

Modelo de Predicción Material Particulado (PM2.5) en Bogotá

Particulate Matter Prediction Model (PM2.5) in Bogota city

Luis Fernando Salas Núñez¹

¹<https://orcid.org/0000-0002-7335-3523>, Corporación Universitaria Minuto de Dios, Bogotá D.C Colombia, luis.salas.nu@uniminuto.edu.co

Fecha de recepción: 26/2022

Fecha de aceptación del artículo: 08/11/2022



Esta obra está bajo una licencia de Creative Commons Reconocimiento-No comercial-SinObraDerivada 4.0 internacional.

DOI: <https://doi.org/10.18041/1794-4953/avances.2.8574>

Cómo citar: Salas Núñez, L. F. (2022). Modelo de Predicción Material Particulado (PM2.5) en la ciudad de Bogotá. Avances Investigación En Ingeniería, 19(2). <https://doi.org/10.18041/1794-4953/avances.2.8574>.

Resumen

La contaminación del aire ha sido ampliamente estudiada debido a la gran amenaza que representa tanto para el medio ambiente como para la salud humana. Principalmente, las grandes ciudades se han visto afectadas por el problema de la contaminación del aire debido a diversos factores, como lo son la industrialización y la sobrepoblación. Estos últimos, tienen una implicación directa en la atmósfera mediante el aumento de las emisiones de gases de efecto invernadero y esmog, lo cual contribuye al agotamiento de la capa de ozono. La contaminación atmosférica trae consigo graves implicaciones en la salud humana, que afectan directamente el sistema respiratorio y generan consecuencias como irritación en la garganta, empeoramiento de afecciones como asma, bronquitis y hasta daños pulmonares permanentes. Sin embargo, teniendo en cuenta la gran importancia que implica la detección de agentes contaminantes, como lo es el material particulado (PM), es posible plantear alternativas para predecir su aparición en determinada localización. En este estudio, se propone el uso de herramientas computacionales y regresión lineal múltiple para predecir el comportamiento del material particulado PM2.5 de la localidad Kennedy, Bogotá. Esta predicción se realizó con la implementación de un algoritmo, que toma como base datos históricos de material particulado suspendido en el aire, Ozono, NO, CO, dirección y velocidad del viento durante el año 2018. Los análisis realizados fueron comparados con datos reales, con el propósito de determinar la veracidad de las predicciones hechas por el algoritmo. Los resultados revelaron que el nivel del PM2.5 es consecuencia de las emisiones de CO, lo cual está relacionado con el tráfico de vehículos, ya que son los mayores productores de este gas. Además, la predicción de los valores de PM2.5 evidenció una media de error del 5.6 PM2.5, lo que significa que el modelo logra una predicción precisa de la calidad del aire. Teniendo en cuenta estos resultados, este estudio representa una alternativa prometedora para la predicción del material particulado, de tal manera que se puedan establecer rutas de acción para mitigar el daño causado por estos contaminantes en el medio ambiente.

Palabras clave: Contaminación del aire, Machine Learning, Material particulado, Regresión lineal, Series de tiempo.

Abstract

Air pollution has been widely studied due to the great threat it poses to both the environment and human health. Mainly, the big cities have been affected by the problem of air pollution due to various factors,

such as industrialization and overpopulation. The latter have a direct implication in the atmosphere by increasing greenhouse gas emissions and smog, which contributes to the depletion of the ozone layer. Air pollution brings serious implications for human health, directly affecting the respiratory system and generating consequences such as throat irritation, worsening conditions such as asthma, bronchitis and even permanent lung damage. However, considering the great importance of detecting polluting agents, such as particulate matter (PM), it is possible to propose alternatives to predict their appearance in a certain location. In this study, we propose the use of computational tools and multiple linear regression to predict the behavior of PM_{2.5} particulate matter from the Kennedy locality, Bogotá. This prediction was made with the implementation of an algorithm, which is based on historical data of particulate matter suspended in the air, Ozone, NO, CO, wind direction and speed during the year 2018. The analyzes carried out were compared with real data, with the purpose of determining the veracity of the predictions made by the algorithm. The results revealed the level of PM_{2.5} is a consequence of CO emissions, which is related to vehicle traffic, since they are the largest producers of this gas. In addition, the prediction of PM_{2.5} values showed a mean error of 5.6 PM_{2.5}, which means that the model achieves an accurate prediction of air quality. Considering these results, this study represents a promising alternative for the prediction of particulate matter, in such a way that action routes can be established to mitigate the damage caused by these pollutants in the environment.

