

Clasificación y Análisis de emisiones de Gases de Efecto Invernadero

Helena Patricia Carrillo Soto*

Abstract

The abstract is a brief (usually one paragraph) summary of the whole paper, including the problem, the method for solving it (when not obvious), the results, and the conclusions suggested or drawn. Do not write the abstract as a hasty afterthought. Profesor si está leyendo esto aun no escribo un abstract porque aun no tengo terminado el análisis.

Palabras clave: Gases de Efecto Invernadero, Aprendizaje No Supervisado, Machine Learning

1 Introducción

Los Gases de Efecto Invernadero (GEI) son los componentes de la atmósfera que absorben la radiación emitida por la superficie de la Tierra y la emiten de regreso.

Los principales GEI son el vapor de agua (H_2O), el dióxido de carbono (CO_2), el óxido nitroso (N_2O), el metano (CH_4) y el ozono (O_3).

También se encuentran GEI creados en su totalidad por el ser humano, como los halocarbonos (CFCs, HCFCs, HFCs y PFCs) los cuales contienen cloro, bromo o flúor y carbono. Estos compuestos son una de las causas del agotamiento de la capa de ozono en la atmósfera. (Ballesteros and Aristizabal, 2007) .

Los Gases de Efecto Invernadero están directamente relacionados con el calentamiento global y este a su vez con el cambio climático.

Este tema ha cobrado relevancia en los último años con el rápido aumento de la temperatura media global, los intentos fallidos de distintos gobiernos de reducir sus emisiones y la llamada de científicos alrededor del mundo de tomar medidas urgentes para controlar el cambio climático.

2 Metodología y Datos

En el presente trabajo se utilizarán métodos de Machine Learning para analizar los datos correspondientes al Inventario Nacional de Emisiones de Gases y Compuestos de Efecto Invernadero del INECC que se puede encontrar en la siguiente liga .

Los datos consisten en las emisiones anuales totales, y desglosadas por industria, de los diferentes gases de efecto invernadero para los años de 1990 a 2021.

*Facultad de Ciencias Físico Matemáticas, Universidad Autónoma de Nuevo León (UANL). Email: helenacarrilloso@uanl.edu.mx

3 Aprendizaje No Supervisado

El aprendizaje no supervisado, usa algoritmos de machine learning para analizar y clasificar datos no etiquetados.

Se utilizan principalmente para encontrar patrones ocultos en los datos.

3.1 K-Means

El algoritmo K-means es un método popular de clustering por aprendizaje no supervisado.

El algoritmo se basa en agrupar los datos en k grupos en base a su media.

En él asumimos que se tiene un set de objetos $D = \{p_1, p_2, \dots, p_n\}$ donde cada objeto $p_i = \{p_{i1}, p_{i2}, \dots, p_{in}\}$ representa un punto en el espacio R^m donde m es el número de atributos o dimensiones de los objetos del set D .

Se define k como el número de clusters y m_k la media del cluster C_k . El objetivo general del método es minimizar la función criterio $e(k)$.

Normalmente la función criterio que se minimiza es la suma de cuadrados del error:

$$e(k) = \sum_{i=1}^k \sum_{p_j \in C_i} \text{dist}(p_j, m_i)^2$$

donde

p_j - punto en el espacio R^m que **representa** el punto p_j .

m_i - media del cluster C_i .

$\text{dist}(p_j, m_i)$ - distancia Euclidiana del punto p_j a la media (centro) del cluster más cercano C_i .

(Kijewska and Bluszczy, 2016)

3.2 K-means en el análisis de GEI

El algoritmo de K-means se ha usado para el análisis de emisiones de Gases de Efecto invernadero en múltiples ocasiones.

