

Análisis de la Calidad de las Asociaciones de Contaminantes y Variables Meteorológicas por Estacionalidad

Analysis of the Quality of the Associations of Pollutants and Meteorological Variables by Seasonality

Jorge Luis Zambrano-Martinez¹ <https://orcid.org/0000-0002-5339-7860>, Marcos Orellana¹ <https://orcid.org/0000-0002-3671-9362>, Agustín Ferrari¹ <https://orcid.org/0009-0009-5696-8954>, Alex Coro¹ <https://orcid.org/0009-0009-6020-3751>, Sebastian Heras¹ <https://orcid.org/0009-0001-9148-0139>

¹Laboratorio de Investigación y Desarrollo en Informática (LIDI)
Universidad del Azuay, Cuenca, Ecuador
jorge.zambrano@uazuay.edu.ec, marore@uazuay.edu.ec,
daferrarif99@es.uazuay.edu.ec,
alex.coro@es.uazuay.edu.ec,
sebastianheras28@es.uazuay.edu.ec



Esta obra está bajo una licencia internacional
Creative Commons Atribución-NoComercial 4.0

Enviado: 2023/07/14

Aceptado: 2023/07/29

Publicado: 2023/10/15

Resumen

Para lograr una evaluación más precisa de la calidad del aire, es necesario conocer la relación que existe entre las variables meteorológicas y los distintos contaminantes atmosféricos; lo que también tendrá la finalidad de evitar los riesgos presentes tanto en el ecosistema como en la salud de los humanos en un futuro próximo. El problema radica en encontrar una asociación entre los contaminantes atmosféricos y las variables meteorológicas, que existe en los modelos y métodos de categorización que pueden ser empleados. Debido a esta razón, el objetivo de este artículo es analizar la calidad de las asociaciones de contaminantes y variables meteorológicas por estacionalidad utilizando árboles de decisión para encontrar conocimiento que permita localizar patrones que serán importantes para el análisis ambiental. Por consiguiente, al aplicar árboles de decisión se pudo lograr un control periódico de la calidad de las asociaciones de contaminantes y variables meteorológicas, cuya validación del nivel de confianza de las reglas de asociación es mayor al 70% en los meses estudiados.

Sumario: Introducción, Trabajos Relacionados, Metodología, Resultados y Conclusiones.

Como citar: Zambrano-Martinez, J. L., Orellana, M., Ferrari, A., Coro, A. & Heras, S. (2023). Análisis de la Calidad de las Asociaciones de Contaminantes y Variables Meteorológicas por Estacionalidad. *Revista Tecnológica - Espol*, 35(2), 50-60.

Recuperado a partir de <http://www.rte.espol.edu.ec/index.php/tecnologica/article/view/1052>

Palabras clave: Árboles de decisión, discretización, inteligencia artificial, minería de datos, reglas de asociación.

Abstract

To achieve a more accurate assessment of air quality, it is necessary to know the relationship between meteorological variables and the different air pollutants; this will also be aimed at avoiding the risks present both in the ecosystem and in human health in the future. The problem begins with finding an association between air pollutants and meteorological variables in the models and categorization methods that can be used. Due to this reason, the objective of this article is to analyze the quality of the associations between pollutants and meteorological variables by seasonality using decision trees to find knowledge that allows locating patterns that will be important for environmental analysis. Therefore, it was possible to achieve a periodic control of the quality of the associations of pollutants and meteorological variables, whose validation of the confidence level of the association rules is greater than 70% in the months studied.

Keywords: Decision trees, discretization, artificial intelligence, data mining, association rules.

Introducción

Actualmente, la contaminación atmosférica es uno de los problemas ambientales más graves del mundo, debido a la gran cantidad de contaminantes emitidos por los humanos, entre ellos destaca el material particulado, menor a 2.5 micrómetros ($PM_{2.5}$); el material particulado, menor a 10 micrómetros (PM_{10}); el dióxido de azufre (SO_2); el monóxido de carbono (CO); el ozono (O_3); y el dióxido de nitrógeno (NO_2). Estos contaminantes presentan un efecto dañino a la salud debido a que provocan diferentes afecciones pulmonares, cerebrales y cardiovasculares, lo que reduce la esperanza de vida (Manisalidis et al., 2020). En las zonas rurales persiste un menor índice de contaminación en comparación con las zonas urbanas, debido a la alta densidad demográfica, conjuntamente con el desarrollo de la industria y el transporte. Las consecuencias que conlleva están relacionadas con problemas de salud, que traen consigo preocupación en los gobiernos. Así, los gobernantes tienen que enfocarse en una adecuada gestión de la calidad del aire, utilizando diferentes estrategias y herramientas que permitan una correcta interpretación de los datos obtenidos en la atmósfera.

