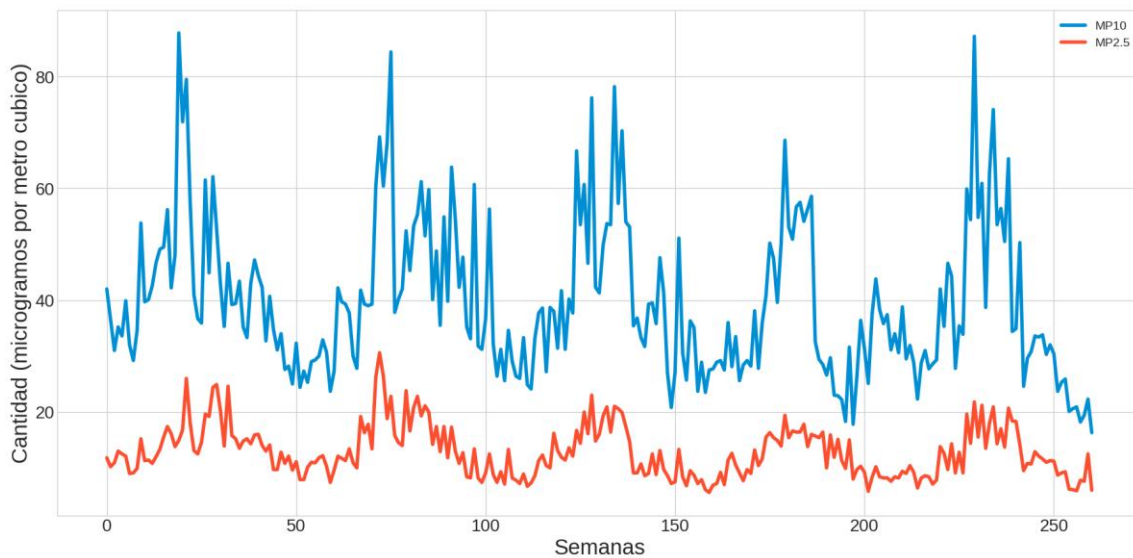


Fig. 3. Gráfico lineal Variables contaminantes.

Además, se destaca un notable aumento en los diagnósticos de IRA durante ciertos periodos, ver Figura 4, confirmando la confiabilidad de los datos utilizados. Estos hallazgos fueron verificados comparando los datos con trabajos externos confirmando la confiabilidad de la información utilizada en el estudio.

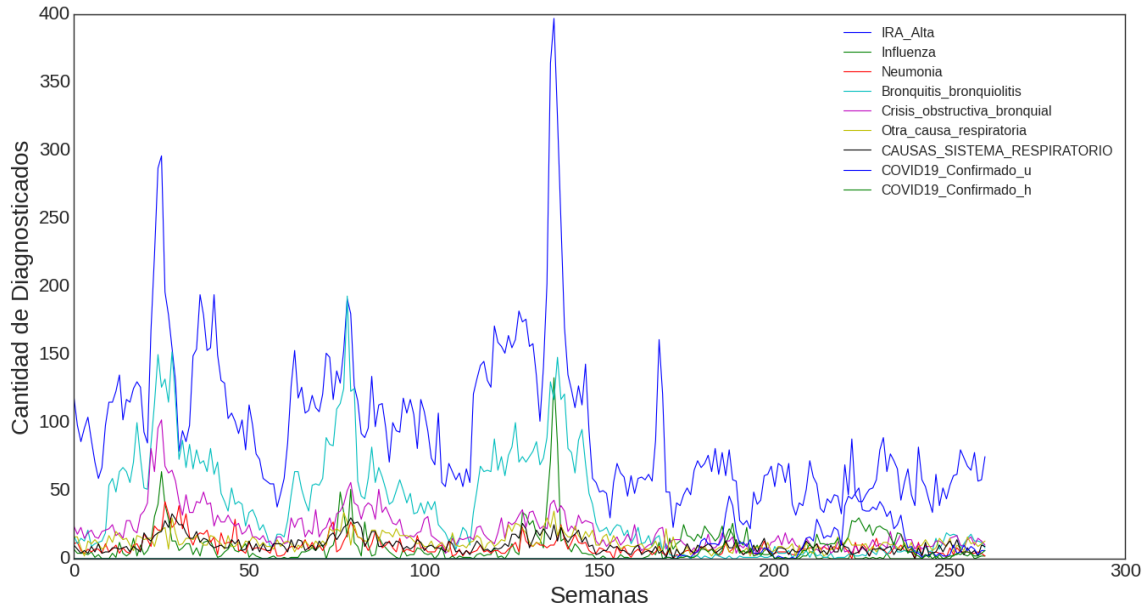


Fig. 4. Gráfico lineal Enfermedades respiratorias.

