5-5-2025

Jorge Burgos Ortega / Antonio Sánchez Sánchez

UCLM - ESI

MEMORIA

BigData

INDICE

[**1. Introducción** 2](#_Toc197281825)

[**Objetivos del proyecto:** 2](#_Toc197281826)

[**Requisitos generales:** 2](#_Toc197281827)

[**2. Tecnología utilizada** 3](#_Toc197281828)

[**MongoDB:** 3](#_Toc197281829)

[**Python y librerías analíticas** 3](#_Toc197281830)

[**Random Forest para predicción de demanda** 4](#_Toc197281831)

[**3. Análisis DAFO** 4](#_Toc197281832)

[**4. DEMO** 5](#_Toc197281833)

[**Objetivo general** 5](#_Toc197281834)

[**Escenario 1: Análisis de tendencias por ubicación y tipo de producto** 5](#_Toc197281835)

[**Escenario 2: Reposición automatizada de inventario** 6](#_Toc197281836)

[**Escenario 3: Predicción de demanda futura** 6](#_Toc197281837)

[**5. Conclusiones** 7](#_Toc197281838)

# **1. Introducción**

En la industria de la moda, la capacidad de adaptarse rápidamente a las tendencias del En la industria de la moda, la capacidad de adaptarse rápidamente a las tendencias del consumidor es clave para mantener la competitividad. Empresas como Zara han revolucionado este sector utilizando Big Data para analizar comportamientos de compra, optimizar su cadena de suministro y diseñar nuevas colecciones basadas en datos reales y actualizados.

Nosotros proponemos **SmartRetail**, una plataforma basada en tecnologías Big Data y bases de datos NoSQL que permite a tiendas de ropa aprovechar beneficios como el análisis en tiempo real de tendencias, predicción de demanda y gestión inteligente del inventario.

Las empresas pequeñas y medianas no cuentan con herramientas avanzadas para analizar de forma eficiente las preferencias de sus clientes, lo que resulta en una mala gestión de inventario, baja rotación de productos y pérdidas económicas.

## **Objetivos del proyecto:**

* Implementar un sistema de análisis de datos de ventas y comportamiento de usuarios.
* Facilitar la toma de decisiones basada en datos para el diseño y reposición de productos.
* Ofrecer recomendaciones automatizadas de stock por tienda, región y temporada.
* Reducir tiempos de reacción ante nuevas tendencias del mercado.

## **Requisitos generales:**

* Gestión de grandes volúmenes de datos.
* Almacenamiento flexible y escalable (MongoDB como base NoSQL).
* Análisis de datos en tiempo real.
* Visualización interactiva de resultados (Streamlit).

# **2. Tecnología utilizada**

## **MongoDB:**

Base de datos NoSQL orientada a documentos, ideal para entornos donde los datos son diversos, no estructurados o cambian con frecuencia. A diferencia de las bases de datos relacionales tradicionales, permite almacenar información en formato JSON flexible, lo cual encaja perfectamente con el dinamismo del sector retail.

En el contexto de SmartRetail, MongoDB se emplea para:

* Registrar transacciones de venta por producto, tienda y fecha.
* Gestionar el inventario en tiempo real.
* Almacenar datos semiestructurados como categorías, ubicación o tipo de prenda.

Ventajas principales:

* ***Escalabilidad horizontal***: permite crecer fácilmente al aumentar la carga de datos.
* ***Esquema dinámico***: se adapta a cambios en el modelo de datos sin necesidad de migraciones complejas.
* ***Integración sencilla con Python***: mediante PyMongo y otras librerías.

## **Python y librerías analíticas**

Python es uno de los lenguajes más utilizados para el análisis de datos y ciencia de datos por su simplicidad, comunidad activa y rica variedad de herramientas.

En este proyecto, Python permite desde la conexión con la base de datos hasta el análisis predictivo y la visualización de resultados.

Librerías utilizadas:

* **Pandas**: para procesar y analizar estructuras de datos tabulares.
* **Scikit-learn:** para entrenar modelos de aprendizaje automático, como regresores y clasificadores.
* **Plotly:** para generar gráficos interactivos de alta calidad.
* **Streamlit:** para construir una interfaz web limpia e intuitiva con filtros, pestañas y botones interactivos.

