

## **Taller Machine learning**

### **Presentado por:**

Alexander Rendón Londoño.

Braian Estrada

### **Presentado a:**

Juan Camilo

### **Programa:**

Ingeniería Informática

Universidad Autónoma Latinoamericana

**(UNAULA)**

2023-2

## Interpretaciones:

### Regresión Lineal:

**Coefficiente  $R^2$  de Regresión Lineal:** El coeficiente  $R^2$  es una medida de la bondad del ajuste del modelo de regresión lineal. En este caso, el valor es aproximadamente 0.948. Esto significa que el modelo de regresión lineal explica la variabilidad en los datos de "Energy indust" de manera efectiva, ya que  $R^2$  está cerca de 1, lo que indica una fuerte correlación entre las variables.

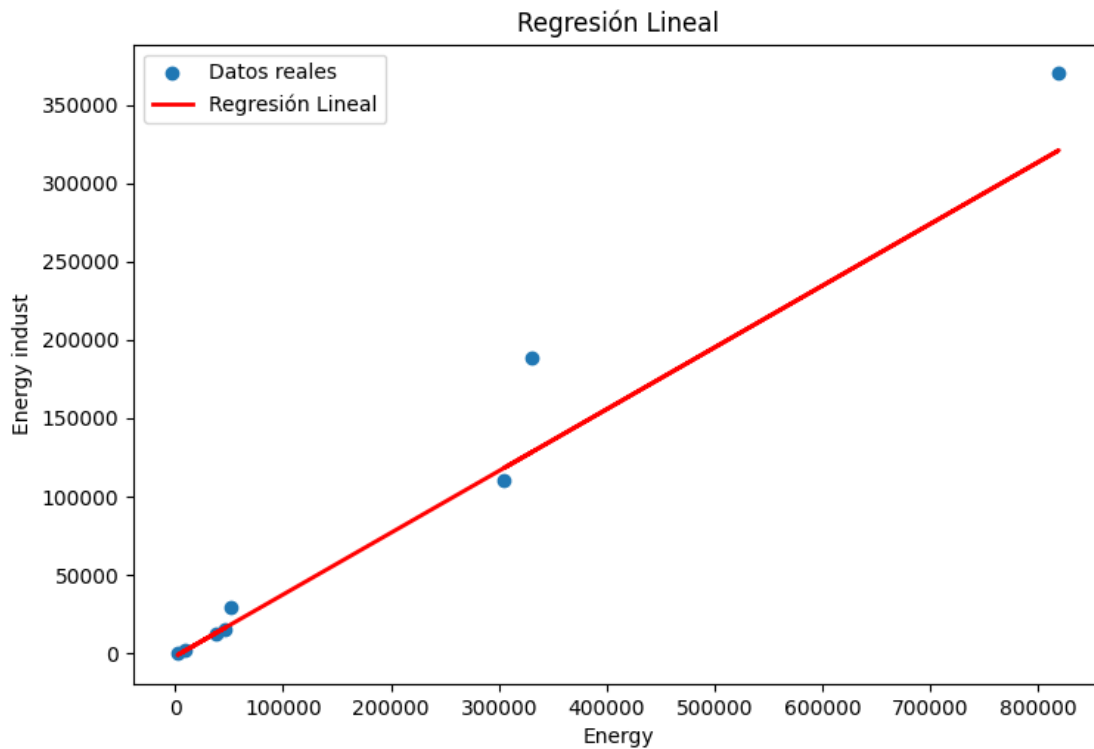
### Parámetros del modelo de regresión lineal:

- **Coefficientes:** El coeficiente asociado a la variable "Energy" es aproximadamente 0.394, lo que sugiere que un aumento en "Energy" se asocia con un aumento en "Energy indust", y el coeficiente es positivo.
- **Ordenada al origen:** La ordenada al origen (intercepto) es aproximadamente -1995.11. Esto significa que, si "Energy" es igual a cero, se espera que "Energy indust" sea aproximadamente -1995.11.
- **Tiempo de ejecución:** El tiempo de ejecución del modelo de regresión lineal fue de aproximadamente 11.73 segundos.

```
Seleccione un modelo:
1. Regresión Lineal
2. Árboles de Decisión
3. K-Means
4. Salir
Ingrese el número de la opción que desea: 1
Coeficiente  $R^2$  de Regresión Lineal: 0.947957925416499
Parámetros del modelo de regresión lineal:
Coeficientes: [[0.39444802]]
Ordenada al origen: [-1995.10659028]
Tiempo de ejecución: 11.726954698562622 segundos
```

**Gráfico de Regresión Lineal:** El gráfico muestra los datos reales (puntos dispersos) y la línea de regresión lineal (en rojo). La fuerte correlación positiva entre "Energy" y "Energy indust" se refleja en la pendiente positiva de la línea de regresión. La alta proximidad de los puntos de datos a la línea indica que el modelo se ajusta muy bien a los datos y puede predecir "Energy indust" a partir de "Energy" de manera efectiva. La alta calidad del ajuste se refleja en el alto valor del coeficiente

$R^2$  (cercano a 1), lo que significa que el modelo explica la mayoría de la variabilidad en "Energy indust".