Keywords: Air pollution, linear regression, machine learning, particulate matter, time series.

1. Introducción

La contaminación del aire es un problema ambiental creciente a lo largo de los años y tiene muchas causas, entre las que se encuentran el aumento de la población y su concentración en grandes ciudades, la industrialización sin procesos amigables con el medio ambiente, la deforestación y quema de bosques, entre otras [1, 2].

La contaminación del aire es una mezcla de partículas sólidas y gases producto de las emisiones de la quema de combustible fósil de automóviles y fábricas, polvo, polen y esporas de moho [1, 2]. El ozono, un gas, es un componente fundamental de la contaminación del aire en las ciudades, el cual se da a nivel de suelo debido a la reacción de la luz del sol con gases contaminantes. Cuando el ozono forma parte de la contaminación del aire también se denomina esmog; respirar este provoca tos, irritación en la garganta, empeoramiento de afecciones como asma, bronquitis y enfisema y hasta daños pulmonares permanentes, si la exposición a este es habitual [3].

Los contaminantes atmosféricos que más impactan en la salud humana son conocidos como material particulado (PM), cuyo tamaño oscila entre los 10 micrómetros. Estos son producto de la quema de combustibles fósiles y su composición puede variar según la fuente que los produce. Además, se ha determinado que estos pueden afectar al ser humano cuando ingresan por el sistema respiratorio e interactúan con este, lo que provoca daños en el organismo como muerte celular o cáncer [4, 5].

En el 2018, la exposición a aire contaminado cobró la vida de 8,7 millones de personas en el mundo [4]. Es así como la estimación de la exposición que tenemos los seres humanos a este fenómeno es de gran importancia para detectar la contaminación presente en el ambiente, para permitir tomar acciones correctivas y no preventivas. Desde el 2017 existe el Índice Bogotano de Calidad del Aire (IBOCA) que se encarga de comunicar el pronóstico de material particulado PM₁₀, PM_{2.5} y Ozono; esta estimación se realiza por medio de estaciones de medida como la red de monitoreo de la calidad del aire de Bogotá (RMCAB) [6, 7].

1.1 Calidad del aire

Bogotá ha pasado por varias emergencias ambientales a causa de la gran cantidad de material particulado suspendido en el aire en algunas zonas, lo cual provoca graves enfermedades, especialmente en la población infantil [8,6]. Bogotá cuenta con una Red de Monitoreo de la Calidad del Aire (RMCAB), que permite recolectar información sobre la concentración de material particulado (PM10, PST, PM2.5), de gases contaminantes (SO₂, NO₂, CO, O₃) y de las variables meteorológicas de precipitación, velocidad y dirección del viento, temperatura, radiación solar, humedad relativa y presión barométrica, en forma continua y permanente [7].

Además de esto, la ciudad se encuentra dentro de un plan decenal de corto y largo plazo para la disminución de los contaminantes aéreos; con dicho plan se da cumplimiento a los tratados firmados por Colombia como el Protocolo de Kyoto, el Acuerdo de París y acoge al Decreto 335 por medio del cual se adopta la estrategia para la actualización del Plan Decenal de Descontaminación del Aire para Bogotá – PDDAB [6].

Aunque el modelo se ha centrado en Bogotá, es necesario destacar que son muchas ciudades nacionales e internacionales que sufren de este mismo efecto, dada la alta cantidad de vehículos e industrias que funcionan en ellas.



Figura 1. Estaciones de medición de la Red de Monitoreo de la Calidad del Aire en Bogotá. Fuente: [6]

Según la resolución 2254 del 1 de noviembre de 2017, emanada por el Ministerio de Ambiente y Desarrollo Sostenible de Colombia en su capítulo I en el artículo 2 muestra los niveles máximos permisibles de contaminantes y en su tabla No. 1 se resumen algunos de los contaminantes y su tiempo de exposición; en el capítulo II se enuncian también los niveles de prevención, alerta o emergencia. En el párrafo 1, del artículo 2 se dispone un cambio en la concentración máxima permisible a un tiempo de exposición de 24 horas para el material particulado PM2.5, el cual, a partir del 1 de julio de 2018, quedó en un máximo permisible de 37µg/m³ [9].