Revisión de conocimiento de dominio

Esta etapa implica la comprensión del contexto y las particularidades del problema a resolver mediante un estudio comparativo de metodologías y procesos utilizados en investigaciones similares (Ministerio de Salud, 2019; Kang & Wagacha, 2014). A partir de lo anterior, el proceso de discretización es fundamental en este trabajo, ya que transforma variables continuas en categóricas, facilitando la aplicación de algoritmos A priori (Kang & Wagacha, 2014). En este estudio, se utilizó la discretización por intervalos de igual ancho (De la Cruz Ruiz et al., 2022). Esta técnica divide el rango de valores de una variable en intervalos de igual tamaño, asignando a cada observación una etiqueta de intervalo correspondiente (Dougherty et al., 1995). Por ejemplo, la variable de temperatura se discretizó en intervalos específicos como "baja", "media" y "alta", lo cual permite una mejor identificación de patrones en los datos. Esta técnica asegura que las categorías resultantes sean útiles para el análisis, manteniendo la integridad y la variabilidad de los datos originales.

Creación del Dataset

Para este estudio, se creó un dataset mediante la técnica de discretización por intervalos de igual ancho (De la Cruz Ruiz et al., 2022). Cada variable fue transformada en tres nuevas variables booleanas (bajo, medio y alto) para facilitar la lectura y la interpretación de los resultados, como se muestra en la Tabla 1:

Tabla 1. Conversión A Variables Categóricas

Variable original	Formato original	Variable nueva	Formato nuevo
Dato_Variable	Float	Dato_Variable_Bajo	Boolean
		Dato_Variable_Medio	Boolean
		Dato_Variable_Alto	Boolean

Además, por recomendación de las expertas que colaboraron en la investigación, se determinó agrupar los datos por periodos de análisis de dos meses (bimestres): enero-febrero, marzo-abril, mayo-junio, julio-agosto, septiembre-octubre y noviembre-diciembre. Esto permite, por un lado, evitar la dispersión de algunas variables que podría causar que ciertos intervalos acaparen la mayoría de los registros. Por ejemplo, en la Figura 4, el valor máximo de la variable IRA es significativamente más alto en comparación con el resto de los registros, lo que genera intervalos cuyos máximos no son aplicables en todos los periodos de tiempo. Por otro lado, esta agrupación permite realizar un análisis más detallado y coherente, ya que las condiciones climáticas y medioambientales varían considerablemente entre enero y julio. El resultado fue un conjunto de datos denominado “dataset model”, utilizado como input para el modelo propuesto en este proyecto.

Diseño e Implementación del Modelo

El diseño e implementación del modelo se realizó utilizando un script en Python, empleando librerías para la manipulación y análisis de datos. Las librerías utilizadas incluyeron Pandas para la manipulación de datos, Sklearn para el análisis y modelado, Mlxtend para la implementación del algoritmo Apriori, y Matplotlib para la visualización de los resultados (Stančin & Jović, 2019).

Aplicación del Modelo

Esta etapa, se centra en la extracción de las reglas de asociación más interesantes aplicando el algoritmo Apriori sobre el dataset model. Para evaluar la cantidad y la calidad de las reglas, se analizaron aquellas con soportes entre el 15% y el 30%, ordenándolas por confianza de mayor a menor. Posteriormente, se estableció un soporte mínimo del 20% y un lift de 2 para garantizar que las reglas sean frecuentes y con asociaciones fuertes. Esta decisión se debe básicamente a que el soporte mínimo depende de los objetivos del proyecto y las consideraciones más relevantes (Ari & Ustazhanov, 2014). Por último, se ordenaron las reglas por confianza de mayor a menor, estableciendo un umbral de confianza del 50%. Esto garantiza que las reglas ocurran con una frecuencia bastante alta.

