



ESCUELA DE NEGOCIOS

MAESTRÍA EN INTELIGENCIA DE NEGOCIOS Y CIENCIA DE DATOS

**SISTEMA DE PREDICCIÓN DE DEMANDA DE PRODUCTOS UTILIZANDO
ALGORITMOS DE ENSEMBLE EN UNA BASE DE DATOS DE VENTAS,
CON EL FIN DE MEJORAR LA PRECISIÓN PREDICTIVA Y OPTIMIZAR LA
GESTIÓN DE INVENTARIO PARA LA PANIFICADORA “MÉXICO” EN EL
AÑO 2024**

**Profesor
Mario Salvador González**

**Autor
Michelle Rebeca Urdiales Cuenca
Edwin Santiago Bonilla Moreno**

2024

RESUMEN

Gestionar las ventas de productos es fundamental para las pequeñas y medianas empresas, pero a menudo resulta desafiante y genera preocupaciones para los dueños y gerentes. Muchas empresas todavía dependen de métodos empíricos o incluso de hojas de cálculo manuales para monitorear las ventas diarias, lo que limita su capacidad para tomar decisiones informadas. Sin embargo, contar con una visión objetiva y basada en datos sobre las ventas y la participación de productos puede proporcionar ventajas competitivas. Esto nos permite realizar análisis detallados que pueden potenciar las ventas y mejorar la eficiencia del negocio.

En este contexto, se estudia el caso de Panificadora México en Quito, con el objetivo de explorar las ventajas de sistematizar el seguimiento de ventas y de implementar modelos de machine learning para predecir los niveles de ventas y realizar proyecciones. Estas proyecciones son cruciales para relacionarlas con los niveles de producción necesarios en el negocio.

El uso de modelos de machine learning en la producción de alimentos ha ganado relevancia en años recientes debido a su capacidad para mejorar la eficiencia y la rentabilidad. Especialmente para pequeñas empresas de alimentos, estas tecnologías pueden optimizar procesos y apoyar la toma de decisiones basadas en datos. Panificadora México, una pequeña empresa de panificación, ha dependido principalmente de métodos empíricos para gestionar sus ventas y producción. La falta de un sistema estructurado ha causado dificultades en la planificación de la producción y la proyección de ventas, lo que ha llevado a problemas de inventario y eficiencia operativa.

Además, se desarrollarán modelos de machine learning para analizar estos datos y generar predicciones precisas de la demanda futura. Estas predicciones serán fundamentales para planificar la producción y gestionar el inventario de manera eficiente. La implementación de estos modelos y sistemas de gestión de datos no solo mejorará los procesos operativos, sino que también reducirá los costos asociados con el exceso de inventario y la pérdida de ventas.

ABSTRACT

Managing product sales is essential for small and medium-sized enterprises, but it often proves challenging and raises concerns for owners and managers. Many businesses still rely on empirical methods or even manual spreadsheet tracking to monitor daily sales, limiting their ability to make informed decisions. However, having an objective, data-driven view of sales and product performance can provide competitive advantages. This allows us to conduct detailed analyses that can boost sales and improve business efficiency.

In this context, the case of a traditional bakery in Quito is examined, aiming to explore the benefits of systematizing sales tracking and implementing machine learning models to predict sales levels and make projections. These projections are crucial for aligning them with the necessary production levels in the business.

The use of machine learning models in food production has gained relevance in recent years due to their ability to enhance efficiency and profitability. Especially for small food businesses, these technologies can optimize processes and support data-driven decision-making. Panificadora Mexico, a small bakery, has primarily relied on empirical methods to manage its sales and production. The lack of a structured system has caused difficulties in production planning and sales forecasting, leading to inventory problems and operational inefficiencies. To address these challenges, Panificadora Mexico plans to implement a data management system integrating sales, inventory, and production information in real-time.

Additionally, machine learning models will be developed to analyze this data and generate accurate predictions of future demand. These predictions will be crucial for production planning and efficient inventory management. The

implementation of these models and data management systems will not only improve operational processes but also reduce costs associated with excess inventory and lost sales. Furthermore, it will enable the company to introduce new products more effectively, based on precise data about customer preferences and market trends.

ÍNDICE DEL CONTENIDO

CARÁTULA	1
RESUMEN	2
ABSTRACT	4
ÍNDICE DEL CONTENIDO	6
ÍNDICE DEL TABLAS	7
ÍNDICE DEL FIGURAS	8
1. INTRODUCCIÓN	10
2. REVISIÓN DE LITERATURA	13
3. IDENTIFICACIÓN DEL OBJETO DE ESTUDIO	19
4. PLANTEAMIENTO DEL PROBLEMA	21
5. OBJETIVO GENERAL.....	23
6. OBJETIVOS ESPECÍFICOS	24
7. JUSTIFICACIÓN Y METODOLOGÍA A IMPLEMENTAR	25
8. RESULTADOS DEL MODELO.....	36
9. DISCUSIÓN DE LOS RESULTADOS Y PROPUESTA DE SOLUCIÓN	46
10. CONCLUSIONES.....	47
11. RECOMENDACIONES	49
12. REFERENCIAS BIBLIOGRÁFICAS	51
13. ANEXOS	54

ÍNDICE DEL TABLAS

Tabla 1 Descripción de variables	25
Tabla 2 Descripción de los grupos de productos	29
Tabla 3 Predicciones de la variable Grupo.....	42

ÍNDICE DEL FIGURAS

Figura 1	Matriz de variables.....	25
Figura 2	Heatmap de las variables	28
Figura 3	Ventas Hojaldre diarias, mensuales y anual	31
Figura 4	Ventas Postres diario, mensual y anual.....	32
Figura 5	Ventas Pan diario, mensual y anual	33
Figura 6	Ventas Pizza diario, mensual y anual.....	34
Figura 7	Productos de temporada en difuntos.....	35
Figura 8	Resultado de modelos para el Grupo HO.....	37
Figura 9	Resultado de modelos para el Grupo PO	38
Figura 10	Resultado de modelos para el Grupo PA	39
Figura 11	Resultado de modelos para el Grupo PI.....	40
Figura 12	Resultado de modelos para el Grupo FI	41
Figura 13	Resultados de errores obtenidos	43

1. INTRODUCCIÓN

La gestión eficiente de las ventas de productos es un desafío crítico para las empresas, especialmente para aquellas que operan en mercados altamente competitivos y dinámicos. En este contexto, la implementación de técnicas avanzadas de análisis de datos y predicción se ha convertido en una necesidad imperiosa para mejorar la toma de decisiones y optimizar los procesos comerciales. Los algoritmos de ensemble en el ámbito del aprendizaje automático han emergido como una solución prometedora, ofreciendo mejoras significativas en la precisión predictiva y la robustez de los modelos utilizados para el análisis de bases de datos de ventas.

Los algoritmos de ensemble combinan múltiples modelos de aprendizaje automático para crear un sistema más fuerte y preciso que cualquier modelo individual. Esta característica es particularmente ventajosa cuando se trata de predecir la demanda de productos, un aspecto crucial para la planificación y gestión de inventarios en una base de datos de ventas. A través de la combinación de diferentes modelos, los algoritmos de ensemble pueden capturar de manera más efectiva los patrones subyacentes en los datos, reduciendo así los errores de predicción. Breiman (2001) destaca que los enfoques de ensemble, como Random Forests y Gradient Boosting, mejoran la precisión al mitigar las limitaciones de los modelos individuales, logrando resultados más robustos y fiables.

Además de la mejora en la precisión, los algoritmos de ensemble también son efectivos en la reducción del sobreajuste, un problema común en el aprendizaje automático donde un modelo puede ajustar demasiado sus

predicciones a los datos de entrenamiento, perdiendo la capacidad de generalizar bien con datos nuevos. Al combinar modelos con diferentes sesgos y varianzas, los algoritmos de ensemble logran un equilibrio que minimiza el riesgo de sobreajuste. Dietterich (2000) argumenta que esta característica es crítica en el análisis de bases de datos de ventas, donde los datos suelen contener tanto patrones complejos como ruido. La capacidad de estos algoritmos para discernir entre la señal y el ruido permite una modelación más precisa y confiable.

La robustez frente a datos ruidosos o atípicos es otra ventaja clave de los algoritmos de ensemble. En una base de datos de ventas, los datos pueden verse influenciados por factores externos impredecibles, como cambios en las tendencias del mercado, eventos estacionales, o incluso situaciones excepcionales como desastres naturales o crisis económicas. Los errores de un modelo individual pueden ser compensados por otros modelos en el ensemble, proporcionando una mayor estabilidad en las predicciones. Zhou (2012) resalta que esta robustez es invaluable en contextos donde la precisión y la fiabilidad de las predicciones pueden impactar significativamente en la toma de decisiones estratégicas.

Otra ventaja notable de los algoritmos de ensemble es su capacidad para capturar relaciones complejas entre variables. En una base de datos de ventas, las relaciones entre diferentes variables pueden ser no lineales y difíciles de modelar con un solo algoritmo. Al combinar diferentes tipos de modelos, incluidos modelos lineales y no lineales, los algoritmos de ensemble pueden representar mejor estas relaciones complejas. Kotsiantis et al. (2007) subrayan que esta

capacidad es esencial para desarrollar modelos predictivos que reflejen con mayor precisión la realidad de los datos, mejorando así la calidad de las decisiones basadas en estos modelos.

Finalmente, la adaptabilidad de los algoritmos de ensemble es crucial en un entorno dinámico como el de las ventas de productos. Las tendencias y patrones de ventas pueden cambiar con el tiempo, y los modelos predictivos deben ser capaces de adaptarse rápidamente a estos cambios. Los algoritmos de ensemble pueden ser actualizados fácilmente con nuevos datos o reentrenados con diferentes configuraciones, lo que asegura que las predicciones se mantengan precisas y relevantes. Sagi y Rokach (2018) enfatizan que esta flexibilidad es fundamental para las empresas que buscan mantener una ventaja competitiva en un mercado en constante evolución.

En resumen, la implementación de algoritmos de ensemble en una base de datos de ventas ofrece múltiples beneficios que incluyen una mejora significativa en la precisión predictiva, reducción del sobreajuste, robustez frente a datos ruidosos, capacidad para capturar relaciones complejas entre variables y adaptabilidad a cambios en los datos. Este proyecto tiene como objetivo explorar y facilitar la aplicación de estos algoritmos en el contexto de la gestión de ventas, proporcionando a las organizaciones herramientas efectivas para mejorar su nivel de predicción y, por ende, su rendimiento comercial.