Para el análisis de este tipo de estudios es necesaria la recolección de los datos provenientes de las estaciones de monitoreo meteorológico continuo, con la finalidad de conocer la calidad del aire y a su vez encontrar los problemas de contaminación atmosférica en una zona. Además, la relación de los contaminantes atmosféricos con las variables meteorológicas, como la humedad, temperatura, velocidad del aire, entre otros, tienden a afectar el clima del entorno, y dan como resultado el deterioro de la salud de las personas (Whiteman et al., 2014). Del mismo modo, Ahmed et al. (2020) sostienen que las concentraciones de contaminación se enfocan en zonas específicas de interés o parques industriales. Huang et al. (2018) determinó que, si no se toman las debidas precauciones, como el cambio de combustibles a más limpios o la depuración de buses anticuados, la contaminación atmosférica continuará en aumento de manera considerable.

Siendo así, los gobiernos locales se encuentran recolectando de manera constante los datos de las variables meteorológicas y contaminantes atmosféricos, a través de estaciones meteorológicas equipadas con sensores, que recopilan toda la información del ambiente para

luego ser procesada y brindar conocimiento. Se utilizan técnicas de minería de datos para el tratamiento de la información obtenida y el descubrimiento de patrones de comportamiento como árboles de decisión, regresiones logísticas, algoritmos genéticos, entre otros.

En este estudio, para realizar el análisis de los patrones de comportamiento y crear modelos de predicción con los datos de los contaminantes atmosféricos y variables meteorológicas en la ciudad de Cuenca – Ecuador, se utilizan los árboles de decisión con la finalidad de validar el grado de confianza de las reglas de asociación obtenidas en un previo estudio (Orellana et al., 2021).

Este artículo está estructurado de la siguiente manera: La Sección 2 presenta los trabajos relacionados con métodos semejantes; la Sección 3 expone la metodología utilizada para llevar a cabo esta investigación; la Sección 4 explica los resultados que se han obtenido tras la aplicación de la metodología; y la Sección 5 presenta las conclusiones obtenidas y los trabajos futuros.

Trabajos Relacionados

Existen investigaciones que estudian a las variables meteorológicas y a los contaminantes atmosféricos y permiten obtener información para distintos propósitos, como predecir nuevo conocimiento para determinar el nivel de calidad de la atmósfera y su posible gestión. A continuación, se exponen las investigaciones que están vinculadas a técnicas de minería de datos que extraen los conocimientos más importantes en este campo.

En el estudio realizado por García et al. (2020) se utilizan métodos de minería de datos para cuantificar el impacto ambiental producido por una empresa Courier en Lima y Callao, en Perú. Como punto inicial, los autores aplican métodos de clasificación a las variables y clusterización (OneRule y K-means), lo cual permite detectar los de mayor demanda de despachos (4 grupos) y determinar el nivel de emisiones de KgCO_2 diarias, separadas por una temporalidad mensual y por tipo de vehículo utilizado. En cuanto a la investigación expuesta por Gayathri et al. (2020), propone la utilización de un modelo basado en técnicas de minería de datos para la predicción de la contaminación del aire, específicamente la aplicación de algoritmo de árbol de decisión *C4.5*. Los autores descomponen dicho modelo en cinco etapas: recolección de datos, preprocesamiento de los datos, árbol de decisión, datos de prueba, predicción, con lo cual concluyen que el sistema propuesto ayudará a mejorar la predicción.

