



**Universidad
Europea**

UNIVERSIDAD EUROPEA DE MADRID

**ESCUELA DE ARQUITECTURA, INGENIERÍA Y DISEÑO
GRADO EN MATEMÁTICAS APLICADAS AL ANÁLISIS
DE DATOS**

PROYECTO DE BIG DATA 1

Pobreza Energética en Asturias

Paula Gómez Lucas

CURSO 2024-2025

Índice

Capítulo 1. Descripción del proyecto	4
1.1 Contexto y justificación	4
1.2 Objetivos del proyecto	4
Capítulo 2. Tecnologías	5
Capítulo 3. Desarrollo del proyecto	6
3.1 Datos	6
3.2 Descripción de la solución, metodologías y herramientas utilizadas	6
Capítulo 4. Infraestructura real	8
4.1 Diagrama de infraestructura	8
4.2 Flujo de datos	8
Capítulo 5. Conclusiones y trabajo futuro	10
5.1 Resultados obtenidos	10
5.2 Conclusiones del trabajo	10
5.3 Trabajo futuro	10
Referencias bibliográficas.....	12

Capítulo 1. Descripción del proyecto

1.1 Contexto y justificación

La pobreza energética es un problema creciente que afecta entre el 10 % y el 17 % de los hogares en España, dependiendo de la región. Este fenómeno tiene implicaciones económicas y sociales significativas, especialmente en áreas rurales, zonas urbanas con infraestructuras antiguas y en el sur del país. Las poblaciones más afectadas incluyen personas mayores, hogares de bajos ingresos, familias monoparentales y comunidades marginadas.

En respuesta, se han implementado medidas como el Bono Social Eléctrico y Térmico, la Estrategia Nacional de la Pobreza Energética, y políticas de protección contra la desconexión. Sin embargo, la efectividad de estas intervenciones requiere análisis más profundos, basados en datos reales, para determinar si realmente logran su propósito.

El presente proyecto se centra en el análisis del uso de la red eléctrica como herramienta para evaluar la eficacia de estas medidas. Los datos obtenidos de fuentes como la Red Eléctrica Española, el Instituto Nacional de Estadística y Datadis permitirán identificar patrones de consumo y necesidades no cubiertas en comunidades vulnerables.

1.2 Objetivos del proyecto

El objetivo principal de este proyecto es determinar cómo el análisis de datos provenientes de la red eléctrica puede evidenciar la efectividad de las medidas contra la pobreza energética. Para ello, se busca:

- Analizar patrones de consumo eléctrico en diferentes regiones afectadas.
- Identificar correlaciones entre el acceso a los recursos energéticos y los niveles de pobreza energética.
- Evaluar el impacto de las políticas actuales mediante indicadores basados en datos reales.
- Proponer mejoras o nuevas estrategias para reducir la incidencia de la pobreza energética en las comunidades más vulnerables.

Capítulo 2. Tecnologías

El desarrollo de este proyecto se sustenta en el uso de tecnologías modernas y herramientas especializadas que facilitan el análisis, visualización y gestión de datos. A continuación, se describen las principales tecnologías empleadas:

Docker

Docker se utilizó para crear y gestionar contenedores que permitieron un entorno de desarrollo consistente y escalable. Esta herramienta facilitó la integración y despliegue de las aplicaciones necesarias para procesar y analizar los datos, asegurando portabilidad entre diferentes sistemas y minimizando problemas de configuración.

Elasticsearch

Elasticsearch, un motor de búsqueda y análisis distribuido, fue empleado para gestionar grandes volúmenes de datos relacionados con el consumo energético. Su capacidad para realizar búsquedas rápidas y consultas complejas permitió explorar patrones y correlaciones relevantes en los datos de la red eléctrica.

Python y sus bibliotecas

Python fue el lenguaje principal utilizado en el proyecto debido a su versatilidad y amplia gama de bibliotecas específicas para el análisis de datos. Entre las bibliotecas empleadas destacan:

- **Pandas:** para la manipulación y análisis estructurado de datos.
- **NumPy:** para operaciones numéricas y gestión de arrays multidimensionales.
- **Matplotlib y Seaborn:** para la visualización de datos.
- **Scikit-learn:** para tareas de modelado predictivo y análisis de patrones.