El objetivo de este estudio es diseñar e implementar un algoritmo capaz de predecir el comportamiento del material particulado PM2.5 de la localidad Kennedy en la ciudad de Bogotá. Este algoritmo usará los datos de material particulado suspendido en el aire, Ozono, NO, CO, dirección y velocidad del viento reportados por la estación RMCAB durante el año 2018.

En este documento se presenta un algoritmo que, basado en series de tiempo y regresión lineal, usa los datos proporcionados por la RMCAB, para predecir el comportamiento del material particulado PM2.5 en Bogotá.

2. Metodología

2.1 Series de tiempo

Los pronósticos son muy usados en muchos tipos de industrias debido a la importancia de tomar acciones antes de que los eventos ocurran o prevenir que estos sucedan. Sectores como el transporte, la ingeniería, la medicina, la agricultura, la física, entre otras disciplinas, usan distintos métodos para reconstruir un modelo a partir de variables que representan una correlación entre el futuro, presente y pasado. De esta manera se puede hacer una caracterización estadística

a dichas variables y generar extrapolaciones que pueden generar pronósticos.

En el caso de las predicciones de tipo climático se cuenta con mucha información histórica al respecto, como la que proporciona el IDEAM (Instituto de Hidrología, Meteorología y Estudios Ambientales), el RMCAB (Red de Monitoreo de la Calidad del Aire) que son medidas de material particulado suspendido en el aire, Ozono, NO, CO, precipitaciones, temperatura, dirección y velocidad del viento que sirven como datos de entrenamiento para muchas metodologías como lo son: regresiones lineales, regresión dinámica, función de transferencia y redes neuronales. El objetivo es encontrar la predicción que mejor se aproxime a la realidad, y por ello se usan comparaciones de predicciones contra la realidad y así ejercer un criterio correcto de cuál herramienta es de mayor utilidad [10].

Las variaciones de una serie de tiempo se clasifican en sistemáticas y aleatorias; las sistemáticas ocurren con regularidad y se pueden modelar; las aleatorias son causadas por situaciones aisladas como terremotos, huelgas y, en consecuencia, son difíciles de modelar. Para este caso en particular se usarán como métodos de aprendizaje para el modelo variaciones sistemáticas, ya que se conoce que las variaciones aleatorias son insuficientes dentro del modelo y el histórico de datos [10-12].

2.2 Modelos de regresión

El objetivo de una regresión es el de relacionar los datos asociados en una relación funcional que se supone existe entre el valor esperado de la variable dependiente y las variables independientes, minimizando la suma de los cuadrados de los errores de una muestra de n observaciones [11, 12]. De esta manera, en este estudio se aplica el modelo de regresión lineal simple utilizando el método de los mínimos cuadrados.

$$y(t) = b_0 + b_1 x(t-1) + b_2 x(t-2) + \dots + b_n x(t-n) + e(t) \quad (\text{EC1})$$

También se deben tener en cuenta algunos supuestos:

- Las variables independientes del modelo son determinísticas.
- Los errores son variables aleatorias que cumplen con:
 - La media es constante e igual a cero
 - La varianza es constante y es igual a σ^2 .
 - No posee covarianza, $\text{cov}(E_i, E_j) = 0$ donde $i \neq j$
 - $E_j \sim N(0, \sigma^2)$ donde $j = 1, 2, \dots, n$.

Los datos anteriores nos permiten saber que el proceso es estacionario; adicionalmente, el ruido blanco en sentido estricto presenta un valor en cada instante del tiempo que no depende de cuál haya sido su valor en los instantes precedentes y que no ejerce influencia alguna en sus valores futuros [10-12].

3. Procedimientos y resultados

Para realizar la predicción de la calidad del aire (PM_{2.5}) por medio de series de tiempo usamos los datos de la estación de medición de la calidad de aire de Kennedy. Las observaciones que se usaron fueron las del año 2018 [13].