Acorde a Siwek et al. (2016), se realizan los modelos de predicción a diferentes contaminantes como PM_{10} , $\text{PM}_{2.5}$, NO_2 , SO_2 y O_3 . Sin embargo, uno de los contaminantes más nocivos para la salud humana es $\text{PM}_{2.5}$, debido a la acumulación en el sistema respiratorio, disminuyendo el correcto funcionamiento pulmonar y aumentando las enfermedades respiratorias. El método que utilizan para predecir el valor de $\text{PM}_{2.5}$ es el de bosques aleatorios y los resultados demuestran que el modelo presentado es viable para aplicarlo en cualquier ciudad contaminada. Es importante considerar el índice de calidad del aire (AQI) utilizado por los gobiernos para determinar el nivel de calidad que se presenta en la atmósfera relacionados con los contaminantes. Asimismo, el trabajo realizado por Huang et al. (2018) utiliza algoritmos de redes neuronales, basados en el índice $\text{PM}_{2.5}$, para llegar a pronosticar la contaminación del ambiente. La conclusión obtenida en este estudio es que la única variable atmosférica vinculada con el contaminante $\text{PM}_{2.5}$ es la humedad, debido a que ayuda a la disipación del contaminante.

Otro contaminante altamente perjudicial es el PM_{10} , de acuerdo con Althuwaynee et al. (2020), debido a que analizan las correlaciones que existen entre PM_{10} y otros contaminantes

como el O_3 y SO_2 . Los datos son obtenidos en Kuala Lumpur, Malasia, y en su análisis se observa que existe una relación lineal directa y positiva entre PM_{10} y SO_2 , mientras que existe una relación semi lineal entre PM_{10} y O_3 .

Continuando con las investigaciones relacionadas con redes neuronales, está la realizada por Athira et al. (2018), que utilizó el conjunto de datos de AirNet, de pronóstico de contaminación del aire de cinco días del Centro Nacional de Monitoreo Ambiental de China (CNEMC) y datos meteorológicos del sistema global de predicción (GFS), donde se aplican modelos de aprendizaje profundo, como red neuronal recurrente (RNN), redes de gran memoria a corto plazo (LSTM) y unidades recurrentes cerradas (GRU). En otra investigación, realizada por Birant (2011), se presentan los resultados de comparación de distintos algoritmos de árboles de decisión como *C4.5*, *CART*, *NBTree*, *BFTree*, *LADTree*, *REPTree*, árbol aleatorio, bosques aleatorios, y árbol modelo logístico (LMT), para clasificar y predecir los niveles de emisión del SO_x , utilizando los datos recopilados de alrededor de 800 instalaciones industriales en Izmir, Turquía. Los autores compararon individualmente cada nivel de emisión del SO_x (bajo, medio, alto y muy alto), obteniendo los mejores resultados en la categoría de bajo nivel. Así, se pudo establecer que los niveles de emisión pueden clasificarse y predecirse con éxito en el 82.4% de los casos.

Por otro lado, Martínez-España et al. (2018), analizaron diferentes técnicas de aprendizaje automático para predecir los niveles del O_3 en la región de Murcia, España. Las técnicas analizadas son de bosques aleatorios, árboles de decisión y K vecinos más próximos (*KNN*); la técnica de bosques aleatorios fue la que más se ajustó. En los resultados obtenidos se puede señalar que entre los parámetros que más influyen en la predicción del ozono se encuentran variables climáticas relacionadas con la temperatura, la humedad y el viento.

Así pues, en este estudio se realiza un análisis de los patrones de comportamiento y la creación de modelos de predicción con los datos obtenidos de los contaminantes atmosféricos y variables meteorológicas en la ciudad de Cuenca - Ecuador, a través de árboles de decisión para ratificar el nivel de confianza de las reglas de asociación previamente establecidas en trabajos previos.

Metodología

En la presente investigación, la metodología que se utilizó para llevar a cabo la experimentación se denomina ADDIE (Análisis, Diseño, Desarrollo, Implementación y Evaluación) y se encuentra representada en la Figura 1. Este modelo es considerado uno de los más utilizados y sus cinco fases comprenden un camino o proceso de aprendizaje.