Anna Kijewska y Anna Bluszczy hicieron uso de él en 2016 para analizar los niveles de emisiones de GEI variantes en países de Europa. (Kijewska and Bluszczy, 2016)

En 2018, Lozza y Bellini lo utilizaron para realizar la clasificación de sistemas ganaderos para estimaciones de GEI. (Lozza and Bellini Saibene, 2018)

En esta ocasión se usará el algoritmo para categorizar las actividades que aportan al total de emisiones de Gases de Efecto Invernadero en México de los años de 1990 a 2021.

3.3 Resultados análisis K-means

Para establecer el número de clusters a utilizar se utilizó el método del codo en el cuál se calcula la suma de las distancias al cuadrado para diferentes valores de k y se busca el valor de k donde la disminución en la suma se realentiza o se acerca a su límite.

El resultado obtenido se puede ver en la **Figura 1**:

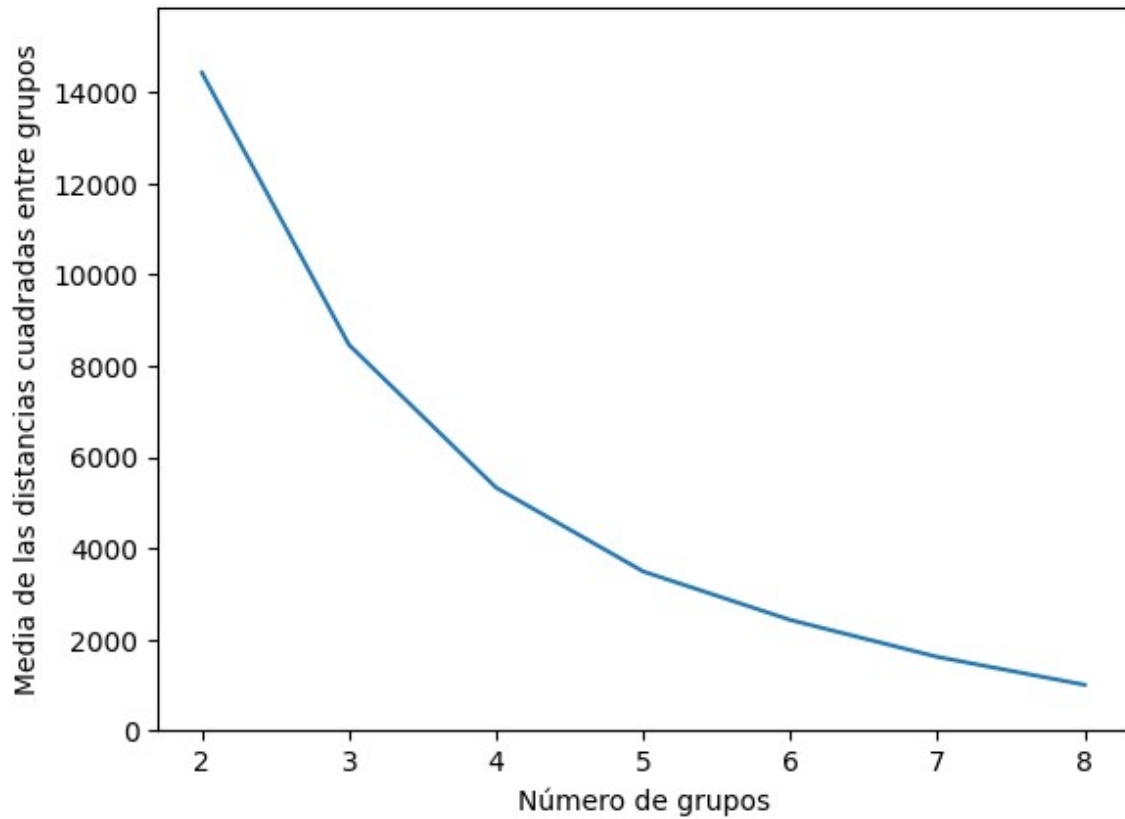


FIGURE 1: Método del codo

Para tener un criterio más claro se puede calcular la distancia entre la recta creada por el mínimo y máximo número de clusters y sabemos que se optimiza k en el número que tenga una mayor distancia a la recta. Podemos observarlo en la Figura 2.

