



**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

ECOPENGUIN

Luis de la Rubia Postigo

Angel Arroyo Gomez

Jhosua Callejas Ramos

Juan de Dios Encinas Duran

CURSO 2024-2025

Índice

Introducción	4
Objetivo principal	4
Características destacadas	4
➤ Análisis histórico del consumo energético	4
➤ Predicciones energéticas	4
➤ Segmentación personalizada mediante clustering	4
➤ Recomendaciones personalizadas	5
➤ Promoción de la sostenibilidad	5
Beneficios para los usuarios	5
Diferenciadores clave	5
Arquitectura de la solución	6
Gestión de datos con Elasticsearch	6
Consultas avanzadas	7
• Filtros por provincia y rango de fechas	7
• Recuperación masiva de documentos	7
• Búsquedas personalizadas	7
Optimización de Elasticsearch	8
• Eliminación de índices antiguos	8
• Configuración de mapeos personalizados	8
• Ajustes en shards y replicación	8
Despliegue y orquestación con Docker	8
Componentes contenerizados:	8
Docker Compose	9
Ventajas del uso de Docker	9
Flujo de trabajo del proyecto	9
Fase 1: Conceptualización	9
Fase 2: Procesamiento de datos	10
- Extracción de datos:	10
- Transformación de datos:	10
Fase 3: Desarrollo del backend	11
Fase 4: Desarrollo del frontend	11
Fase 5: Despliegue con Docker	11
Fase 6: Optimización y pruebas	12
Futuras mejoras	12
Conclusión	12

Introducción

EcoPenguin es una plataforma digital que busca cambiar por completo la forma en que gestionamos el consumo energético en hogares y organizaciones. Su meta principal es ayudar a los usuarios a reducir costos, adoptar prácticas sostenibles y tomar decisiones mejor informadas, todo con el apoyo de tecnologías avanzadas. Combina herramientas como análisis de datos, predicciones basadas en inteligencia artificial, segmentación personalizada y visualizaciones interactivas dentro de un entorno intuitivo y accesible para todos. Desde que se concibió, el proyecto ha seguido un enfoque iterativo que pone énfasis en la escalabilidad, el rendimiento y, sobre todo, en ofrecer una excelente experiencia para el usuario. En este informe se detallan los avances del proyecto, las mejoras realizadas, la arquitectura técnica y las principales funcionalidades que lo hacen destacar.

Objetivo principal

La misión de EcoPenguin es dar a los usuarios las herramientas y la información necesarias para que puedan reducir su consumo energético y los costos que este implica. También busca fomentar prácticas sostenibles, como incorporar tecnologías renovables y más eficientes, y ayudar a que las personas tomen decisiones mejor informadas mediante recomendaciones personalizadas basadas en sus hábitos energéticos. No se trata solo de ahorrar dinero, sino de impulsar un cambio hacia un modelo de consumo energético más sostenible y consciente, algo esencial en el contexto de los desafíos globales como la eficiencia energética y la reducción de la huella de carbono.

Características destacadas

EcoPenguin es un sistema completo diseñado para cubrir diferentes necesidades energéticas:

➤ **Análisis histórico del consumo energético**

Permite a los usuarios explorar cómo ha sido su consumo energético en el pasado y compararlo con su comportamiento actual. Además, ofrece estadísticas detalladas como la media, la moda y la varianza, lo que facilita la identificación de patrones o áreas que necesitan optimización.

➤ **Predicciones energéticas**

Gracias a modelos avanzados como Prophet, la plataforma es capaz de predecir el consumo energético de los próximos 12 meses. Esto les permite a los usuarios anticiparse a posibles aumentos de consumo y ajustar sus hábitos para evitar gastos innecesarios.