2. REVISIÓN DE LITERATURA

La gestión de abastecimiento e inventarios es una de las principales actividades dentro de las organizaciones, sin embargo, es uno de los puntos donde se presentan mayores oportunidades de mejora, pues principalmente las pequeñas y medianas empresas tienen dificultades en conocer los abastecimientos óptimos y se ven inmersas en problemas como sobre abastecimiento o falta de abastecimiento (Carreño, Amaya, Ruíz y Tiboche, 2019).

Los procesos de control de inventarios y producción para la mayor parte de negocios en el grupo de PYMES, se manejan con base en la experiencia basándose en técnicas cualitativas sin que tengamos mayor injerencia de procesos técnicos o estadísticos, lo cual evidencia una tendencia a no manejar datos exactos de niveles de ventas y pérdidas que le permita medir los niveles óptimos de producto a generar, lo cual puede ser mejorado en gran medida, siempre y cuando las PYMES se orienten a la toma de decisiones basadas en datos (Gutiérrez y Rodríguez, 2008).

Al respecto, Coopers y Lybrand (1997), indican que existen muchas confusiones al considerar estos procesos pues “aunque los componentes son aplicables a todas las empresas, las pequeñas y medianas pueden implantarlos de forma distinta que las grandes. Aunque sus sistemas de control pueden ser menos formales y estructurados, una pequeña empresa también puede tener un control interno eficaz; los componentes se derivan en; 1) Entorno de control; 2) Evaluación de los riesgos; 3) Actividades de control; 4) Información y comunicación y 5) Supervisión”.

Una de las opciones mejor adaptables es conocer los niveles de ventas que puedan tener las empresas, si bien es cierto los inventarios pueden no estar normalizados o registrados, las ventas son un registro que se exige normativamente a las PYMES en su mayoría. Pavón et. al. (2019) indica que es necesario llevar un registro detallado de los ingresos de ventas y los costos de los bienes vendidos, así como mantener un registro completo de los activos y pasivos de la empresa. Además, los gerentes y empleados requieren información contable más detallada que la que se presenta en los estados financieros, especialmente al tratar con aspectos como las cuentas por cobrar a los clientes.

Con los registros antes indicados podremos iniciar y tener información para poder aplicar modelos que permitan predecir de cierta forma los niveles de producción necesarios para evitar los problemas de abastecimiento y sobre abastecimiento en la empresa que tomaremos como punto de investigación.

En esta ejemplificación de proyecto, se utilizará una base de datos del nivel de ventas de Panificadora México, para poder utilizar estos datos a través de métodos estadísticos y predictivos proporcionados por la ciencia de datos, que permita poder proyectar los niveles necesarios de inventario y abastecimiento de producto dentro de esta empresa que pertenece al sector de PYME.

Al investigar cómo poder manejar de mejor forma un sistema de inventario aplicando ciencia de datos se contempló al modelo ensemble que es una técnica de aprendizaje automático que combina múltiples modelos de predicción para mejorar la precisión y el rendimiento general del sistema. En lugar de depender de un solo modelo para realizar predicciones, el modelo ensemble utiliza la

sabiduría colectiva de varios modelos más simples para producir predicciones más sólidas y generalizables (Lovato & Carvalho, 2021).

Estas oportunidades de mejora pueden ser potenciadas a través de la adaptación de las organizaciones a la tecnología considerando la revolución tecnológica y la orientación a datos informáticos, pues esto permite que las PYMES puedan tomar decisiones en cuanto a las grandes cantidades de información (Muñoz-Hernández, m et. al, 2016).

A partir de ello es importante considerar que los Sistemas de Información hoy en día se han vuelto un aliado estratégico de las organizaciones, y más aún en las PYMES, donde en la mayoría de las ocasiones cuentan con información que no se logra procesar de forma adecuada, cómo Roo-Huerta & Boscán-Romero (2012) indican “es el almacenaje de diferentes temas o información para ser analizada de forma conjunta o independiente y facilita efectuar consultas, análisis y presentación de información concreta”.

Estos sistemas de información consolidan datos y las organizaciones pueden tomar decisiones, después de procesarlos, al respecto, esto es lo que hoy en día se conoce como inteligencia de negocios. Esto busca que los procesos internos puedan gestionarse de una forma más ágil y que, procesos como el control de inventarios, abastecimiento, entre otros, puedan tener una objetividad de gestión, simplemente gestionando e infiriendo de los datos que las organizaciones ya las tienen establecidas (Coronel, Morris & Rob, 2011).

Con la implementación de inteligencia de negocios, se acompañan varias estrategias y herramientas que permiten procesar información y potenciar procesos como el concepto de ensemble que ha sido ampliamente estudiado y

utilizado en la comunidad de aprendizaje automático. Algunos autores relevantes describen y discuten el modelo ensemble:

Bishop (2016) menciona que el concepto de ensemble combina las predicciones de varios modelos de aprendizaje débil para mejorar la capacidad predictiva. Esta estrategia puede generar modelos más precisos que los modelos individuales por separado.

Zhi-Hua Zhou (2012) explica que el modelo ensemble es una estrategia poderosa para mejorar la precisión predictiva en el aprendizaje automático. Al combinar múltiples modelos, logra una precisión mayor que la de cualquier modelo individual. Esta técnica ha sido exitosa en diversas aplicaciones de aprendizaje automático y minería de datos.

Caruana y Niculescu-Mizil (2006) argumentan que el ensemble, al unir varios modelos de aprendizaje débil, es una técnica comprobada para mejorar la capacidad predictiva en diferentes tipos de tareas de aprendizaje supervisado. Esto se logra al capturar una gama más amplia de patrones en los datos, lo que mejora la capacidad de generalización del modelo.

Tarrés (2019) menciona que los ensemble combinan múltiples algoritmos para obtener una predicción más precisa que la que cada algoritmo individualmente podría proporcionar. Hay diferentes tipos de ensembles, como el hard voting y el soft voting.

Hard Voting: Este método toma las predicciones de cada modelo y elige la clase más común como resultado del ensemble. En caso de empate, se prioriza el modelo con la mejor precisión.

Soft Voting: Este método considera las probabilidades de cada clase en lugar de las predicciones. Se calcula la media de las probabilidades y se elige la clase con la media más alta. También se puede asignar pesos a las probabilidades de los diferentes modelos para dar más importancia a ciertos modelos.

Para construir un ensemble, es necesario tener diferentes clasificadores. Algunos de los clasificadores más utilizados y efectivos incluyen:

SVM (Support Vector Machines): Un método que busca encontrar un hiperplano óptimo que separe los datos de las diferentes clases.

Decision Tree: Un método que crea un modelo predictivo basado en reglas de decisión aprendidas de las características de los datos.

KNeighbors: Un método que predice las etiquetas basándose en las muestras más cercanas al nuevo punto.

Random Forest: Un conjunto de árboles de decisión que se ajustan en subconjuntos del conjunto de datos y se promedian para mejorar la precisión predictiva.

Naive Bayes: Un conjunto de métodos de aprendizaje supervisado que aplican el teorema de Bayes asumiendo independencia entre las características.

Estos clasificadores son ampliamente utilizados debido a su facilidad de implementación y sus buenos resultados en una variedad de problemas de aprendizaje automático.

Para el proyecto ha sido contemplado hacer uso de modelos estadísticos como son Random Forest y XGB Regressor ya que nos permiten manejar bases de datos complejos y grandes lo que puede traducirse en predicciones más

precisas sobre la demanda futura del inventario. Así mismo, ambos modelos pueden identificar características relevantes que afectan el inventario como la estacionalidad de la demanda, tendencias históricas y a comprender mejor los patrones detrás de la demanda y la gestión del inventario. En conclusión nos ayuda a reducir los costos asociados con el almacenamiento y el exceso de inventario.

3. IDENTIFICACIÓN DEL OBJETO DE ESTUDIO

La implementación de algoritmos de ensemble en una base de datos de ventas de productos llega a ser muy beneficiosa por varias razones, la primera es poder mejorar la precisión predictiva ya que los algoritmos de ensemble combinan múltiples modelos de aprendizaje automático para mejorar la precisión predictiva en comparación con un único modelo. Esto puede ser especialmente útil en una base de datos de ventas donde se necesita predecir la demanda de productos con precisión. Segundo, se puede aplicar una reducción del sobreajuste al combinar modelos que tienen diferentes sesgos y varianzas. Esto puede ser crítico en una base de datos de ventas donde se pueden encontrar patrones complejos, pero también ruido en los datos. Tercero, permite crear robustez frente a datos ruidosos o atípicos, ya que los errores de un modelo pueden ser compensados por otros modelos en el ensemble. En una base de datos de ventas, donde los datos pueden ser afectados por factores externos impredecibles, esta robustez es valiosa. Cuarto, permite la captura de complejas relaciones entre variables, esto al combinar diferentes modelos, incluidos modelos lineales y no lineales. Esto es útil en una base de datos de ventas donde las relaciones entre variables pueden ser no lineales y difíciles de modelar con un solo algoritmo. Y, por último, se adapta a los cambios en los datos, ya que pueden ser actualizados fácilmente con nuevos datos o reentrenados con diferentes configuraciones. En una base de datos de ventas, donde las tendencias y patrones pueden cambiar con el tiempo, esta capacidad de adaptación es esencial.

Es por ello que el presente proyecto permite facilitar la información de cómo aplicar los diferentes algoritmos de ensemble en una base de datos de ventas, y, por ende, permitir a las organizaciones mejorar su nivel de predicción.

4. PLANTEAMIENTO DEL PROBLEMA

Gestionar las ventas de productos es fundamental para las pequeñas y medianas empresas, pero a menudo resulta desafiante y genera preocupaciones para los dueños y gerentes. Muchas empresas todavía dependen de métodos empíricos o incluso de hojas de cálculo manuales para monitorear las ventas diarias, lo que limita su capacidad para tomar decisiones informadas.

Sin embargo, contar con una visión objetiva y basada en datos sobre las ventas y la participación de productos puede proporcionar ventajas competitivas. Esto nos permite realizar análisis detallados que pueden potenciar las ventas y mejorar la eficiencia del negocio.

En este contexto, se estudia el caso de una panificadora tradicional en Quito, con el objetivo de explorar las ventajas de sistematizar el seguimiento de ventas y de implementar modelos de machine learning para predecir los niveles de ventas y realizar proyecciones. Estas proyecciones son cruciales para relacionarlas con los niveles de producción necesarios en el negocio.

El uso de modelos de machine learning en la producción de alimentos ha ganado relevancia en años recientes debido a su capacidad para mejorar la eficiencia y la rentabilidad. Especialmente para pequeñas empresas de alimentos, estas tecnologías pueden optimizar procesos y apoyar la toma de decisiones basadas en datos.

