

# **SISTEMA INTELIGENTE DE PREDICCIÓN DE INVENTARIOS CON BIG DATA PARA EL MERCADO DE CONSTRUCCIÓN**

**Jimmy Sumba, Freddy Montalvan, Eduardo Mendieta**

**Cuenca 14 de febrero de 2023**

## **1. Resumen**

### **1.1 Antecedentes:**

La creciente demanda de productos de construcción ha llevado a las distribuidoras a buscar soluciones tecnológicas que optimicen la gestión de inventarios y mejoren la satisfacción del cliente. El uso de herramientas de Big Data y machine learning se ha vuelto esencial para predecir tendencias y comportamientos de compra en este sector.

### **1.2 Problema a Resolver**

La distribuidora MQ enfrenta dificultades para prever la demanda de productos, lo que resulta en desabastecimiento o exceso de inventario. Esto afecta tanto la eficiencia operativa como la satisfacción del cliente, generando pérdidas económicas y oportunidades de negocio no aprovechadas.

### **1.3 Objetivo**

El objetivo general es predecir la demanda de productos de construcción mediante un sistema inteligente que permita un análisis de ventas más efectivo y una mejor satisfacción del cliente, ampliando así el negocio de la distribuidora MQ.

### **1.4 Metodología a Aplicar**

**Se utilizará un enfoque de análisis de datos que incluye:**

- **Captación de Datos:** Recolección de información a partir de facturas y registros físicos, organizados en un archivo de Excel.
- **Procesamiento y Preprocesamiento:** Limpieza y transformación de datos utilizando Python y bibliotecas como Pandas.
- **Exploración de Datos:** Análisis estadístico y visualización de patrones y tendencias en las ventas.
- **Modelado Predictivo:** Implementación de algoritmos de machine learning para predecir la demanda de productos.

## **1.5 Resultados Esperados**

**Se espera obtener un modelo predictivo que permita:**

- Mejorar la precisión en la gestión de inventarios.
- Aumentar la satisfacción del cliente mediante un suministro más eficiente de productos.
- Generar informes interactivos que faciliten la toma de decisiones estratégicas en la distribuidora MQ.

Este enfoque integral no solo optimiza los procesos internos, sino que también posiciona a la distribuidora como líder en el uso de tecnología en el sector de la construcción.

## **2. Introducción**

En el competitivo mercado de la construcción, las distribuidoras enfrentan retos significativos en la gestión de inventarios y la satisfacción del cliente. Prever la demanda de productos se ha vuelto crucial para garantizar la eficiencia operativa y maximizar oportunidades de negocio. Este artículo presenta un proyecto de la distribuidora MQ que busca establecer un sistema inteligente de predicción de inventarios mediante el uso de Big Data y machine learning.

El proceso comienza con la captación de datos a partir de facturas y registros físicos, organizados en un archivo de Excel. A través del procesamiento y preprocesamiento de datos, se abordan inconsistencias y se prepara la información para su análisis. La exploración de datos permite identificar patrones en las ventas, sentando las bases para el modelado predictivo. Este enfoque integral no solo optimiza el inventario, sino que también mejora la satisfacción del cliente, posicionando a la distribuidora MQ como líder en el uso de tecnología en el sector.

## **3. Investigación**

El sector de la construcción ha crecido notablemente, lo que ha llevado a las distribuidoras a enfrentar desafíos en la gestión de inventarios y la satisfacción del cliente. Prever la demanda de productos es esencial para optimizar operaciones; según McKinsey & Company (2020), las empresas que implementan soluciones de Big Data pueden mejorar su eficiencia operativa hasta en un 20%. El proyecto de la distribuidora MQ se centra en establecer un sistema inteligente de predicción de inventarios, comenzando con la captura y procesamiento de datos de ventas. La exploración de estos datos permite identificar patrones críticos en el comportamiento de compra. Deloitte (2021) señala que el uso de análisis predictivo mejora la satisfacción del cliente y reduce costos de inventario, posicionando a la distribuidora MQ como un referente en tecnología en el sector.