En base a esto se usan 4 clústers.

Los resultados arrojan los centros encontrados en la Tabla 1:

TABLE 1: Centros: Puntos donde se encuentran los centros de cada uno de los clusters C_k .

C_k	x	y	z	w	v	u	r	s
1	-0.0907	-0.0822	-0.0089	0.0020	-0.0469	0.0019	0.0019	-0.1110
2	-0.1175	10.6739	-0.1918	-0.0760	-0.0551	-0.0724	-0.0724	3.0160
3	5.7762	-0.1407	0.6824	-0.0760	-0.0551	-0.0724	-0.0724	5.4903
4	-0.1129	-0.1810	-0.1918	-0.0760	19.3721	-0.0724	-0.0724	-0.1575

También se obtienen los tamaños de grupo mostrados en la Tabla 2:

Debido a que el espacio es de 8 dimensiones solo se muestran una de las caras en la Figura 3.

4 Apendizaje Supervisado

A diferencia del aprendizaje no supervisado, el aprendizaje supervisado utiliza datos etiquetados. Este utiliza datos de entrenamiento para enseñar a los modelos a generar la salida deseada.

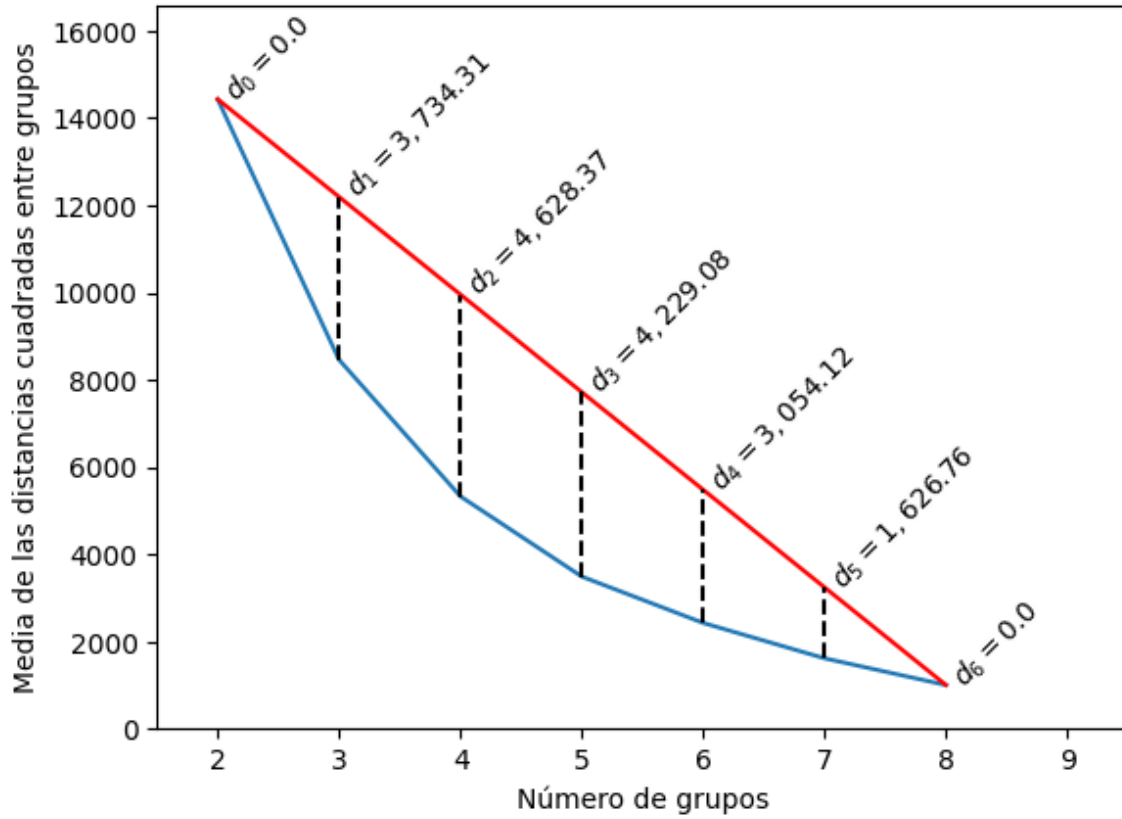


FIGURE 2: Método del codo con distancias

 TABLE 2: Tamaño: Tamaño de cada uno de los clusters C_k .