Panificadora México, una pequeña empresa de panificación, ha dependido principalmente de métodos empíricos para gestionar sus ventas y producción. La falta de un sistema estructurado ha causado dificultades en la

planificación de la producción y la proyección de ventas, lo que ha llevado a problemas de inventario y eficiencia operativa.

Para abordar estos desafíos, Panificadora México planea implementar un sistema de gestión de datos que integre información de ventas, inventario y producción en tiempo real. Además, se desarrollarán modelos de machine learning para analizar estos datos y generar predicciones precisas de la demanda futura. Estas predicciones serán fundamentales para planificar la producción y gestionar el inventario de manera eficiente.

La implementación de estos modelos y sistemas de gestión de datos no solo mejorará los procesos operativos, sino que también reducirá los costos asociados con el exceso de inventario y la pérdida de ventas. Además, permitirá a la empresa introducir nuevos productos de manera más efectiva, basándose en datos precisos sobre las preferencias del cliente y las tendencias del mercado.

5. OBJETIVO GENERAL

Diseñar un sistema de predicción de demanda de productos utilizando algoritmos de ensemble en una base de datos de ventas, con el fin de mejorar la precisión predictiva y optimizar la gestión de inventario.

6. OBJETIVOS ESPECÍFICOS

Realizar un análisis exhaustivo de la base de datos de ventas para comprender la estructura de los datos, identificar posibles fuentes de ruido o sesgo, y determinar las variables relevantes para la predicción de la demanda.

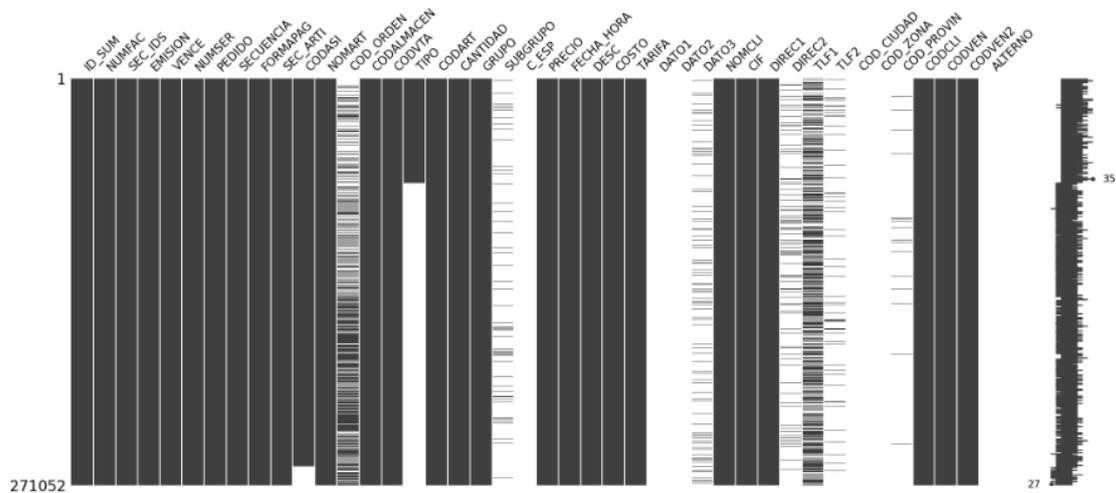
Investigar y evaluar el modelo de algoritmo de ensemble Random Forest y XGB Regressor para determinar su utilidad y aplicabilidad en el problema de predicción de demanda de Panificadora México.

Construir un modelo para la predicción de inventario más útil con el que se pueda evaluar su utilidad y eficacia en la planificación y la toma de decisiones comerciales.

7. JUSTIFICACIÓN Y METODOLOGÍA A IMPLEMENTAR

La base de datos obtenida para realizar el siguiente proyecto proviene de la empresa Panificadora México. Dado que el puesto de trabajo en el que se labora es de Gerente Administrativo Comercial, ha permitido observar la necesidad de crear procesos que permitan minimizar los costos, mejorar los inventarios y crear proyecciones que permitan obtener la cantidad recomendable de producción. Es por ello que se recolectó la información de las ventas del año 2023 que proporcionará la información necesaria para el respectivo análisis.

Figura 1
Matriz de variables



Fuente: Elaboración propia

A continuación, se detallan las variables del dataset:

Tabla 1
Descripción de variables

Variabes	Fuente De Datos	Naturaleza/ Tipo De Datos	Descripción
Id_Sum	Base De Datos	Numérico	Identificador Único De La Suma

Numfac	Base De Datos	Numérico	Número De Factura
Sec_Ids	Base De Datos	Numérico	Sección De Identificación
Emision	Base De Datos	Fecha/Hora	Fecha De Emisión De La Factura
Vence	Base De Datos	Fecha/Hora	Fecha De Vencimiento De La Factura
Numser	Base De Datos	Numérico	Número De Serie
Pedido	Base De Datos	Numérico	Número De Pedido
Secuencia	Base De Datos	Numérico	Secuencia De La Factura
Formapag	Base De Datos	Categorico	Forma De Pago De La Factura
Sec_Arti	Base De Datos	Numérico	Sección Del Artículo
Codasi	Base De Datos	Numérico	Código De Asignación
Nomart	Base De Datos	Texto	Nombre Del Artículo
Cod_Orden	Base De Datos	Numérico	Código De Orden
Codalmacen	Base De Datos	Numérico	Código De Almacén
Codvta	Base De Datos	Numérico	Código De Venta
Tipo	Base De Datos	Categorico	Tipo De Factura
Codart	Base De Datos	Numérico	Código Del Artículo
Cantidad	Base De Datos	Numérico	Cantidad De Artículo
Grupo	Base De Datos	Texto	Grupo De Artículo
Subgrupo	Base De Datos	Texto	Subgrupo Del Artículo
C_Esp	Base De Datos	Numérico	Código Especial
Precio	Base De Datos	Numérico	Precio Del Artículo
Fecha_Hora	Base De Datos	Fecha/Hora	Fecha Y Hora De La Transacción
Desc	Base De Datos	Numérico	Descuento Aplicado
Costo	Base De Datos	Numérico	Costo Del Artículo
Tarifa	Base De Datos	Numérico	Tarifa De La Factura

Dato1	Base De Datos	Texto	Datos Adicionales 1
Dato2	Base De Datos	Texto	Datos Adicionales 2
Dato3	Base De Datos	Texto	Datos Adicionales 3
Nomcli	Base De Datos	Texto	Nombre Del Cliente
Cif	Base De Datos	Texto	Código De Identificación Fiscal Del Cliente
Direc1	Base De Datos	Texto	Dirección Del Cliente 1
Direc2	Base De Datos	Texto	Dirección Del Cliente 2
Tlf1	Base De Datos	Texto	Teléfono Del Cliente 1
Tlf2	Base De Datos	Texto	Teléfono Del Cliente 2
Cod_Ciudad	Base De Datos	Numérico	Código De La Ciudad Del Cliente
Cod_Zona	Base De Datos	Numérico	Código De La Zona Del Cliente
Cod_Provin	Base De Datos	Numérico	Código De La Provincia Del Cliente
Codcli	Base De Datos	Numérico	Código Del Cliente
Codven	Base De Datos	Numérico	Código Del Vendedor Principal
Codven2	Base De Datos	Numérico	Código Del Segundo Vendedor
Alternos	Base De Datos	Numérico	Código De Artículo Alternativo

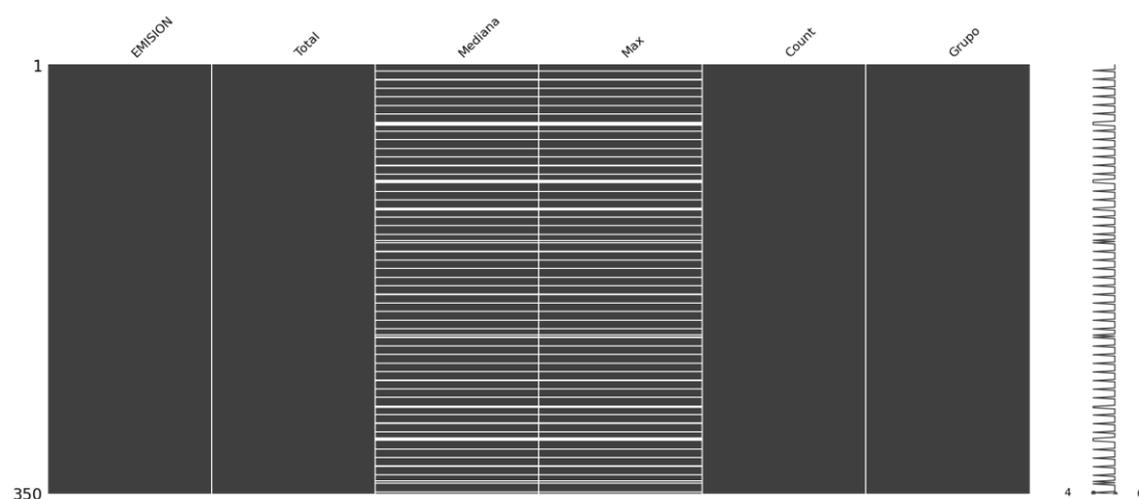
Fuente: Elaboración propia.

Para el uso de las variables a implementar en el modelo de Random Forest y XGB Regressor, se ha hecho uso de un enfoque univariado ya que permite que el análisis de series temporales sea utilizado cuando solo se considera una única variable en el análisis, como la demanda de un producto en un sistema de inventario. Este enfoque es adecuado cuando las variables son estáticas, es decir, no cambian con el tiempo, o cuando se asume que los efectos de otras variables son insignificantes para el propósito del análisis. Es así que el interés de este proyecto es predecir la demanda de un producto

específico en función del tiempo, sin considerar otras variables que puedan afectar esta demanda, como el precio, la temporada, la promoción, etc., entonces un enfoque univariado es el apropiado. Sin embargo, si se desea tener en cuenta múltiples variables que pueden influir en la demanda, entonces se requeriría un enfoque multivariado en el análisis de series temporales (Box et. al, 2015).

En este caso hemos definido sólo las variables de Emisión, Grupo y Cantidad para el estudio de los modelos.

Figura 2
Heatmap de las variables



Fuente: Elaboración propia

El heatmap de la figura 2 revela que el conjunto de datos tiene problemas significativos de valores faltantes en las columnas "Mediana" y "Max", mientras que las demás columnas ("EMISION", "Total", "Count" y "Grupo") están completas. Este tipo de visualización es fundamental, ya que permite identificar rápidamente la distribución y la frecuencia de los valores faltantes. Esto es crucial para el preprocesamiento de datos, facilitando la imputación de valores faltantes y asegurando que los datos sean adecuados para el análisis estadístico o la

modelización posterior. En el caso de este dataset, es importante considerar que se maneja la columna Count, ya que es importante contar con los totales de ventas, más no con promedios ni conocer los valores máximos.