Según McKinsey & Company (2020):

"Las empresas que implementan soluciones de Big Data pueden mejorar su eficiencia operativa hasta en un 20%, lo que resalta la importancia de adoptar tecnologías avanzadas para optimizar las operaciones en el sector de la construcción" (p. 15).

Deloitte (2021) señala que las organizaciones que utilizan análisis predictivo en sus operaciones no solo logran una mejora significativa en la satisfacción del cliente, sino que también experimentan una reducción en los costos asociados al inventario.

#### **4. Metodología CRISP-DM**

Se ha adoptado la metodología CRISP-DM (Cross-Industry Standard Process for Data Mining), que se estructura en seis etapas (Cortina, 2015). Esta metodología es beneficiosa por su clara estructura y su enfoque en la calidad de los datos y la alineación con los objetivos empresariales.

##### **Fases de la Metodología**

- **Comprensión del negocio:** Se reunió al equipo de desarrollo con la empresa MQ Distribuidora para entender el negocio y establecer reglas para el modelo. Se identificó el objetivo del sistema de predicción (Sara, 2016).
- **Comprensión de los datos:** Se construyó un conjunto de datos en Excel a partir de registros físicos y se realizó un análisis preliminar de las variables. Posteriormente, se exploraron los datos con Pandas en Python (Sara, 2016).
- **Preparación de los datos:** Se limpió y trató el conjunto de datos, transformando la información a su formato correcto en Pandas, preparándola para el modelado (Cortina, 2015).
- **Modelado:** Se seleccionaron tres modelos de aprendizaje supervisado: Regresión Lineal, Árboles de Decisión para Regresiones y Random Forest, para comparar los resultados (Cortina, 2015).
- **Evaluación:** Se aplicaron técnicas como el Error Cuadrático Medio (MSE) y la Visualización de Predicciones para evaluar si el modelo lograba predicciones satisfactorias (Cortina, 2015).
- **Despliegue:** Aunque no se realizó un despliegue del sistema, se utilizó Power BI para presentar los resultados mediante un tablero de control (Cortina, 2015).

Esta metodología permite iteraciones flexibles y mejora la comunicación entre los miembros del equipo, contribuyendo al éxito del proyecto (Moine, Haedo, & Gordillo, 2011).

#### **4.1 DESARROLLO**

##### **Fase 1: Comprensión del negocio de MQ Distribuidora de Ladrillos y Tejas**

En la Fase 1, el equipo de desarrollo se reunió con el gerente de MQ Distribuidora para entender el negocio y sus necesidades. Se establecieron reglas para crear un modelo de base de datos alineado con los objetivos empresariales y se identificó el propósito del sistema de predicción (Sara, 2016).

##### **Definición de Requisitos:**

Para crear la base de datos, se realizó un levantamiento de información en las instalaciones de MQ Distribuidora, guiados por un experto en las operaciones. Se establecieron reglas de negocio que indican que cada producto se almacena en una única bodega, que las bodegas

pueden contener diferentes productos y que es crucial conocer el estado y la capacidad de almacenamiento de cada una para planificar mantenimientos periódicos.

### **Diseño Conceptual**

Se elaboró un esquema Entidad-Relación (ER) basado en las reglas de negocio de MQ Distribuidora, que representa gráficamente las entidades y sus interacciones, simplificando la complejidad de los datos. Un diagrama ER ayuda a identificar redundancias y asegura que se capturen todos los requisitos necesarios, actuando como un plano para la construcción de la base de datos. Las entidades clave identificadas incluyen Cliente, Compra, Producto, Proveedor, Despacho, Bodega, Empleado y Vehículo, con relaciones y atributos definidos para reflejar el funcionamiento del negocio.

### **Diseño Lógico**

El diseño lógico transforma el diagrama entidad-relación en un Modelo Relacional, organizando los datos en tablas que reflejan las entidades y sus relaciones a través de claves primarias y foráneas. Este modelo es crucial para garantizar la integridad y coherencia de la información, facilitando la gestión y recuperación de datos. Además, se realiza un proceso de normalización hasta la tercera forma normal (3NF) para reducir redundancias y mejorar la calidad de los datos, asegurando que cada tabla contenga atributos atómicos y que las dependencias estén correctamente definidas. El resultado es un Modelo Relacional Normalizado que optimiza la estructura de la base de datos.