## **Random Forest para predicción de demanda**

Random Forest es un algoritmo de aprendizaje automático basado en la combinación de múltiples árboles de decisión. Cada árbol es entrenado con una muestra distinta de los datos y la predicción final se obtiene por consenso (promedio en regresión o votación en clasificación).

Usamos Random Forest para predecir cuántas unidades se venderán la próxima semana. El modelo tiene en cuenta variables como:

* Día del año
* Categoría del producto
* Ciudad
* Día de la semana
* Mes

Esto permite generar estimaciones precisas incluso ante datos complejos o con ruido, lo que lo convierte en una opción robusta y adecuada para el retail de moda.

# **3. Análisis DAFO**

El análisis DAFO es una herramienta estratégica que se utiliza para entender la situación de una empresa en el momento de llevar a cabo dicho análisis.

|  |  |
| --- | --- |
| **Fortalezas** | **Debilidades** |
| Tecnología escalable y flexible | Necesita conocimientos técnicos específicos |
| Adaptación rápida | Gestión más compleja que bases SQL tradicionales |
| Integración sencilla con herramientas de análisis | Rendimiento puede variar sin diseño optimizado |
| **Oportunidades** | **Amenazas** |
| Alta demanda de soluciones Big Data | Cambio constante en tendencias tecnológicas |
| Posibilidad de aplicar el modelo a otras industrias | Competencia con soluciones propietarias ya existentes |
| Mejora continua del ecosistema NoSQL, como MongoDB | Necesidad de asegurar privacidad de datos sensibles |

# **4. DEMO**

## **Objetivo general**

Demostrar, mediante una aplicación interactiva, cómo el análisis de datos y el aprendizaje automático pueden ayudar a una cadena de tiendas de moda a tomar mejores decisiones sobre inventario, ventas y tendencias.

La demo está construida con tecnologías reales y escalables, y refleja el funcionamiento de una herramienta de retail inteligente aplicable en escenarios reales.

**Componentes de la demo**

* **MongoDB**: base de datos NoSQL donde se almacenan registros de ventas, productos, inventario y ubicaciones.
* **Python (backend)**: se encarga de procesar los datos, agruparlos, aplicar filtros y entrenar el modelo de predicción.
* **Streamlit (interfaz visual)**: permite a cualquier usuario (sin conocimientos técnicos) explorar los datos mediante filtros, pestañas y gráficas interactivas.

## **Escenario 1: Análisis de tendencias por ubicación y tipo de producto**

**Descripción:**  
El sistema muestra las unidades vendidas por ciudad, permitiendo comparar regiones entre sí. Además, permite visualizar cómo se comporta cada tipo de producto en cada zona geográfica.

**Datos utilizados:**

* Ubicación de la tienda
* Categoría del producto
* Unidades vendidas

**Resultados esperados:**

* Gráfico de barras con volumen de ventas por ciudad
* Gráfico agrupado con ventas por categoría y ciudad
* Identificación de tendencias locales

## **Escenario 2: Reposición automatizada de inventario**

**Descripción:**  
La aplicación analiza el inventario actual y señala automáticamente los productos con menos de 50 unidades disponibles, sugiriendo su reposición.

**Datos utilizados:**

* Unidades en stock por producto y tienda
* Unidades vendidas en el periodo reciente

**Resultados esperados:**

* Tabla con productos en riesgo de quiebre de stock
* Botón para exportar estos datos como CSV
* Filtro por ciudad y categoría

## **Escenario 3: Predicción de demanda futura**

**Descripción:**  
Se entrena un modelo predictivo con datos históricos para estimar la demanda de la próxima semana. El modelo tiene en cuenta variables como la ubicación, categoría, día del año, mes y día de la semana.

**Datos utilizados:**

* Histórico de ventas
* Variables temporales (día, mes, día de la semana)
* Ubicación y categoría del producto

**Resultados esperados:**

* Gráfico de líneas con la predicción de demanda
* Métricas del modelo (R², MAE, RMSE)
* Línea histórica de demanda real agregada por día

# **5. Conclusiones**

El uso de tecnologías Big Data como MongoDB y herramientas analíticas en Python permite transformar cómo las cadenas de moda gestionan su inventario y responden al mercado. Con esta propuesta hemos conseguidos varios beneficios claves que pueden marcar el destino de una empresa respecto a la competencia:

* Mejora en la **eficiencia operativa**.
* Aumento de la **satisfacción del cliente**.
* Se reduce el **exceso o falta de stock**.