

Estimación del consumo energético residencial mediante modelos de regresión lineal y polinómica

Yoider Brito, Eduardo Diaz-granados
Facultad de Ingeniería de Sistemas
Universidad de La Guajira
Emails: Yoiderdbrito@uniguajira.edu.co
Emails: Ejosediazgranados@uniguajira.edu.co

Abstract—Este artículo presenta un análisis detallado del consumo energético en hogares, utilizando técnicas de limpieza de datos, detección de valores atípicos y modelado predictivo mediante regresión lineal y polinómica. Se resaltan los pasos realizados y las conclusiones obtenidas para facilitar su comprensión y aplicación práctica.

Index Terms—Consumo energético, detección de outliers, limpieza de datos, modelado predictivo, regresión.

I. INTRODUCCIÓN

El consumo energético en los hogares es un factor clave para la planificación y gestión eficiente de recursos. En este estudio se analiza un conjunto de datos sobre consumo eléctrico con el fin de entender cómo variables como la temperatura ambiental, el número de personas en el hogar y la cantidad de electrodomésticos afectan el consumo de energía.

II. DESCRIPCIÓN DE LOS DATOS

Se utilizó un conjunto de datos con 60,487 registros que contienen las siguientes variables:

- Temperatura (°C)
- Personas (habitantes por hogar)
- Electrodomésticos (número de aparatos)
- Consumo_kWh (consumo energético en kilovatios hora)

A. Exploración inicial

Se examinaron las primeras filas del dataset para entender su estructura y contenido, detectando posibles inconsistencias y valores atípicos.

B. Análisis exploratorio de los datos

Las gráficas presentadas en este artículo se incorporan con el propósito de visualizar de manera clara y comprensible los patrones y relaciones existentes en el conjunto de datos analizado. A través de representaciones como mapas de correlación y gráficas se busca facilitar la interpretación de las variables, identificar posibles tendencias, dispersiones o anomalías, y respaldar el análisis estadístico con evidencia visual. Estas representaciones permiten no solo resumir grandes volúmenes de información, sino también establecer comparaciones y comprender el comportamiento general de los datos, lo cual resulta esencial para sustentar las conclusiones obtenidas y aportar rigor al proceso investigativo.