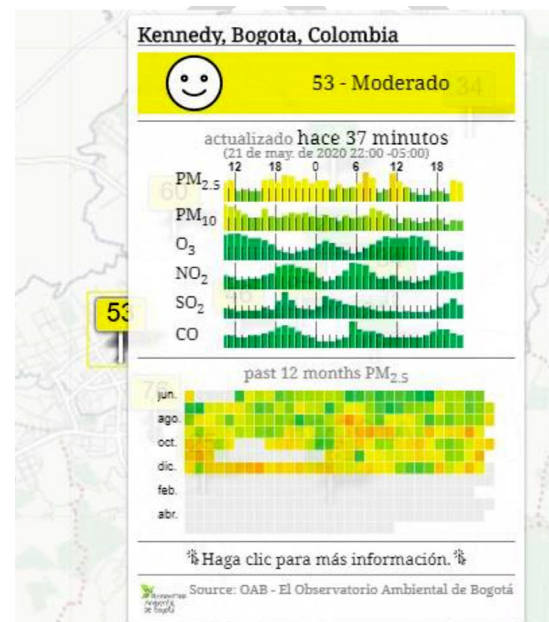


Figura 2. Localización de estación de medición. Fuente: [7].

3.1 Preprocesamiento

Lo primero que se hizo fue ordenar los datos para visualizarlos de una manera sencilla. Lo siguiente fue seleccionar un rango de tiempo donde no faltaran tantos datos, ya que se identificó que había temporadas en las que algún sensor estaba desconectado y todos los datos de la variable estaban en blanco.

Aunque se eligió un rango de tiempo en el cual no faltaran tantos datos, igual, ciertos valores estaban en blanco. Para resolver este problema se utilizó una interpolación lineal para incluir los datos faltantes.

Finalmente, como último paso, se filtraron los datos con un filtro de media móvil con una ventana de 3 datos con el objetivo de reducir la volatilidad de los datos y que estos representen un comportamiento más general de la variable por predecir [14, 15].

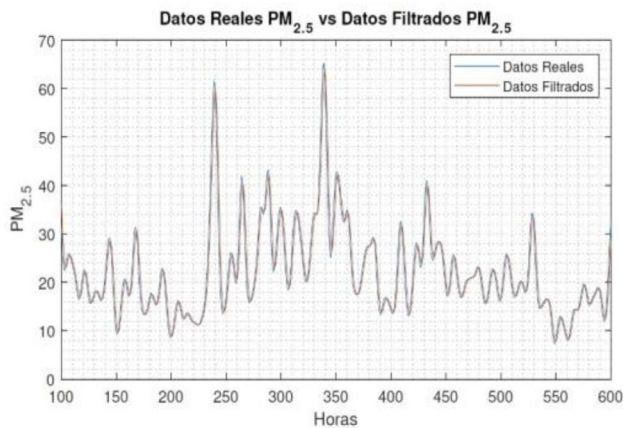


Figura 3. Filtrado de datos de PM_{2.5}, este filtrado se aplicó a todos los datos de la estación de monitoreo. Fuente: [Autor].

3.2 Modelo

Una vez realizado el preprocesamiento de los datos se planteó el modelo de predicción. Primero los datos por predecir van a ser el PM_{2.5}:

$$y(t)=PM_{2,5} \quad (EC2)$$

Ahora, para los datos de entrada del modelo se eligieron de las horas anteriores de la misma variable por predecir, se incluyó el dato de la misma hora, pero del día anterior y se agregó el dato del promedio de PM_{2.5} a esa hora de todos los datos de entrenamiento. Además de los datos de tiempo también se incluyeron otros factores como la velocidad, y la dirección del viento, la cantidad de CO, la cantidad de NO, la cantidad de ozono, la hora del día y la fecha [16, 17].

