**Análisis Predictivo del Consumo Eléctrico: Comparación de Modelos Lineales y Polinómicos  
Autor:** Jorge Patiño **Fecha:** 20 de octubre de 2025

*(Nota: Los valores numéricos exactos en este informe se basan en los resultados de las celdas anteriores (.ipynb). Estos pueden variar ligeramente cada vez que se ejecuta la Celda 2 debido a la semilla aleatoria, pero las conclusiones generales serán consistentes).*

1. **Introducción**

El objetivo de este estudio es predecir el consumo eléctrico en hogares (**Consumo\_kWh**) basándose en tres variables predictoras: temperatura exterior (**°C**), número de personas en el hogar y la cantidad de electrodomésticos en uso. Para lograr esto, se aplicó un proceso de limpieza de datos y se comparó el rendimiento de un modelo de **Regresión Lineal Múltiple (MLR)** con modelos de **Regresión Polinómica (Grado 2 y 3)**.

1. **Metodología**
   1. **Generación y Carga de Datos**

El conjunto de datos se generó localmente utilizando el script **generate\_csv.py**, creando el archivo **consumo\_hogar.csv**. Este dataset simulado contenía entre 60,000 y 65,000 registros, e incluía valores faltantes (NaN) y valores atípicos (outliers) para simular un escenario de datos real.

* 1. **Preprocesamiento y Limpieza de Datos**

El análisis inicial (**.info()**, **.describe()**)confirmó la presencia de nulos y la distorsión estadística causada por outliers. El proceso de limpieza fue el siguiente:

1. **Tratamiento de Nulos:** Se optó por la **imputación por la mediana** para rellenar todos los valores faltantes (NaN). Esta estrategia se eligió sobre la media (promedio) por ser más robusta frente a los valores atípicos.
2. **Eliminación de Outliers:** Siguiendo la solicitud del ejercicio, se **eliminaron los outliers** de la variable objetivo (**Consumo\_kWh**). Se utilizó el método **Z-Score**, eliminando todas las filas donde el consumo se desviaba más de 3 desviaciones estándar de la media. Este proceso filtró eficazmente los valores extremos, como se validó visualmente con gráficos **boxplot**.
   1. **Análisis Exploratorio y Modelado**

Los gráficos de dispersión (**scatterplot**) entre las variables independientes y el consumo mostraron relaciones predominantemente lineales. Los datos limpios se dividieron en conjuntos de entrenamiento (**80%**) y prueba (**20%**). Se aplicó **StandardScaler** a las variables predictoras (**X**) para estandarizar sus escalas.

Se entrenaron tres modelos:

1. **Regresión Lineal Múltiple (MLR)**
2. **Regresión Polinómica (Grado 2)**
3. **Regresión Polinómica (Grado 3)**
4. **Resultados y Evaluación**

Los modelos se evaluaron en el conjunto de prueba, comparando su Error Cuadrático Medio (MSE) y su Coeficiente de Determinación ().

**Tabla 1: Comparación de Métricas de Rendimiento**

|  |  |  |
| --- | --- | --- |
| **Modelo** | **MSE** |  |
| Regresión Lineal Múltiple | 18.0358 | 0.8257 |
| Regresión Polinómica (Grado 2) | 18.0345 | 0.8257 |
| Regresión Polinómica (Grado 3) | 18.0324 | 0.8257 |

*(La semilla aleatoria afectará los decimales, pero no la conclusión)*

Como se observa en la tabla, el rendimiento de los tres modelos es prácticamente idéntico. El Coeficiente de Determinación () es de **0.8257** para los tres, lo que indica que todos los modelos explican aproximadamente el **82.6%** de la variabilidad en el consumo eléctrico.

El MSE muestra una reducción *marginal* (en el tercer y cuarto decimal) a medida que aumenta el grado polinómico. Esta mejora es insignificante y no se refleja en un aumento del

1. **Conclusiones**

**¿Qué modelo se ajusta mejor y por qué?**

El modelo que mejor se ajusta, bajo el principio de parsimonia (preferir el modelo más simple), es la Regresión Lineal Múltiple (MLR).

Aunque los modelos polinómicos (Grado 2 y 3) tienen un MSE *ligeramente* inferior, la ganancia en rendimiento es despreciable. La complejidad añadida (de 3 variables a 9 o 19 características polinómicas) no se justifica. Los datos fueron generados con una fórmula fundamentalmente lineal, por lo que el modelo MLR captura la relación casi a la perfección.

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

Analizando los coeficientes del modelo MLR (que fueron calculados sobre datos escalados, permitiendo una comparación directa de su impacto):

* **Intercepto (Media del Consumo):** 49.56
* **Personas:** 4.18
* **Electrodomesticos:** 7.28
* **Temperatura:** 3.98

El factor con el **mayor impacto** en el consumo eléctrico es ahora la **Cantidad de Electrodomésticos** (coeficiente de 7.28). En segundo lugar, y muy cerca en impacto, se encuentra el **Número de Personas** (4.18). La **Temperatura Exterior** (3.98) tiene el impacto más bajo de las tres variables.

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

**Sí, se presenta sobreajuste (overfitting) de manera leve.** El modelo de Grado 3 es el más complejo (con el mayor número de características), pero no logra un mejor rendimiento () que el modelo lineal simple en el conjunto de prueba. Esto sugiere que el modelo de Grado 3 está empezando a ajustarse al ruido específico del conjunto de entrenamiento en lugar de generalizar mejor la tendencia de los datos.

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

1. **Enfocar la Eficiencia en Aparatos:** Dado que la **Cantidad de Electrodomésticos** es el predictor principal, la empresa debe priorizar programas de **"Plan Renove"** y subsidios para la adquisición de equipos de alta eficiencia energética, buscando reducir el consumo unitario de cada aparato.
2. **Incentivos por Ocupación:** El **Número de Personas** sigue siendo un factor crucial. La empresa debe diseñar programas y tarifas especiales dirigidos a hogares con alta ocupación para fomentar el ahorro en el consumo per cápita.
3. **Gestión Inteligente de la Demanda:** La **Temperatura** sigue siendo un predictor significativo. La empresa puede promover el uso de tecnologías de domótica (hogares inteligentes) y termostatos programables para optimizar el uso de calefacción y refrigeración, ayudando a estabilizar la demanda en la red.