

MODELO DE MACHINE LEARNING PARA LA PREDICCIÓN DEL COMPORTAMIENTO DE COMPRAS DE CLIENTES EN LA INDUSTRIA DEL CEMENTO (UNACEM ECUADOR S.A.)

Autoras:

Diana Carolina Cabezas Falcón

Stephanie Salomé Mejía Vera

Profesor

Mario Salvador González

2024



01

CONTEXTO

La industria del cemento es crucial para el desarrollo económico de cualquier país, especialmente en sectores como la construcción e infraestructura.

02

FUENTE DE DATOS

Para el desarrollo de este proyecto, se utilizarán los datos históricos de ventas de UNACEM ECUADOR S.A., recolectados durante los últimos tres años

03

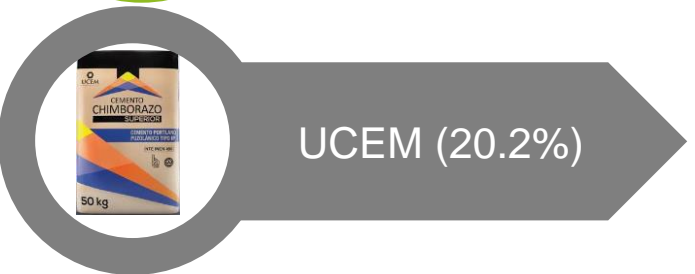
RELEVANCIA

La capacidad de predecir la demanda de cemento es vital para cualquier empresa en esta industria debido a la volatilidad y naturaleza cíclica del mercado.



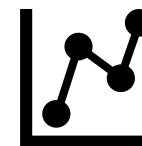
Industria del Cemento

La industria del cemento es un sector clave a nivel global, especialmente en países en desarrollo donde el crecimiento de la infraestructura es un pilar fundamental para el desarrollo económico. En Ecuador este mercado es dominado por 3 grandes empresas.



El modelo de Stackelberg

En Ecuador, la industria del cemento ha experimentado un desarrollo significativo impulsado por la creciente demanda, la competencia no solo se basa en precio sino en capacidad de producción y distribución.





Unacem Ecuador S.A., enfrenta un desafío crítico, desarrollar una herramienta que le permita predecir el comportamiento de compra para contar con el stock suficiente para satisfacer la demanda del mercado, en reiteradas ocasiones una falta de predicción ha resultado en la pérdida de clientes en un mercado competitivo.



Al poder pronosticar de manera informada la demanda, Unacem puede planificar sus procesos de producción y logística llevándolos a ser más eficientes.



Al garantizar que el producto esté disponible cuando los clientes lo necesitan, Unacem mejoraría significativamente la satisfacción del cliente, fortalecimiento de marca.



Con un modelo de predicción, Unacem puede minimizar la migración de sus clientes por falta de stock. Esto ayudará a la empresa a retener a sus clientes actuales y a evitar la pérdida de participación en el mercado



Optimizar la cadena de suministro y reducir los costos asociados

MODELO DE MACHINE LEARNING PARA LA PREDICCIÓN DEL COMPORTAMIENTO DE COMPRAS DE CLIENTES
EN LA INDUSTRIA DEL CEMENTO (UNACEM ECUADOR S.A.)





OBJETIVO GENERAL



Objetivo

El objetivo general del proyecto es desarrollar un modelo de machine learning que permita predecir con precisión la demanda de cemento por parte de los clientes de **Unacem Ecuador S.A.**

Este modelo debe ser capaz de anticipar cuándo y cuánto cemento solicitarán los clientes, basándose en el análisis de datos históricos de ventas.

La implementación exitosa del modelo permitirá a la empresa mejorar su capacidad de respuesta a la demanda, optimizar la producción y la distribución, y reducir los costos operativos, lo que fortalecerá su posición competitiva en el mercado.

Recopilar Datos Históricos

Recolectar datos de ventas de los últimos 3 años con información primordial para análisis

Desarrollo Modelo Predictivo

Desarrollo y evaluación de modelos machine learning para predecir la demanda de cemento

Implementación Modelo

Implementar modelo predictivo y desarrollar un plan de acción para utilizar las predicciones del modelo para toma de decisiones