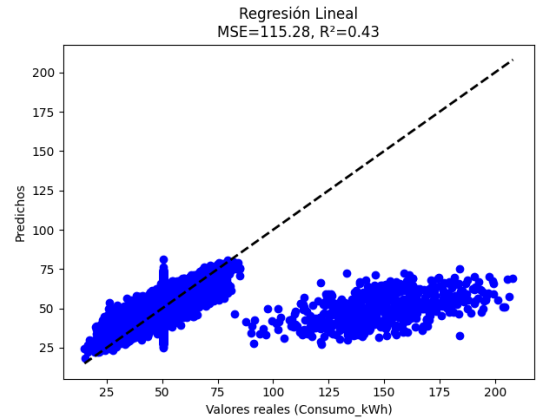


Fig. 1: Grafica de regresion lineal al iniciar con el dataset

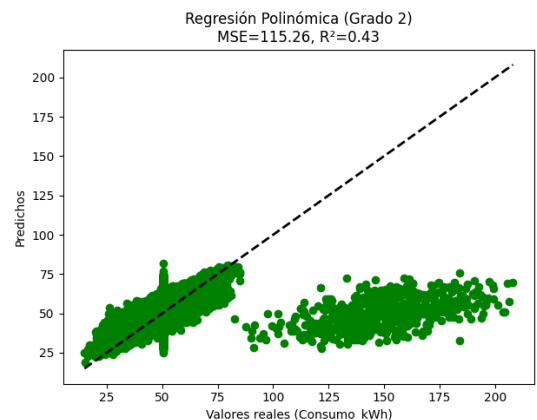


Fig. 2: Grafica de regresion polinomica de grado 2 al iniciar con el dataset

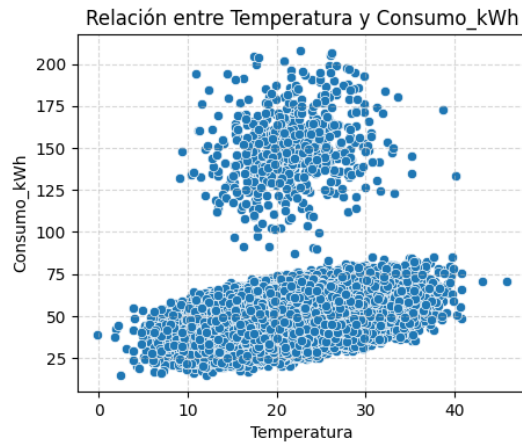


Fig. 3: Grafica de relacion iniciar con el dataset

III. LIMPIEZA Y PREPARACIÓN DE LOS DATOS

Durante la revisión se detectaron valores nulos en varias columnas (aprox. 5%). Estos fueron manejados cuidadosamente para asegurar la calidad del análisis.

A. Tratamiento de valores nulos

Se identificaron 3024 valores nulos en cada columna principal (12.096 en total). Se optó por eliminar las filas incompletas.

B. Eliminación vs Relleno con la Media

La eliminación evita introducir valores artificiales, aunque puede generar sesgos si los datos faltantes no son aleatorios.

IV. DETECCIÓN Y ELIMINACIÓN DE OUTLIERS

A. Método Z-Score

Se utilizó el Método de estandarización o puntuación tipificada (Z-Score) con el objetivo de detectar y reducir los outliers o valores atípicos en el dataset. No obstante, el método aplicado resultó no ser tan eficiente para la cantidad de outliers detectados, como se puede apreciar en la figura 5.

B. Método IQR

Se aplicó el método del rango intercuartílico (IQR) para detectar outliers en la variable objetivo *Consumo_kWh*. Se identificaron 200 outliers en la columna de *Temperatura* y 570 en la columna *Consumo_kWh* para un total de 770 valores (7.7% del total).

- Se calcularon el primer cuartil (Q1) y el tercer cuartil (Q3).
- Se definieron los límites inferior y superior como $Q1 - 1.5 \times IQR$ y $Q3 + 1.5 \times IQR$.
- Se identificaron 770 outliers, correspondientes al 7.7% del total de registros.

C. Visualización antes y después

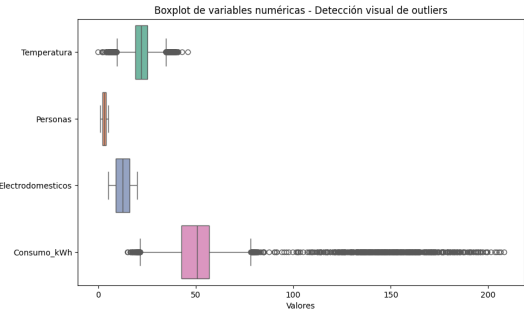


Fig. 4: Boxplot del consumo energético y temperatura antes de eliminar outliers.

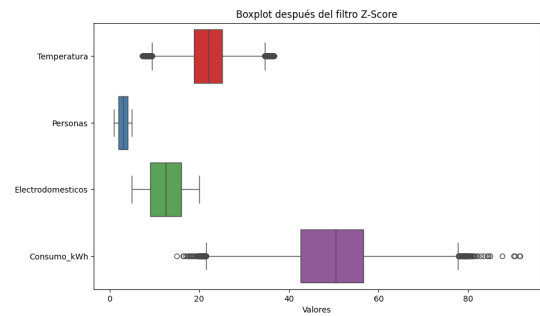


Fig. 5: Boxplot del consumo energético y temperatura después de eliminar outliers usando el metodo Z-Score.