Todo esto se organizó en un vector de la siguiente manera:

$$x(t)=b_1 x(t-1)+b_2 x(t-2)+b_3 x(t-3)+b_4 x(t-4)+b_5 x(t-5)+b_6 x(t-6)+b_7 x(t-7)+b_8 x(t-8)+b_9 x(t-9)+b_{10} x(t-24)+b_{11} x_{prom}+b_{12} N(t-1) +b_{13} C(t-1)+b_{14} C(t-2)+b_{15} C(t-3)+b_{16} O(t-1)+ b_{17} V(t-1)+b_{18} H(t)+b_{19} D(t)+b_{20} DV(t-1) \quad (EC3)$$

Donde:

- $x(t)$, es la cantidad de material particulado de 2.5 en el instante.
- $(t-j)$, son las horas anteriores a la medición actual, $j \in \mathbb{Z}$
- x_{prom} , es la cantidad de material particulado de 2.5 promedio en la estación de acuerdo con el día anterior.
- $N(t-1)$, es la cantidad de Monóxido de nitrógeno (NO) hace una hora.
- $C(t-1)$, es la cantidad de Monóxido de carbono (CO) hace una hora.
- $O(t-1)$, es la cantidad de Ozono hace una hora.
- $V(t-1)$, es la velocidad del viento hace una hora.
- $H(t)$, es la hora del día.
- $D(t)$, es el día de la semana.
- $DV(t-1)$, es la dirección del viento.
- b_i , son los coeficientes de la regresión lineal, donde $i=1,2,\dots,20$.

Se hace X un vector que contenga las variables $x(t-i), x_{prom}, N(t-1), C(t-i), O(t-1), V(t-1), H, D$ y $DV(t-1), y=x(t)$ y θ un vector que contenga los coeficientes b_i , ahora se expresa la regresión en términos de X, y y θ :

$$y=X^T \theta \quad (EC4)$$

Se conocen los valores de X y y se necesita estimar θ , por lo cual, planteamos el siguiente problema de minimización:

$$\operatorname{argmin}_{\theta} J(\theta) = \sum_{i=1}^r (y^{(i)} - X^{(i)T} \theta)^2 \quad (EC5)$$

donde r: N° datos

Nótese que para estimar los valores de θ se usa el criterio del error cuadrático, pero también se puede usar el criterio de error; ahora se reescribe esta ecuación en términos de unas nuevas variables Y y H (EC6), para luego llevarlos a la forma estándar de un problema de regresión lineal, se define:

$$Y = [y^{(1)} \ y^{(2)} \ \dots \ y^{(r)}]^T \quad H = \begin{bmatrix} X^{(1)T} \\ X^{(2)T} \\ \vdots \\ X^{(r)T} \end{bmatrix} \quad (EC6)$$

Si se expande la sumatoria del problema de minimización (EC5) se nota que es la norma L₂ (euclidiana) al cuadrado; por tanto se puede expresar el problema como:

$$\operatorname{argmin}_{\theta} J(\theta) = \|Y - H\theta\|_2^2 \quad (EC7)$$

Analizando J(θ) en (EC7) se encuentra que su Hessiana es 2H^TH y esta es una matriz positiva semidefinida, por lo cual para encontrar el minimizador de la función se aplica descenso de gradiente [11, 12][14, 15]:

$$\begin{aligned} \nabla_{\theta} J(\theta^*) &= 0 \\ \nabla_{\theta} J(\theta^*) &= \nabla_{\theta} [\|Y - H\theta\|_2^2] \\ \nabla_{\theta} [Y^T Y - Y^T H\theta - \theta^T H^T Y + \theta^T H H^T \theta] \\ &\quad - 2H^T Y + 2H^T H\theta \\ \nabla_{\theta} J(\theta^*) = 0 &= -2H^T Y + 2H^T H\theta^* \\ \theta^* &= (H^T H)^{-1} H^T Y \quad (EC8) \end{aligned}$$

Nótese que en (EC8) se debe garantizar que (H^TH) > 0, es decir, positiva definida para que se pueda invertir; esto se cumple si la cantidad de variables $\theta(n)$ es menor que el

número de datos (r) para hallar el modelo lo cual aplica en nuestro caso [11, 12].

3.3 Entrenamiento modelo

Para realizar el entrenamiento del modelo se usó el 80 % de los datos disponibles ya que se deben dejar reservados ciertos datos para la validación. También se tuvo en cuenta separar los datos de entrenamiento y de validación lo suficiente para que al realizar la verificación del modelo no se utilicen datos con los que fue entrenado.