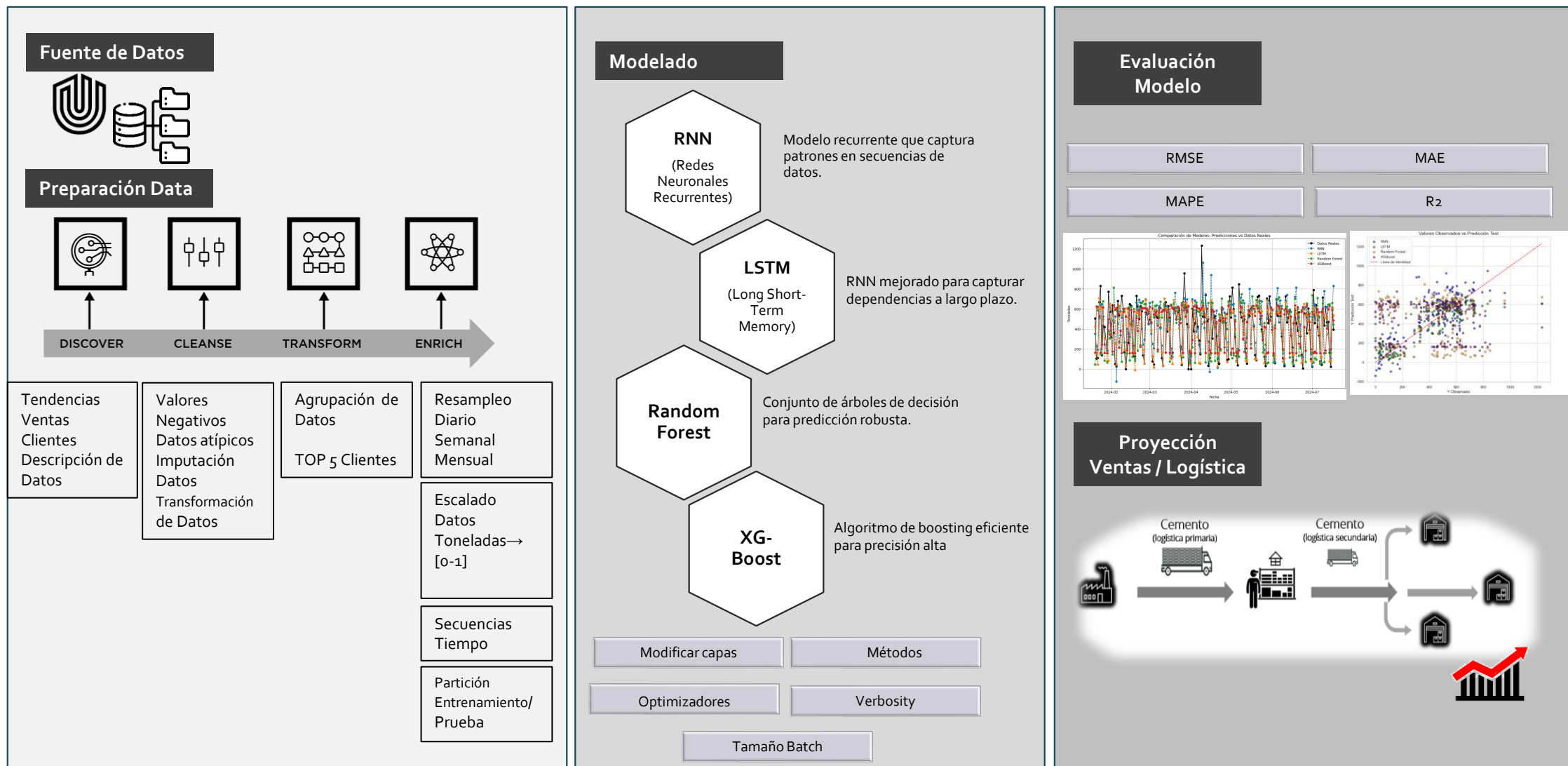
Monitoreo y Evaluación Modelo

Seguimiento y ajuste continuo del modelo para precisión, eficacia e impacto.