### **Diseño Físico y Implementación**

El diseño físico de la base de datos se realiza mediante SQL, que permite crear tablas, definir relaciones y establecer restricciones de integridad, facilitando la gestión de datos en sistemas de gestión de bases de datos (SGBD). Para la implementación, se eligió PostgreSQL, un sistema de código abierto que soporta transacciones ACID y maneja grandes volúmenes de datos con alto rendimiento. Su robustez y capacidad de escalabilidad lo hacen ideal para aplicaciones empresariales críticas, asegurando una gestión eficiente y segura de la información.

### **Fase 2: Comprensión de los datos**

En la fase de comprensión de los datos, se construyó un dataset en Excel a partir de facturas y registros, acompañado de un diccionario de datos para asegurar la calidad. Posteriormente, se utilizó Google Colab para el procesamiento y preprocesamiento de datos, abordando errores y transformando variables con Python y bibliotecas como Pandas. Se realizó un análisis exploratorio para identificar patrones y tendencias, seguido por el modelado utilizando Regresión Lineal, Árboles de Decisión y Random Forest para predecir el stock mensual. Finalmente, se evaluaron los modelos mediante métricas como MAE y  $R^2$ , y se desarrolló un dashboard en Power BI para visualizar los resultados y facilitar la toma de decisiones estratégicas.

## 5. RESULTADOS

Se recopiló un dataset de 22 columnas en MQ Distribuidora para entrenar modelos de predicción de stock y analizar la satisfacción del cliente. Los modelos evaluados (Regresión Lineal, Árboles de Decisión y Random Forest) mostraron ajustes imperfectos y algunos errores extremos. En Power BI, se destacó que la zona de Cuenca Norte presenta la mayor satisfacción, con un 83.17% de clientes satisfechos siendo hombres.

## 6. CONCLUSIONES

A partir de los resultados obtenidos, se pueden extraer las siguientes conclusiones:

- La calidad del dataset es crucial para el éxito del modelo; datos incompletos o erróneos pueden comprometer la precisión de las predicciones.
- La insuficiencia de registros para ciertos productos limita la capacidad de los modelos para generalizar, lo que puede llevar a resultados menos confiables.
- Ampliar la base de datos con más registros y variables relevantes es esencial para mejorar el rendimiento y la efectividad de los modelos de predicción.

## 7. BIBLIOGRAFÍA

- Flores Taquiri, G. A., & Montalvo Celis, J. S. (2024). Machine learning para la predicción en la gestión de inventario dirigida a PYMES de venta de productos
- McKinsey & Company. (2020). The future of construction: A global forecast for construction to 2030.
- Deloitte. (2021). Data-driven decision making in construction: The role of predictive analytics.
- Rodríguez Rodríguez, J. E. (2010). Fundamentos de minería de datos. Editorial Universidad Distrital Francisco José de Caldas. Editorial UD.  
[https://books.google.es/books?hl=es&lr=lang\\_es&id=t7zvEAAQBAJ&oi=fnd&pg=PT7&dq=preprocesamiento+de+datos+en+minería+de+datos&ots=8B4fr30eET&sig=2hPBYZTh\\_ufexTuZzx1c3ptlvBw#v=onepage&q=preprocesamiento%20de%20datos%20en%20minería%20de%20datos&f=false](https://books.google.es/books?hl=es&lr=lang_es&id=t7zvEAAQBAJ&oi=fnd&pg=PT7&dq=preprocesamiento+de+datos+en+minería+de+datos&ots=8B4fr30eET&sig=2hPBYZTh_ufexTuZzx1c3ptlvBw#v=onepage&q=preprocesamiento%20de%20datos%20en%20minería%20de%20datos&f=false)
- Cajahuanca, J. E. V., Raymundo, Á. F. N., Franco, A. C. L., & Flores, J. D. J. (2022). Deserción universitaria: Evaluación de diferentes algoritmos de Machine Learning para su predicción. Revista de ciencias sociales, 28(3), 362-375.