Power BI

Power BI fue empleado para la visualización avanzada de datos. Con esta herramienta se desarrollaron dashboards interactivos que facilitaron la presentación de los resultados del análisis de consumo energético. Esto permitió a los stakeholders interpretar fácilmente las tendencias y evaluar la efectividad de las medidas implementadas contra la pobreza energética.

Estas tecnologías, combinadas, brindaron un enfoque robusto y escalable para abordar el problema, permitiendo un análisis eficiente y preciso de la información disponible.

Capítulo 3. Desarrollo del proyecto

3.1 Datos

Para abordar el problema de la pobreza energética, se recopilaron y analizaron datos de diversas fuentes confiables, que permitieron una visión integral del consumo energético en España. Las principales fuentes utilizadas fueron:

- **Gobierno Nacional de España:** Datos estadísticos y normativos sobre consumo y pobreza energéticos.
- **Gobierno Regional de Asturias:** Información específica de una de las regiones con alta incidencia de pobreza energética.
- **Instituto Nacional de Estadística (INE):** Datos socioeconómicos relevantes que permitieron establecer correlaciones con el consumo energético.
- **API de Redeia (Red Eléctrica Española):** Información en tiempo real sobre el uso de la red eléctrica, patrones de consumo y disponibilidad de recursos.
- **Datadis:** Datos energéticos proporcionados por empresas del sector, como Iberdrola, Naturgy y Endesa, utilizados para comprender el consumo residencial.

Los datos fueron limpiados, transformados y analizados para identificar patrones significativos, asegurando su integridad y calidad mediante técnicas avanzadas de preprocesamiento.

3.2 Descripción de la solución, metodologías y herramientas utilizadas

El desarrollo del proyecto siguió una metodología **Scrum**, un enfoque ágil que permitió organizar el trabajo en iteraciones cortas, priorizando entregables funcionales en cada sprint. Este método facilitó la integración de los datos y herramientas mientras se adaptaban las decisiones a las necesidades emergentes del análisis.

La solución propuesta se centró en la integración y análisis de los datos a través de las siguientes herramientas:

- **Docker:** Para garantizar entornos de desarrollo consistentes y desplegar aplicaciones de manera eficiente.
- **Elasticsearch:** Para gestionar y consultar grandes volúmenes de datos de consumo energético.
- **Python y sus bibliotecas:** El análisis y modelado de datos se realizó utilizando bibliotecas como Pandas, NumPy, y Scikit-learn, mientras que Matplotlib y Seaborn se emplearon para generar visualizaciones.
- **Power BI:** La solución incluye dashboards interactivos que presentan las tendencias y hallazgos clave del análisis de consumo energético.
- **Visual Studio:** Utilizado como entorno de desarrollo integrado (IDE) principal para gestionar el código del proyecto y los scripts de análisis.

El proceso incluyó las siguientes etapas clave:

1. **Recolección de datos:** Obtención de datos de las fuentes mencionadas, incluyendo la API de Redeia y Datadis.
2. **Preprocesamiento:** Limpieza, transformación y normalización de los datos para garantizar su calidad y comparabilidad.
3. **Análisis exploratorio:** Identificación de patrones de consumo energético mediante estadísticas descriptivas y visualizaciones.
4. **Modelado y validación:** Uso de técnicas analíticas y modelos predictivos para evaluar el impacto de las medidas contra la pobreza energética.
5. **Visualización:** Creación de dashboards en Power BI para comunicar los resultados a los stakeholders de manera efectiva.

Este enfoque permitió estructurar un análisis riguroso, proporcionando una solución robusta para evaluar la efectividad de las medidas existentes y proponer nuevas estrategias basadas en los datos obtenidos.