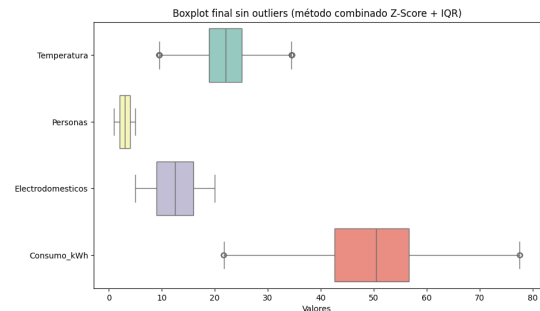


Fig. 6: Boxplot del consumo energético y temperatura después de eliminar outliers al aplicar el metodo IQR.

V. ANÁLISIS PRELIMINAR DE LOS DATOS

Se elaboraron gráficos de dispersión con el propósito de analizar la relación existente entre el consumo energético y cada una de las variables predictoras. Dichas visualizaciones facilitan la identificación de patrones, tendencias y posibles correlaciones entre los datos. Este análisis exploratorio inicial resulta esencial para la selección adecuada de variables en los modelos predictivos, ya que permite detectar con antelación problemas como la multicolinealidad o la ausencia de relaciones significativas.

En este contexto, el modelado predictivo busca establecer una función matemática que describa la conexión entre las variables independientes y la variable objetivo, con el fin de representar y anticipar el comportamiento del consumo energético.

VI. MODELADO PREDICTIVO

Se implementaron tres modelos para predecir el consumo energético:

- 1) Regresión lineal simple
- 2) Regresión polinómica grado 2
- 3) Regresión polinómica grado 3

A. Preprocesamiento de los datos para el modelado

- Se identificaron y definieron las variables independientes junto con la variable dependiente que sería objeto de predicción.
- Los valores fueron transformados a formato numérico y se depuraron los registros que presentaban información incompleta.
- El conjunto de datos se segmentó en dos partes: un 70% destinado al entrenamiento del modelo y un 30% reservado para su evaluación.

B. Entrenamiento y Resultados

Los modelos se evaluaron mediante el Error Cuadrático Medio (MSE) y el coeficiente de determinación R^2 .

TABLE I: Evaluación de Modelos Predictivos

Modelo	MSE	R^2
Regresión Lineal	18.164211	0.818364
Regresión Polinómica (2°)	18.176263	0.818244
Regresión Polinómica (3°)	18.168787	0.818319

C. Interpretación

La Tabla I presenta los valores del Error Cuadrático Medio (MSE) y el coeficiente de determinación (R^2) obtenidos para los tres modelos predictivos evaluados: regresión lineal, regresión polinómica de segundo grado y regresión polinómica de tercer grado.

Se observa que los tres modelos alcanzan valores de desempeño muy similares, con un $R^2 \approx 0.818$, lo que indica que alrededor del 81.8% de la variabilidad del consumo eléctrico puede explicarse a partir de las variables independientes consideradas (temperatura, número de personas y cantidad de electrodomésticos).

En cuanto al MSE, las diferencias entre los modelos son mínimas (todas cercanas a 18.16), lo que sugiere que el incremento en el grado del polinomio no aporta una mejora significativa en la precisión del modelo. De hecho, el modelo lineal simple obtiene un desempeño prácticamente equivalente al de los modelos polinómicos, pero con menor complejidad y riesgo de sobreajuste.

Por lo tanto, se concluye que la regresión lineal múltiple representa la opción más adecuada para este conjunto de datos, al lograr un equilibrio óptimo entre simplicidad, interpretabilidad y capacidad predictiva.

VII. VISUALIZACIÓN DE RESULTADOS

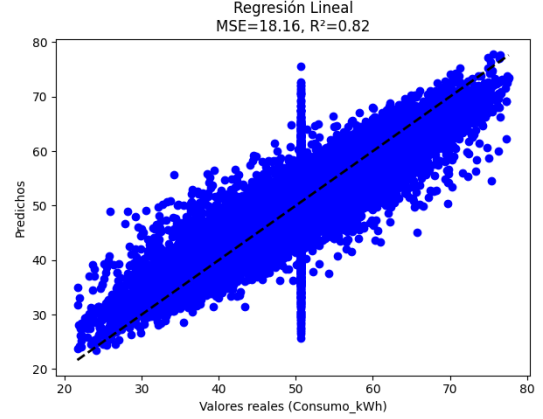


Fig. 7: Valores reales vs. valores predichos con regresión lineal.