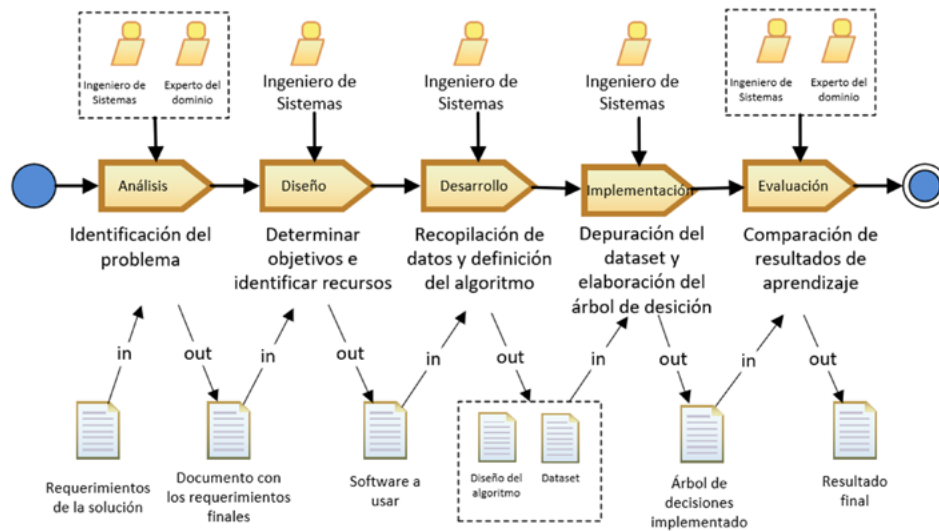
A continuación, se presentan, con más detalle, las fases de la metodología ADDIE aplicadas a esta investigación:

- **Análisis:** La propuesta de esta investigación es realizar un análisis de la calidad de las asociaciones de contaminantes y variables meteorológicas por estacionalidad, aplicando árboles de decisión a fin de verificar si las reglas de asociación expuestas en la investigación de Orellana et al. (2021) se cumplen con su calidad.
- **Diseño:** Considerando el análisis de la calidad de las asociaciones de contaminantes y variables meteorológicas, el principal objetivo de este análisis es el cumplimiento de la calidad de las reglas de asociación y así definir el comportamiento de las reglas de asociación correspondientes a los meses específicos. Es por esa razón que, con

este objetivo en mente, se ha definido trabajar con la herramienta de minería de datos aplicando árboles de decisión, que es una técnica de inteligencia artificial utilizando el software *RapidMiner*. Esta herramienta reduce el uso del código para el modelado de los datos, agilizando el análisis y el procesamiento de grandes cantidades de datos.

- **Desarrollo:** Para la selección de los parámetros necesarios, se procede en función de la recopilación de los datos para definir un algoritmo con la mayor precisión posible.
- **Implementación:** Para realizar la implementación, se obtuvo un conjunto de datos de los contaminantes atmosféricos tomados entre el mes de enero a diciembre del año 2018 en la ciudad de Cuenca, Ecuador. Para realizar el preprocesamiento de los datos, donde se realiza la depuración y limpieza, se utilizó una plataforma de ciencia de datos denominada *RapidMiner*; luego se procedió a realizar las pruebas correspondientes.
- **Evaluación:** Mediante el ingreso de los datos al software mencionado se determinaron los parámetros más importantes del conjunto de datos, para generar un árbol de decisión correspondiente a cada mes y verificar que las reglas de asociación obtenidas se cumplan o no en ciertos meses.

Figura 1
Diseño de la metodología ADDIE



Preprocesamiento de datos

Como se conoce, al recolectar los datos por medio de los sensores de las estaciones meteorológicas, estos pueden presentar anomalías o valores atípicos, ya sea por diversas razones como interrupciones en la fuente energética o errores de calibración en los sensores que están utilizando. Por lo tanto, la recolección de datos de los contaminantes atmosféricos es tomada cada diez minutos y las variables meteorológicas con una frecuencia de cada minuto. La recolección de los datos entre los meses de enero a diciembre del 2018 fue de 52,559 registros.

Este conjunto de datos contiene atributos que son dependientes de otros, en otras palabras, los valores finales de una variable dependen de la mezcla de cualidades que sean partícipes en una determinada dimensión del estudio.

Cabe recalcar que se implementó la depuración de valores no encontrados, con la finalidad de descartar los datos que contienen valores vacíos o nulos. Luego de realizar el preprocesamiento de los datos, se obtuvieron un total de 39,507 registros y se eliminaron un total de 8,497 datos, para obtener valores consistentes de comportamiento.

Aplicación de técnicas de discretización

La discretización es el procedimiento en el cual se divide el rango del atributo continuo en intervalos. Cada uno de los intervalos se etiqueta como un valor discreto, para luego a los datos originales asignarlos a los valores discretos (Hemada & Vijaya Lakshmi, 2013).