En Panificadora México existen 591 productos que se comercializan y a su vez estos pertenecen a grupos de producción, por ejemplo, los productos milhoja, cachitos de jamón, empanadas etc, pertenecen al grupo de HO u Hojaldre puesto que la producción para estos se genera a partir de la elaboración de una sola masa. De igual manera se manejan las producciones para todos los grupos de productos los cuales detallamos a continuación:

Tabla 2
Descripción de los grupos de productos

Grupo	Cantidad Vendida 2023	Descripción
HO	73789	Productos elaborados de masa de hojaldre
PO	62680	Productos elaborados con mezclas de postres
PA	50061	Productos de masa para pan
PI	20950	Productos de masa de pizza
TI	18476	Productos complementarios (bebidas entre otros)
TO	17098	Productos que utilizan masa de tortas
GA	9799	Productos de galletería
CH	5762	Productos de chocolate
FI	5234	Productos de temporada en noviembre (día de difuntos)

DE	4663	Productos utilizados para decoración
VA	1939	Utensilios para envolver los productos vendidos
RE	572	Productos de repostería
T1	29	Productos obsequiados de muestra

Fuente: Elaboración propia

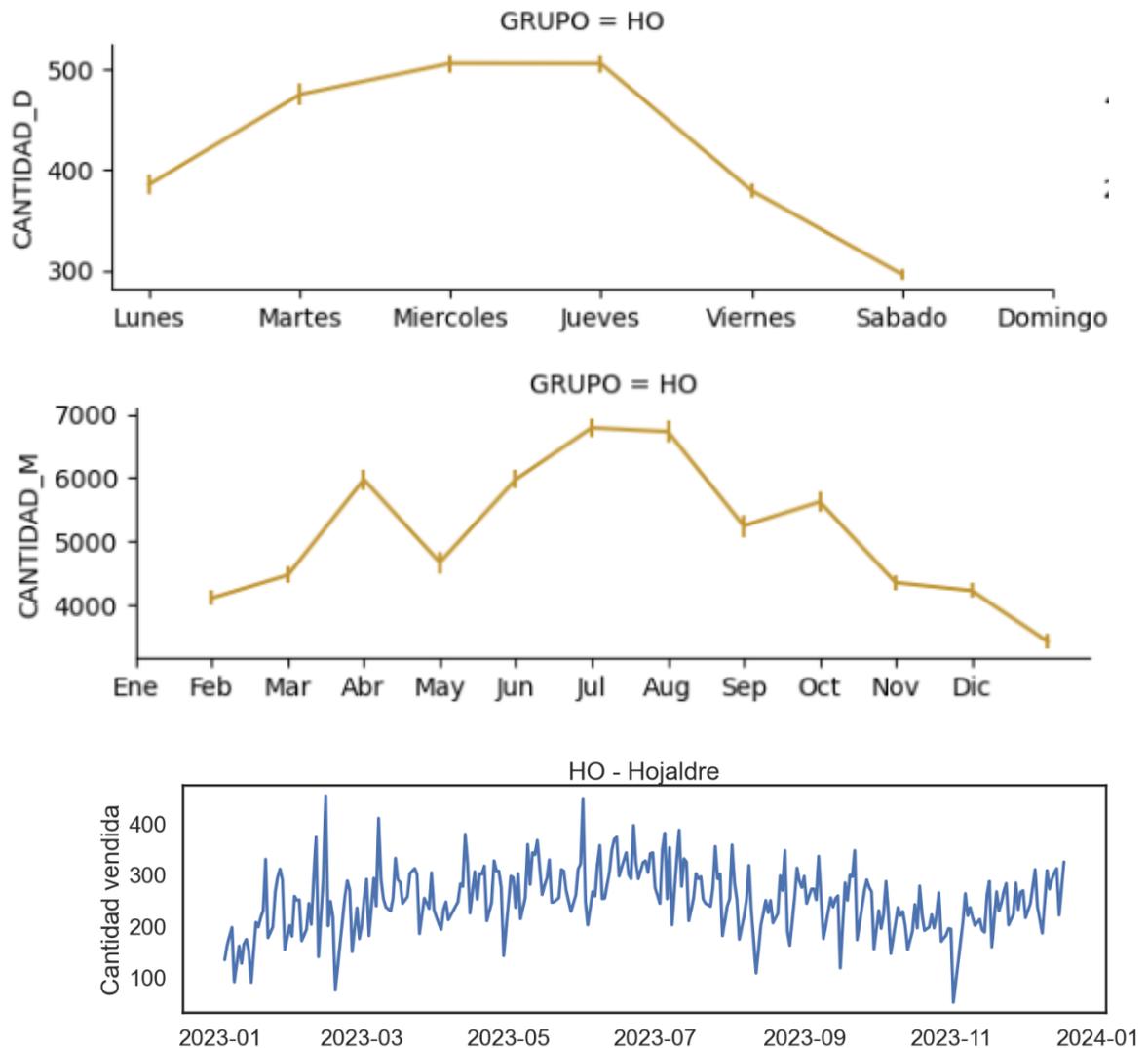
Es importante entender que se tomó como variable a grupos ya que en ellos abarcan subgrupos elaborados con los mismos materiales pero vendidos en diferentes presentaciones, así que para poder conocer de mejor forma cómo manejar de mejor forma el inventario se tomará como referencia a cada grupo de productos.

Luego de conocer el dataset y las variables que lo conforman procedemos a escoger el modelo que se usará para el respectivo análisis, en este caso el método estadístico seleccionado para este análisis es Random Forest, una técnica de aprendizaje automático ampliamente utilizada que aplica algoritmos de tipo ensemble basados en la construcción de múltiples árboles de decisión, este enfoque ofrece una mayor estabilidad y precisión en comparación con un solo árbol de decisión, lo que lo convierte en una elección ideal para nuestro análisis de predicción de ventas en Panificadora México.

Para la aplicación de este modelo necesitamos conocer la relación que existen entre las variables predictoras y la objetivo, en este caso nuestra variable objetivo son los Grupos y la Emisión y la variable predictora es la cantidad, aplicando la librería seaborn podemos obtener la relación entre estas variables con la función de relplot se pudo crear las gráficas que determinan la evolución de las ventas a lo largo del año 2023.

A continuación, observamos los gráficos de los grupos HO, PO, PA, PI y FI, los otros gráficos de los demás grupos se lo puede observar de igual manera.

Figura 3
Ventas Hojaldre diarias, mensuales y anual

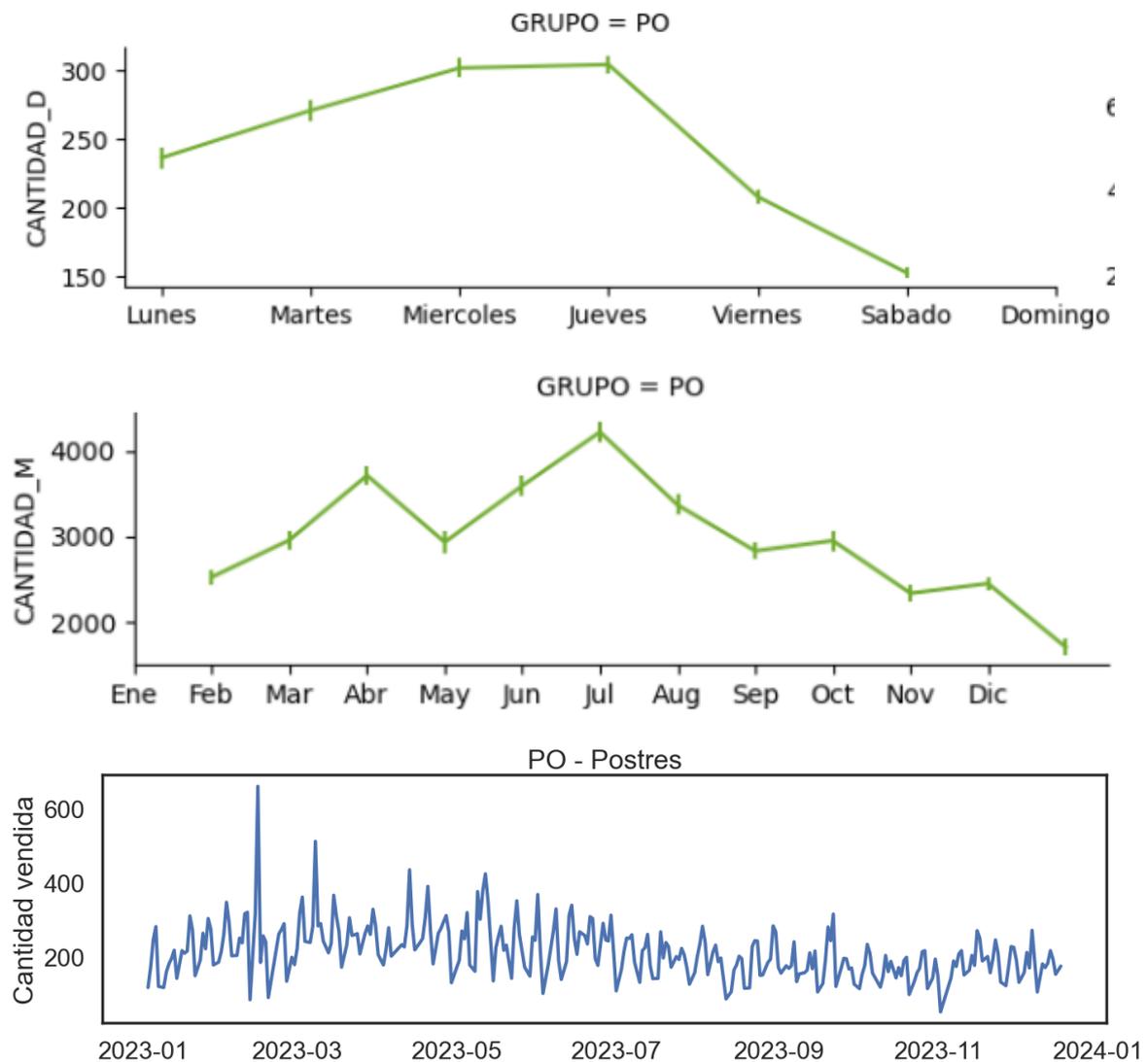


Fuente: Elaboración propia

Se puede observar en la figura 3 cómo han ido evolucionando las ventas del producto HO a lo largo de los días, meses y año, se observa la tendencia diaria como los días jueves son picos altos esto se debe a que las ventas ese día se toman pedidos para entregas posteriores, a su vez se nota que la tendencia es casi lineal sin muchos picos altos excepto por los meses de abril, julio y agosto

que son fechas en donde hay mayor demanda por temas de grados, bautizos, confirmaciones. Y así mismo se observa una bajada en el mes de noviembre ya que en esas fechas el producto estrella son las guaguas de pan por la temporada de los difuntos.