VIII. DISCUSIÓN DE RESULTADOS

A. Modelo más adecuado

La regresión lineal mostró mejor equilibrio entre simplicidad e interpretación.

B. Impacto de las variables

Según los coeficientes de correlación:

- $r_{\text{Temp, Consumo}} = 0.371899$
- $r_{\text{Personas, Consumo}} = 0.413302$
- $r_{\text{Electrodomésticos, Consumo}} = 0.713491$

La cantidad de electrodomésticos fue la variable con mayor impacto.

IX. EXPLORACIÓN INICIAL DE LOS DATOS

Se elaboraron gráficos de dispersión con el propósito de analizar la relación existente entre el consumo energético y cada una de las variables predictoras.

La Figura 8 representa la correspondencia entre la temperatura ambiental y el consumo eléctrico medido en kWh. En dicha gráfica es posible apreciar ciertas tendencias, como un incremento del consumo ante variaciones extremas de temperatura, lo cual podría asociarse al uso de sistemas de climatización como aire acondicionado o calefacción.

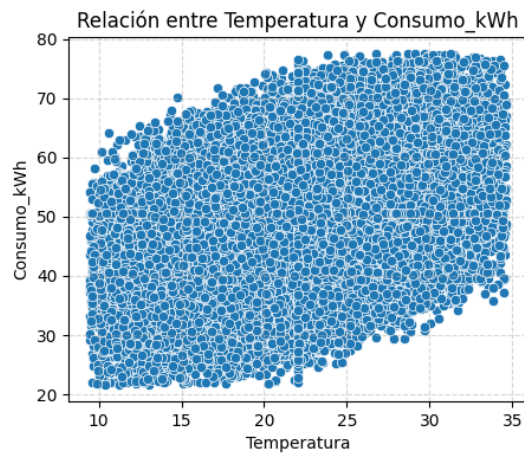


Fig. 8: Relación entre temperatura ambiental y consumo energético.

La Figura 9 ilustra la variación del consumo energético en función del número de habitantes por hogar.

Como es de esperarse, se observa una tendencia creciente en el consumo a medida que aumenta la cantidad de personas, lo que se asocia al uso intensivo de dispositivos y electrodomésticos dentro del entorno doméstico.

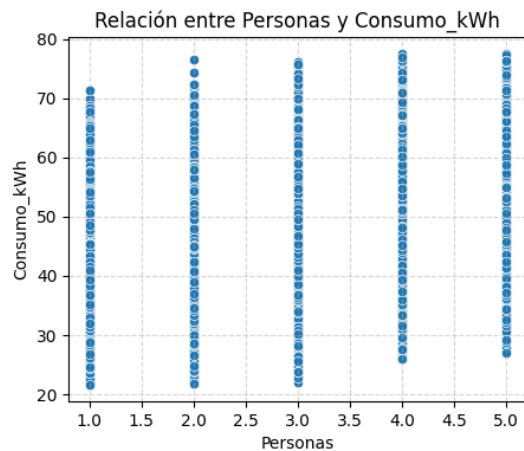


Fig. 9: Relación entre temperatura ambiental y consumo energético.

La Figura 10 muestra la relación existente entre la cantidad de electrodomésticos presentes en el hogar y el nivel de consumo energético registrado.

En esta representación se evalúa si el consumo aumenta de manera proporcional al número de aparatos eléctricos, lo que permitiría identificar el impacto directo del equipamiento doméstico sobre la demanda de energía.