Cluster	nk
1	4022
2	32
3	64
4	10

4.1 Modelos de aprendizaje supervisado en el análisis de GEI

Distintos modelos de aprendizaje supervisado se han usado a lo largo del tiempo para analizar emisiones de Gases de Efecto Invernadero.

Zewei Jiang y Shihong Yang junto con otros investigadores usaron los modelos de Random Forest (RF), regresiones K-Nearest Neighbor (KNN), regresiones Gradient Boosting(GBR), y regresiones Lineares (LR), así como modelos stacking, para modelar emisiones de GEI en diferentes escalas de tiempo de campos de arroz. (Jiang et al., 2023) .

Por otro lado, Sabrina Hempel y Julian Adolph junto con su equipo usaron regresiones Gradient Boosting, Random Forests y regresiones lineares múltiples, regresiones Ridge regression (Ridge), una red neuronal con una capa oculta así como Support Vector Machines (SVM) y procesos Gaussianos para evaluar las emisiones de metano de un edificio lechero con ventilación natural. (Hempel et al., 2020)

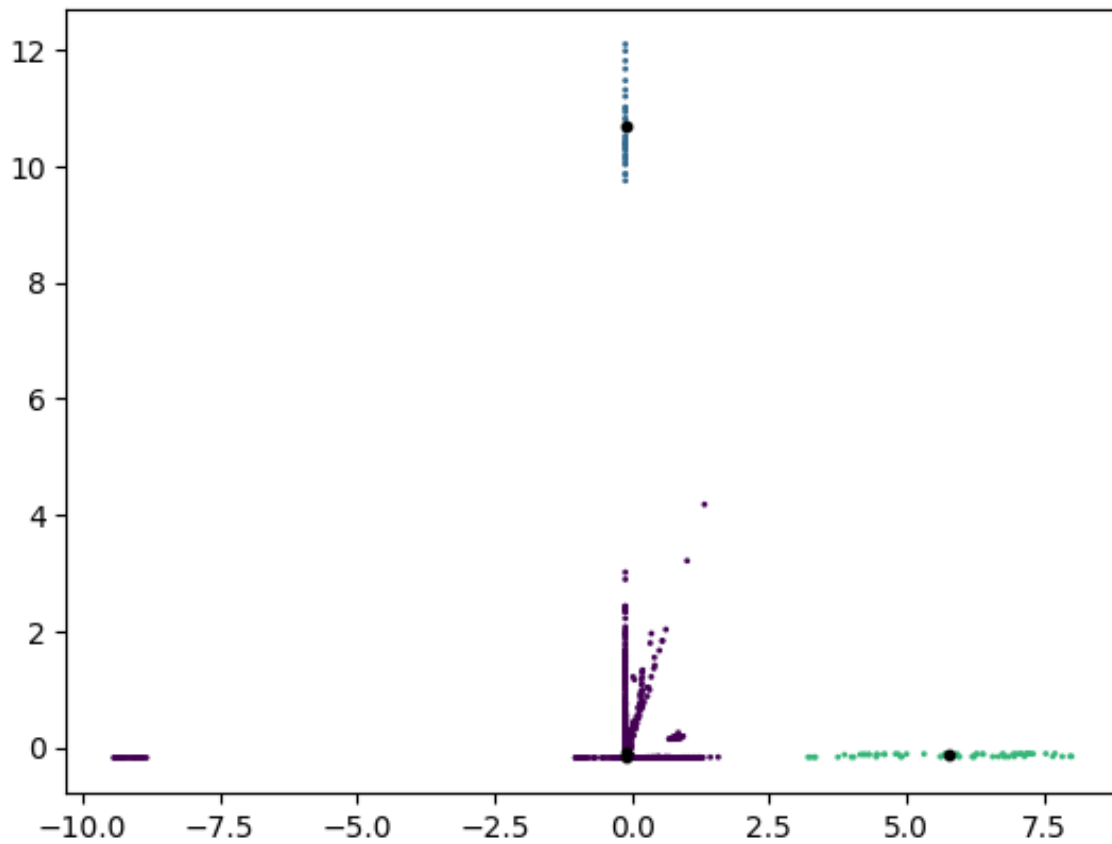


FIGURE 3: Clusters de CO2 vs CH4

4.2 Análisis y Resultados

Para este estudio primero se analizaron diferentes modelos de los encontrados en la literatura para evaluar cuál de ellos era el de mejor desempeño. Se usó RMSE (Root Mean Squared Error) para compararlos.

Los modelos usados fueron:

- Bayesian Ridge (BR)
- Gradient Boosting (GBR)
- Regresión Linear (LR)
- Random Forest (RF)
- K-Nearest Neighbor (KNN)
- Regresor Stacking usando GBR y RF

Todos se revisaron un total de 10 veces y los mejores 20 resultados fueron los presentados en la Tabla 3:

TABLE 3: Mejores 20 resultados de pruebas de regresiones por RMSE

	Modelo	RMSE	Tiempo
1	LinearReg	8.717637e-16	0.014209
2	LinearReg	1.206522e-15	0.002382