➤ Segmentación personalizada mediante clustering

Con un modelo K-Means, EcoPenguin clasifica a los usuarios en grupos según características como su consumo energético, el número de residentes en el hogar y la potencia contratada. A partir de esto, se generan recomendaciones específicas para cada perfil. Por ejemplo:

- Las familias de consumo medio-alto (Cluster 0) reciben sugerencias para optimizar aún más su energía.
- Los hogares con consumo alto y pocos residentes (Cluster 1) obtienen consejos para reducir el uso intensivo.
- Las familias grandes con buen manejo energético (Cluster 2) son motivadas a adoptar energías renovables.
- Las viviendas eficientes con consumo bajo (Cluster 3) ven resaltada su eficiencia, incentivando a replicar sus prácticas.

➤ Recomendaciones personalizadas

La plataforma genera alertas e ideas concretas para que los usuarios mejoren su eficiencia energética. Esto incluye propuestas como cambiar tarifas según el patrón de uso, instalar tecnologías sostenibles o ajustar la potencia contratada para ahorrar en costos.

➤ Promoción de la sostenibilidad

EcoPenguin va más allá de los ahorros económicos. Busca también fomentar prácticas responsables, ofreciendo información sobre subvenciones para la adopción de tecnologías renovables. En el futuro, planea incluir herramientas para realizar análisis de retorno de inversión (ROI) en estas tecnologías.

Beneficios para los usuarios

EcoPenguin está pensado para personas y organizaciones de cualquier nivel técnico, ofreciendo una experiencia que combina accesibilidad con funcionalidad. Para familias y hogares, ayuda a entender cómo sus hábitos afectan el consumo energético y les ofrece predicciones junto con recomendaciones prácticas para ahorrar tanto energía como dinero. Por otro lado, las empresas y organizaciones encuentran en EcoPenguin una herramienta poderosa para realizar análisis detallados que les permiten implementar políticas energéticas sostenibles y evaluar el impacto económico de adoptar tecnologías renovables.

Diferenciadores clave

EcoPenguin sobresale por su enfoque integral y personalizado. Desde el punto de vista tecnológico, utiliza herramientas modernas como Elasticsearch, Docker, React y FastAPI para garantizar que el sistema sea escalable y de alto rendimiento. Además, su diseño modular facilita la incorporación de nuevas funcionalidades, como análisis de ROI o acceso a bases de datos de subvenciones. Pero lo que realmente lo hace único es su énfasis en la sostenibilidad: más allá del ahorro económico, busca promover prácticas responsables y sostenibles. Todo esto se complementa con una interfaz sencilla y potente que asegura que cualquier usuario, sin importar su nivel técnico, pueda aprovechar al máximo la plataforma.

Arquitectura de la solución

Estructura general EcoPenguin se basa en una arquitectura moderna y modular, diseñada para garantizar que el sistema sea escalable, fácil de mantener y completamente portable.

En el **frontend**, desarrollado con React.js, se prioriza una interfaz amigable y sencilla que permite a los usuarios realizar predicciones, analizar su cluster energético y consultar estadísticas o datos históricos. Es el punto de interacción directo con la plataforma y está diseñado para que cualquiera pueda usarlo sin problemas.

El **backend**, construido con FastAPI, actúa como intermediario entre el frontend y Elasticsearch. Este gestiona todas las solicitudes relacionadas con datos históricos y predicciones, además de encargarse del procesamiento de los modelos de clustering y las recomendaciones personalizadas. Entre sus principales características están la seguridad, garantizada mediante autenticación básica y manejo de CORS, y su modularidad, que permite agregar nuevas funcionalidades sin comprometer el rendimiento del sistema.

En cuanto a **Elasticsearch**, este es el núcleo de almacenamiento y análisis de datos. Está diseñado para manejar grandes volúmenes de información y realizar búsquedas complejas de manera rápida. Gracias a sus índices personalizados, permite búsquedas avanzadas por rangos de fechas, provincias y patrones de consumo, proporcionando respuestas en tiempo real que aseguran una experiencia fluida para el usuario.

Finalmente, **Docker** juega un papel clave al facilitar el despliegue y ejecución de EcoPenguin en cualquier entorno. Cada componente del sistema (frontend, backend y Elasticsearch) se encapsula en un contenedor independiente, y Docker Compose se encarga de orquestar su interacción para que todo funcione de forma coordinada. Esto no solo asegura portabilidad, sino también escalabilidad y consistencia, ya que el sistema puede ejecutarse de manera uniforme en cualquier entorno compatible con Docker.