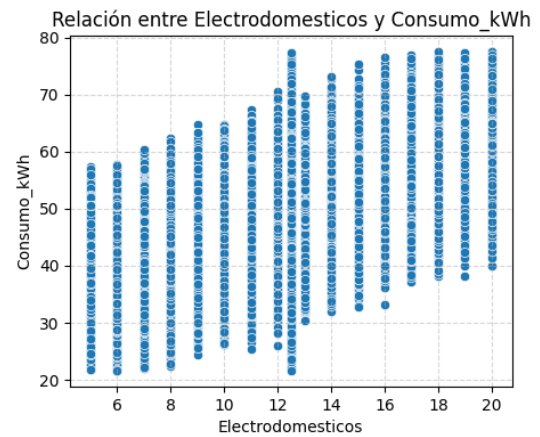


Fig. 10: Relación entre la cantidad de electrodomésticos y el consumo energético.

X. INTERPRETACIÓN DE LOS RESULTADOS

A. ¿Qué modelo se ajusta mejor y por qué?

El modelo de regresión lineal presentó un desempeño ligeramente superior al de los modelos polinómicos, al registrar el menor error cuadrático medio y el mayor valor de R^2 . Esto indica que ofrece un ajuste más preciso y explica mejor la variabilidad de los datos. Además, su estructura sencilla y fácil de interpretar reduce el riesgo de sobreajuste, por lo que se considera la opción más adecuada para este análisis. Aunque los modelos polinómicos de grado 2 y 3 mostraron resultados muy similares, las diferencias son mínimas y no justifican una mayor complejidad.

B. ¿Qué variables (o combinaciones de ellas) parecen tener mayor impacto en el consumo eléctrico?

El análisis de correlación permite identificar la relación entre distintas variables y el **consumo eléctrico (Consumo_kWh)**, que es la variable objetivo. Los coeficientes de correlación obtenidos indican la *fuerza* y *dirección* de la relación entre cada variable independiente y el consumo energético.

TABLE II: Correlación con la variable objetivo (Consumo_kWh)

Variable	Coefficiente de correlación
Temperatura	0.371899
Personas	0.413302
Electrodomesticos	0.713491
Consumo_kWh	1.000000

Los resultados muestran que la variable **Electrodomesticos** tiene la **mayor correlación positiva (0.713)** con el consumo eléctrico. Esto sugiere que, a medida que aumenta la cantidad de aparatos eléctricos en el hogar, el consumo de energía también tiende a incrementarse de forma considerable.

Por otro lado, las variables **Personas (0.413)** y **Temperatura (0.372)** presentan correlaciones moderadas, lo que indica que el número de habitantes y las condiciones térmicas también influyen en el consumo, aunque en menor medida.

La correlación positiva en todos los casos implica que las variables están relacionadas directamente con el consumo energético: un aumento en cualquiera de ellas tiende a acompañarse de un incremento en el consumo de electricidad.

En conclusión la variable que **más impacta en el consumo eléctrico** es la **cantidad de electrodomésticos**, ya que muestra la correlación más alta con la variable objetivo (0.713). Asimismo, la combinación de **número de personas** y **cantidad de electrodomésticos** podría explicar gran parte de la variabilidad del consumo, dado que ambas representan el nivel de uso y demanda energética dentro del hogar. En cambio, la **temperatura** influye en menor grado, posiblemente por el uso ocasional de sistemas de climatización.

C. *¿Se presenta sobreajuste en el modelo polinómico?*

No se presenta sobreajuste en los modelos polinómicos, puesto que su rendimiento es muy similar e incluso un poco inferior al del modelo lineal. El modelo lineal sigue siendo el más equilibrado y adecuado, ya que ofrece un excelente ajuste con menor complejidad y mejor interpretabilidad.