Este proceso tiene un enfoque importante en el preprocesamiento de datos para su utilización en técnicas de inteligencia artificial, ya que con la aplicación de un método eficaz de discretización no solo se puede reducir la demanda de memoria del sistema y mejorar la eficiencia de la minería de datos, sino que también hace que el conocimiento extraído del conjunto de datos sea más sólido y fácil de entender (Hemada & Vijaya Lakshmi, 2013).

Para este estudio, es necesario aplicar la discretización con cantidades similares de datos en cada grupo para utilizar las reglas de asociación presentadas por Orellana et al. (2021) y así analizarlas mediante árboles de decisión para obtener sus respectivas medidas de rendimiento.

Aplicación de árboles de decisión

El algoritmo de árboles de decisión es un modelo lógico parecido a un árbol binario construido con base en un conjunto de datos de entrenamiento. Este ayuda a predecir el valor de una variable objetivo mediante el uso de variables predictivas. Este es el método ideal para la clasificación de información y su posterior evaluación en sus diferentes escenarios (Alsagheer et al., 2017).

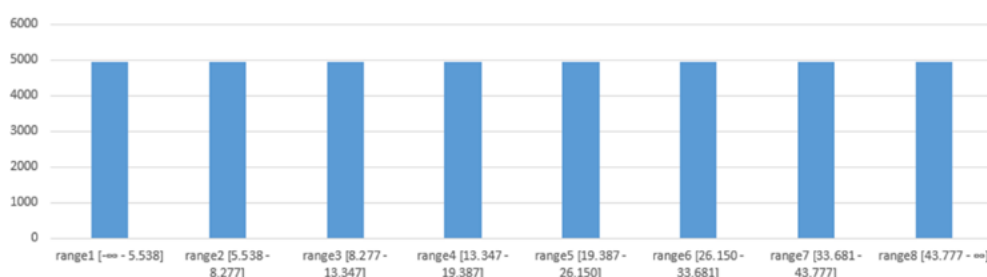
Resultados

Al momento de tener un conjunto de datos amplio, lo óptimo y recomendable es agrupar los datos que representan una misma unidad. Por lo tanto, se debe tomar en cuenta que la técnica sea fiable por el hecho de obtener patrones para observar con más detalle el comportamiento de los datos.

Técnica de discretización

Para tener un mejor enfoque en la investigación, se decidió crear varios conjuntos con una cantidad de datos equilibrada, como se puede observar en la Figura 2. A través de esta técnica de asociación, presentada por Orellana et al. (2021), se crearon ocho conjuntos con distintos rangos de datos para finalmente verificar las reglas de asociación.

Figura 2
Discretización por frecuencia

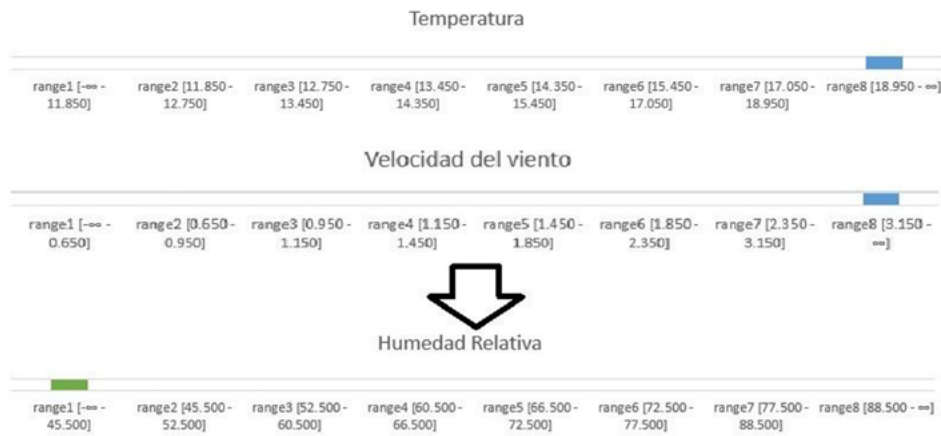


Reglas de asociación

Para este estudio se contó con tres reglas de asociación previamente definidas por Orellana et al. (2021) y mediante la discretización se obtuvieron las mismas reglas con el conjunto de datos.