3	LinearReg	1.242156e-15	0.007742
4	LinearReg	1.269252e-15	0.003934
5	LinearReg	1.518410e-15	0.004000
6	LinearReg	1.524309e-15	0.002510
7	LinearReg	1.551162e-15	0.003907
8	LinearReg	1.640076e-15	0.002428
9	LinearReg	1.782332e-15	0.002427
10	LinearReg	2.923362e-15	0.008730
11	BayesRidge	9.344281e-08	0.007534
12	BayesRidge	9.438717e-08	0.002169
13	BayesRidge	1.188605e-07	0.003024
14	BayesRidge	1.465989e-07	0.002062
15	BayesRidge	1.574630e-07	0.003200
16	BayesRidge	1.631492e-07	0.003612
17	BayesRidge	1.832583e-07	0.010631
18	BayesRidge	2.056415e-07	0.003067
19	BayesRidge	2.430986e-07	0.002257
20	BayesRidge	4.581878e-07	0.003622

References

- Ballesteros, H. B. and Aristizabal, G. L. (2007). Información técnica sobre gases de efecto invernadero y el cambio climático. *Instituto de Hidrología, Meteorología y Estudios Ambientales-IDEAM. Subdirección de Meteorología (Bogotá, Colombia)*. 96p.
- Hempel, S., Adolphs, J., Landwehr, N., Willink, D., Janke, D., and Amon, T. (2020). Supervised machine learning to assess methane emissions of a dairy building with natural ventilation. *Applied Sciences*, 10(19).
- Jiang, Z., Yang, S., Smith, P., and Pang, Q. (2023). Ensemble machine learning for modeling greenhouse gas emissions at different time scales from irrigated paddy fields. *Field Crops Research*, 292:108821.
- Kijewska, A. and Bluszcz, A. (2016). Research of varying levels of greenhouse gas emissions in european countries using the k-means method. *Atmospheric Pollution Research*, 7(5):935–944.
- Lozza, A. and Bellini Saibene, Y. (2018). Clasificación de sistemas ganaderos para estimación de gases de efecto invernadero. In *Conferencia Latinoamericana sobre Uso de R en Investigación+ Desarrollo (LatinR 2018)-JAIIO 47 (CABA, 2018)*.