

Documentación Técnica

Proyecto: Desarrollo e implementación de un modelo de predicción de la demanda para la industria del sector retail.

1 Metodología

Para el proyecto mencionado se ha definido 5 pasos cruciales en el proceso del desarrollo del modelo los cuales se puede evidenciar en el gráfico mostrado a continuación:

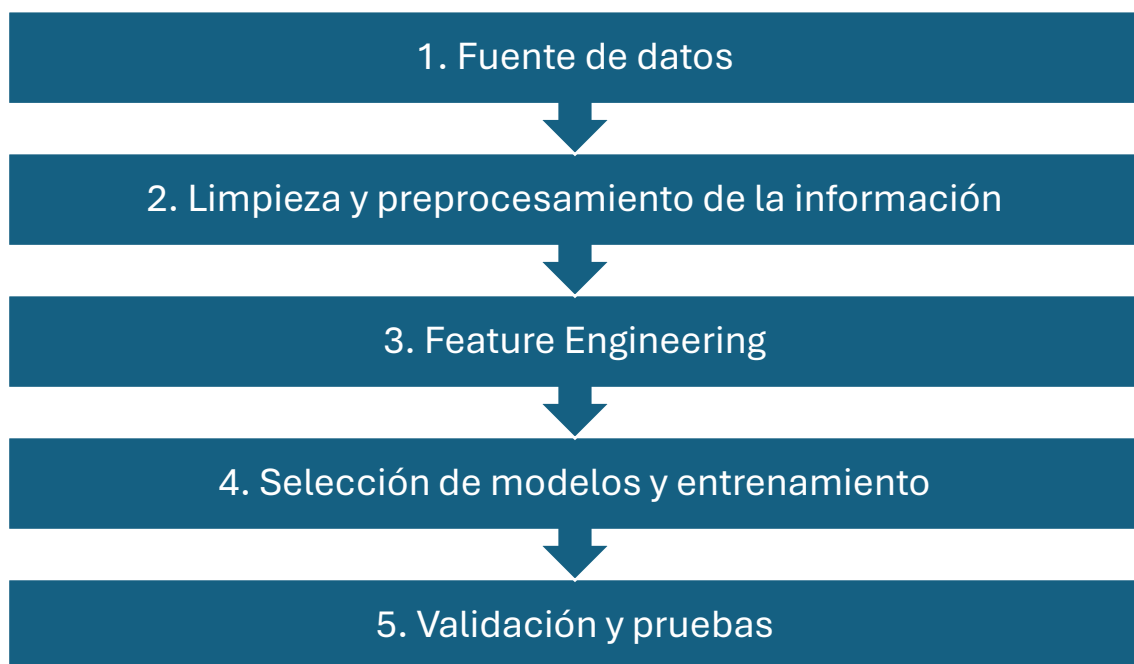


Ilustración 1: Proceso de desarrollo del modelo.

1.1 Fuente de datos:

- Fuente: ventas históricas de 2023–2025 (base de datos Pycca).
- Frecuencia: diaria.
- Variables consideradas: categoría de producto, tienda, dimensiones de tienda, ciudad, estacionalidad, festivales, fines de semana, venta

1.2 Limpieza y preprocesamiento:

- Imputación de valores faltantes mediante relación con otras columnas.
- Conversión de variables categóricas a formato numérico mediante OneHotEncoding.
- Normalización/Estandarización de variables numéricas.

1.3 Feature Engineering:

- Creación de variables rezagadas (lags) de ventas de 1 y 7 días
- Ventanas móviles para capturar tendencias y estacionalidad (rolling mean).
- Codificación de variables temporales (mes, semana, día de la semana, festivos, fin de semana).

1.4 Selección de modelo y entrenamiento:

- Algoritmos: XGBoost , LightGBM , Ridge, Lasso, Sarima (Md Sabbirul Haque, 2023).
- Validación: división temporal (train:2023–2024, test:2025 hasta julio).
- Métrica de evaluación: MAPE, RMSE.

1.5 Validación y prueba:

- Evaluación de desempeño sobre datos de prueba.
- Comparación con modelos base (promedio histórico).

2 Supuestos del modelo

Para el desarrollo del modelo se consideraron los siguientes supuestos:

1. La relación entre variables históricas y ventas futuras se mantiene estable en el corto plazo.
2. No se esperan cambios drásticos en la estrategia comercial, precios o remates.
3. El patrón de estacionalidad de años anteriores se repetirá en el futuro.
4. La información utilizada está completa y libre de errores significativos.
5. Factores externos como crisis o condiciones energéticas no tendrán variaciones radicales durante el horizonte de predicción.
6. En la predicción de demanda no se consideran las devoluciones realizadas.