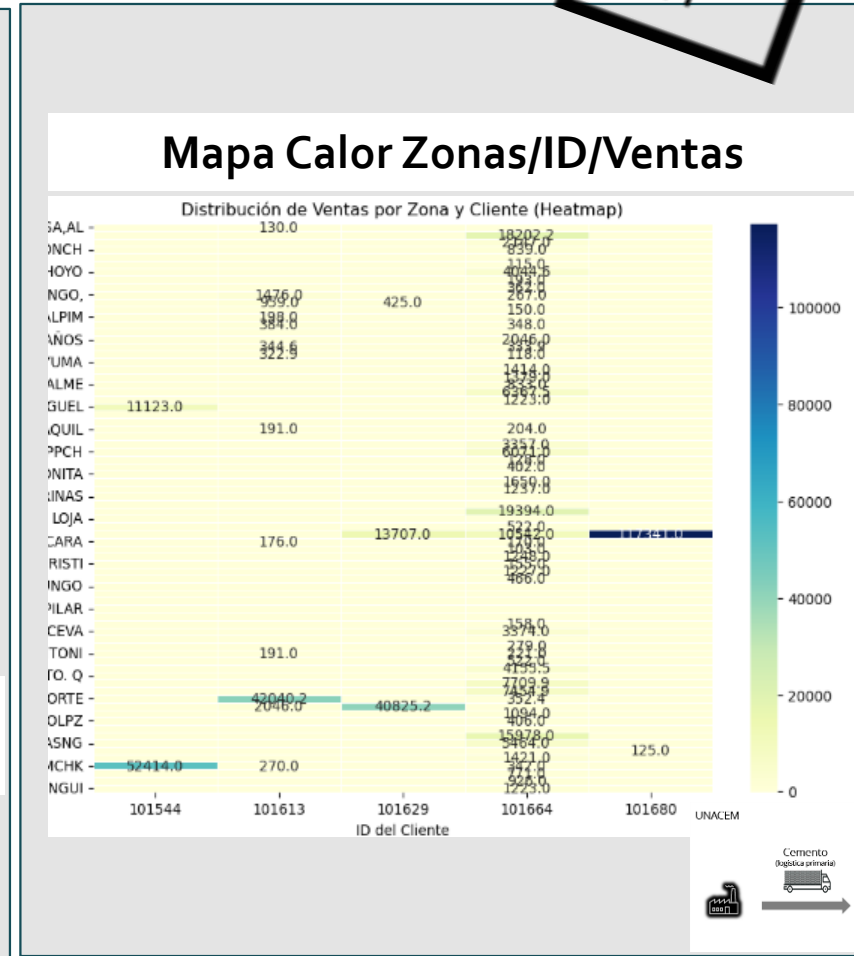
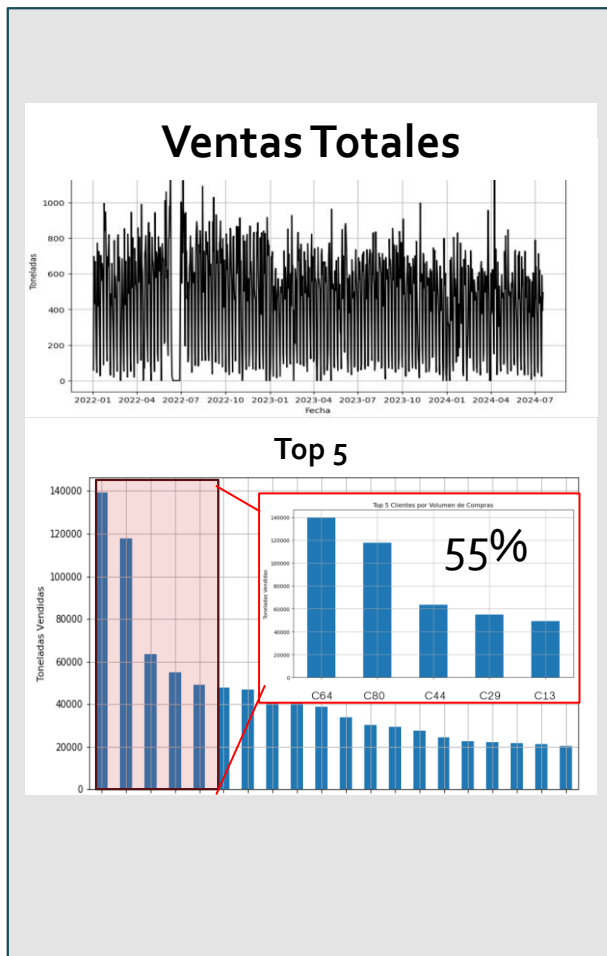


OBJETIVOS ESPECÍFICOS

Flujo de Trabajo



Preparación Data



¿Discretización?