Al realizar la regresión con los datos de entrenamiento se obtuvieron los siguientes coeficientes:

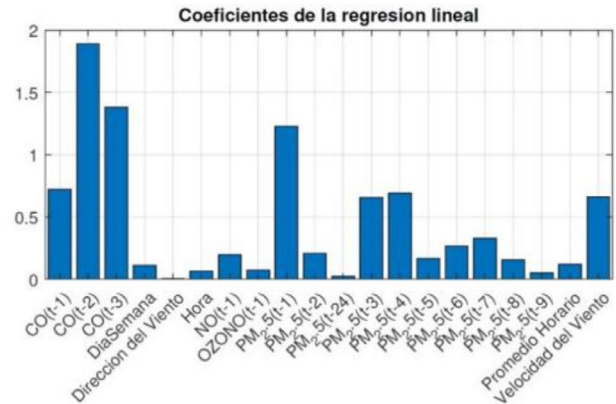


Figura 5. Coeficientes de regresión lineal. Fuente: [Autor]

Estos coeficientes dan una noción general de la importancia de cada variable en relación con la variable que se quiere predecir. Es decir, que las variables que presentan mayor magnitud en el modelo son el CO en el presente y en la hora anterior, la velocidad del viento y el promedio horario del PM2.5. Aunque las otras variables no presentan un coeficiente alto no se pueden ignorar pues en conjunto estas variables pueden aportar un valor importante al modelo.

Analizando los coeficientes de la regresión, se puede inferir que el nivel del PM2.5 está muy relacionado con los niveles de CO

y por tanto con el tráfico de la zona, pues los mayores productores de este gas son los vehículos; también que su concentración depende de la velocidad del viento, es decir, si la velocidad es alta el material particulado tiende a dispersarse.

3.4 Validación modelo

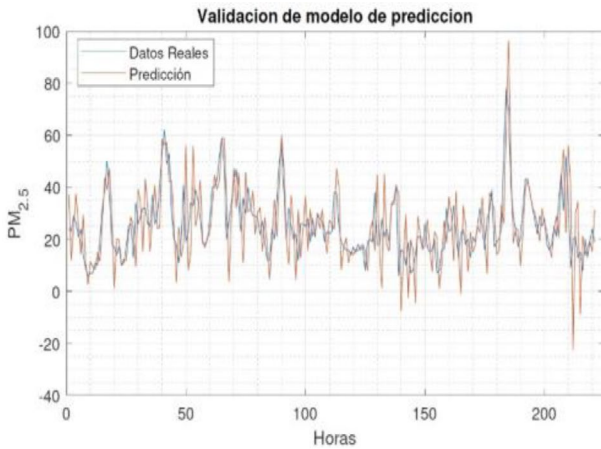


Figura 6. Validación de la predicción 1 hora después de la medición, comparada con datos de [13]. Fuente: [Autor].

Una vez calculados los parámetros del modelo lineal, se prosigue a realizar su validación. Se utilizan los datos sobrantes para realizar una predicción y se comparan con los datos reales de PM2.5 (Ver Figura 6).

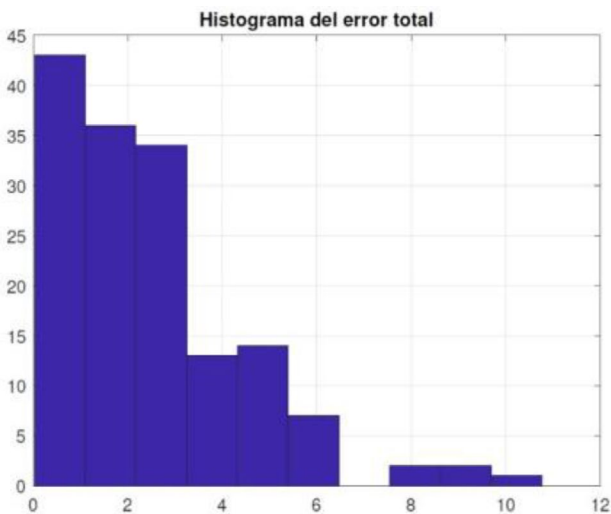


Figura 7. Histograma del error total. Fuente: [Autor].