3 Limitaciones

Para el correcto funcionamiento del modelo se deben de considerar las siguientes limitaciones:

1. El modelo no contempla eventos imprevistos como rupturas de stock, cambios en proveedores o picos de demanda excepcionales no presentes en el histórico.
2. La calidad de la predicción depende directamente de la calidad y granularidad de los datos.

3. Las predicciones a largo plazo (más de 6–12 meses) son menos precisas debido a la acumulación de incertidumbre y necesitarían un análisis más profundo y mayor poder computacional.
4. Posible pérdida de precisión en categorías con pocos datos históricos.
5. Limitada capacidad de generalizar a tiendas o productos nuevos sin histórico.
6. Cambios en el formato y estructura de la información recopilada.

4 Plan de implementación

El plan de implementación de modelo consta de dos puntos clave:

- Integración del modelo
- Gobernanza de datos

4.1 Integración del modelo

4.1.1 Implementación del modelo

El modelo se integrará como una herramienta de soporte en la planificación de pedidos a proveedores, considerando pronósticos por sucursal y período ya sea este diario, semanal o mensual. Se comparará el pronóstico con el stock actual y un stock de seguridad para calcular necesidades de reposición. Además, se generará una tabla con sugerencias de compra que incluirá producto, sucursal/bodega, cantidad sugerida, proveedor asociado. La integración con el sistema que maneja la empresa para gestionar las compras permitirá identificar los stocks próximos a agotarse en las sucursales más visitadas y coordinar su pronto reabastecimiento.

4.1.2 Proceso de automatización

- **Automatización del pipeline de datos:** Extracción automática de los datos de forma diaria o semanal seguido de un preprocesamiento automatizado en donde se obtengan las features de interés y finalmente obtener el formato preciso para utilizarlo en el modelo, todo esto mediante herramientas como AWS Glue para el proceso ETL, además de Sagemaker para la tarea de machine Learning.
- **Ejecución programada del modelo:** Tareas programadas para el entrenamiento y actualización del modelo con su posterior pronóstico.
- **Dashboard interactivo:** Visualización en tiempo real de demanda, stock y alertas para su reposición.

4.1.3 Monitoreo del rendimiento

- **KPI de evaluación:** Monitoreo de MAPE e impacto en el stock.
- **Monitoreo continuo:** Comparación semanal de pronóstico vs real.
- **Alertas automáticas:** Definición de umbrales y alertas para tomar acción ante cualquier eventualidad.

- **Versionado:** Almacenamiento de versiones y control de cambios de cada actualización.

4.2 Gobernanza de datos

4.2.1 Calidad de los datos

- Con el propósito de garantizar que los datos utilizados por el modelo sean correctos y precisos se aplicarán las validaciones respectivas en cuanto a tipos de datos correctos en la carga, estandarización de formatos, manejo de datos faltantes y atípicos.
- Se debe asegurar que los datos provengan de fuentes confiables ya sea de un ERP o del sistema propio de la empresa.
- Se debe realizar un monitoreo continuo con sus respectivos reportes y alertas ante anomalías.
- Todo esto obedece a un ciclo PDCA (Plan, Check, Do, Act) en donde se definen métricas, se aplican las validaciones, se revisan los reportes y se corrigen los fallos o novedades.

4.2.2 Actualización de modelos

- El modelo será reentrenado de forma periódica ya sea mensual o trimestral, además de ser ajustado inmediatamente ante cualquier evento extraordinario.
- Se evaluará periódicamente el modelo usando el MAPE y MAE procediendo a realizar una actualización anticipada ante cualquier degradación presentada en la evaluación.
- Cada nueva versión del modelo será documentada y evaluada antes de su salida a producción.

4.2.3 Definición de responsabilidades

Los roles y responsabilidades para esta implementación están dados de la siguiente forma usando un mapa **RACI**:

R – Responsable (Responsable de ejecutar)

A – Accountable (Responsable final / Autoridad)

C – Consulted (Consultado)

I – Informed (Informado)

Actividad	Ingeniero de datos	Data scientist	Equipo de compras	TI	Gobernanza de datos	Gerencia
Limpieza,validación y mantenimiento	R		C	C	C	A
Desarrollo,entrenamiento,validacion y optimizacion	C	R	C	C	C	A
Uso de resultados	C	C	R	C	C	A
Soporte de servidores	C	C	C	R	C	A
Políticas y control de calidad	C	C	I	C	R	A
Aprobacion de despliegue	I	I	I	I	I	R

Tabla 4.1: Definicion de roles y responsabilidades

5 Referencias

Md Sabbirul Haque, M. S. (2023). *Retail Demand Forecasting: A Comparative Study for Multivariate Time Series*.