Gestión de datos con Elasticsearch

Elasticsearch es el motor central que da vida a EcoPenguin, encargado de manejar grandes volúmenes de datos y realizar consultas avanzadas en tiempo real. Esta herramienta es esencial para gestionar, analizar y buscar información dentro del proyecto.

Índices clave, EcoPenguin utiliza tres índices principales que funcionan como bases de datos especializadas para cubrir diferentes necesidades:

1. **Consumo_energetico_viviendas_historico_completo**, Este índice almacena todos los datos históricos sobre el consumo energético procesado. Incluye detalles como la fecha y hora del consumo, ubicación geográfica (provincia), consumo energético en kWh/m² y potencia contratada. Gracias a este índice, los usuarios pueden consultar un historial completo que les permite analizar cómo han evolucionado sus patrones de consumo a lo largo del tiempo.
2. **Predicciones**, aquí se guardan los resultados generados por el modelo Prophet, encargado de predecir el consumo energético futuro. Cada documento contiene la fecha de la predicción, el valor estimado (yhat) y los intervalos de confianza (yhat_lower y yhat_upper). Este índice resulta clave para que los usuarios puedan planificar su consumo energético y anticiparse a posibles variaciones.
3. **Clusters**, este índice contiene las clasificaciones generadas por el modelo K-Means. En él se almacenan datos como el cluster asignado a cada usuario, las variables clave que determinaron su clasificación (consumo promedio, número de residentes, potencia contratada) y las recomendaciones específicas para cada grupo. Esto permite personalizar al máximo la experiencia del usuario, ofreciéndole alertas y sugerencias basadas en su perfil energético.

Consultas avanzadas

Una de las grandes fortalezas de Elasticsearch es su capacidad para realizar búsquedas rápidas y complejas, lo que en EcoPenguin se traduce en funcionalidades clave:

- **Filtros por provincia y rango de fechas**

Los usuarios pueden consultar su consumo energético en un período específico o incluso comparar cómo varía el comportamiento entre diferentes provincias. Las consultas por fecha están diseñadas con el tipo range, lo que permite definir de forma precisa los rangos de inicio y fin para obtener datos específicos.

- **Recuperación masiva de documentos**

Al aprovechar el parámetro size en las consultas, el sistema es capaz de recuperar grandes volúmenes de datos para procesarlos y analizarlos en tiempo real, sin afectar el rendimiento.

- **Búsquedas personalizadas**

Mediante queries como match, los usuarios pueden localizar consumos o perfiles específicos que cumplan ciertos criterios. Esto hace que la segmentación y la personalización de resultados sean mucho más precisas y relevantes.

Optimización de Elasticsearch

Para garantizar que Elasticsearch funcione de forma eficiente, se implementaron varias mejoras durante el desarrollo del sistema:

- **Eliminación de índices antiguos**

Antes de cargar nuevos datos, se eliminan automáticamente los índices obsoletos para evitar duplicados y conflictos. Esto asegura que la información almacenada sea siempre consistente y esté actualizada.

- **Configuración de mapeos personalizados**

Se definieron estructuras específicas para cada índice, asignando tipos de datos como date, float y keyword. Esto no solo optimiza las búsquedas, sino que también reduce considerablemente los tiempos de respuesta.

- **Ajustes en shards y replicación**

Se ajustó la distribución de shards (fragmentos de datos) y su replicación para equilibrar la carga del sistema, mejorando tanto su rendimiento como su resiliencia ante posibles fallos.

Despliegue y orquestación con Docker

El uso de Docker fue una decisión estratégica que garantiza la portabilidad, escalabilidad y facilidad de mantenimiento de EcoPenguin. Cada componente del sistema está empaquetado en un contenedor independiente, lo que asegura que funcione de forma aislada pero perfectamente integrada.