Validación del Modelo

La validación del modelo se realizó mediante el método del juicio de expertos, que consiste en obtener opiniones argumentadas de personas con trayectoria reconocida en el tema específico (Piatetsky-Shapiro & Steingold, 2000). Se efectuó un primer filtrado del conjunto de reglas de asociación, obteniendo 211 reglas divididas por periodos de análisis (bimestres). Posteriormente, las expertas Mg. María Paola Vieytes Carrizo y la licenciada en medicina Sonia Ibaceta Lorca evaluaron estas reglas. Basándose en su criterio y experiencia, seleccionaron 27 reglas representativas del contexto de la investigación. Este segundo filtrado se realizó a ciegas, sin proporcionar las métricas a las expertas, para que su selección dependiera exclusivamente de sus conocimientos y experiencia.

Análisis y Validación de los Resultados

En esta última etapa. Tras la selección realizada por las expertas en la sección anterior (segundo filtro, que dio como resultado 27 reglas), se realizó un tercer filtrado. A partir de aquí, se seleccionaron únicamente las reglas con una confianza superior al 80%. Como resultado, se obtuvieron 6 reglas, divididas en los diferentes periodos de análisis (bimestres), ver Tabla 2.

Tabla 2. Resultados reglas de asociación

#	Antecedentes		Consecuentes	Periodo	Rango
1	{'Influenza ', 'Bronquitis'}	->	{'Neumonía ', 'IRA Alta'}	May-Jun	5 a 14
2	{'ROCIO', 'Influenza', 'HUMEDAD', 'Temperatura mínima'}	->	{'IRA Alta ', 'TEMPERATURA'}	May-Jun	15 a 64
3	{'Otra causa respiratoria ', 'Temperatura mínima'}	->	{'IRA Alta'}	May-Jun	15 a 64
4	{'TEMPERATURA', 'Temperatura mínima'}	->	{'Causas sistema respiratorio '}	May-Jun	65 y más

5	{'Neumonía', 'MP10'}	->	{'Otra causa respiratoria'}	Sep-Oct	65 y más
6	{'MP2.5', 'TEMPERATURA'}	->	{'ROCIO', 'Neumonía', 'Temperatura máxima'}	Nov-Dic	Todos

- Regla 1: La influenza y la bronquitis son enfermedades respiratorias que pueden derivar en neumonía e infecciones respiratorias agudas (IRA), especialmente en niños, según (Garrote & del Carmen Rojas, 2015; Borrell & Segura, 2016). Ambas pueden tener un origen viral y afectar las vías respiratorias inferiores, con síntomas como tos seca, dificultad respiratoria, fiebre y sibilancias. Estas infecciones presentan un patrón estacional, siendo más frecuentes en meses fríos.
- Regla 2: La combinación de variables climáticas, como temperatura de rocío entre 7 y 11 °C, humedad relativa del 60 al 70 %, y temperaturas bajas entre 15 y 18 °C, puede aumentar el riesgo de IRA en adultos durante mayo y junio (Cifuentes Martínez et al., 2020). La influenza también puede contribuir a este aumento de riesgo.
- Regla 3: Existe una conexión entre enfermedades respiratorias y temperaturas mínimas entre 6 y 9 °C, que pueden debilitar el sistema inmunológico, aumentando la susceptibilidad a IRA, especialmente en personas con enfermedades respiratorias subyacentes (Dünner et al., 2020; Arduzzo et al., 2019).
- Regla 4: Temperaturas bajas entre 5 y 13 °C pueden aumentar el riesgo de enfermedades respiratorias generales. Las fluctuaciones de temperatura pueden provocar acumulaciones de polen en el aire, aumentando la exposición para personas alérgicas (Cifuentes Martínez et al., 2020; Dünner et al., 2020; Pino et al., 2015)
- Regla 5: A finales de invierno y comienzos de primavera, concentraciones de material particulado de 10 micras (MP10) entre 17 y 34 µg/m³ pueden aumentar el riesgo de bronquitis, rinitis, faringitis y amigdalitis. El MP10 también puede causar inflamación de las vías respiratorias y enfermedades pulmonares crónicas como la neumoconiosis y la silicosis (Cifuentes Martínez et al., 2020; Dünner et al., 2020; Pino et al., 2015).
- Regla 6: A finales de año, concentraciones de material particulado de 2.5 micras (MP2.5) entre 6 y 9 µg/m³ y temperaturas elevadas entre 18 y 29 °C pueden aumentar el riesgo de neumonía. La combinación de contaminación y altas temperaturas crea condiciones ideales para la propagación de virus responsables de la neumonía (Borrell & Segura, 2016; Cifuentes Martínez et al., 2020; Dünner et al., 2020). Por ejemplo, un estudio en China encontró mayor riesgo de neumonía en áreas con altos niveles de MP2.5 y altas temperaturas (Tian et al., 2019).

