

Predicción de Consumo Energético en Ciudades Inteligentes

Proyecto de Procesamiento de Grandes Volúmenes de Datos

Autores:

[Amalia Beatriz Valiente Hinojosa]

[Noel Pérez Calvo]

Fecha: 17 de octubre de 2025

1. ¿Qué se pretende lograr?

El objetivo central del proyecto es **predecir el consumo energético en distintas zonas de una ciudad inteligente**, utilizando técnicas de procesamiento de grandes volúmenes de datos mediante las plataformas *Hadoop* y *Apache Spark*. A través del análisis de datos históricos de consumo eléctrico, se busca anticipar la demanda en las próximas horas, detectar picos de consumo y evaluar el impacto de las condiciones climáticas sobre el uso energético urbano.

2. Dataset seleccionado

Nombre: *Household Electric Power Consumption Dataset*

Fuente: Kaggle / UCI Machine Learning Repository

Formato: CSV (valores separados por punto y coma “;”)

URL: <https://www.kaggle.com/datasets/uciml/electric-power-consumption-data-set>

3. Justificación del dataset

Volumen

El conjunto de datos contiene más de **dos millones de registros**, correspondientes a mediciones minuto a minuto del consumo eléctrico de un hogar durante casi **cuatro años** (2006–2010). Su tamaño aproximado de 120 MB en formato CSV, y el incremento que supone su almacenamiento distribuido en *HDFS*, permiten simular un entorno realista de **procesamiento de grandes volúmenes de datos**, adecuado para la aplicación de tecnologías como *Hadoop* y *Spark*.

Características

El dataset incluye variables como:

- Fecha y hora de cada registro.
- Potencia activa y reactiva global.
- Voltaje y corriente.
- Energía submedida en tres zonas del hogar (*Sub-metering-1*, *2* y *3*).

Estas variables conforman una **serie temporal multivariable**, adecuada para tareas de predicción y análisis de patrones de consumo. Además, los datos presentan una variabilidad natural y cierto nivel de ruido, lo que refleja condiciones reales de consumo y permite evaluar técnicas de limpieza y modelado robustas.

Pertinencia

El conjunto de datos resulta altamente pertinente para los objetivos del proyecto, ya que:

- Contiene **registros reales de consumo energético**, directamente relacionados con la meta de predecir la demanda eléctrica.
- Su granularidad temporal (minuto a minuto) facilita el análisis de **patrones horarios, diarios y estacionales**.
- Puede combinarse con datos meteorológicos (temperatura, humedad, precipitaciones) para analizar el **impacto del clima** en la demanda, fortaleciendo el enfoque de ciudades inteligentes.

Conclusión

El dataset seleccionado constituye una base sólida para el desarrollo del proyecto, ya que ofrece volumen, calidad y relevancia suficientes para implementar un sistema de predicción energética escalable mediante herramientas de procesamiento distribuido. De esta manera, se busca contribuir al diseño de estrategias de eficiencia y sostenibilidad energética en el contexto de las ciudades inteligentes.