Figura 4
Ventas Postres diario, mensual y anual

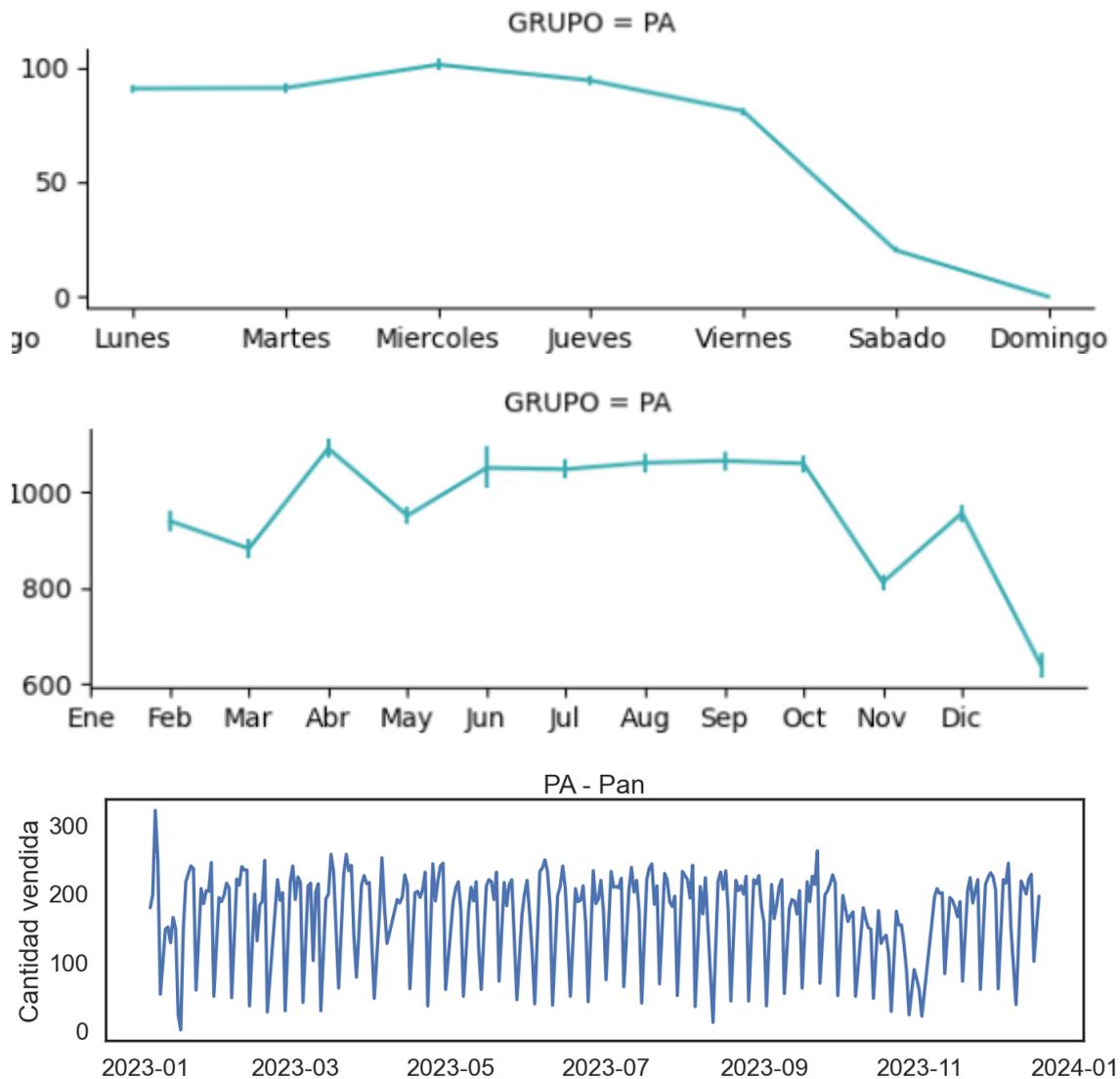


Fuente: Elaboración propia

En la figura 4 se observa igualmente una tendencia lineal del mismo modo en el mes de marzo hay un alza de ventas debido a la festividad de Semana Santa que se realiza en esas fechas, de igual forma que la figura 3 el por que se

elabora más los días es jueves es porque lo realizan bajo pedido y los pedidos se receptan ese día.

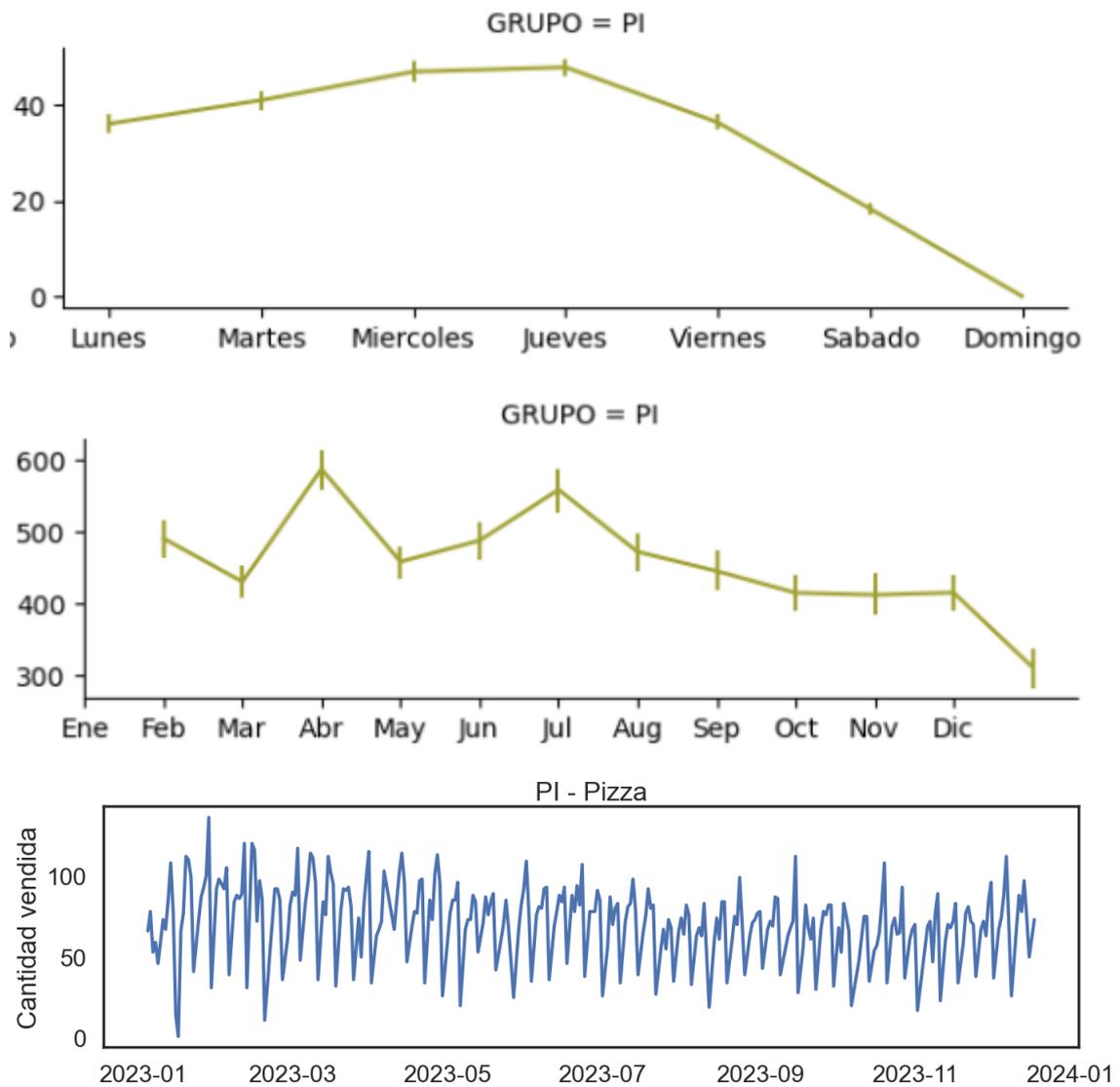
Figura 5
Ventas Pan diario, mensual y anual



Fuente: Elaboración propia

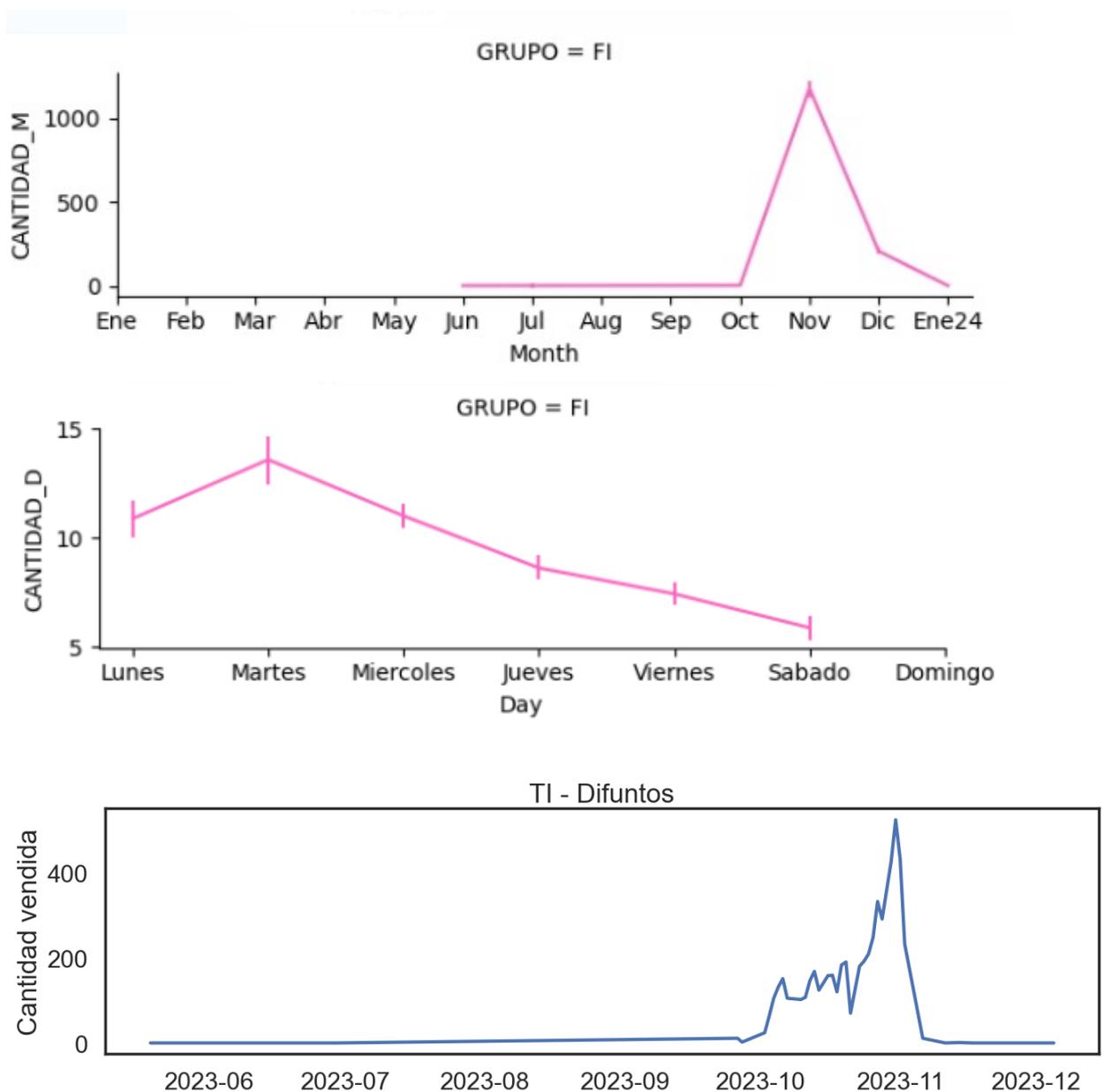
En el caso de la figura 5 de pan vemos una tendencia muy lineal, en este caso el producto lo adquieren por días bastante y por otros no son tan demandados.