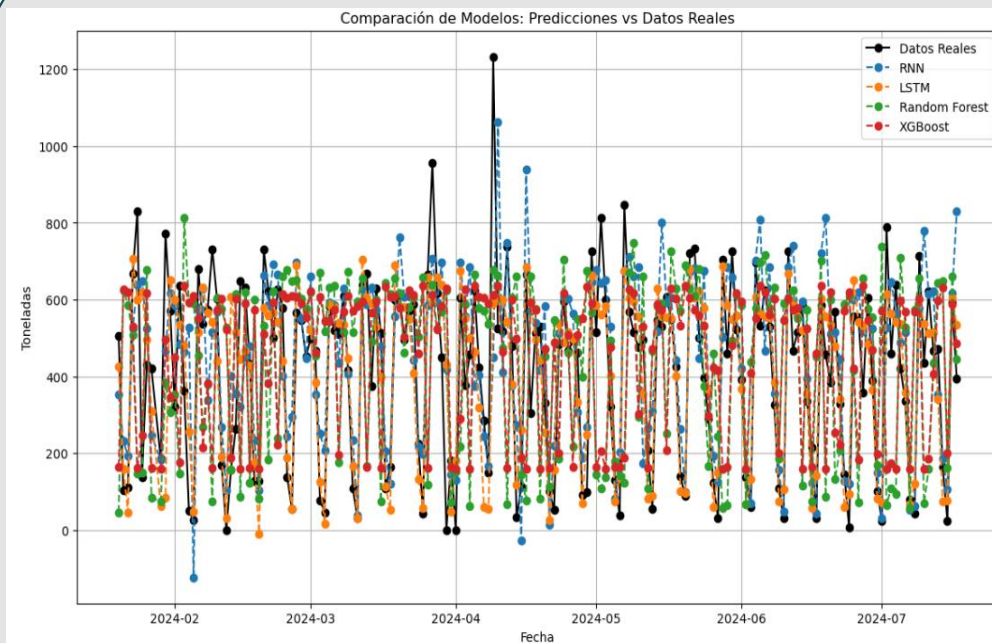
Modelado general suma del top 5

Modelado por cliente en el top 5



Modelado

Evaluación Modelo/Retroalimentación

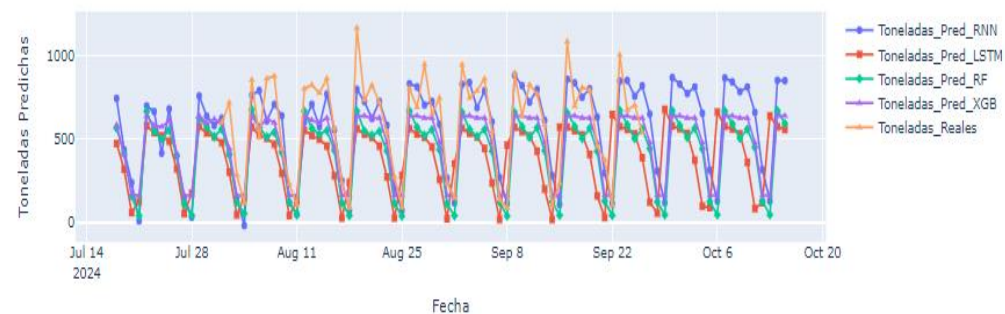


RNN: MAE = 0.1106, MAPE = 239.97%, $R^2 = 0.4070$
 LSTM: MAE = 0.0855, MAPE = 209.25%, $R^2 = 0.6339$
 Random Forest: MAE = 0.0848, MAPE = 43.44%, $R^2 = 0.7490$
 XGBoost: MAE = 0.0950, MAPE = 57.37%, $R^2 = 0.6986$

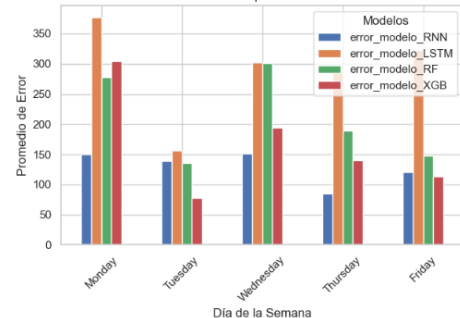
Data Fecha Max -
Julio 2024

Retroalimentación
Agosto-Septiembre

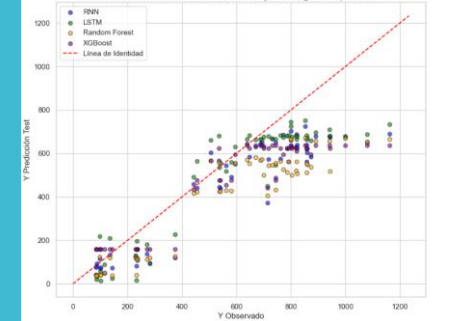
Comparación de Predicciones de Toneladas por Modelo



Promedio de Error por Día de la Semana



Valores Observados vs Predicción Proyección Agosto Septiembre



Discusión del Desempeño de Modelos basado en métricas de rendimiento



Mejor desempeño: Random Forest

MAE: 0.08, MAPE:
43%, R^2 : 0.75
Captura relaciones no
lineales con alta
precisión



LSTM: Buen rendimiento con Alta Dispersión

MAE: 0.09, MAPE:
209%, R^2 :0.63
Dificultades con
valores extremos



XGBoost:Desempeño Competitivo

MAE: 0.1, MAPE:
57%, R^2 :0.7
Tiene precisión en
pico y valles



Peor Desempeño: RNN

MAE: 0.1, MAPE:
240%, R^2 : 0.4
Presenta dificultades
capturando la
estructura de los
datos