Componentes contenerizados:

- Frontend: El contenedor de React incluye todas las dependencias necesarias para ejecutar la interfaz gráfica, garantizando que la experiencia del usuario sea consistente, sin importar el entorno.
- Backend: El backend de FastAPI está empaquetado con todas sus dependencias (como Elasticsearch Python SDK, Prophet y scikit-learn), lo que asegura que los procesos de predicción y clustering sean reproducibles en cualquier máquina.
- Elasticsearch: Este componente se implementa en su propio contenedor, configurado para manejar grandes volúmenes de datos y múltiples consultas simultáneas. Además aseguramos que los datos no se pierdan incluso si el contenedor se reinicia.

Docker Compose

Docker Compose es una herramienta clave que orquesta cómo interactúan los diferentes contenedores. Dentro de EcoPenguin, su uso simplifica tanto la configuración como la ejecución del sistema:

- Automatización: Con un único archivo (docker-compose.yml), se definen todos los servicios (frontend, backend y Elasticsearch). Un solo comando (docker-compose up) es suficiente para poner en marcha todo el sistema.
- Redes internas: Compose crea redes privadas entre los contenedores, permitiendo que se comuniquen directamente mediante nombres como "backend" o "elasticsearch", sin necesidad de usar direcciones IP complicadas.
- Escalabilidad: En momentos de alta demanda, Docker Compose permite aumentar el número de instancias de cualquier servicio, asegurando que el sistema mantenga un rendimiento óptimo incluso con un incremento en el tráfico o los datos.

Ventajas del uso de Docker

Docker ofrece múltiples beneficios que hacen de EcoPenguin un sistema robusto y eficiente:

- Portabilidad: Todo el sistema puede ejecutarse en cualquier máquina compatible con Docker, sin importar el sistema operativo.
- Consistencia: Al utilizar contenedores, se garantiza que las versiones de las dependencias sean siempre las mismas en cualquier entorno, evitando problemas de compatibilidad.
- Mantenimiento sencillo: Actualizar cualquier componente, como el modelo Prophet en el backend, es tan simple como reconstruir su imagen correspondiente.
- Eficiencia operativa: Docker optimiza los recursos, asignando a cada contenedor solo lo que necesita para ejecutarse de manera efectiva.

Flujo de trabajo del proyecto

El desarrollo de EcoPenguin ha seguido un enfoque estructurado y con iteraciones continuas. Esto ha permitido construir cada fase sobre la anterior, asegurando que el resultado final sea un producto sólido y de alta calidad. Además de cumplir los objetivos iniciales, este enfoque ayudó a identificar mejoras clave durante el proceso.

Fase 1: Conceptualización

El proyecto comenzó con una visión clara: ayudar a las familias a gestionar su consumo energético, reducir costos y fomentar prácticas sostenibles. Para lograrlo, se definieron objetivos clave como:

- Crear una herramienta fácil de usar pero con análisis avanzado de datos.
- Ofrecer predicciones confiables y personalizadas que faciliten la planificación energética.
- Diseñar un sistema que segmente a los usuarios según su perfil energético, generando recomendaciones prácticas y relevantes.

Para alcanzar estos objetivos, se optó por tecnologías modernas que garantizaran escalabilidad, modularidad y portabilidad. Estas incluyeron Elasticsearch como motor de almacenamiento y búsqueda, FastAPI para un backend eficiente, React.js para un frontend interactivo y Docker para contenerizar y orquestar todos los componentes.

Fase 2: Procesamiento de datos

- Extracción de datos:

El equipo desarrolló el script `extractData1.py`, encargado de automatizar la descarga de información desde APIs externas como Red Eléctrica de España (REE) y AEMET. Este script facilita la recopilación de datos históricos de consumo energético, balances y precios de mercado, almacenándolos en formato CSV con una estructura uniforme lista para ser procesada.

- Transformación de datos:

Una vez extraídos, los datos pasan por una limpieza y transformación utilizando el script `edaData1.py`.

- Limpieza: Se eliminan duplicados, se gestionan valores nulos y se corrigen formatos inconsistentes.
- Estandarización: Se ajustan columnas y formatos de fecha para garantizar compatibilidad con Elasticsearch.