La primera regla de asociación establece que al tener una temperatura que se encuentra en un rango superior a 18,950 °C, junto a un valor de la velocidad del viento superior a 3,150 m/h, produce como resultado una humedad relativa baja; en otras palabras, los datos de la humedad relativa se ubican menor igual a 45,5%. En la Figura 3, se puede observar la primera regla de asociación.

Figura 3
Primera regla de asociación



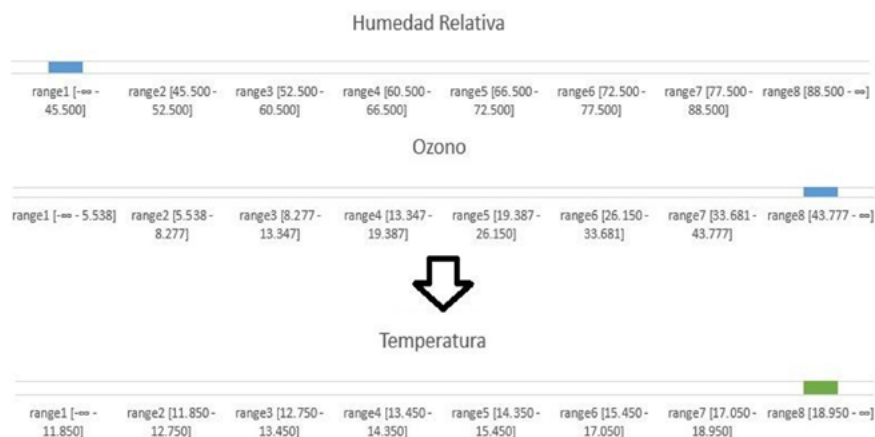
Por otro lado, la segunda regla de asociación propone que al presentarse una medición baja de O₃, menor a 5,538 µg/m³ y una humedad relativa alta, mayor a 88,5%, da como resultado valores altos del punto de rocío. Esto se puede apreciar en la Figura 4.

Figura 4
Segunda regla de asociación



Por último, la tercera regla de asociación plantea que, al presentarse una humedad relativa baja, menor a 45,5%, y una medida del O₃ alta, que supere el valor de 45 µg/m³, el resultado que se presenta es una temperatura alta, que puede sobrepasar los 18,950 °C. como se puede observar en la Figura 5.

Figura 5
Tercera regla de asociación



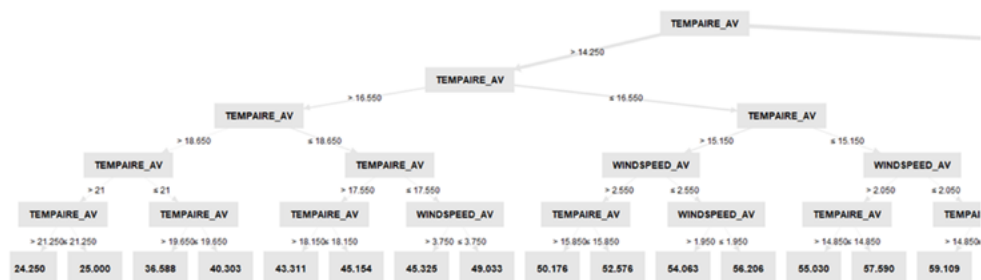
Árboles de decisión

La finalidad de aplicar los árboles de decisión es validar el nivel de confianza de las reglas de asociación presentadas en la subsección anterior. Por lo tanto, se realizó la validación del nivel de confianza con árboles de decisión mensualmente, desde enero hasta diciembre de 2018, y se observó la frecuencia con la que se cumplen dichas reglas de asociación.

Se aplicó este método a la primera regla y se obtuvo lo siguiente:

- La primera regla no se cumplió en 3 de los 12 meses analizados, por lo que representa el 25%, siendo los meses de junio, julio y agosto.
- Existieron dos meses en los que se pudo observar que la velocidad del viento no influyó para representar los niveles bajos propuestos de humedad relativa como se puede observar en la Figura 6.