### Árboles de Decisión:

**Coefficiente  $R^2$  de Árboles de Decisión:** El coeficiente  $R^2$  es una medida de la bondad del ajuste del modelo de Árboles de Decisión. En este caso, el valor es aproximadamente 0.634. Esto significa que el modelo de Árboles de Decisión explica una parte significativa de la variabilidad en los datos de "Energy indust", pero no tan efectivamente como el modelo de Regresión Lineal, ya que  $R^2$  es menor que 1. Sin embargo, sigue siendo un valor bastante razonable y sugiere una buena capacidad del modelo para predecir "Energy indust".

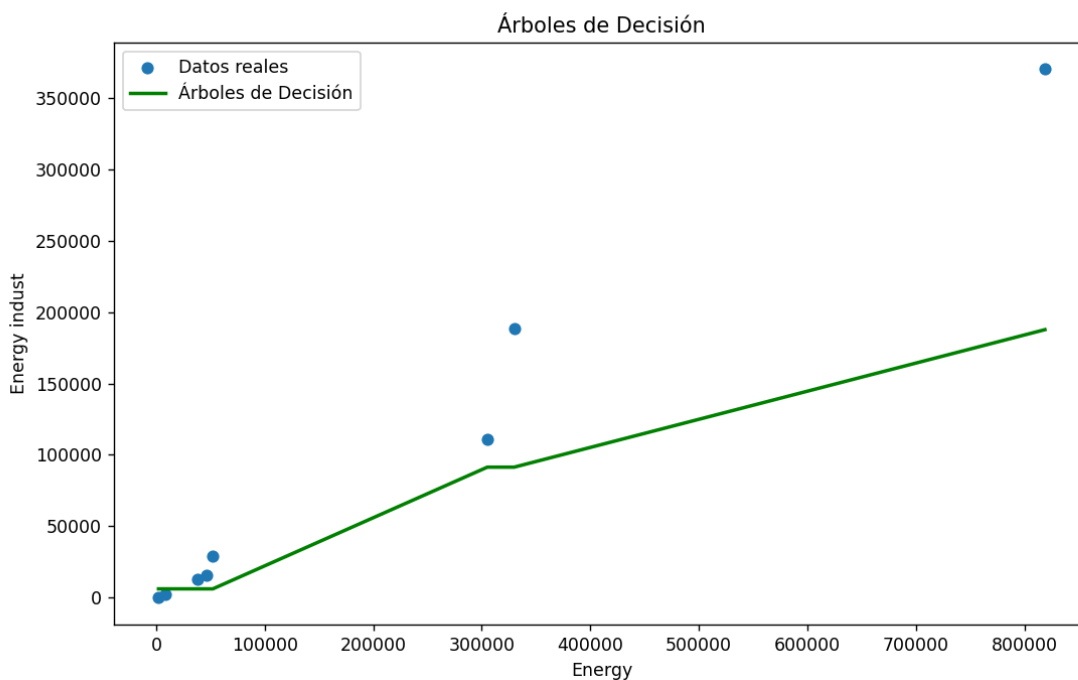
**Profundidad máxima del árbol:** La profundidad máxima del árbol en este modelo es 5. Esto indica que el árbol de decisión se ha construido con una profundidad limitada, lo que significa que no es muy profundo. La profundidad máxima limitada puede ayudar a evitar el sobreajuste del modelo a los datos.

**Tiempo de ejecución:** El tiempo de ejecución del modelo de Árboles de Decisión fue de aproximadamente 6.06 segundos.

```

Seleccione un modelo:
1. Regresión Lineal
2. Árboles de Decisión
3. K-Means
4. Salir
Ingrese el número de la opción que desea: 2
Coeficiente R^2 de Árboles de Decisión: 0.6338700507256245
Profundidad máxima del árbol: 5
Tiempo de ejecución: 6.0613977909088135 segundos
  
```

En el gráfico de Árboles de Decisión, los datos reales se muestran como puntos dispersos, y la línea representa la predicción del modelo. Aunque el modelo no se ajusta tan estrechamente a los datos como el modelo de Regresión Lineal, todavía proporciona una buena aproximación a la relación entre "Energy" y "Energy indust". La profundidad limitada del árbol sugiere que el modelo está equilibrado entre la complejidad y la capacidad de generalización, evitando el sobreajuste.