- Validación: Se verifican los datos para asegurar su integridad y evitar errores en pasos posteriores.

Esta fase es crucial, ya que la calidad de los datos afecta directamente la precisión de los modelos predictivos y las recomendaciones generadas.

Fase 3: Desarrollo del backend

El backend, desarrollado con FastAPI, gestiona las solicitudes de los usuarios, interactúa con Elasticsearch y procesa modelos de aprendizaje automático. Entre sus funciones principales destacan:

- Predicciones energéticas: Usando el modelo Prophet, genera estimaciones precisas basadas en datos históricos. Estas predicciones incluyen valores esperados y rangos de confianza, proporcionando información detallada y útil.
- Análisis de clustering: El backend emplea un modelo preentrenado de K-Means para clasificar a los usuarios en clusters según su perfil energético. A cada grupo se le asignan recomendaciones específicas, como ajustes en tarifas o la adopción de tecnologías sostenibles.
- Estadísticas generales: Calcula métricas clave (media, moda, varianza) para ayudar a los usuarios a entender sus patrones de consumo y detectar áreas de mejora.
- Seguridad y escalabilidad: Se implementaron autenticación básica y manejo de CORS para garantizar que el backend sea seguro y accesible desde el frontend.

Fase 4: Desarrollo del frontend

El frontend, desarrollado en React.js, ofrece una interfaz intuitiva y fácil de usar, con varios componentes clave:

- Home.js: La página principal que conecta a los usuarios con las funcionalidades del sistema, como predicciones, clustering y estadísticas.
- Predict.js: Permite ingresar la provincia y un rango de fechas para generar predicciones energéticas detalladas.
- ClusterAnalysis.js: Facilita la entrada de datos como consumo promedio, número de residentes y potencia contratada para obtener resultados de clustering y recomendaciones personalizadas.
- Statistics.js: Muestra estadísticas clave del consumo energético, ayudando a los usuarios a entender su historial y tomar decisiones informadas.

Fase 5: Despliegue con Docker

Docker es la base del despliegue de EcoPenguin. Cada componente del sistema se conteneriza de forma independiente, lo que garantiza compatibilidad y facilidad de mantenimiento.

- Frontend: Ejecutado en su propio contenedor para asegurar una experiencia de usuario consistente.
- Backend: Contenerizado con todas sus dependencias, lo que simplifica el mantenimiento.
- Elasticsearch: Desplegado como un servicio independiente, con almacenamiento persistente para proteger los datos críticos.

Fase 6: Optimización y pruebas

Durante el desarrollo, se realizaron iteraciones para resolver limitaciones y optimizar el sistema:

- Optimización de Elasticsearch: Se configuraron índices personalizados y se eliminaron índices antiguos automáticamente para evitar duplicados.
- Pruebas del backend y frontend: Se implementaron pruebas unitarias e integrales utilizando herramientas como Postman y Axios para validar las rutas y las interacciones.
- Refinamiento de modelos: Los modelos Prophet y K-Means se ajustaron y validaron con datos reales para asegurar su precisión.

Futuras mejoras

El equipo de EcoPenguin planea incorporar nuevas funcionalidades, como:

1. Análisis de ROI: Herramientas para evaluar el retorno de inversión en tecnologías sostenibles.
2. Foro de subvenciones: Un módulo que conecte a los usuarios con incentivos financieros según su ubicación y necesidades.
3. Visualizaciones mejoradas: Gráficos más interactivos para explorar datos en mayor profundidad.
4. Expansión geográfica: Integración de datos de otras regiones y países para aumentar el alcance de la plataforma.

Conclusión

EcoPenguin es una solución innovadora que combina tecnología avanzada con un enfoque centrado en el usuario. Su capacidad para analizar, predecir y optimizar el consumo energético lo convierte en una herramienta indispensable para hogares y organizaciones. Con un fuerte compromiso con la sostenibilidad y la accesibilidad, EcoPenguin está en camino de consolidarse como líder en el mercado de soluciones energéticas inteligentes.