Figura 6
Ventas Pizza diario, mensual y anual



En el caso de la figura 6 igualmente la tendencia es muy lineal ya que la demanda del producto es por temporadas.

Figura 7
Productos de temporada en difuntos



Fuente: Elaboración propia

En la figura 7 es bastante inusual su demanda ya que desde el mes de octubre a noviembre observamos un alza de ventas, esto se debe a que los productos que se ofrecen son netamente direccionados para la festividad del día de los difuntos. Esto permite determinar que solo en estas fechas se obtiene una gran demanda de productos.

8. RESULTADOS DEL MODELO

Una vez aplicado el modelo de random forest, procedemos a tener los resultados de las regresiones y predicciones para conocer las cantidades óptimas de producto, para ello a continuación describiremos los resultados a partir del entrenamiento del modelo y posterior análisis de los datos obtenidos.

Este análisis predictivo se centra en utilizar el modelo Random Forest para predecir los niveles de demanda de cada grupo de productos en función de las variables seleccionadas. Esto permite anticipar y planificar de manera más efectiva su producción para satisfacer la demanda del mercado.

El procesamiento de las variables utilizadas, Grupo (grupo de productos), Count (número de productos vendidos), y EMISIÓN (fecha de venta). Los datos se segmentan en grupos y se calcula la media, Cantidad y máximo diario de Count.

Se aplican diferencias con retrasos de 1 y 2 ($Differences([1,2])$) para hacer la serie temporal más estacionaria. Esto elimina tendencias y patrones estacionales, mejorando la capacidad del modelo para capturar relaciones subyacentes (Hyndman & Athanasopoulos, 2018).

Para el entrenamiento del modelo se utiliza MLForecast con modelos de Random Forest y XGBoost. Se consideran lags (rezagos) de 4, 8 y 10 días, así como transformaciones de las características mediante medidas expansivas y medias móviles. Una vez aplicado lo anterior, el modelo se entrena en los datos históricos y se realizan predicciones para un horizonte temporal de 10 días. En la siguiente descripción de gráficos es importante conocer que se trabajara con dos modelos:

Random Forest Regressor línea verde

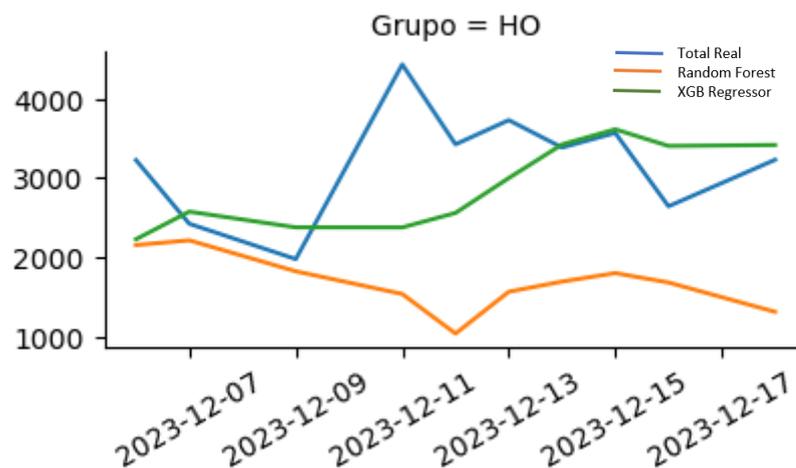
XGBRegressor - línea naranja

Valor actual - línea azul

Estos indicadores se detallan con el fin de compararlos e identificar cual nos refleja una mejor proyección para niveles de producción.

En el grupo de HO - Hojaldre podemos observar que existe una diferencia entre los modelos, donde en los días de mayor demanda tienden a tener una diferencia en la proyección de producción, ante ello se define que el modelo con mejor predicción sería XGB Regressor.

Figura 8
Resultado de modelos para el Grupo HO

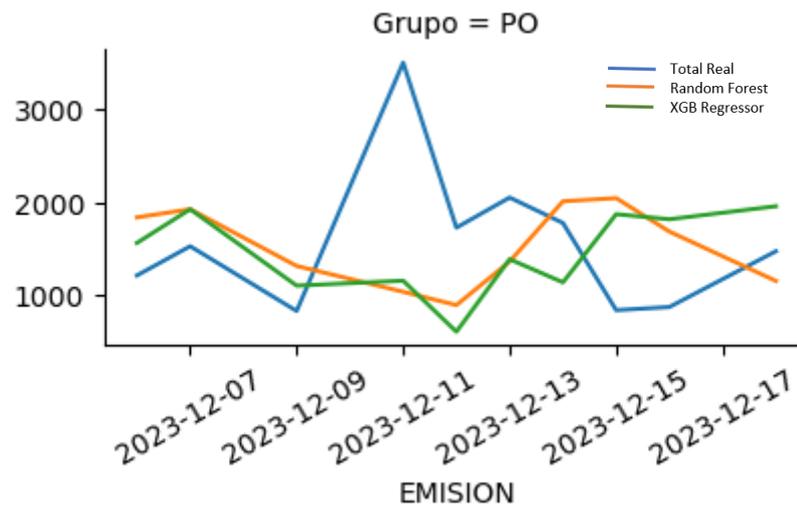


Fuente: Elaboración propia

En este grupo podemos definir que la producción en los días finales debe aumentar y tener mayor capacidad de reacción ante la demanda.

En el grupo de PO - Postres identificamos que los modelos tienen una tendencia muy similar de predicción por lo cual podríamos aplicar cualquiera de estos en este resultado.

Figura 9
Resultado de modelos para el Grupo PO

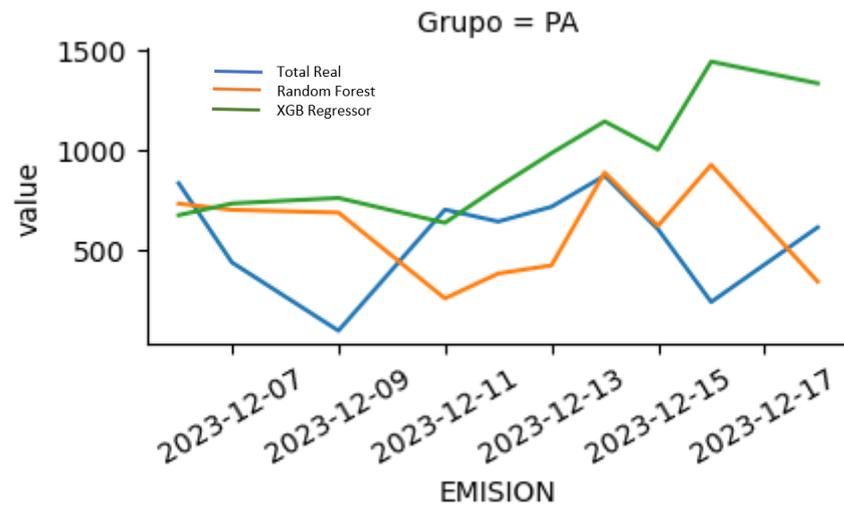


Fuente: Elaboración propia

Aunque existe un detalle importante a considerar, donde podemos ver que los días donde en el grupo HO teníamos mayor demanda, en este grupo descende, esto nos ayuda a identificar esta variabilidad y tomar acciones correctivas dentro de las unidades respectivas.

En el grupo de Pa - Panadería existe una notable diferencia entre los resultados de ambos modelos, donde en el modelo de Random Forest Regresor tiende a predecir niveles de producción a la baja mientras que el modelo XGB Regresor proyecta la predicción hacia mayor demanda y por ende mayor nivel de producción.

Figura 10
Resultado de modelos para el Grupo PA

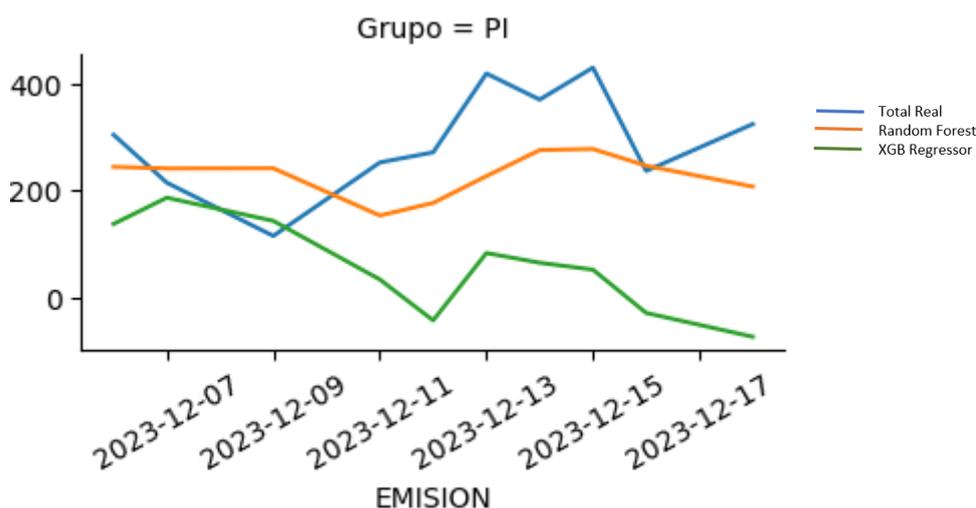


Fuente: Elaboración propia

La decisión de tomar el modelo XGB Regressor se basa en una tendencia más optimista para Panificadora México.