Figura 8. % de error vs. horas de anticipación. Fuente: [Autor].

En las anteriores figuras (7 y 8) se puede evidenciar que el modelo predice los valores de PM2.5 de una manera bastante precisa, con una media de error de 5.6032 PM2.5 que equivale a un error del 21.8681%. Aunque esta predicción fue bastante acertada, existe la posibilidad de que se requiera una predicción con mayor tiempo de anticipación. A continuación, se realizará un análisis del error con respecto a las horas de anticipación de la predicción.

Al observar la figura 8 se puede decir que conforme se aumentan las horas de la predicción mayor es el error que tiene el modelo y que al acercarse a la misma que se encuentra el modelo, pero del día siguiente el error comienza a decaer de nuevo.

Esta investigación no tuvo financiamiento de ninguna clase.

4. Conclusiones

Este estudio presentó el diseño detallado de un algoritmo útil para la predicción del PM2.5, utilizando datos específicos de la localidad de Kennedy en Bogotá. Los resultados

permitieron determinar la importancia de conocer de manera temprana el comportamiento del material particulado en cierto espacio, con el fin de prevenir problemas asociados a la salud humana y afectaciones al medio ambiente. Asimismo, este estudio representa un primer paso para el uso de herramientas computacionales en la detección de agentes contaminantes, como el PM. Teniendo en cuenta esto, fue posible concluir que:

- Cada componente evaluado contiene distintas relaciones con el aumento y disminución de ciertos elementos en el aire; pero se notó que es importante tenerlos en cuenta a todos para disminuir el error causado por estos componentes ya que en conjunto pueden representar una variable importante.
- El modelo implementado logra ser muy acertado en la estimación futura de la contaminación o concentración de distintas sustancias en el aire.
- El error es relativamente bajo al tener en cuenta que contiene muchos valores aleatorios dentro de los datos de análisis y entrenamiento.
- Es necesario comparar el modelo implementado con modelos basados en redes neuronales o funciones de transferencia, para ver qué tan efectiva y precisa fue la predicción respecto a estas.
- El problema de contaminación en la ciudad requiere medidas y normatividades que están por fuera del alcance del proyecto; sin embargo, esas medidas pueden desarrollarse, así como pueden prevenirse emergencias a raíz de conocer cómo será el comportamiento futuro de la contaminación en la ciudad. De esta manera llegar a tomar decisiones preventivas y no correctivas.
- Es útil aplicar esta metodología con mayor cantidad de variables como tráfico, lluvia, radiación UV y temperatura, para intentar reducir el error tanto como sea posible.