ALCANCE Y LIMITACIONES

Para esta investigación, se consideraron únicamente el material particulado de 10 micras (MP10) y el material particulado fino de menos de 2.5 micras (MP2.5) como variables contaminantes. Esta decisión se tomó debido a la falta de registros de otras variables en el repositorio correspondiente para el período de 2017-2021. Además, los datos de enfermedades respiratorias provienen exclusivamente del Hospital Regional de Copiapó San José del Carmen, sin incluir diagnósticos de otras instituciones médicas. Y con respecto a las limitaciones. Primero, la falta de datos sobre otras variables ambientales impide comprender completamente las asociaciones entre contaminantes y enfermedades respiratorias. Segundo, la ausencia de información sobre el sexo de los pacientes y categorías diagnósticas detalladas limita la generalización de los hallazgos. Además, el análisis de rangos etarios amplios (15 a 64 años) no permite identificar grupos particularmente vulnerables. Estas limitaciones dificultan sacar conclusiones definitivas y precisas sobre el impacto individual de cada variable ambiental y su interacción en la salud respiratoria. Aun así, este estudio contribuye al conocimiento del comportamiento de este tipo de enfermedades.

CONCLUSIONES

Este trabajo presenta un método novedoso para identificar patrones entre enfermedades respiratorias y variables medioambientales en Copiapó. La metodología del estudio permite identificar asociaciones, no causalidades, entre la contaminación ambiental y las enfermedades respiratorias. En otoño, la infección respiratoria aguda (IRA) en Copiapó está asociada principalmente con influenza y bronquitis. En primavera, la contaminación y condiciones climáticas causan bronquitis, rinitis, faringitis y amigdalitis. Los niveles de MP2.5 y MP10 están relacionados con consultas respiratorias, afectando a todas las edades. Es necesario ampliar la investigación sobre la relación entre la contaminación ambiental y las enfermedades respiratorias a otras ciudades del país. La falta de estaciones de monitoreo de calidad del

aire es una limitación importante. Los resultados ofrecen una base para futuras investigaciones y sugieren considerar el impacto del COVID-19, ya que no fue incluido en este estudio por ser un evento inesperado dentro del período de captura de datos de cinco años.

AGRADECIMIENTOS

Este trabajo fue financiado por el fondo de investigación de la Universidad de Atacama, proyecto DIUDA 22390. De igual manera agradecer a las expertas Mg. María Paola Vieytes Carrizo y la licenciada en medicina Sonia Ibaceta Lorca, quienes colaboraron en el desarrollo de esta investigación.