La evaluación de los modelos de series temporales revela que Random Forest ofrece el mejor desempeño general, con un bajo MAE y una alta capacidad para capturar relaciones no lineales entre variables. LSTM, a pesar de ser un modelo avanzado, muestra dificultades en la predicción de valores extremos, como reflejado en su alto MAPE. XGBoost se posiciona como un competidor fuerte, especialmente en su capacidad para predecir con precisión picos y valles en los datos. RNN, en cambio, presenta el peor rendimiento por su alta dispersión en los errores y su incapacidad para capturar la estructura temporal.





Propuesta de Solución

Implementación de Random Forest y XGBoost

Del análisis realizado se ha concluido que ambos modelos son precisos en predicción de demanda y en relación a reducción de costos de stock

Integración en el modelo de gestión

Los modelos se integrarán en el sistema de planificación existente (ERP, BI), generando informes automáticos que permitirán al equipo tomar decisiones estratégicas.

Monitoreo y Ajuste Continuo

A medida que se vaya recolectando datos los modelos deben ser recalibrados periodicamente para mantener su precision.

Colaboración y expansión

El enfoque colaborativo no solo garantizará una implementación efectiva, sino que también facilitará la identificación de oportunidades de mejora y la adaptación del modelo a las diversas condiciones del mercado en otras regiones o segmentos.



Medidas contingentes para Ajustes del Modelo

Ajustes en la planificación

Random Forest para demanda moderada, XGBoost para alta demanda, usar ensamble para minimizar errores

Monitoreo en tiempo real

Revisar predicciones y ventas semanalmente, recalibrar si el error MAPE supera el 60%

Stock contingente

Mantener un 10-15% de stock adicional para cubrir posibles fluctuaciones de demanda.





Conclusiones

Cada fase, desde la recolección y limpieza de datos hasta la selección y evaluación de modelos, fue ejecutada de acuerdo con las mejores prácticas en analítica de datos. Los resultados obtenidos se detallan a continuación:



Random Forest y XGBoost destacan como los mejores modelos para predecir la demanda de los principales clientes de la cementera, debido a su precisión y capacidad para manejar la variabilidad en la demanda. Ambos modelos permiten una mejor planificación de inventarios y producción.

LSTM y RNN enfrentan dificultades para predecir picos de demanda o variaciones extremas, mostrando errores elevados en escenarios de alta volatilidad. Aunque son avanzados, su rendimiento es menos fiable en comparación con Random Forest y XGBoost.

Los errores de predicción no son uniformes a lo largo de la semana, afectando especialmente días como lunes y jueves, lo que requiere ajustar las estrategias logísticas y de inventario según los patrones diarios de demanda.

La optimización de inventarios y producción basada en las predicciones de demanda ha reducido costos operativos, mejorando la capacidad de respuesta de la empresa frente a las fluctuaciones del mercado, aunque es necesario un monitoreo constante para recalibrar los modelos y **ajustar contingentes**.





Recomendaciones Basadas en Innovación



Sistema de Gestión

1. Implementación de un sistema de gestión para ajustes de niveles de stock de acuerdo a la estacionalidad de los días, con especial afectación en los días con mayor demanda.



Recalibrar continuamente los modelos

1. Considerar la inclusión de nuevas variantes como indicadores macroeconómicos para recalibrar los modelos de machine learning y mejorar la precisión en los mismos.



Sistema de alertas automáticas

1. Desarrollar sistema de alertas automáticas para notificar cuando se presenten variaciones importantes entre la predicción y los datos reales para realizar los ajustes respectivos.



Incremento de Stock

1. Analizar la variación diaria en base a los resultados obtenidos para mantener un stock en piso adicional al requerido para abastecer el mercado en caso de que varíe la competitividad del mismo.



Capacidad de logística y distribución

1. Plantear propuestas de mejora en cuanto a la capacidad de logística y distribución para poder reforzar la misma y cumplir la demanda de predicción en los distintos segmentos y puntos de venta.



Expandir modelos predictivos

1. La expansión de los modelos predictivos a otros segmentos de clientes y áreas geográficas representa una oportunidad clave para mejorar la eficiencia operativa a gran escala, avanzar de forma progresiva, recalibrando continuamente los modelos y monitoreando su precisión en nuevos segmentos, lo que permitirá una expansión controlada y efectiva, minimizando riesgos y maximizando el impacto positivo de la predicción.



¡GRACIAS!