Referencias bibliográficas

- [1] Apte, J., Marshall, J., Cohen, A., and Brauer, M. "Addressing Global Mortality from Ambient PM 2.5". *Environ. Sci. Technol.* 2015, 49, 13, 8057–8066, June 16, 2015. DOI: <https://doi.org/10.1021/acs.est.5b01236>
- [2] Weagle, C., Snider, G., Li, C., Donkelaar, A., Philip, S., Bissonnette, P., Burke, J., Jackson, J., Latimer, R., Stone, E., Abboud, I., Akoshile, C., Anh, N., Brook, J., Cohen, A., Dong, J., Gibson, M., Griffith, D., He, K., and Martin, R. "Global Sources of Fine Particulate Matter: Interpretation of PM 2.5 Chemical Composition Observed by SPARTAN using a Global Chemical Transport Model". *Environ. Sci. Technol.* Vol. 52, 20, 11670–11681 (2018). DOI: 10.1021/acs.est.8b01658.
- [3] Atkinson, R., Butland, B., Dimitroulopoulou, C., Heal, M., Stedman, J., Carslaw, N., & Jarvis, D., Heaviside, C., Vardoulakis, S., Walton, H. and Anderson, H. "Long-term exposure to ambient ozone and mortality: A quantitative systematic review and meta-analysis of evidence from cohort studies". *BMJ Open.* 6. e009493. DOI: [10.1136/bmjopen-2015-009493](https://doi.org/10.1136/bmjopen-2015-009493).
- [4] Vohra, K., Vodonos, A., Schwartz, J., Marais, E. A., Sulprizio, M. P., and Mickley, L. J. "Global mortality from outdoor fine particle pollution generated by fossil fuel combustion: Results from GEOS-Chem". *Environ. Res.*, vol. 195, no. 110754. (2021). DOI: <https://doi.org/10.1016/j.envres.2021.110754>
- [5] Li, X., Jin, L., and H. Kan, "Air pollution: a global problem needs local fixes". *Nature.*, 570, 437-439 (2019). DOI: <https://doi.org/10.1038/d41586-019-01960-7>
- [6] Secretaria distrital de Ambiente. Calidad del aire en Bogotá, Alcaldía de Bogotá, Bogotá, 2018. Disponible en: <https://www.ambientebogota.gov.co/calidad-del-aire>.
- [7] Red de Monitoreo de Calidad del Aire de Bogotá – RMCAB, Estaciones RMCAB, Alcaldía de Bogotá (2019). Disponible en: <http://rmcab.ambientebogota.gov.co/home/map>.
- [8] Mann, J. K., Lutzker, L., Holm, S. M., Margolis, H. G., Neophytou, A. M., Eisen, E. A., Costello, S., Tyner, T., Holland, N., Tindula, G., Prunicki, M., Nadeau, K., Noth, E. M., Lurmann, F., Hammond, S. K., and Balmes, J. R. "Traffic-related air pollution is associated with glucose dysregulation, blood pressure, and oxidative stress in children". *Environ Res.* 2021;195:110870. DOI: [10.1016/j.envres.2021.110870](https://doi.org/10.1016/j.envres.2021.110870).
- [9] Ministerio de Ambiente y Desarrollo Sostenible de Colombia. "Resolución 2254 del 01 de noviembre de 2017" (2017). Disponible en: https://www.icbf.gov.co/cargues/avance/docs/resolucion_minambienteds_2254_2017.htm
- [10] Alvin C. and Bruce G. "Linear Models In Statistics" 2nd Edition, New Jersey. John Wiley & Sons (2007).
- [11] Darlington, R. and Hayes, A. "Regression Analysis and Linear Models: Concepts, Applications, and Implementation", New York, The Guilford Press (2017).
- [12] Montero, R. "Modelos de regresión lineal múltiple", Documentos de Trabajo en Economía Aplicada. Universidad de Granada. España. Disponible en: https://www.ugr.es/~montero/matematicas/regresion_lineal.pdf.
- [13] Datos Abiertos, "https://www.datos.gov.co/" Medición de la calidad del aire en Bogotá, 2018. [Online]. Available: <https://www.datos.gov.co/Ambiente-y-Desarrollo-Sostenible/Medicin-de-la-calidad-del-aire-en-Bogota-> [Accessed 2019].
- [14] Forsyth F. and Ponce J. "Computer Vision: A Modern Approach" 2nd Edition, New Jersey. Pearson (2012).
- [15] Bertsekas, D. "Nonlinear Programming" 2nd Edition. Belmont. Athena Scientific (1999).
- [16] Jin, L., XIE, J., Wong, C., Chan, S., Abbaszade, G., Schnelle-Kreis, J., Zimmermann, R., Li, J., Zhang, G., Fu, P., and Li, X. "Contributions of City-Specific Fine Particulate Matter (PM2.5) to Differential In Vitro Oxidative Stress and Toxicity Implications between Beijing and Guangzhou of China". *Environ. Sci. Technol.* vol. 53, issue 5, pp. 2881-289 (2019). DOI: 10.1021/acs.est.9b00449.
- [17] Shaddick, G., Thomas, M., Jobling, A., Brauer, M., Donkelaar, A., Burnett, R., Chang, H., Cohen, A., Van Dingenen, R., Dora, C., Gumy, S., Liu, Y., Martin, R., Waller, L., West, J., Zidek, J., and Prüss-Ustün, A. "Data Integration Model for Air Quality: A Hierarchical Approach to the Global Estimation of Exposures to Ambient Air Pollution". *J R Stat Soc Series B Stat Methodol* vol. 67(1), pages 231-253, (2016). DOI: 10.1111/rssc.12227.