Dentro del grupo PI - Pizzas el panorama entre los dos modelos es distinto a la predicción anterior, donde el modelo Random Forest Regressor, indica un modelo más optimista en cuanto a la producción a realizar, mientras que el otro modelo presenta niveles de producción con una tendencia a producir menor cantidad.

Figura 11
Resultado de modelos para el Grupo PI

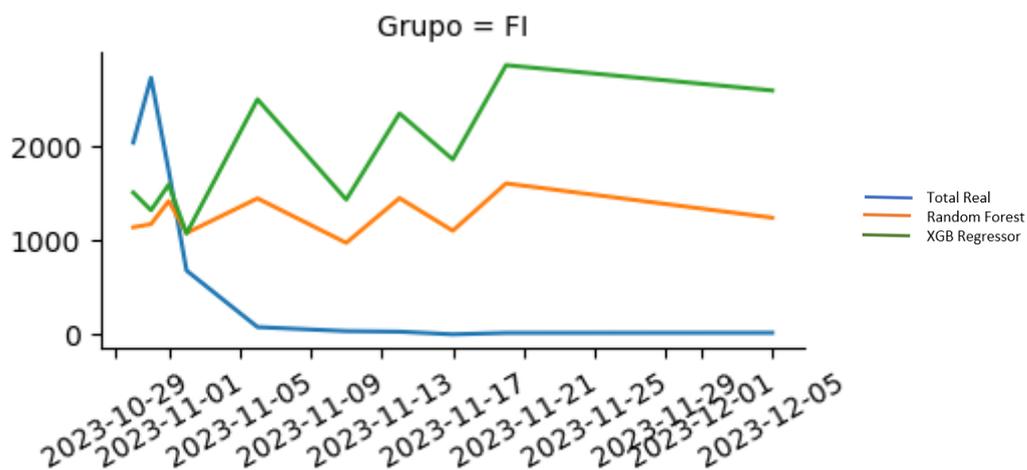


Fuente: Elaboración propia

Los resultados para este grupo de igual manera indican que al inicio del periodo existe mayor demanda que al final del periodo lo cual es un indicador que permite optimizar los niveles de producción.

En el grupo de FI - Productos de difuntos, inicialmente vemos que son productos estacionales, mismos que sólo se producen una vez al año y que a su vez, generan ingresos importantes en dicha fecha. En este caso el Modelo Random Forest Regressor presenta resultados con mayor optimismo en cuanto a la predicción de producción, e incluso se muestra una línea con una estabilidad en los niveles de demanda, lo que no sucede en el modelo XGB Regressor.

Figura 12
Resultado de modelos para el Grupo FI



Fuente: Elaboración propia

De lo contrario, la tendencia del modelo XGB Regressor adicionalmente muestra una tendencia a la baja en la demanda.

Hay una particularidad en todas las gráficas expuestas, donde se identifica que hay una baja en la demanda en los días viernes, sábado y domingos, pues al respecto es importante entender, que el giro de negocio permite que las personas separen sus pedidos de bocaditos y producto entre semana, donde los jueves tiende a tener una mayor demanda puesto que realizan los pedidos de producto para los eventos de los días viernes y sábado, por lo contrario, el día domingo no se abre el establecimiento.

El número de productos por grupo se ha definido en las predicciones conforme a los siguientes resultados:

Tabla 3
Predicciones de la variable Grupo

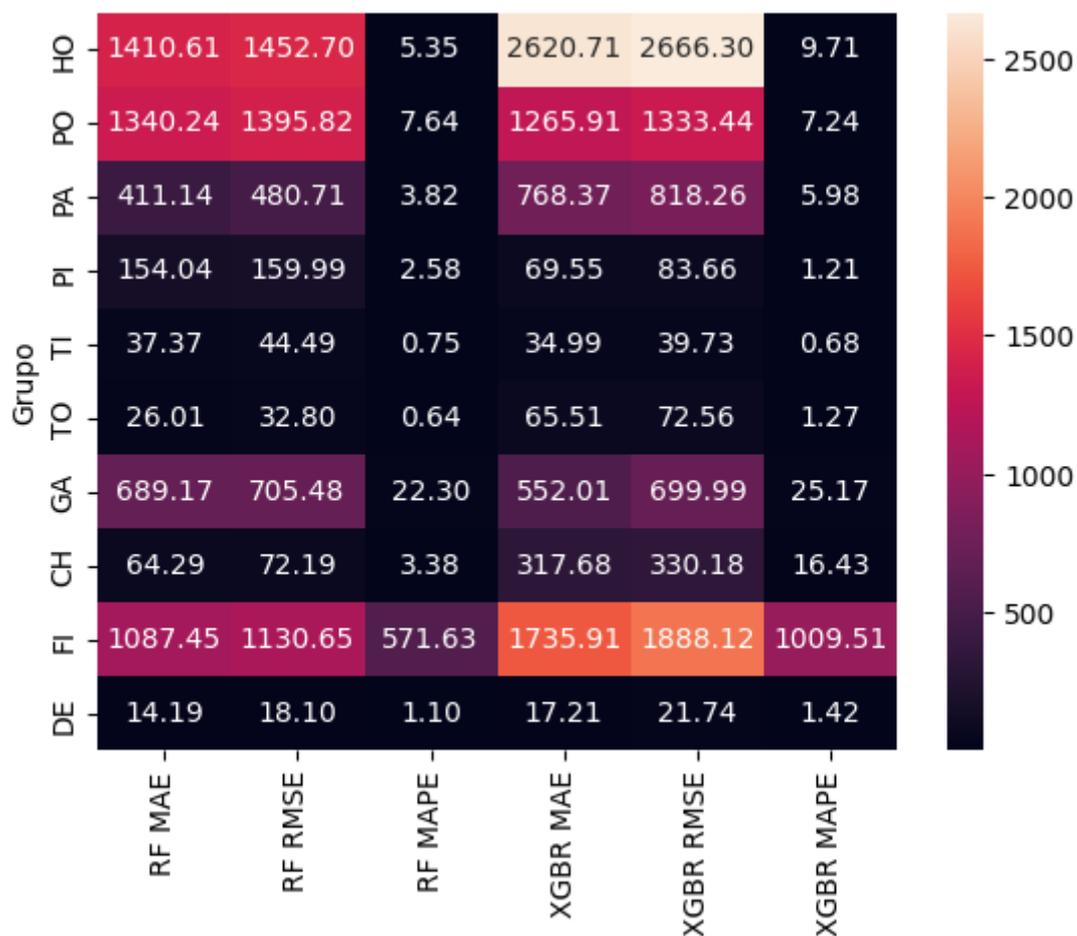
Número	Grupo	RandomForestRegressor	XGBRegressor
1	HO	3482.812109	38992.5117
2	GA	1245.488.086	14184.1377
3	PO	1169.291113	14011.5918
4	PI	4564.459961	36181.5503
5	CH	3506.800049	4630.4668
6	PA	1983.670166	1154.64966
7	TO	328.799988	675.286499
8	TI	287.060089	195.549713
9	DE	-16119.993	-14.910514
10	FI	-1094.689063	-18052.873

Fuente: Elaboración propia

Podemos identificar que las predicciones de ambos modelos para los 10 días de producción son muy similares, aunque el modelo de XGB Regressor tiene resultados con mayor optimismo que el Random Forest Regressor. Es importante indicar que estos resultados pueden cambiar conforme se vaya alimentando el modelo con los meses nuevos de ventas.

Para evaluar el desempeño de los modelos Random Forest Regressor (RF) y XGB Regressor (XGBR) en la predicción de las ventas de grupos de productos, hemos calculado varias métricas de error para cada uno de los mismos. A continuación, se presentan las características estadísticas de cada modelo:

Figura 13
Resultados de errores obtenidos



Fuente: Elaboración propia

Para determinar el modelo con mejor desempeño, se han analizado diversas métricas de error: el Mean Absolute Error (MAE), el Root Mean Squared Error (RMSE) y el Mean Absolute Percentage Error (MAPE).

Al ser los niveles de producción valores elevados con relación a los errores presentados en la figura 17, podemos definir que los valores establecidos como error no generan mayor impacto en cuanto a los niveles de producción, por ejemplo, en el grupo de hojaldre donde XGB Regressor tiene un valor de error de 2620 y el total de producción anual es de 73789, esto representa el 4% de

error en cuanto a los niveles totales de producción. Esta tendencia es similar en todos los grupos.

El modelo de proyección se realiza en base a 10 días considerando que es un tiempo prudente para generar las producciones por la razón principal que es la durabilidad de las masas en frío, aunque existen masas que tiene duraciones menores, en el giro de negocio se cuida la calidad y la rotación de las masas.

Mean Absolute Error (MAE):

El modelo Random Forest Regressor (RF) presenta generalmente valores más bajos de MAE en comparación con XGB Regressor (XGBR), excepto en los grupos PO y GA.

Root Mean Squared Error (RMSE):

RF también muestra valores menores de RMSE en la mayoría de los casos, indicando menor variabilidad en los errores.

Mean Absolute Percentage Error (MAPE):

RF tiene valores de MAPE considerablemente más bajos, especialmente en el grupo FI, donde el MAPE de XGBR es extremadamente alto.

Es importante definir la comparación para los grupos que hemos elegido en la explicación del modelo, por lo cual las describimos a continuación.

- En el grupo HO, RF tiene un MAE de 1410.61 y un MAPE de 5.35%, mientras que XGBR tiene un MAE de 2620.71 y un MAPE de 9.71%, indicando un mejor desempeño de RF.

- En el grupo PO, RF presenta un MAE de 1340.24 y un MAPE de 7.64%, mientras que XGBR tiene un MAE de 1265.91 y un MAPE de 7.24%. Aquí, ambos modelos muestran un desempeño similar, pero RF tiene un MAPE ligeramente más alto.
- En el grupo PA, RF muestra un MAE de 411.14 y un MAPE de 3.82%, mientras que XGBR tiene un MAE de 768.37 y un MAPE de 5.98%. En este caso, RF muestra un mejor desempeño con valores de error más bajos.
- En el grupo PI, RF tiene un MAE de 154.04 y un MAPE de 2.58%, mientras que XGBR muestra un MAE de 69.55 y un MAPE de 1.21%, donde XGBR tiene un mejor rendimiento.
- En el grupo FI, RF muestra un MAE de 1087.45 y un MAPE de 571.63%, mientras que XGBR tiene un MAE de 1735.91 y un MAPE de 1009.51%. Aunque ambos modelos tienen valores de error altos, RF muestra un mejor desempeño en este caso.