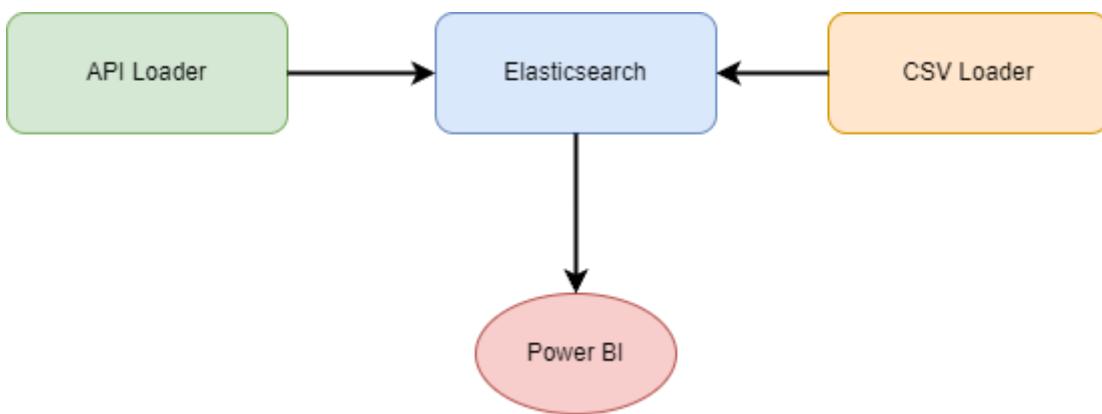
Capítulo 4. Infraestructura real

4.1 Diagrama de infraestructura

La infraestructura del proyecto se basa en una arquitectura modular y escalable creada mediante Docker y Docker Compose, lo que permite aislar y coordinar los diferentes componentes necesarios para el análisis de datos. Se implementaron tres contenedores principales:

1. **Contenedor de Elasticsearch:** Responsable de almacenar y gestionar los datos energéticos recuperados.
2. **Contenedor de API Loader:** Encargado de acceder a las APIs de fuentes como Redeia y Datadis para obtener datos relevantes.
3. **Contenedor de CSV Loader:** Diseñado para procesar archivos CSV provenientes de otras fuentes de datos, como el INE y fuentes regionales.

La comunicación entre los contenedores está gestionada mediante Docker Compose, garantizando que los datos fluyan adecuadamente entre las diferentes partes del sistema. Visual Studio se utilizó para desarrollar los scripts y archivos de configuración, lo que permitió un entorno controlado y eficiente.

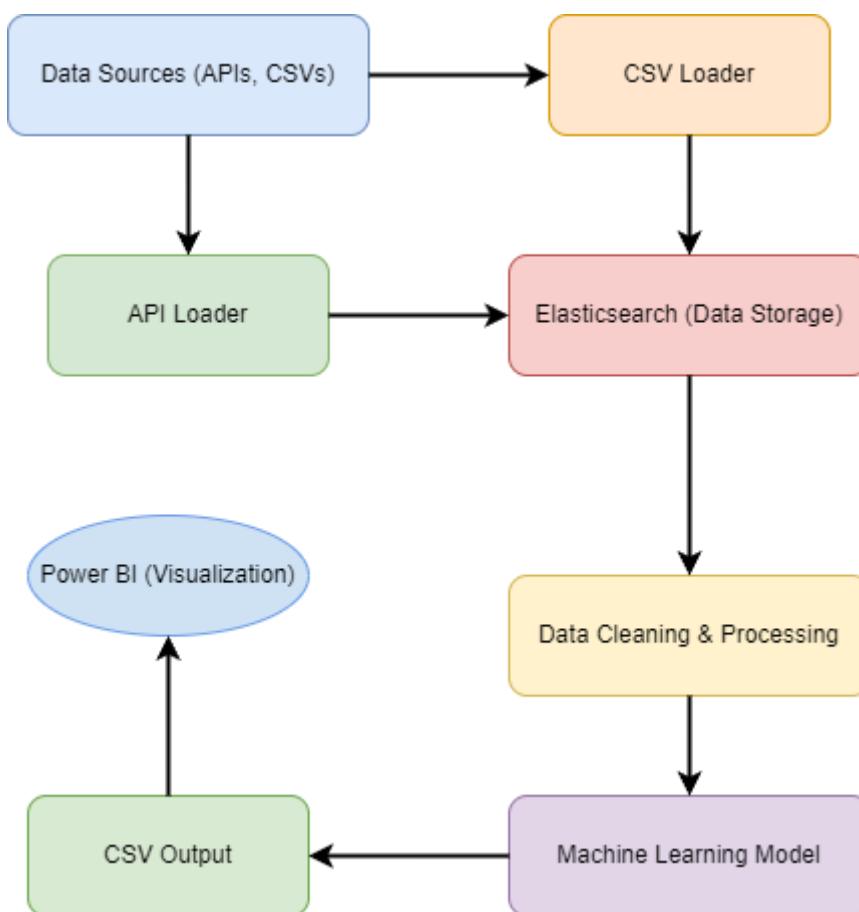


4.2 Flujo de datos

El flujo de datos sigue un proceso claro y estructurado:

1. **Obtención de datos:**
 - El contenedor **API Loader** se conecta a las APIs (Redeia, Datadis) y descarga datos directamente.
 - El contenedor **CSV Loader** procesa archivos CSV de fuentes externas (Gobierno de España, INE).