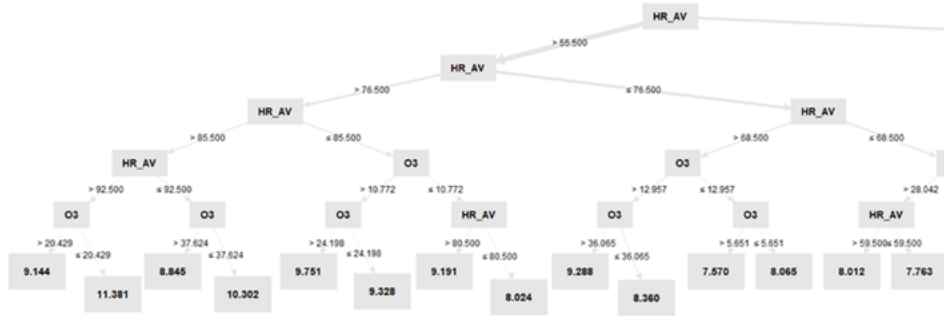
Figura 6
Árbol de decisión aplicado en la primera regla de asociación



En cuanto a la segunda regla de asociación, los resultados obtenidos fueron:

- Su medida de rendimiento fue del 66.7%; es decir, la regla llegó a cumplirse en ocho de los 12 meses. Sin embargo, cabe recalcar que en algunos casos la medición obtenida de O₃ es demasiado elevada, por lo que no se puede llegar a establecer con claridad lo que sucedió en el mes de enero.
- También existe el caso en el que los valores de O₃ llegan a ser los esperados pero la humedad relativa es baja y no llega a superar el 88.5%, lo cual influyó en la obtención de la medida de rendimiento de la regla, como se puede observar en la Figura 7.

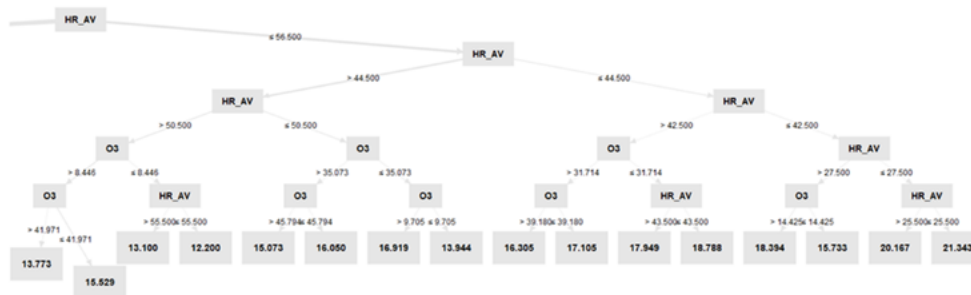
Figura 7
Árbol de decisión aplicado en la segunda regla de asociación



Por último, los resultados obtenidos en la tercera regla de asociación, como se puede observar en la Figura 8, fueron:

- En esta última regla se trabaja con O₃, humedad relativa y temperatura, cumpliéndose en el 75% de los meses excepto los meses de enero, julio y octubre.
- Igual existen casos problemáticos con el contaminante O₃, debido a que los valores medidos tienden a ser bajos, pero se alcanza la temperatura y humedad relativa planteada por la regla de asociación.

Figura 8
Árbol de decisión aplicado en la tercera regla de asociación



Conclusiones

A lo largo de la investigación se aplicaron técnicas de minería de datos, siendo los árboles de decisión los que permitieron realizar un análisis del comportamiento de las variables meteorológicas y contaminantes atmosféricos, conjuntamente con las reglas de asociación de forma mensualizada. Así, los datos analizados fueron los recolectados por las estaciones meteorológicas en la ciudad de Cuenca, Ecuador, del año 2018. De esa manera, una de las etapas primordiales que se realizaron en esta investigación fue la discretización de los datos, debido a que fue necesario realizar una correcta modulación de los datos obtenidos para aplicar el algoritmo y obtener resultados eficientes.

Para concluir esta investigación es importante tener en cuenta que los árboles de decisión son algoritmos de aprendizaje supervisado no paramétricos que permiten predecir el valor de una variable objetivo mediante el uso de variables predictivas. Por lo tanto, al utilizar este método en esta investigación hizo posible un control periódico de la calidad de las asociaciones de contaminantes y variables meteorológicas por estacionalidad, siendo así la comprobación de las reglas de asociación en la mayoría de los meses estudiados cumpliéndose más de 70% de ellas.

