

# ESTUDIO ESTADÍSTICO DEL CO2 EN EL AULA

Colaboradores: García Mata Daniela, Mateo Fernández Argimiro F, Medina Talavera Javier, Moralejo Márquez Blanca, Ruiz Álvarez David, Valenzuela Jiménez Juan C. Profesor coordinador: D<sup>a</sup> M<sup>a</sup> Antonia Mateos Camacho

Colegio La Inmaculada Algeciras: C/ Misioneras Concepcionistas 1, 11205 Algeciras, Cádiz [www.lainmaculadaalgeciras.com](http://www.lainmaculadaalgeciras.com) [mantoniamateos@gmail.com](mailto:mantoniamateos@gmail.com)

## INTRODUCCIÓN

Este trabajo surge por el interés de introducir a los alumnos/as en la programación y en técnicas estadísticas avanzadas. Teniendo en cuenta el avance en Big Data, Machine Learning y Deep Learning, que hoy día está presente en cualquier ámbito de nuestra vida, se hace necesario transmitir conocimiento más profundo acerca de cómo realmente se tratan los datos. Aprovechando el momento que se está viviendo en las clases debido a la pandemia y la recopilación de datos del aire en el aula se planteó investigar un modelo matemático que predijera los niveles de contaminación de CO<sub>2</sub>, primero mediante las técnicas propias del currículum de matemáticas de bachillerato, para luego contrastarlas con métodos más avanzados. Demostrar a los alumnos/as de forma experimental la importancia de la selección y estudio de la relevancia de las variables es primordial para que comprendan en profundidad el significado del coeficiente de correlación y la bondad de ajuste de un modelo estadístico. Aprender a manejar un programa como R, los paquetes específicos para el estudio de la relevancia de las variables fundamentados en el teorema de Bayes, los que permiten ejecutar algoritmos y comparar los modelos y sus gráficas, es fundamental para el futuro tanto académico como profesional de estas generaciones.



## FIGURAS



Figura 1. Correlaciones variables

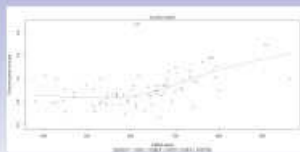


Figura 2. Modelo de regresión múltiple

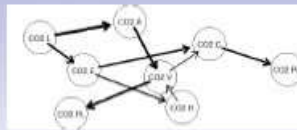


Figura 3. Estructura Bayesiana

Conditional density: CO2.C | CO2.E + CO2.V  
Coefficients:  
(Intercept) CO2.E CO2.V  
-769.17617 78.19248 -22.79721

Figura 4. Modelo matemático



CO<sub>2</sub>

## RESULTADOS

Se estudió la matriz de correlación, figura 1, y se observa poca correlación entre las variables.

Se crearon modelos lineales de regresión simple y múltiple atendiendo a la correlación. Se fueron introduciendo variables con distintos modelos lineales, e incluso se creó uno con todas las variables estudiadas. Como resultados se ha obtenido que estos modelos matemáticos no son buenos, ya que su coeficiente de correlación y de determinación múltiple, respectivamente son cercanos a 0. Figura 2

Por ello, se planteó un estudio más profundo de la relevancia de las variables y sus relaciones ocultas a través de algoritmos como Hill Climbing y otros algoritmos híbridos. Se crearon distintas estructuras y se compararon obteniendo la red de la figura 3. Los nodos son las variables, y los arcos nos da la relación entre las variables. Se puede apreciar el arco azul que indica que hay enlaces entre variables ocultas, y la relación entre ellas.

A partir de esta estructura, se utilizó el algoritmo para hallar la probabilidad entre los arcos, todo ello basado en el Teorema de Bayes. Y se obtuvo un modelo matemático que predice los niveles de CO<sub>2</sub>, basándose en las variables que son relevantes para su inferencia. Figura 4

## CONCLUSIONES

Como conclusiones hemos obtenido:

Se ha demostrado que los modelos lineales no ajustan bien los datos.

De forma experimental se ha investigado la importancia de la selección y estudio de la relevancia de las variables, y se ha constatado que es primordial para comprender en profundidad el significado del coeficiente de correlación y la bondad de ajuste de un modelo estadístico.

Se ha obtenido mediante un algoritmo, basado en una técnica de optimización matemática, un modelo que permite predecir la cantidad de CO<sub>2</sub> en el aula.

Aprender a manejar un programa como R, los paquetes específicos para el estudio de la relevancia de las variables fundamentados en el teorema de Bayes, los que permiten ejecutar algoritmos y comparar los modelos, es fundamental para el futuro tanto académico como profesional del alumnado.