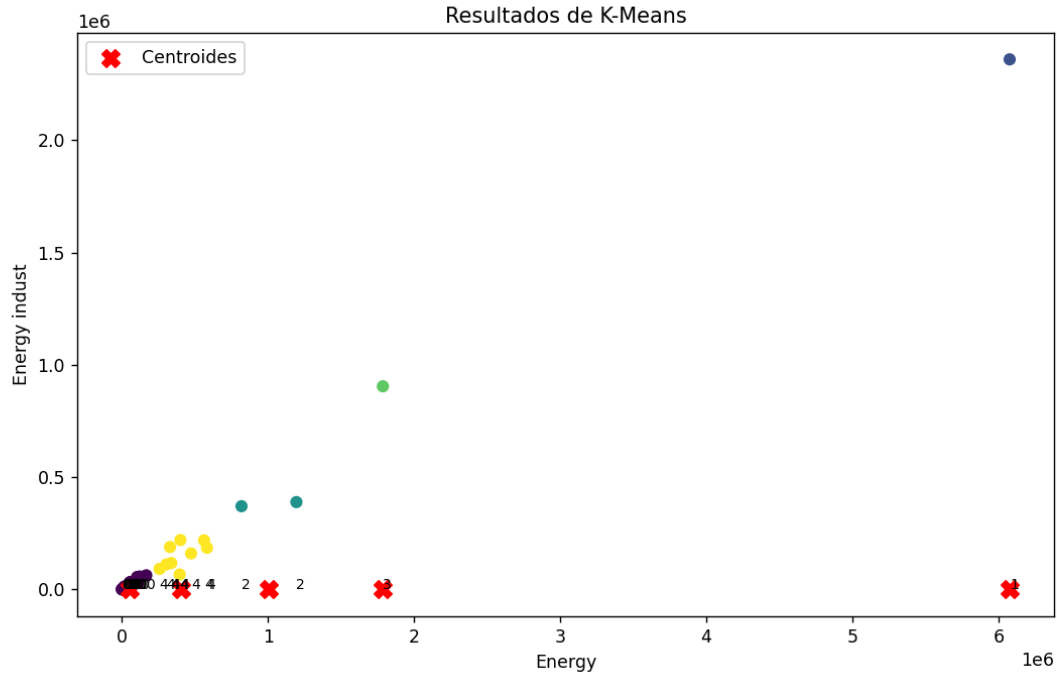


## K-Means:

**Puntuación de silueta de K-Means:** La puntuación de silueta es una medida de cuán bien definidos están los clusters en el modelo de K-Means. En este caso, la puntuación de silueta es aproximadamente 0.724, lo que indica que los clusters son bastante bien definidos y separados. Una puntuación de silueta cercana a 1 sugiere que los puntos dentro de un cluster están cerca entre sí y bien separados de otros clusters.

```
Seleccione un modelo:
1. Regresión Lineal
2. Árboles de Decisión
3. K-Means
4. Salir
Ingrese el número de la opción que desea: 3
Puntuación de silueta de K-Means: 0.7238118311838251
Cluster 4: [0, 5, 11, 17, 27, 33, 36, 37, 38]
Cluster 0: [1, 2, 3, 4, 6, 7, 8, 9, 10, 13, 14, 15, 16, 19, 20, 21, 22, 23, 24, 25, 26, 28, 29, 31, 32, 34, 35]
Cluster 2: [12, 18]
Cluster 3: [30]
Cluster 1: [39]
Tiempo de ejecución: 14.027527570724487 segundos
```

- **Clusters:** El modelo de K-Means ha agrupado los datos en cinco clusters diferentes. A continuación, se muestra una descripción de los elementos en cada cluster:
- **Cluster 0:** Este cluster contiene un grupo grande de países. Estos países tienen características similares en términos de emisiones de gases de efecto invernadero. Son más homogéneos en comparación con otros clusters.
- **Cluster 1:** Este cluster contiene un solo país, que se diferencia significativamente de los demás en términos de emisiones de gases de efecto invernadero. Este país es un "outlier" en el conjunto de datos.
- **Cluster 2:** Este cluster contiene dos países que también son diferentes de los demás en términos de emisiones. Pueden tener características específicas que los diferencian de otros países.
- **Cluster 3:** Este cluster contiene un solo país, que es un "outlier" similar al Cluster 1.
- **Cluster 4:** Este cluster contiene varios países que tienen características similares en términos de emisiones de gases de efecto invernadero. Son más homogéneos en comparación con otros clusters.



## Conclusión:

Estos modelos ofrecen diferentes perspectivas para analizar el conjunto de datos de emisiones. La elección del modelo depende de los objetivos y las necesidades específicas del análisis. La Regresión Lineal es adecuada para predecir relaciones lineales entre variables, los Árboles de Decisión ofrecen flexibilidad y capacidad de generalización, mientras que K-Means es útil para agrupar datos en clusters basados en similitudes. La comprensión de las características de cada modelo permite tomar decisiones informadas al abordar problemas de Machine Learning en el futuro.