Como un trabajo futuro, al tener un conjunto de datos atmosféricos más amplio se puede llegar a predecir mediante la aplicación de árboles de decisión el comportamiento de los contaminantes atmosféricos y presentar un tratamiento efectivo y preventivo ante factores de riesgo que puedan afectar negativamente la salud de la población.

Reconocimientos

Los autores desean agradecer al Vicerrectorado de Investigaciones de la Universidad del Azuay por el apoyo financiero y académico, así como a todo el personal de la escuela de Ingeniería de Ciencias de la Computación, y el Laboratorio de Investigación y Desarrollo en Informática (LIDI).

Referencias

- Ahmed, H. W., & Alamire, D. J. H. (2020). Labelling Data for Correlation Pollution Dataset by Using Machine Learning. *Journal of Physics: Conference Series*, 1530(1), 012028. <https://doi.org/10.1088/1742-6596/1530/1/012028>
- Alsagheer, R. H. A., Alharan, A. F. H., & Al-Haboobi, A. S. A. (2017). Popular Decision Tree Algorithms of Data Mining Techniques: A Review. *International Journal of Computer Science and Mobile Computing*, 6(6), 133-142.
- Althuwaynee, O. F., Balogun, A., & Al Madhoun, W. (2020). Air pollution hazard assessment using decision tree algorithms and bivariate probability cluster polar function: evaluating inter-correlation clusters of PM10 and other air pollutants. *GIScience & Remote Sensing*, 57(2), 207-226. <https://doi.org/10.1080/15481603.2020.1712064>
- Athira, V., Geetha, P., Vinayakumar, R., & Soman, K. P. (2018). DeepAirNet: Applying Recurrent Networks for Air Quality Prediction. *Procedia Computer Science*, 132, 1394-1403. <https://doi.org/10.1016/j.procs.2018.05.068>
- Birant, D. (2011). Comparison of Decision Tree Algorithms for Predicting Potential Air Pollutant Emissions with Data Mining Models. *Journal of Environmental Informatics*, 17(1), 46-53. <https://doi.org/10.3808/jei.201100186>
- García-Ojeda, J. C., Alvites, J. A., & Puello, P. (2020). Cuantificación de la contaminación por dióxido de carbono producida por empresas courier usando minería de datos: una mirada a las provincias de Lima y Constitucional del Callao (Colombia). *Revista ESPACIOS*, 41(41), 1015.
- Gayathri, M., Shankar, R., & Duraisamy, S. (2020). Air pollution prediction using data mining technique. *International Research Journal of Engineering and Technology (IRJET)*, 7(8), 4292-4297.
- Hemada, B., & Vijaya Lakshmi, K. S. (2013). A Study On Discretization Techniques. *International Journal of Engineering Research & Technology*, 2(8).
- Huang, C.-J., & Kuo, P.-H. (2018). A Deep CNN-LSTM Model for Particulate Matter (PM2.5) Forecasting in Smart Cities. *Sensors*, 18(7), 2220. <https://doi.org/10.3390/s18072220>
- Manisalidis, I., Stavropoulou, E., Stavropoulos, A., & Bezirtzoglou, E. (2020). Environmental and health impacts of air pollution: a review. *Frontiers in public health*, 8, 14.
- Martínez-España, R., Bueno-Crespo, A., Soto, J., Muñoz, A., & Cecilia, J. (2018). Air-Pollution Prediction in Smart Cities through Machine Learning Methods: A Case of Study in Murcia, Spain. *Journal of Universal Computer Science*, 24(3), 261-276.

- Orellana, M., Salto, J., & Cedillo, P. (2021). Behavior Analysis of Atmospheric Components and Meteorological Variables Applying Data Mining Association Techniques. En *Proceedings of the 2021 Future of Information and Communication Conference (FICC)* (Vol. 2, pp. 192-204). Springer, Cham. https://doi.org/10.1007/978-3-030-73103-8_12
- Siwek, K., & Osowski, S. (2016). Data mining methods for prediction of air pollution. *International Journal of Applied Mathematics and Computer Science*, 26(2), 467-478. <https://doi.org/10.1515/amcs-2016-0033>
- Whiteman, C. D., Hoch, S. W., Horel, J. D., & Charland, A. (2014). Relationship between particulate air pollution and meteorological variables in Utah's Salt Lake Valley. *Atmospheric Environment*, 94, 742-753. <https://doi.org/10.1016/j.atmosenv.2014.06.012>