REFERENCIAS

- Allison, P. D. (2001) Missing data. University of Pennsylvania, USA: Quantitative Applications in the Social Sciences, Volume 136.
- Ari, N., & Ustazhanov, M. (2014) 'Matplotlib in Python', in 2014 11th International Conference on Electronics, Computer and Computation (ICECCO). IEEE, pp. 1–6.
- Arduso, L. R., Neffen, H. E., Fernández-Caldas, E., Saranz, R. J., Parisi, C. A., Tolcachier, A., & Marino, D. (2019) 'Intervención ambiental en las enfermedades respiratorias', *Medicina (Buenos Aires)*, 79(2), pp. 123–136.
- Aroca, P. R., García, C. L., & López, J. J. G. (2009) 'Estadística descriptiva e inferencial', *Revista El Auge de la Estadística en el Siglo XX*, 22, pp. 165–176.
- Asociación Latinoamericana de Tórax (2017) 'Foro de las Sociedades Respiratorias Internacionales: El impacto mundial de la Enfermedad Respiratoria', México.
- Barros Monge, M. (2005) 'Sociedad Chilena de Enfermedades Respiratorias: 75 años de historia', *Revista Chilena de Enfermedades Respiratorias*, 21(1).
- Bellinger, C., Mohomed Jabbar, M., Zaïane, O., & Osornio-Vargas, A. (2017) 'A systematic review of data mining and machine learning for air pollution epidemiology', *BMC Public Health*, 17(1), 907. doi: 10.1186/s12889-017-4914-3.
- Borrell, L. S., & Segura, M. C. (2016) 'Neumonía y neumonía recurrente', *Pediatría Integral*, 20(1), pp. 38–42.
- Chávez Chong, C. O., Sánchez García, J. E., & DelaCerde, J. (2015) 'Análisis de componentes principales funcionales en series de tiempo económicas', *GECONTEC: Revista Internacional de Gestión del Conocimiento y la Tecnología*, 3(2).
- Cifuentes Martínez, P., Rodríguez-Fernández, A., Luengo, C., & Tapia, L. (2020) 'Relación entre contaminación atmosférica y consultas por enfermedades respiratorias en atención primaria de urgencia', *Revista Chilena de Enfermedades Respiratorias*, 36(4), pp. 260–267. doi: 10.4067/S0717-73482020000400260.
- Cuesta Cambra, U., Niño González, J. I., Martínez Martínez, L., & Díez Bielsa, C. (2019) 'Nuevas herramientas de información en campañas de concienciación social. El "Big Data" aplicado a la Sequía y Salud Pública', *Revistas.ucm.es*.
- Dagnino, J. (2014) 'Datos faltantes (missing values)', *Revista Chilena de Anestesia*, 43, pp. 332–334. Available at: <https://revistachilenadeanestesia.cl/PII/revchilanestv43n04.13.pdf>.
- De la Cruz Ruiz, F., Ruiz, C., & Canul-Reich, J. (2022) 'Reglas de asociación para el estudio de la vaginosis bacteriana', *Komputer Sapiens*, 26/88.
- De León, J. M., Acosta, D., Lorduy, F. A., De la Cruz Pinzón, C., Arrieta, J. M. E., Jaramillo, C., & Pinzón, H. (1997) *Infección respiratoria aguda*. Instituto de Seguros Sociales.
- Dougherty, J., Kohavi, R., & Sahami, M. (1995) 'Supervised and unsupervised discretization of continuous features', in *Machine Learning Proceedings*. Morgan Kaufmann, pp. 194–202.
- Dünner, M. A., Puentes, R., Vaquero, A., & Díaz, J. (2020) 'Enfermedades no transmisibles y clima en Chile: un resumen de evidencia para el período 1990-2019', *Revista del Instituto de Salud Pública de Chile*, 4(2).
- García Pérez, P., & Aguilar, C. (2013) 'Vigilancia epidemiológica en salud', *Revista Archivo Médico de Camagüey*, pp. 121–128.
- Garrote, P. R., & del Carmen Rojas, M. (2015) 'La validación por juicio de expertos: dos investigaciones cualitativas en Lingüística aplicada', *Revista Nebrija de Lingüística Aplicada a la Enseñanza de Lenguas*, (18), pp. 124–139.