2. **Almacenamiento inicial:** Los datos obtenidos se almacenan en el contenedor de Elasticsearch para facilitar consultas rápidas y manejo de grandes volúmenes de información.
3. **Procesamiento y limpieza:**
 - A través del script `elasticsearch_data_loader.py`, se extraen los datos de Elasticsearch.
 - En esta etapa, los datos son limpiados y normalizados según sea necesario.
4. **Modelado y análisis:** Los datos limpios se utilizan para entrenar modelos de inteligencia artificial que identifican patrones y generan indicadores clave sobre pobreza energética.
5. **Exportación y visualización:** Finalmente, los datos procesados y resultados se exportan como archivos CSV. Estos son cargados en Power BI, donde se crean dashboards y visualizaciones interactivas para interpretar y comunicar los resultados.



Capítulo 5. Conclusiones y trabajo futuro

5.1 Resultados obtenidos

El desarrollo del proyecto permitió obtener un análisis inicial sobre la efectividad de las medidas implementadas para combatir la pobreza energética, tomando como referencia los datos de consumo eléctrico en la región de Asturias. Los dashboards interactivos y los modelos predictivos diseñados permitieron identificar patrones relevantes en el consumo energético y correlaciones con factores socioeconómicos.

Aunque estos resultados constituyen un avance significativo, deben considerarse como pruebas de concepto debido a las limitaciones actuales en la cobertura de datos y en la madurez de los modelos analíticos.

5.2 Conclusiones del trabajo

El análisis realizado evidencia que el uso de datos energéticos y socioeconómicos permite obtener información valiosa sobre la incidencia de la pobreza energética y la efectividad de las políticas públicas. Sin embargo, las conclusiones derivadas de este trabajo están limitadas por la restricción geográfica de los datos (centrados exclusivamente en Asturias) y por la necesidad de refinar los modelos analíticos empleados.

El proyecto pone de manifiesto el potencial del análisis de datos para mejorar la toma de decisiones en el ámbito de la pobreza energética, subrayando la importancia de contar con un enfoque escalable y con datos representativos a nivel nacional.

5.3 Trabajo futuro

El trabajo futuro debería centrarse en las siguientes áreas clave:

- Ampliación de la cobertura geográfica:** Incorporar datos de consumo energético y socioeconómico de otras comunidades autónomas y, eventualmente, de toda España para obtener una visión más completa y representativa.
- Mejoras en los modelos analíticos:** Refinar los modelos de análisis predictivo y los dashboards desarrollados, incorporando técnicas avanzadas de machine learning y herramientas de visualización más robustas para garantizar mayor precisión y usabilidad.
- Integración de nuevas fuentes de datos:** Explorar la posibilidad de incluir datos adicionales, como indicadores climáticos, precios energéticos y características demográficas específicas, para enriquecer el análisis y aumentar su valor predictivo.

-
4. **Validación de las políticas existentes:** Evaluar de manera más exhaustiva la efectividad de las medidas actuales (como el Bono Social Eléctrico) en otras regiones, a fin de proponer recomendaciones informadas para optimizar su impacto.

Estos pasos serán esenciales para transformar las pruebas de concepto desarrolladas en herramientas completas y escalables que contribuyan de manera significativa a la lucha contra la pobreza energética en España.

Referencias bibliográficas

1. Estrategia Nacional Contra la Pobreza Energética 2019-2024. (2019). Ministerio para la Transición Ecológica y el Reto Demográfico, Gobierno de España.
2. Actualización de Indicadores de la Estrategia Nacional Contra la Pobreza Energética. (2020). Ministerio para la Transición Ecológica y el Reto Demográfico.
3. Actualización de Indicadores de la Estrategia Nacional Contra la Pobreza Energética (2019-2024). (2022). Ministerio para la Transición Ecológica y el Reto Demográfico.
4. Instituto Nacional de Estadística (INE). Encuesta de Presupuestos Familiares (EPF) y Encuesta de Condiciones de Vida (ECV). Datos sobre el gasto energético en España.
5. Red Eléctrica de España (Redeia). Datos energéticos en tiempo real.
6. Datadis. Plataforma de datos energéticos.
7. Datos.gob.es. Longitud de los límites geográficos de Asturias. Catálogo de datos abiertos de España.