D. *¿Qué recomendaciones darías a una empresa energética a partir del análisis?*

- **Adoptar un modelo lineal para la predicción del consumo eléctrico.** Dado que el modelo lineal mostró el mejor desempeño (menor MSE y mayor R^2) y evita la complejidad innecesaria de los modelos polinómicos, se recomienda utilizarlo para estimar el consumo energético a partir de variables como número de personas, cantidad de electrodomésticos y temperatura. Su simplicidad facilita la interpretación, la planificación y la toma de decisiones efectivas.
- **Promover el uso responsable y consciente de la energía.** En lugar de centrarse solo en los aparatos eléctricos, la empresa puede desarrollar programas educativos comunitarios que fomenten hábitos sostenibles, como apagar los dispositivos en desuso, aprovechar la luz natural o realizar mantenimientos preventivos. Esto no solo reduce el consumo, sino que también fortalece la conciencia ambiental entre los usuarios.
- **Considerar el tamaño del hogar y hábitos de uso.** Aunque en menor medida, el número de personas también influye en el consumo. La empresa podría segmentar sus usuarios por tamaño de hogar y ofrecer planes tarifarios personalizados o recomendaciones de ahorro ajustadas a sus patrones de consumo y nivel de demanda.
- **Incentivar el uso de energías renovables en los hogares.** Aprovechando los resultados del análisis, la empresa puede promover la instalación de paneles solares residenciales o sistemas híbridos de energía limpia, ofreciendo descuentos o alianzas con proveedores. Esto contribuye directamente a la reducción de la huella de carbono y al bienestar ambiental de la comunidad.
- **Desarrollar modelos predictivos simples y escalables.** Dado que aumentar la complejidad del modelo no mejora

significativamente el rendimiento, la empresa debería priorizar modelos interpretables, robustos y sostenibles, que puedan integrarse en plataformas de monitoreo energético. Esto permitirá anticipar picos de consumo y planificar mejor la distribución de energía sin desperdicio.

XI. CONCLUSIONES

El presente estudio permitió estimar el consumo energético residencial a partir de variables clave como la temperatura, el número de personas y la cantidad de electrodomésticos en el hogar, empleando modelos de regresión lineal y polinómica. A partir del análisis y los resultados obtenidos, se concluye lo siguiente:

- La **limpieza de datos y la eliminación de outliers** mediante los métodos Z-Score e IQR mejoraron significativamente la calidad del conjunto de datos, permitiendo obtener modelos más estables y representativos de la realidad observada.
- Entre las variables analizadas, la **cantidad de electrodomésticos** mostró la mayor correlación con el consumo eléctrico, evidenciando su papel determinante en la demanda energética de los hogares. Las variables de temperatura y número de personas también influyen, aunque en menor medida.
- Los resultados del modelado mostraron que los tres enfoques, regresión lineal, polinómica de grado 2 y polinómica de grado 3 ofrecen desempeños muy similares, con un $R^2 \approx 0.818$. Sin embargo, el **modelo lineal** se destaca como la opción más adecuada al equilibrar precisión, simplicidad e interpretabilidad.
- No se evidenció **sobreajuste** en los modelos polinómicos; no obstante, su mayor complejidad no aporta mejoras sustanciales, por lo que se recomienda priorizar modelos más simples y generalizables.
- Desde una perspectiva práctica, los hallazgos respaldan el uso de modelos predictivos lineales para la gestión energética residencial, facilitando la toma de decisiones en planificación, optimización del consumo y desarrollo de estrategias sostenibles por parte de las empresas del sector.

En síntesis, este trabajo demuestra que, con una adecuada preparación de los datos y el uso de técnicas estadísticas apropiadas, es posible estimar el consumo energético con alta precisión y ofrecer información valiosa para promover un uso más eficiente y responsable de la energía en los hogares.

REFERENCES

- [1] **Código fuente en:** <https://github.com/EdiazMejia/Parcial-Programacion-Avanzada>
- [2] H. S. Hippert, C. E. Pedreira, and R. C. Souza, "Neural networks for short-term load forecasting: A review and evaluation," *IEEE Transactions on Power Systems*, vol. 16, no. 1, pp. 44–55, 2001.
- [3] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statistics and Computing*, vol. 14, no. 3, pp. 199–222, 2004.