- Gobierno de Chile (2010) Estrategia Nacional de Salud para el cumplimiento de los Objetivos Sanitarios de la Década. Available at: https://extranet.who.int/nutrition/gina/sites/default/filesstore/CHL_2011%20Estrategia%20Nacional%20de%20Salud.pdf.
- Gobierno de Chile (2024) Ministerio de Salud - Gobierno de Chile. Available at: <https://www.minsal.cl/>.
- Gómez Sarria, N. (2014) Climatología urbana de Copiapó como ciudad localizada en un medio ambiente árido. Available at: <https://repositorio.uchile.cl/handle/2250/130424>.
- González, R., Pinto, R., & Álvarez, J. P. (2017) 'Las enfermedades respiratorias en Chile: un reflejo de nuestra historia', *Revista Médica Clínica Las Condes*, 28(1), pp. 152–154. doi: 10.1016/S0716-8640(17)30020-8.
- Infante Amunátegui, A. (2015) Hoya n° 302 Copiapó. Catastro de pozos al 31 de mayo de 1971.
- Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM) (2019) Glosario Meteorológico. Colombia, p. 286.
- Instituto Nacional de Estadísticas (INE) (2016) Estadísticas Vitales Informe Anual 2016. Chile.
- Kang, S. M., & Wagacha, P. W. (2014) 'Extracting diagnosis patterns in electronic medical records using association rule mining', *International Journal of Computer Applications*, 108(15).
- Karamizadeh, S., Abdullah, S. M., Manaf, A. A., Zamani, M., & Hooman, A. (2013) 'An overview of principal component analysis', *Journal of Signal and Information Processing*, 4(3B), p. 173.
- Linoff, G., & Berry, M. J. (1997) *Data Mining Techniques: For Marketing, Sales, and Customer Support*. Hoboken.
- Mejía, J. C. G. (2019) 'Aplicación de la técnica regresión logística de la minería de datos en el proceso de descubrimiento de conocimiento (KDD) en bases de datos operativas o transaccionales', *Perspectiv@s*, 14(13), pp. 51–55.
- Ministerio de Salud (2019) Boletín Epidemiológico Trimestral: Influenza, Se 1–39. Available at: https://epi.minsal.cl/wpcontent/uploads/2019/11/BET_INFLUENZA_OCTUBRE_2019.pdf.
- Moine, J. M., Haedo, A. S., & Gordillo, S. E. (2011) 'Estudio comparativo de metodologías para minería de datos', in XIII Workshop de Investigadores en Ciencias de la Computación. Available at: <http://sedici.unlp.edu.ar/handle/10915/20034>.
- Moreno, J. (2014) 'Una aproximación a Big Data', *Revista de Derecho de la UNED*, 471. doi: 10.5944/rduned.14.2014.13303.
- Moya Amaris, M. E., & Rodríguez Rodríguez, J. E. (2003) 'La contribución de las reglas de asociación a la minería de datos', *Tecnura*, 7(13), pp. 94–109. doi: 10.14483/22487638.6175.
- Piatetsky-Shapiro, G., & Steingold, S. (2000) 'Measuring lift quality in database marketing', *ACM SIGKDD Explorations Newsletter*, 2(2), pp. 76–80.
- Pino, P., Iglesias, V., Garreaud, R., Cortés, S., Canals, M., Folch, W., ... and Steenland, K. (2015) 'Chile confronts its environmental health future after 25 years of accelerated growth', *Annals of Global Health*, 81(3), pp. 354–367.
- Riquelme Santos, J. C., Ruiz, R., & Gilbert, K. (2006) 'Minería de Datos: Conceptos y Tendencias', *Revista Iberoamericana de Inteligencia Artificial*, (29), pp. 11–18.
- Rivera, M., Díaz, C. A., Céspedes, R. A., & Kalergis, P. F. (2016) 'Virus Respiratorio Sincicial: un desafío para la salud pública a nivel mundial', *Revista de la Sociedad Española de Bioquímica y Biología Molecular*, 26.
- Rodríguez Jiménez, R. M., Benito Capa, Á., & Portela Lozano, A. (2004) *Meteorología y Climatología*. Fundación Española para la Ciencia y la Tecnología (FECYT), pp. 12–33.
- Romero, C. (2003) *Aplicación de técnicas de adquisición de conocimiento para la mejora de cursos hipermedia adaptativos basados en Web*. Tesis Doctoral, Universidad de Granada.
- Rosas, J. F. M., & Verdejo, E. Á. (2009) 'Métodos de imputación para el tratamiento de datos faltantes: aplicación mediante R/Splus', *Revista de Métodos Cuantitativos para la Economía y la Empresa*, 7, pp. 3–30.
- Shafique, U., & Qaiser, H. (2014) 'A comparative study of data mining process models (KDD, CRISP-DM and SEMMA)', *International Journal of Innovation and Scientific Research*, 12(1), pp. 217–222.
- Siebes, A. (2000) 'Data Mining and Statistics: A Systems Point of View', in *Computational Intelligence in Data Mining*. Vienna: Springer Vienna, pp. 1–38.

- Sidow Osman, A. (2019) 'Data Mining Techniques: Review'. Available at: <http://ojs.mediu.edu.my/index.php/IJDSR/article/view/1841/717>.
- Silverstein, C., Brin, S., & Motwani, R. (1998) 'Beyond market baskets: Generalizing association rules to dependence rules', *Data Mining and Knowledge Discovery*, 2, pp. 39–68.
- Stančín, I., & Jović, A. (2019) 'An overview and comparison of free Python libraries for data mining and big data analysis', in 2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO). IEEE, pp. 977–982.
- Tian, Y., Liu, H., Wu, Y., Si, Y., Li, M., Wu, Y., & Hu, Y. (2019) 'Ambient particulate matter pollution and adult hospital admissions for pneumonia in urban China: a national study', *Journal of Environmental Science*.
- Vega, M. Á., Mora, L. M. Q., & Badilla, M. V. C. (2020) 'Inteligencia artificial y aprendizaje automático en medicina', *Revista Médica Sinergia*, 5(8), e557-e557.
- World Health Organization (2015) 'OMS, Infecciones del tracto respiratorio'. Available at: https://www.who.int/topics/respiratory_tract_diseases/es/
- Zaki, M. J., & Wong, L. (2004) 'Data Mining Techniques', *Lecture Notes Series, Institute for Mathematical Sciences, National University of Singapore*, pp. 125–163.