En general, el modelo Random Forest Regressor (RF) parece ser superior al XGB Regressor (XGBR) para la mayoría de los grupos de productos, mostrando menores valores de MAE y RMSE, así como valores de MAPE significativamente más bajos en la mayoría de los casos. Esto sugiere que RF proporciona predicciones más precisas y consistentes en el análisis de esta base de datos de ventas.

No obstante, aunque RF muestra un mejor rendimiento global, la elección final del modelo puede depender de factores específicos del contexto

empresarial y las características particulares de los datos. Se recomienda realizar una evaluación más profunda y pruebas adicionales para confirmar esta elección en entornos de producción.

9. DISCUSIÓN DE LOS RESULTADOS Y PROPUESTA DE SOLUCIÓN

Utilizar un modelo entrenado como Random Forest y XGB Regressor para predecir la demanda futura de productos basándose en datos históricos de ventas, ha permitido planificar de mejor forma las compras futuras y evitar escasez o excedentes de inventario, tal como los gráficos expuestos anteriormente pueden demostrar. Así mismo se ha podido ajustar los niveles de inventario en función de las predicciones de demanda, minimizando así los costos asociados con el exceso de inventario o la falta de existencias.

De igual forma, los resultados que arrojaron los modelos ayudó a identificar los productos que tienen una alta probabilidad de venta en un período de tiempo específico, lo que permite priorizar la reposición de inventario para los productos más populares y maximizar las ventas, a su vez que permite detectar anomalías como fluctuaciones inesperadas en las ventas.

Es por ello que como propuesta de solución para la gestión de inventarios se realizará lo siguiente:

- A. Optimizar los inventarios: se ajustará los niveles de inventarios en base a las predicciones de la demanda.
- B. Política de stock de seguridad: con ella se permitirá gestionar las fluctuaciones en la demanda y evitar rupturas de inventario, especialmente en una panificadora donde los productos son perecederos.

C. Rotación de productos: implementar el sistema FIFO para minimizar el desperdicio de productos perecederos

El sistema FIFO (First In, First Out) es un método de gestión de inventarios que asegura que los productos que se ingresan primero en el inventario sean también los primeros en salir. Este sistema es útil para gestionar productos perecederos de la panificadora, donde es crucial minimizar el desperdicio asegurando que los productos más antiguos se vendan antes de que se vuelvan obsoletos o caduquen. Por ello es necesario: primero, tener organización en el almacén con sus productos bien etiquetados y fechados, disponiendo siempre de estantería en donde se colocará los productos de acuerdo al orden de llegada, los más antiguos primero y los nuevos atrás. segundo, capacitar al personal sobre estas practicas y tener politicas de almacenamiento en donde se indique que los productos deben ser seleccionados y vendidos siguiendo el orden FIFO y tercero, tener un software para la gestión de inventarios que faciliten el sistema FIFO y a su vez implementar código de barras para facilitar la identificación y rastreo de productos.

Es importante tener en cuenta que no se puede implementar el sistema FIFO sin mejorar la precisión de las predicciones de demanda ya que calcular el stock óptimo necesario permitirá mantener el inventario dentro de los niveles deseados.

10. CONCLUSIONES

De acuerdo con los datos obtenidos se puede concluir lo siguiente.

- Analizar exhaustivamente un dataset de ventas permite desarrollar modelos predictivos precisos y efectivos para la demanda. Al comprender la estructura de los datos y las relaciones entre variables, y al identificar y tratar fuentes de ruido y sesgo, se pueden mejorar significativamente las predicciones de demanda.
- Conocer los diferentes métodos que estudia la Ciencia de Datos para aplicar un sistema de inventarios permite mejorar la precisión de las predicciones de demanda, optimizar el nivel de inventario, reducir los costos y mejorar la satisfacción del cliente. Al adoptar un enfoque basado en datos, Panificadora México puede no solo sobrevivir, sino prosperar en un mercado competitivo, asegurando que sus operaciones sean eficientes, sostenibles y orientadas al cliente.
- Integrar modelos predictivos en el sistema de inventarios en Panificadora México es una estrategia esencial para optimizar recursos y reducir costos. La precisión en la previsión de la demanda, combinada con la capacidad de adaptarse rápidamente a las condiciones cambiantes del mercado, no solo mejora la eficiencia operativa, sino que también proporciona una base sólida para la toma de decisiones estratégicas. En un mercado competitivo, la adopción de tecnologías avanzadas y la utilización de datos para guiar las operaciones son cruciales para mantener y mejorar la competitividad y el éxito a largo plazo de la empresa.

- Crear estrategias permiten a Panificadora México tener un sistema de inventarios automatizado lo que provoca estar en una mejora continua, ser pionera en el uso de la Ciencia de Datos para la demanda de inventarios produce una ventaja competitiva enorme y tendrá una respuesta rápida a los cambios de la demanda, adaptándose a las condiciones de mercado en tiempo real.

11. RECOMENDACIONES

- Explorar en detalle los beneficios de la implementación de algoritmos de ensemble en una base de datos de ventas de productos, así como examinar cómo estos algoritmos pueden mejorar la precisión predictiva, reducir el sobreajuste, crear robustez frente a datos ruidosos, capturar relaciones complejas entre variables y adaptarse a cambios en los datos.
- Realizar una evaluación exhaustiva de diferentes modelos de algoritmos de ensemble, como Random Forest y XGBRegressor aplicado a los productos para tener mayor objetividad.
- Proporcionar análisis comparativos detallados, destacando las fortalezas y debilidades de cada modelo en diferentes contextos y grupos de productos.
- Examinar cómo la implementación de estos modelos puede influir en la eficiencia operativa y la rentabilidad de las empresas, especialmente en términos de planificación de producción, gestión de inventarios y satisfacción de la demanda del mercado.

- Proporcionar recomendaciones prácticas para la implementación exitosa de algoritmos de ensemble en entornos empresariales, incluyendo consideraciones técnicas, recursos necesarios y mejores prácticas para la integración de estos modelos en los procesos existentes.
- Identificar posibles áreas de investigación futuras, como la optimización de algoritmos de ensemble para casos específicos de negocio, la exploración de nuevas técnicas de modelado y la integración de datos adicionales para mejorar la precisión predictiva.

12. REFERENCIAS BIBLIOGRÁFICAS

- Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time Series Analysis: Forecasting and Control* (5th ed.). Wiley
- Carreño Dueñas, D. A., Amaya González, L. F., Ruiz Orjuela, E. T., & Tiboche, F. J. (2019). Diseño de un sistema para la gestión de inventarios de las pymes en el sector alimentario. *Industrial Data*, 22(1). <https://www.redalyc.org/journal/816/81661270007/html/>.
- Coopers & Lybrand (1997), Los nuevos conceptos de control interno (Informe COSO). Madrid, España: Ediciones Diaz de Santos.
- Coronel, C., Morris , S., & Rob, P. (2011). Database systems: Design, implementation and management. [Sistema de base de datos: Diseño, implementación y gestión.]. Recuperado de <https://n9.cl/i9a58>
- Farooq, S., Sattar, A., & Mukhtar, H. (2018). Machine Learning Applications in the Food Industry. En **Intelligent Systems in Production Engineering and Maintenance** (pp. 315-335). Springer.
- Gutiérrez, V. y Rodríguez, L. (2008). Diagnóstico regional de gestión de inventarios en la industria de producción y distribución de bienes. *Revista Facultad de Ingeniería Uniersidad de. Antioquía*, (45), 157-171.
- Li, Q., Hu, Y., & Guo, Y. (2022). Case Studies on the Application of Machine Learning in Small Food Production Businesses. **Journal of Food Engineering*, 316*, 110880.
- Lobato, T. da C., & Carvalho, B. do N. (2021). Proposta de um modelo ensemble para credit scoring. *Brazilian Journal of Development*, 7(3), 24280–24297. <https://doi.org/10.34117/bjdv7n3-232>

- Muñoz-Hernández, H., Osorio-Mass, R. C., & Zúñiga-Pérez, L. M. (2016). Inteligencia de los negocios. Clave del éxito en la era de la información. [Business Intelligence: Key to Success in the Information Age]. *Clío América*, 10(20), 194-211. <https://doi.org/10.21676/23897848.1877>
- Osorio García, C. (2008). Modelos para el control de inventarios en las pymes. *PANORAMA*, 2(6), 4-10. <https://www.redalyc.org/pdf/3439/343929218002.pdf>
- Pavón Sierra, D. E., Villa Andrade, L. C., Rueda Manzano, M. C., & Lomas, E. X. (2019). Control interno de inventario como recurso competitivo en una PyME de Guayaquil. *Revista Venezolana de Gerencia*, 24(87), 860-873. Universidad del Zulia. https://www.redalyc.org/journal/290/29060499014/html/#redalyc_29060499014_ref4
- Raschka, S., & Mirjalili, V. (2019). *Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow 2*. Packt Publishing.
- Sagi, O., & Rokach, L. (2018). Ensemble learning: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(4), e1249.
- Tarrés Benet, L. (2019). Clasificación de lesiones en la piel con un ensemble de redes neuronales residuales [Tesis de grado]. Repositorio institucional RiUPC Escola Tècnica Superior d'Enginyeria de Telecomunicació de Barcelona. <http://hdl.handle.net/2117/165984>

- Valdez, Alvarado. A. (2019). Machine Learning para Todos.
[https://www.researchgate.net/publication/338518406_Machine_Learning
_para_Todos](https://www.researchgate.net/publication/338518406_Machine_Learning_para_Todos)
- Wang, J., Zhang, T., Li, C., & Wu, C. (2019). Predicting Consumer Preference for Food Products Using Machine Learning. *Food Research International, 122*, 275-284.
- Zare, M., Tappia, E., & Paliyath, G. (2021). Machine Learning Techniques for Food Quality and Safety: State-of-the-Art and Future Challenges. *Trends in Food Science & Technology, 108*, 214-227.
- Zhang, Y., Zhang, H., Zhang, J., & Liu, X. (2020). Application of Machine Learning in Food Production Process Optimization. *IEEE Access, 8*, 118156-118166.
- Zhou, Z. H. (2021). Machine learning. Springer nature.

13. ANEXOS

https://colab.research.google.com/drive/1JNU5InZaKUPrRvpWOnr054X1QVIXdD6?usp=drive_link